INVESTIGATING MULTILINGUAL ADOLESCENTS' COGNITIVE AND CODESWITCHING PERFORMANCE: AN EXPLORATORY APPLICATION OF LANGUAGE CONTROL MODELS

May Zantout

Lady Margaret Hall

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary Term 2019

DEDICATED TO

Danny and Jenna

ABSTRACT

Different interactional settings require multilingual speakers to switch from one language to another (i.e., to 'codeswitch') or to speak in one language only. This ability to alternate languages, or adhere to one language, involves a language control mechanism, a core element of Green's (1998) Inhibitory Control Model (ICM) which proposes to explain how bilinguals regulate control of their language systems. Recently, Green and colleagues (Green & Abutalebi, 2013; Green & Li, 2014; Green, 2018) have extended this model by putting forth the Adaptive Control Hypothesis (ACH) and the Control Process Model (CPM) which take into account external and internal variables, such as language proficiency and code-switching habits, which may influence language choice. While the ICM has been generally supported, the ACH and CPM have not been extensively investigated. The present study explores the application of the aforementioned models and hypothesis in a multilingual context by examining balanced and unbalanced multilingual speakers as they switch between three languages. It also examines how language proficiency, self-reported code-switching behaviour, and other individual differences influence cognitive and code-switching task performance. Another aim of the project is to assess whether having a high proficiency in a third language confers additional cognitive and/or linguistic benefits. Sixty adolescent participants (aged 17-18), 30 French-immersion students and 30 English-immersion students, were recruited in Lebanon, a multilingual country where French-English-Arabic code-switching is ubiquitous, to perform cognitive and linguistic tasks. A Flanker task, the arrow Simon task, and a backward digit-span task were also administered to examine conflict monitoring, inhibitory control, and working memory, as well as possible correlations between cognitive and linguistic task performance. Participants also performed a picture-naming task and a narration task which allowed the examination of switch costs and fluency, respectively. It was predicted that the French-immersion participants, given their frequent trilingual code-switching and higher proficiency level in their third language, would outperform their English-immersion counterparts, whose use of French is minimal, in both domain-general and linguistic tasks. Results, however, indicated that the two groups performed similarly in cognitive tasks. However, on the linguistic tasks, French-immersion (more balanced) participants exhibited different patterns from their English-immersion counterparts (who were less balanced) on several conditions of the trilingual picture-naming task and on the narration task in the single-language condition. Switch costs differed across the three languages for both groups, with the strongest language(s) requiring increased inhibition, thus lending support to the ICM and CPM. Language experience, including frequency of code-switching and salient-cue detection, was also highly correlated with performance on both linguistic and cognitive tasks, therefore lending support for the ACH. Implications of the role of the third language are discussed in terms of the types of tasks employed in this study, and discussions of the concept of language proficiency and habitual code-switching experiences vis-à-vis the control processes in both models are offered.

ACKNOWLEDGEMENTS

I would like to thank my mentors who have accompanied me throughout my doctoral journey at the University of Oxford. Firstly, I am indebted to my primary supervisor, Prof. Victoria Murphy, who has been a pillar of support. She has offered advice, insights, feedback, and encouragement on the personal and academic levels, and I very grateful for her guidance. Vicki, thank you for being a great mentor and friend. I would also like to thank my second supervisor, Prof. Li Wei, who has very generously devoted his personal time and expertise towards helping me achieve my doctoral degree. Every discussion we have had has been enlightening, and I am very grateful for each comment and feedback you have given. I cannot thank you enough for your guidance, Li Wei. I would also like to thank my former second supervisor, Dr. Catherine Walter, who began this journey with me. I would not have followed my dream of pursuing a doctoral degree if it had not been for you, Catherine. Thank you for always believing in me. To my supervisors—present and past—I dedicate my thesis and success.

I would also like to thank my examiners, Prof. Ernesto Macaro and Dr. Roberto Filippi, who were present at the final stage of my doctoral journey. Thank you for the wonderful discussion and for making the viva voce one of the most beautiful, memorable experiences of my doctoral programme.

I am also very grateful for the most supportive family one could ever wish for. My amazing husband, Houssam, has stood by me through the ups and downs of my studies and supported me every step of the way. He has made many sacrifices and selfless decisions without hesitation; he brought true meaning to the word 'partner', and I am blessed to share my life with him. I would also like to thank my parents, Nada and Jamil, but words fail me. Mom and Dad, I can never say 'thank you' enough for your support, especially during the final stages of my thesis writing. I would also like to thank my brother, Bassam, and sister-in-law, Salwa, for their support and for the beautiful storyboards that Salwa designed for my narration task. Finally, many thanks to my beautiful inspirations, Danny and Jenna, whose multilingual utterances always amaze and amuse me. I love you both, always and forever. This is all for you.

Finally, I would like to thank my dear friends and family who have offered encouragement and support in numerous ways. Special thanks to my cousin, Ribal El Masri, who helped me recruit participants. Many thanks to Anna, Annina, Dr. Daniel Caro, and Dr. Daniel Lunn for the R-iveting statistics discussions. Thanks to all the participants who took part in my study and to the wonderful parents who welcomed me into their homes for data collection. In the spirit of this thesis, 'thanks, merci, and شكراً'. May my knowledge, education, and work be of great benefit to others.

Table of Contents

ABSTRACT ii
LIST OF TABLES ix
LIST OF FIGURES xii
LIST OF APPENDICES xiii
LIST OF ABBREVIATIONS xiv
CHAPTER 1: INTRODUCTION 1
1.1 BACKGROUND OF THE STUDY 1
1.2 AIMS OF THE STUDY 4
1.3 RESEARCH QUESTIONS 5
1.4 CONTEXT OF THE STUDY 6
1.5 SUMMARY OF THE THESIS CHAPTERS 8
CHAPTER 2: OVERVIEW OF THE LITERATURE 10
2.1 INTRODUCTION 10
2.2 BILINGUALISM AND EXECUTIVE FUNCTIONS 10
2.2.1 Defining Executive Functions 10
2.2.2 An Overview of Bilingualism and Cognition Research 13
2.2.3 Bilingualism and Inhibitory Control (IC) 16
2.2.3 Bilingualism and Conflict Monitoring 21
2.2.4 Bilingualism and Working Memory 26
2.2.5 Bilingual vs. Trilingual Advantage 31
2.2.6 Summary of Bilingualism and Executive Functions 36
2.3 LANGUAGE CONTROL AND EXECUTIVE FUNCTIONS 39
2.3.1 Defining Codeswitching 39
2.3.1.1 Mixing, Switching, and Borrowing 39
2.3.1.2 Codeswitching Typology 40
2.3.1.3 Bilingual vs. Trilingual Codeswitching 42
2.3.1.4 Voluntary vs. Cued Codeswitching 45
2.3.2 Measuring Codeswitching 46
2.3.2.1. Switch Costs 47
2.3.2.2 Fluency (Pausing) 48
2.3.3 Language Control Models 50
2.3.3.1 The Inhibitory Control Model 51
2.3.3.2 The Adaptive Control Hypothesis 68

2.3.3.3 The Control Process Model (CPM) 73
2.3.3.4 Summary of Language Control Models 86
2.4 Factors Influencing Cognitive and Linguistic Control 88
2.4.1 Language Proficiency 88
2.4.2 Language Experience 90
2.4.2.1 Codeswitching Frequency 90
2.4.2.2 Language Use 98
2.4.2.3 Language Exposure 99
2.4.2.4 Language Typological Distance 101
2.4.2.5 Age 103
2.4.2.6 Socioeconomic Status 104
2.4.3 Summary of Factors Influencing Task Performance 105
2.5 Rationale and Overarching Hypotheses 106
2.6 Chapter Summary 107
Chapter 3: Methodology 109
3.1 Introduction 109
3.2 Research Design 109
3.2.1 Variables 109
3.2.2 Research Questions 111
3.3 Sampling Frame 123
3.3.1 Sampling Procedure 123
3.3.2 Sample Size 126
3.4 Participants 128
3.5 Materials 131
3.5.1 Language Background Questionnaire (LBQ) 132
3.5.2 Oral Proficiency Tests (OPTs) 133
3.5.3 Verbal Fluency Task 136
3.5.4 The Flanker Task 138
3.5.5 The Arrow Simon Task 143
3.5.6 Backward Digit-Span Task 148
3.5.7 Trilingual Picture-Naming Task (TPNT) 150
3.5.8 Narration Task 153
3.6 Overall Data Collection Procedure 156
3.7 Pilot Study 157

3.8 ETHICAL CONSIDERATIONS 158
3.9 CHAPTER SUMMARY 159
CHAPTER 4. RESULTS AND DISCUSSION CHAPTER I 161
4.1 INTRODUCTION 161
4.2 General Analytical Strategy 161
4.3 Cognitive Task Performance 164
4.3.1 RQ1A - The Flanker Task 165
4.3.1.1 Accuracy 165
4.3.1.2 Reaction Time (RT) 175
4.3.1.3 Monitoring Cost 180
4.3.1.4 Conflict Effect 182
4.3.2 RQ1B – The Simon Task 183
4.3.2.1 Accuracy 184
4.3.2.2 Reaction Time 186
4.3.2.3 Simon Effect 189
4.3.3 RQ1C – BDST 191
4.3.4 RQ1D – Correlations between Working Memory, Conflict Effect, and Simon Effect 193
4.3.5 Summary of Cognitive Task Performance 198
4.4 ANALYSIS OF THE INDEPENDENT VARIABLES 199
4.4.1 Oral Proficiency 199
4.4.1.1 Oral Proficiency Tests (OPT) 199
4.4.1.2 Verbal Fluency Task (VFT) 202
4.4.1.3 OPT vs. VFT scores 207
4.4.2 Language Experience 210
4.5 RQ2-- Predictors of Cognitive Task performance 222
4.5.1 Language Proficiency and Cognitive Performance 223
4.5.1.1. Language Proficiency and the Flanker Task 223
4.5.1.2. Language Proficiency and the Simon Task 228
4.5.1.3 Language Proficiency and Working Memory Task (BDST) 234
4.5.1.4 Summary of Language Proficiency as a Predictor of Cognitive Performance 238
4.5.2 Language Experience as a Predictor of Cognitive Performance 239
4.5.2.1. Language Experience and the Flanker Task 239
4.5.2.2. Language Experience and the Simon Task 243

4.5.2.3 Language Experience and Working Memory (BDST) 247
4.5.2.4 Summary of Language Experience as a Predictor of Cognitive Performances 250
4.6 Chapter Summary 250
Chapter 5: Results and Discussion Chapter 2 252
5.1 Introduction 252
5.2 RQ3—Linguistic Task Performance 252
5.2.1 Trilingual Picture-Naming Task 252
5.2.1.1 General Analytical Strategy 253
5.2.1.2 Accuracy 254
5.2.1.3 Prior Language Effects 258
5.2.1.4 Summary of TPNT Accuracy Analysis 263
5.2.1.5 Reaction Times (RTs) 263
5.2.1.6 Prior Language Effects 268
5.2.1.7 Summary of TPNT Analyses 272
5.2.2 The Narration Task (NT) 273
5.2.2.1 General Analytical Strategy 273
5.2.2.2 Pause Frequency Analysis 276
5.2.2.3 Pause Duration 282
5.2.2.4 Summary of NT Analyses 286
5.2.3 Correlation between NT and TPNT 287
5.3 RQ4- Predictors of Linguistic Task Performance 291
5.3.1 Language Proficiency and Linguistic Performance 292
5.3.1.1 Language Proficiency and Switch Costs 293
5.3.1.2 Language Proficiency and Pause Frequency 300
5.3.1.3 Language Proficiency and Pause Duration 308
5.3.2 Language Experience as a Predictor of Linguistic Task Performance 316
5.3.2.1. Language Experience and Switch Costs 317
5.3.2.2 Language Experience and Pause Frequency 321
5.3.2.3 Language Experience and Pause Duration 327
5.4 RQ5 – Relationship between Cognitive and Linguistic Task Performance 333
5.5 Chapter Summary 340
Chapter 6: General Discussion 341
6.1 Introduction 341

6.2 Summary and Discussion of Overall Findings 341
6.2.1 Executive Functions and the Trilingual Advantage 342
6.2.1.1 The Absence of the Trilingual Advantage 342
6.2.1.2 Cognitive Task Correlations 343
6.2.1.3 Inhibitory Control, Conflict Monitoring, and Working Memory 344
6.2.2 Linguistic Task Performance and Control Model Applications 348
6.2.2.1 Switch Costs 348
6.2.2.2 Fluency 352
6.2.2.3 Switch Cost and Fluency Overlap 356
6.2.3 Applicability of the Theoretical Models 357
6.3 Methodological and Analytical Limitations 361
6.4 Chapter Summary 363
Chapter 7: Conclusion 364
7.1 Introduction 364
7.2 Conclusions and Pedagogical Implications 364
7.3 Contributions to the field 368
7.4 Suggestions for future research 370
References 372
Appendix A 404
Appendix B 420
Appendix C 422
Appendix D 424
Appendix E 446
Appendix F 468
Appendix G 473
Appendix H 477

LIST OF TABLES

Table	Table Title	Page
Table 2.1	Switches vs. Loans Distinction	40
Table 2.2	Muysken's Typology of Codeswitching (Codemixing)	41
Table 2.3	ACH: Influence of Different Contexts on Task Schemas and Control Processes	69
Table 2.4	ACH Cognitive Processes Description	71
Table 3.1	Description of Variables	110
Table 3.2	A Priori Sample Size Calculations for Repeated Measures ANOVA (F-Tests)	127
Table 3.3	A Priori Sample Size Calculations for Correlations (T-Tests	128
Table 3.4	Participant Profile Summaries by Immersion Type	129
Table 3.5	Verbal Fluency Task Categories	137
Table 3.6	Flanker Task Stimuli Conditions	139
Table 3.7	Flanker Task Block Congruency Summary	141
Table 3.8	Flanker Task Block Sequence Counterbalancing Summary	141
Table 3.9	Backward Digit Span Task	150
Table 3.10	Summary of Pilot Study Modifications	158
Table 4.1	Mean and Standard Deviations of Flanker Task Accuracy Rates	166
Table 4.2	GLMM Output for Flanker Accuracy Data	168
Table 4.3	GLMM ANOVA (Type III) Summary for Flanker Accuracy Data	169
Table 4.4	Flanker Accuracy Monitoring x Congruence Contrasts	171
Table 4.5	Mean and Standard Deviations of Flanker Task Reaction Times (RTs)	176
Table 4.6	MEM ANOVA (Type III) Summary for Flanker RTs	177
Table 4.7	GLM ANOVA (Type III) Summary of Block Order (Flanker Accuracy)	179
Table 4.8	GLM ANOVA (Type III) Summary of Block Order (Flanker RTs)	179
Table 4.9	Descriptive Statistics of Monitoring Costs	181
Table 4.10	Mean and Standard Deviations of Simon Task Accuracy Rates	184
Table 4.11	GLMM Output for Simon Task Accuracy Data	185
Table 4.12	Mean and Standard Deviations of Simon Task RTs	186
Table 4.13	GLM Output for Simon RTs	187
Table 4.14	GLMM output for BDST scores	193
Table 4.15	Correlation Coefficient Matrix of Cognitive Tasks	195
Table 4.16	Means and Standard Deviations (in parentheses) of OPT Scores	201
Table 4.17	High Proficiency in Two vs. Three Languages (HPBT)	202
Table 4.18	Means and Standard Deviations (in parentheses) of the VFT Scores	204
Table 4.19	VFT Language-Pair Contrasts (EMI)	205
Table 4.20	VFT Language-Pair Contrasts (FMI)	206
Table 4.21	Language Experience -- Construct Description and Measurement	211
Table 4.22	Language Experience Constructs -- Descriptive Data	212
Table 4.23	Language Use EMI vs. FMI	213
Table 4.24	Language Use Contrasts (EMI)	214
Table 4.25	Language Use Contrasts (FMI)	215
Table 4.26	Language Exposure EMI vs. FMI	216
Table 4.27	Language Exposure Contrasts (EMI)	217
Table 4.28	Language Exposure Contrasts (FMI)	217
Table 4.29	MEM ANOVA (Type III) Summary for Conflict Effect x OPT Scores	224

Table 4.30	MEM ANOVA (Type III) Summary for Conflict Effect by VFT Scores	225
Table 4.31	MEM Output Summary for Conflict Effect by VFT Scores	225
Table 4.32	MEM ANOVA (Type III) Output Summary for Conflict Effect by HPBT and OPT	227
Table 4.33	MEM Output Summary for Conflict Effect by HPBT and VFT scores	228
Table 4.34	MEM ANOVA (Type III) Table of Main Effects and Interactions for Simon Effect scores by OPT	229
Table 4.35	MEM Output Summary for Simon Effect by OPT scores	230
Table 4.36	MEM Output Summary for Simon Effect by VFT scores	231
Table 4.37	MEM Output Summary of Simon Effect by HPBT and OPT	232
Table 4.38	MEM Output Summary for Simon Effect by HPBT and VFT	233
Table 4.39	GLMM Output Summary for BDST by OPT	235
Table 4.40	GLMM Output Summary for BDST and VFT	236
Table 4.41	GLMM Output Summary for BDST by HPBT and OPT	237
Table 4.42	GLMM Output Summary for BDST by HPBT and VFT	238
Table 4.43	GLM Output Summary for Conflict Effect and Language Experience (Model 1)	241
Table 4.44	GLM Output Summary for Conflict Effect scores by Language Experience (Model 2)	243
Table 4.45	GLM Output Summary for Simon Effect by Language Experience (Model 1)	245
Table 4.46	GLM Output Summary for Simon Effect by Language Experience (Model 2)	247
Table 4.47	GLMM Output Summary for BDST by Language Experience (Model 1)	249
Table 4.48	GLMM Output Summary for BDST by Language Experience (Model 2)	249
Table 5.1	TPNT Accuracy Means (in %) and Standard Deviations (EMI and FMI)	255
Table 5.2	GLMM ANOVA (Type III) Summary for TPNT Accuracy Data	255
Table 5.3	GLMM Output Summary for TPNT Accuracy Data	257
Table 5.4	TPNT mean accuracy rates in percentages and Standard Deviations for Language Pairs	259
Table 5.5	TPNT Mean RTs (in ms) and Standard Deviations	264
Table 5.6	MEM ANOVA (Type III) Summary of TPNT RT Data	265
Table 5.7	MEM Output Summary for TPNT RT Data	267
Table 5.8	TPNT mean RTs in milliseconds and Standard Deviations (in parentheses) for Language Pairs	269
Table 5.9	PRAAT Output and Computations	274
Table 5.10	Mean and SD of Pause Frequency Ratios	277
Table 5.11	MEM ANOVA (Type III) Summary of Narration Task -- Pause Frequency	277
Table 5.12	MEM Output for Pause Frequency Ratio	278
Table 5.13	Narration Task Pause Frequency Group Contrasts	279
Table 5.14	NT Pause Frequency Ratios Post-hoc Analyses by Group	280
Table 5.15	Mean and SD of NT Pause Durations	282
Table 5.16	MEM ANOVA (Type III) Summary of Narration Task Pause Duration	282
Table 5.17	MEM Output for Pause Duration	283
Table 5.18	NT Mean Pause Duration -- Group Contrasts	284

Table 5.19	NT Pause Duration Planned-Contrast Analyses by Group	285
Table 5.20	EMI Correlation Analyses – TPNT RTs and NT Pause Durations	289
Table 5.21	FMI Correlation Analyses – TPNT RTs and NT Pause Durations	290
Table 5.22	MEM ANOVA (Type III) Summary for Switch Costs (Model 1)	293
Table 5.23	MEM Output Summary for Switch Costs and OPT (Model 1)	294
Table 5.24	MEM ANOVA (Type III) Summary for Switch Costs and VFT (Model 2)	296
Table 5.25	MEM ANOVA (Type III) Summary for Switch Costs and HPBT (Model 3)	299
Table 5.26	MEM ANOVA (Type III) Summary for Pause Frequency and OPT (Model 1)	301
Table 5.27	MEM Output Summary for Pause Frequency *post-hoc* analyses	302
Table 5.28	MEM Output Summary for Pause Frequency and VFT	304
Table 5.29	MEM ANOVA (Type III) Summary of French-only Condition and French VFT Scores	305
Table 5.30	MEM Output Summary for Pause Frequency and HPBT	307
Table 5.31	MEM ANOVA (Type III) Summary for Pause Duration and OPT	309
Table 5.32	Overall Summary of post-hoc analyses for Pause Duration	310
Table 5.33	MEM ANOVA (Type III) Summary for Pause Duration and VFT (Model 2)	313
Table 5.34	MEM ANOVA (Type III) Summary for Pause Duration and HPBT (Model 3)	313
Table 5.35	MEM Output Summary for Pause Duration and HPBT post-hoc analysis	314
Table 5.36	MEM ANOVA (Type III) Summary of Switch Costs and Language Use (Model 1)	317
Table 5.37	MEM Output Summary for Switch Costs and Language Experience (Model 2)	319
Table 5.38	MEM ANOVA (Type III) Summary for Switch Costs and Language Experience (Model 3)	320
Table 5.39	MEM ANOVA (Type III) Summary for Pause Frequency and Language Use (Model 1)	322
Table 5.40	MEM Output Summary for Pause Frequency and Language Use (Model 1)	323
Table 5.41	MEM Output Summary for Pause Frequency and Language Exposure (Model 2)	325
Table 5.42	MEM ANOVA (Type III) Summary of Pause Frequency, Codeswitching Frequency, and Salient Cue Detection (Model 3)	326
Table 5.43	MEM ANOVA (Type III) Summary of Codeswitching Frequency, Salient Cue Detection, and Trilingual Pause Frequency	327
Table 5.44	MEM ANOVA (Type III) Summary of Pause Duration and Language Use (Model 1)	328
Table 5.45	MEM ANOVA (Type III) Summary of Duration and Language Exposure (Model 2)	329
Table 5.46	MEM ANOVA (Type III) Summary of Pause Duration, Codeswitching Frequency and Salient Cue Detection (Model 3)	331
Table 5.47	GLM ANOVA (Type III) Summary of Switch Costs and Cognitive Tasks	334
Table 5.48	Mean Number and Mean Duration (in seconds) of Trilingual Narration Task Switch Directionality	339
Table 5.49	GLM Summary of Narration Task and Cognitive Predictors	340

LIST OF FIGURES

Figure	Figure Title	Page
Figure 2.1	The Inhibitory Control Model (adapted from Green, 1998)	52
Figure 2.2	Codeswitching Control Process Model (Green & Li, 2014)	74
Figure 3.1	Schematic representation of the events in an experimental trial of the flanker task	143
Figure 3.2	Arrow Simon Task Trials	145
Figure 3.3	Schematic representation of components of a congruent trial in the Arrow Simon Task	148
Figure 3.4	Schematic representation of a switch trial event from Lebanese Arabic to English in the trilingual picture-naming task	152
Figure 4.1	Flanker Accuracy Data (EMI vs. FMI)	167
Figure 4.2	Flanker Accuracy Interaction (Monitoring x Congruence)	169
Figure 4.3	Flanker Accuracy rates	174
Figure 4.4	Flanker RTs	178
Figure 4.5	Simon Accuracy Rates -- Congruence Contrasts	186
Figure 4.6	Simon RTs – Congruence Contrasts	188
Figure 4.7	BDST Raw Scores	193
Figure 4.8	Scatterplot of Cognitive Task Correlations	196
Figure 4.9	OPT scores by group	201
Figure 4.10	Violin plots for VFT distributions	205
Figure 4.11	Correlations between OPT and VFT scores (EMI)	208
Figure 4.12	Correlations between OPT and VFT Scores (FMI)	209
Figure 4.13	Language Use by group	214
Figure 4.14	Density distribution plot of language use (EMI)	215
Figure 4.15	Density distribution plot of language use (FMI)	216
Figure 4.16	Overlapping Distributions of Language Use vs. Language Exposure (EMI vs. FMI)	220
Figure 4.17	Interaction between CSF and SalC in the Simon Task	246
Figure 5.1	TPNT Accuracy Interaction Plot (Group x Language)	256
Figure 5.2	TPNT Mean Accuracy Rates (in %) by Language Pairs	262
Figure 5.3	Violin plots showing the distribution of the TPNT RTs	265
Figure 5.4	Interaction Plot of TPNT RT data	266
Figure 5.5	Schematic representation of RT switch costs across the language pairs	271
Figure 5.6	Pause Frequency x Group x Language Condition Interaction Plot	278
Figure 5.7	Density distribution of Pause Frequency by group	281
Figure 5.8	Pause Duration x Group x Language Condition Interaction plot	283
Figure 5.9	Pause Duration by Group and Language Condition	286
Figure 5.10	Correlations between TPNT and NT (EMI)	289
Figure 5.11	Correlations between TPNT and NT (FMI)	290
Figure 5.12	OPT scores x Switch Costs (per language)	295
Figure 5.13	Switch Costs x VFT scores	298
Figure 5.14	Pause Frequency as predicted by OPT Scores (EMI vs. FMI)	301
Figure 5.15	HPBT x Language Condition Interaction Plot	308
Figure 5.16	Interaction plot for Group x OPT Scores x Pause Duration	310
Figure 5.17	HPBT x Language Condition x Pause Duration Interaction Plot	315
Figure 5.18	Switch Costs x Language Use x Group Interaction Plot	318
Figure 5.19	Language Use x Pause Frequency x Group Interaction plot	322
Figure 5.20	Language Use x Pause Duration x Group Interaction Plot	328
Figure 5.21	Switch Costs by Language x Group x Simon Effect score Interaction plot	335
Figure 5.22	Inhibitory Control x Switch Costs x Language Interaction Plot	336

LIST OF APPENDICES

Appendix	Appendix Name	Page
Appendix A	Language Background Questionnaire	404
Appendix B	Oral Proficiency Test	420
Appendix C	OPT Scoring Rubric	422
Appendix D	Narration Task Sample	424
Appendix E	CUREC, Information Sheets, and Consent Forms	446
Appendix F	Data Collection Schedule	468
Appendix G	Pilot Study Descriptive Analysis	473
Appendix H	Chapter 4 Supplementary Analyses	477
Appendix I	Chapter 5 Supplementary Analyses	493

LIST OF ABBREVIATIONS

ACH	Adaptive Control Hypothesis
BDST	Backward Digit Span Task
CPM	Control Process Model
CSF	Codeswitching Frequency
EMI	English Medium of Instruction
FMI	French Medium of Instruction
HPBT	High Proficiency Bilingualism or Trilingualism
IC	Inhibitory Control
ICM	Inhibitory Control Model
L1	First Language
L2	Second Language
L3	Third Language
LBQ	Language Background Questionnaire
MEM	Mixed-Effects Model
NT	Narration Task
OPT	Oral Proficiency Test
RQ	Research Question
RT	Reaction Times
SalC	Salient Cue Detection
SE	Standard Error
TPNT	Trilingual Picture Naming Task
VFT	Verbal Fluency Task
WM	Working Memory

CHAPTER 1: INTRODUCTION

This chapter offers a brief introduction to the study, outlines its overarching aims and research questions, describes the context of the study, and offers a summary of the thesis chapters.

1.1 BACKGROUND OF THE STUDY

It has been suggested that it is almost impossible to extricate a monolingual speech segment from Lebanese informal conversations (Joseph, 2004). This is because Lebanon is a multilingual society whose speakers typically produce utterances combining English, French, and Lebanese Arabic. The presence of and seamless alternation between the three languages is a staple of Lebanese speech (Grosjean, 1982), yet it is not unique to the Lebanese speech community. In almost every linguistic context where languages are in contact, and depending on different interactional demands, bilingual and multilingual speakers either switch from one language to another or speak in one language only. This ability to alternate languages (i.e., to *codeswitch*) or to adhere to one language consistently is generally driven by the speakers' motivation to ensure communicative success. Despite appearing like a simple task, however, producing the intended language(s) while suppressing interference from the non-intended language(s) (i.e., 'language control') is said to be regulated by a complex cognitive network, as postulated by the Inhibitory Control Model (ICM, Green, 1986, 1998) and its extension the Control Process Model (CPM, Green & Li, 2014; Green, 2018). Numerous studies and theoretical models have explored how bilingual and multilingual speakers are able to control their speech production, with ample evidence in support of the notion that inhibitory control mechanisms facilitate bi/multilingual speech production and

codeswitching (Schwieter & Ferreira, 2016). However, the mechanisms and cognitive processes which regulate language production and mediate the ability to codeswitch are still debated (Green & Li, 2014). This study aims to contribute to the literature on language control by examining the inhibitory mechanisms required to produce one-word utterances and prolonged utterances in three languages and by exploring the overlap in the processes underlying linguistic and cognitive control.

The implications of the relationship between bi/multilingualism and cognitive control extend beyond examining language control. A plethora of studies has purported the notion of 'bilingual advantage' where bilingual speakers are said to have enhanced cognitive performance compared to their monolingual counterparts (Bialystok, Craik & Luk, 2012). This is attributed to bilinguals' experience in constantly controlling and monitoring their two languages at the same time to ensure the target language is selected (Green, 1998). Bilingualism has been said to enhance 'executive functions' which is an overarching term referring to the processes regulating behaviour, including speech (Miyake, Friedman, Emerson, Witzki, Howerter, & Wager, 2000). It is argued that bilinguals must rely on these executive functions to constantly monitor the two languages to ensure that the target language is selected (Green, 1998). As such, bilingualism 'trains' these functions on a daily basis, as increased cognitive demands associated with language control sharpen cognitive control skills. These skills are then deployed in performing non-linguistic tasks, leading to ostensible cognitive advantages in bilinguals (Green & Abutalebi, 2013; Kroll, Bobb, & Hoshino, 2014). However, it is not clear which aspects of the bilingual experience confer cognitive benefits, and whether these benefits are specific to a certain type of cognitive skill.

Indeed, the robustness of the bilingual advantage has been challenged (Paap & Greenberg, 2013), mainly due to the lack of replicability of previous findings

(Duñabeitia, Hernàndez-Cabrera, Estévez, Macizo, Anton, & Fuentes, 2014; Hilchey & Klein, 2011). One attempt to reconcile the inconsistencies in the findings has been proposed by Bak (2016) who suggests that discrepancies may stem from differences in bilingual participants' profiles. By examining individual (and group) differences, such as language proficiency, language exposure, language use, and codeswitching frequency, it is possible to identify the specific aspects of the bilingual experience which may influence executive functions. As such, examining the habitual linguistic practices of bilinguals may provide a better understanding of the relationship between bilingualism and domain-general cognitive processes (Bialystok, 2009; Green & Abutalebi, 2013). Recently, the Adaptive Control Hypothesis (ACH, Green & Abutalebi, 2013) and CPM have been put forth to take into account not only individual differences but also differences in the cognitive demands placed on bilinguals depending on the interactional context they are in. The ACH and CPM suggest that bilinguals who constantly separate their languages differ in terms of cognitive control from bilinguals who codeswitch frequently. Accordingly, it is of interest to break down the notions of 'bilingualism' and 'cognitive control' in order to achieve a better understanding of the precise interaction between the two constructs. The present study, therefore, aims to offer new insights into the advantages of bi/multilingualism by deconstructing the terms—and processes—underpinning the notion of the bilingual advantage. Further, this study investigates whether language proficiency and language experience, or individual differences, influence performance on both linguistic and cognitive tasks.

While the majority of studies examine the bilingual advantage by comparing bilingual groups to monolingual ones, there have been recent calls to examine a 'multilingual advantage'. Multilingualism has been said to statistically supersede monolingualism on a global scale (Edwards, 2019). Accordingly, it is important to

investigate whether learning a third language (L3) confers similar advantages to learning a second language, and whether these advantages are qualitatively different. The dearth of studies examining multilingual contexts may be ascribed to theoretical and practical challenges which researchers face. Much of the literature and the models put forth to explain language control claim that they can extend from bilingual to multilingual contexts (Hoffmann & Stavans, 2007). However, multilingualism involves a larger array of factors and contexts in which different languages are acquired, processed, separated, and combined, thus becoming more complex than the reductionist approach which subsumes multilingualism under bilingualism (Sorace & Serratrice, 2009). Indeed, multilingualism may promote or engender specific aspects of cognition during language processing which are not accessed or involved in bilingual language control. The present study aims to fill this gap in the literature by testing the applicability of the three current proposals of language control, namely the ICM, ACH, and CPM.

1.2 AIMS OF THE STUDY

In order to investigate the overlapping processes of cognitive and language control, as well as to further explore the notion of a 'multilingual' advantage, this project examines factors modulating cognitive and linguistic task performance by trilinguals who vary in their L3 proficiency and language experience. Adopting the ICM, ACH, and CPM as its theoretical framework, the study bases its design on testing executive functions and trilingual codeswitching separately and then correlating results from the different tasks. Specifically, it analyses results obtained from well-established cognitive tasks, namely the flanker task, the Simon task, the backward-digit span task, and a picture-naming task, as well as from a novel instrument which elicits long speech utterances in each language, as well as in a trilingual mode. Accordingly, the present

study's design endeavours to approximate language production in single-word utterances studied in experimental conditions to naturalistic codeswitching behaviour exhibited by multilingual speakers in their everyday interactions. In so doing, the study aims to assess whether the ICM and CPM are supported in terms of the different cognitive demands underlying language control. Performance on non-linguistic measures of executive functions, as well as on linguistic measures, is further explored by examining the influence of modulating factors such as participants' codeswitching frequency, language use, and language exposure. Furthermore, the three typologically distant languages of the selected speech community—Arabic, English, and French—have never been examined together in the literature, adding to the contributions of this study. Discussions and implications of the concepts of language proficiency and different codeswitching experiences vis-à-vis the current models also contribute to the literature. The project thus aims to fill gaps in the literature both on a theoretical and a methodological level.

1.3 RESEARCH QUESTIONS

In order to address the aims of the study, the research questions of this project are introduced here, but they are further elaborated upon in the Methodology chapter (Section 3.2.2).

RQ1. Do English Medium of Instruction (EMI) and French Medium of Instruction (FMI) participants differ in performance across the domain-general cognitive tasks: (A) the flanker task, (B) the Simon task, and (C) the Backward Digit-Span task?

RQ2. Is EMI and FMI participants' performance on non-linguistic, domain general tasks modulated by their language proficiency and/or language experience?

RQ3. Do EMI and FMI participants differ in performance on the linguistic tasks: (A) the trilingual picture-naming and (B) the narration task?

RQ4. Do language proficiency and language experience predict performance within and across EMI and FMI participants in the picture-naming task and narration task?
RQ5. What is the relationship between cognitive mechanisms and temporal switching costs in the linguistic tasks?

1.4 CONTEXT OF THE STUDY

Lebanon, a pluralistic society which espouses multilingualism, has witnessed major historical and sociolinguistic shifts that have created and shaped its diverse linguistic nature. The three main languages spoken in Lebanon are Arabic, English, and French, while minority languages include Armenian, German, Assyrian, and Kurdish (Bizri, 2013; Suleiman, 2003). Although Arabic has always been the main spoken language, numerous colonial powers have tried to dethrone its official use. For almost four centuries (16th-20th), Lebanon was under the rule of the Ottoman Empire which constantly attempted to replace Arabic with Turkish in governmental, educational, and social establishments (Nashef, 2013). French and American missionaries helped the Lebanese people retain their official language and advocated the use of Arabic along with French and English in educational settings (Shaaban & Ghaith, 1999). After the fall of the Ottoman Empire, Lebanon fell under the French mandate which tried to abolish Arabic by perpetuating the dominance of French over Arabic. As the attempt to eradicate Arabic failed, French authorities made both languages the country's official languages. American protestant missionaries, on the other hand, supported the use of Arabic along with English in religious and educational settings. They further established universities which then motivated the French missionaries to establish French universities, leading bilingual education to spread (1999).

After gaining its independence, Lebanon made Arabic the only official language, yet it emphasised the importance of language learning in its schools. Formal instruction in English and French is now the norm, where English Medium-of-Instruction (EMI) or French Medium-of-Instruction (FMI) programmes are adopted in both public (state) and private schools. Children are also formally instructed in Modern Standard Arabic (MSA), along with English/French as a second language (L2) starting at age 3, which is the school-entry age in Lebanon, and French/English as a third language (L3) in Year 1, 4, or 6, depending on the school, and up to Year 12. Children are taught all subjects, except for the languages and social studies, in their L2, making the immersion language their most frequently used language at school (Zantout, 2014). As such, by the age of 17-18, children are expected to have high proficiency in at least two of the three languages and to alternate between Arabic, English, and French inside and outside school.

Indeed, frequent codeswitching between the three languages is ubiquitous in social interactions, yet what makes this speech community unique is that it combines three typologically distant languages: Lebanese Arabic is a Semitic dialect with no standardised script, whereas English is a Germanic language and French is a Romance language (Comrie, 1991). These three-language combinations in utterances and conversations which span a variety of contexts (formal and informal) and mediums (written and spoken) are what distinguish the Lebanese context (Joseph, 2004). The three languages, however, are kept separate in school settings during formal class time yet are frequently codeswitched during non-instructional sessions in the school day, such as during break time and extra-curricular clubs. Furthermore, different interactional settings in which Lebanese children and adolescents converse are typified by different language use and frequent codeswitching depending on their interlocutors. For example, Lebanese children have been observed to use Arabic mainly with their grandparents, whereas they

would codeswitch between Arabic and the language of their immersion programme when conversing with their siblings (Zantout, 2014). Lebanese children and adolescents, therefore, constantly meander between various sociolinguistic contexts which exert different demands in terms of language choice. As such, examining Lebanese adolescents' speech may offer great insights into the mechanisms underlying switching between languages. Furthermore, the two types of immersion programmes in Lebanon allow the study of two types of multilinguals within the same multilingual community.

1.5 SUMMARY OF THE THESIS CHAPTERS

The thesis comprises seven chapters. With the first chapter providing a brief introduction to the aims and context of the study, Chapter 2 presents an in-depth literature review which offers important definitions of the main constructs examined in this study, namely cognitive control and codeswitching. It then offers a detailed review of the research on language and cognition, discussing both empirical and theoretical studies and articulating gaps in the literature. The chapter then discusses the theoretical underpinnings for the current study, with an emphasis on two language production models. The chapter concludes by offering an overview of the factors which modulate cognitive and linguistic task performance, linking these to the rationale and hypotheses of the current project.

Following the literature review, Chapter 3 describes the methodology adopted in the study. It offers a description of the design, the variables, the research questions (with a brief description of the method of analysis for each question), the participants, and the materials. The pilot study and ethical considerations are also briefly discussed.

Chapters 4 and 5 present the results of the present study. Chapter 4 addresses the first two research questions which pertain to participants' cognitive performance on three

tasks and the factors which influence this performance. Chapter 5 examines the results of the two linguistic tasks and examines the predictor variables which modulate this performance. It also examines the relationship between language control and cognitive control. Both results chapters include brief discussions of the findings at the task- (or micro-) level, whereas Chapter 6 offers a macro-level discussion of the overall findings.

Chapter 6, the General Discussion chapter, is divided thematically by integrating relevant research questions, and the findings of this study are discussed in light of the overall themes and hypotheses, as well as in relation to previous research. Discussions of the implications for the theoretical models and notions are offered. The chapter also considers the limitations of the present project.

The concluding chapter, Chapter 7, provides an overall summary of the study, its aims, and its findings, highlighting general trends observed in the data. It then offers a brief summary of the implications of these findings and attempts to place them within a larger context. The chapter then highlights the main contributions of the study and proposes suggestions for future research.

CHAPTER 2: OVERVIEW OF THE LITERATURE

2.1 INTRODUCTION

This chapter discusses relevant empirical and theoretical studies examining cognitive and linguistic control. It starts by offering an overview of cognitive control, followed by a critical discussion of bilingualism and cognitive control research. The chapter gives a definition and operationalisation of 'codeswitching', followed by a critical assessment of the theories and language control models (i.e., the conceptual framework) which are adopted in this study. Factors influencing cognitive and linguistic task performance are discussed, leading to the rationale, which is the final section of this chapter.

2.2 BILINGUALISM AND EXECUTIVE FUNCTIONS

In line with the aims of the study, this section aims to highlight the specific aspects of executive functions which are examined in this project. The section starts by offering a definition of 'executive functions', followed by an overview of bilingualism and cognition research. It then focuses on three cognitive elements, namely inhibitory control, conflict monitoring, and working memory. A discussion of a 'trilingual advantage' is then presented.

2.2.1 Defining Executive Functions

Many studies on bilingualism and, more recently, multilingualism, have focused on specific cognitive advantages that speakers of more than one language seem to have over their monolingual counterparts. One advantage is the ability to ignore distracting stimuli in everyday conversations, such as background sounds or other ambient languages being spoken. To ensure proper comprehension and production in the language/s of the interlocutor(s) in such noisy environments, bilingual speakers must rely on selective

mechanisms which enable them to suppress non-relevant information and monitor their speech for any 'inappropriate' utterances. These mechanisms fall under the term 'executive functions', which involves a network of processes orchestrating goal-oriented behaviour, such as speech production (Titone, Gullifer, Subramaniapillai, Rajah, & Baum, 2017). These processes are generally conceptualised as taking place in a top-down manner, where executive functions coordinate and maintain lower-level cognitive functions by either operating pre-emptively (i.e., 'proactive control') or immediately when cognitive demands arise (i.e., 'reactively') (Miyake & Friedman, 2012; Braver, 2012). This section offers a brief overview of 'executive functions'. As this construct is central to the present study, it is introduced in this section and expanded upon in the subsequent sub-sections.

Much of the research on executive functions and bilingualism adopts Miyake et al.'s (2000) framework of executive functions. According to this model, there are three main components: *inhibitory control*, *shifting*, and *updating*. Though distinct, all three components involve an element of inhibition, which is argued to be the locus of the bilingual advantage (Bialystok, Craik, Grady, Chau, Ishii, Gunj, & Pantex, 2005; Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009; Martin-Rhee & Bialystok, 2008; Verreyt, Woumans, Vandelanotte, Szmalec, & Duyck, 2016). Inhibitory Control (IC) refers to the ability to suppress irrelevant information or unintended behavioural patterns and task schemata and focus attention on relevant information. For instance, in speech production, a bilingual speaker conversing with a monolingual must suppress the non-target language in order to speak consistently in the language of their interlocutor. As such, it is argued that inhibitory control underlies language choice (Green, 1998). Recently, IC has been said to involve high-level monitoring processes, such as 'conflict monitoring' (Costa et al., 2009). 'Conflict monitoring' is a cognitive mechanism which

allows speakers to promptly adapt to changes in a task requirement. This is achieved through managing the co-activated task schemata, or action sequences which control output, including inhibitory schemata (Botvinick, Carter, Braver, Barch & Cohen, 2001). Bilinguals are therefore conceptualised as constantly monitoring external and internal cues and managing language selection through inhibition (Costa et al., 2009).

IC is said to overlap with *shifting* which involves conflict monitoring as well. *Shifting* refers to the switching back and forth between tasks (Miyake et al., 2000). For example, when conversing with two monolinguals, one French and one English, a French-English bilingual will need to speak in French to the French-speaking monolingual and then switch to English when conversing with the English-speaking monolingual. According to Costa et al. (2009) and Green and Abutalebi (2013), this ability to switch is mediated through *conflict monitoring* by activating English inhibitory schemata and deactivating French inhibitory schemata when the French-English bilingual is speaking to a French monolingual interlocutor, for example.

Furthermore, monitoring and inhibition are also said to be involved in the third construct—*updating*—which represents *working memory*. According to Miyake et al. (2000), updating requires monitoring and coding information relevant to a task and replacing (or inhibiting) old, irrelevant information with the new information. As such, updating entails monitoring (and processing in most cases) new incoming elements that are needed for a present task. For instance, in a backward-digit span task, participants are asked to listen to and recall numbers and then reproduce them in the opposite order. Inhibition allows participants to focus on the task at hand by suppressing irrelevant tasks or information, in this case the original sequence of the digits.

The 'unity and diversity' of Miyake et al.'s (2000) framework makes it difficult to distinguish the components. Based on the above, the present study loosely adopts the

framework and uses the construct of 'executive functions' as an umbrella term for *inhibitory control abilities*, *conflict monitoring skills*, and *working memory capacity*. Each of these three aspects of cognitive performance is examined separately, and overlaps are explored and explained when appropriate. Although the study does not examine non-verbal task shifting, references to this component are made in light of the literature on language switching.

2.2.2 An Overview of Bilingualism and Cognition Research

Over the past century, research investigating the influence of bilingualism on cognition has moved away from initial claims that bilingualism negatively impacts intelligence and causes confusion amongst bilingual speakers (Baker & Athanasopoulos, 2011). It has since been argued that these claims were perpetuated by socio-political motivations, especially in the case of immigration (Edwards, 2019). These early studies have also been criticised for numerous methodological shortcomings, such as using ethnocentrically biased measures of IQ, which were administered to bilinguals in English, including English non-native immigrants, and recruiting participants of diverse socioeconomic backgrounds (Hakuta, 1986). In the 1960s, researchers revised the means by which the link between bilingualism and cognition was investigated, resulting in a shift in findings.

The pioneering work of Peal and Lambert (1962) paved the way for more methodologically sound approaches to this field. In their landmark study, the researchers controlled for socioeconomic status by matching monolingual (English and French) and French-English bilingual groups of children (age 10). They also administered a series of verbal and nonverbal tests that tapped into a range of cognitive abilities, instead of using the traditional IQ test adopted in previous studies. Their results indicated that on the majority of tasks, bilinguals significantly outperformed their monolingual counterparts,

especially in tests examining what the researchers referred to as 'mental flexibility' (p.19). The researchers stated that the bilingual group in their study were able to outperform their monolingual counterparts due to the advantages which bilingualism conferred on their intelligence. Specifically, bilinguals were more adept at new 'concept formation' and 'a diversified set of mental abilities', as opposed to monolinguals whose unitary structure of intelligence had to cater for different task type demands (p.20). Despite certain shortcomings in the study, such as the non-objective measure of socioeconomic status and the broad nature of the intelligence tests, Peal and Lambert's study was the first to repudiate previous claims of the detrimental effects of bilingualism on cognition. It has also propelled further research into investigating the precise aspects of cognition which are linked to bilingualism (Bialystok, 2017). The increase in migration and spread of bilingual language education, as well as immersion programmes, have further fuelled interest in this field.

The shift towards a more objective and favourable attitude towards bilingualism and its 'consequences' is encapsulated by the *bilingual advantages hypothesis,* formulated by Bialystok and her colleagues (Bialystok, 1999; Bialystok et al., 2012). Based on Peal and Lambert's (1962) findings of cognitive advantages in bilingual children, this hypothesis was originally put forth to advocate metalinguistic awareness as the source of the bilingual advantage in children. However, with the surge in studies, Bialystok and her colleagues identified additional advantages, such as enhanced ability to selectively monitor, suppress, and manage competing linguistic cues or constraints (Costa, Hernández, & Sebastián-Gallés, 2008; Martin-Rhee & Bialystok, 2008; Yang, Yang, & Lust, 2011). Bialystok further extended the scope of the hypothesis to encompass findings from studies of different populations, including young and older adults and pathological populations (Bialystok, 2017).

Recently, more rigorous research has attempted to identify the specific aspects of executive functions associated with a potential bilingual advantage. This research generally adopts Miyake at al.'s (2000) framework of executive functions. Studies have observed better performance by bilinguals on inhibitory control and monitoring tasks, with inconsistent—yet promising—findings with respect to working memory (Bialystok, 2017; Grundy & Timmer, 2017). While these findings are elaborated upon further in the next section, it is important to note here that task type (verbal vs. nonverbal) and the associated cognitive skills are what typically incur differences between monolingual and bilingual performance (Bialystok, 2009). That is, bilinguals outperform monolinguals on non-verbal executive function performance. Conversely, monolinguals are observed to outperform bilinguals on word retrieval tasks due to interference effects which slow down bilinguals' lexical access. This is attributed to the parallel activation of both languages and interference from competing items which result in an increase in response latencies (Abutalebi & Green, 2007; Green, 1998). However, bilinguals' enhanced performance on executive function tasks are said to stem from their ability to resolve conflict due to this language coactivation. Accordingly, the experience of having to mentally juggle two languages is said to negatively influence lexical access yet simultaneously enhance executive functions. However, these findings are contested.

In the last few years, the veracity of the 'bilingual advantages hypothesis' (Bialystok, 1999) has been the centre of much controversy. Although researchers seem to be divided on whether an advantage truly exists (i.e., whether bilinguals consistently outperform monolinguals), there is overall consensus that bilingualism has detrimental effects as was the belief in the past[1]. The controversy is rooted in the methodological

[1] However, there are still cases in the United Kingdom in which educators have discouraged parents from raising their children as multilinguals (Festman, Poarch, & Dewaele, 2017), with arguments that the

concerns and discrepancy in the replicability of previous findings (Paap, 2014; Paap & Sawi, 2014). Specifically, empirical studies examining the bilingual advantage have been said to generally recruit a small number of participants and report low effect sizes (Paap & Greenberg, 2013). However, given the psycholinguistic nature of these tasks, this has been argued to be standard across all research subdivisions within psychology, including bilingualism and language production (Bak, 2016). Abutalebi and Green (2016) also point out that current research is attempting to tease apart 'executive functions' and explore other variables which may influence findings, such as individual differences among bilinguals, making replicability difficult. Individual variability in the bilingual experience, such as the frequency of codeswitching in which bilinguals engage across linguistic contexts, has been said to be a strong predictor of differential performance on cognitive tasks (Costa et al., 2009; Soveri, Rodriguez-Fornells, & Laine, 2011; Yang & Yang, 2016). Accordingly, current and future research is believed to better demarcate the specific features of bilingualism which pertain to specific aspects of executive functions. The following section offers a breakdown of the cognitive performance and bilingualism research for each component of executive functions.

2.2.3 Bilingualism and Inhibitory Control (IC)

The first component in Miyake et al.'s (2000) framework, inhibitory control, plays a central role in theories of language control. As mentioned earlier, IC refers to the ability to suppress irrelevant information and to focus attention on relevant information. IC may be further divided into two cognitive processes: interference suppression and response inhibition. Interference suppression, which occurs at the conceptual level before

'English as an Additional Language' label stigmatises children as 'special needs' (Filippi et al., 2018) and/or of poor socioeconomic backgrounds.

a behavioural response is generated, refers to the ability to ignore distracting information and direct attention to the target response. Response inhibition refers to inhibition of an already generated (typically automated) behavioural response (Martin-Rhee & Bialystok, 2008). This dichotomy is supported by fMRI evidence where different activation patterns revealed the dissociation between the two processes: while interference suppression results in increased activation in the anterior cingulate cortex, response inhibition increases activity in prefrontal cortex (Abutalebi & Green, 2007). Interference suppression is typically examined through tasks involving bivalent stimuli (stimuli involving two conflicting dimensions), whereas response inhibition is tested through tasks involving univalent stimuli (stimuli which have one dimension and two possible responses). Accordingly, the incongruence of interference suppression stimuli involves suppression of the irrelevant dimension and inhibition of the incorrect response, thereby making these tasks more complex than response-inhibition tasks (Bunge, Dudukovic, Thomason, Vaidya, & Gabrieli, 2002). This distinction is further explained in the following subsections.

2.2.3.1 Interference Suppression

In order to examine differences in IC between monolinguals and bilinguals, empirical studies have employed tasks which tap into interference suppression or response inhibition abilities (or both). To study interference suppression, researchers have used tasks such as the flanker task, the Simon task, and the Stroop task as they present conflicting stimuli which act as distractors. The flanker task was first introduced by Eriksen and Eriksen (1974) and later adapted by Fan, McCandliss, Sommer, Raz, and Posner (2002). It is ostensibly the purest measure of interference suppression, due to its non-complex nature (Costa et al., 2008). In the traditional version of the flanker task, participants are presented with a row of five arrows (→→→→→) and are instructed to

indicate the direction of the central arrow, which is pointing either left or right. The direction of the arrows surrounding the central arrow constitute a conflict if they are facing a direction opposite to that of the central arrow (e.g., →→←→→). Participants must inhibit the distractor stimuli to provide a correct response. Typically, if the central arrow is pointing in the same direction as the distractors (i.e., congruent condition), responses are faster and more accurate than when they are opposite (i.e., incongruent condition). A *conflict effect* is then measured by subtracting response latencies and accuracy rates of congruent trials from incongruent ones. The conflict effect is thus purported to measure interference suppression skills.

In a study by Yang et al. (2011), English-monolingual, Korean-monolingual, and English-Korean bilingual children performed a version of the flanker task (using fish-images instead of arrows). Results indicated that the bilingual children outperformed both groups of monolinguals both in speed and in accuracy. These results have been replicated in numerous studies (e.g., Yang & Yang, 2016; Yoshida, Tran, Benitez, & Kuwabara, 2011, Poarch & van Hell, 2012; Poarch & Bialystok, 2015). While the flanker task was able to detect enhanced interference suppression abilities by children, results from tasks with adults (especially young adults) have not consistently revealed a difference between monolinguals and bilinguals. This has been attributed to the fact that the participants are at their optimum stage of cognitive performance, which possibly masks differences between the two groups (Bialystok, 2017). Alternatively, it is possible that the task is not complex enough to detect differences between monolingual and bilingual adults, as children as young as 7 were observed to respond as quickly and as accurately as adults on this task (Yang & Yang, 2016). Therefore, the use of a more complex version of the flanker task is recommended, along with other complex tasks, such as the Simon task.

As with the flanker task, the Simon task entails congruent and incongruent trials, the latter ones consisting of a conflict via a stimulus appearing in different locations on a screen. In the standard version of the Simon task, a stimulus (a blue or a red square) is presented either on the left or the right side of a computer screen. Each stimulus is associated with a key press, and participants are instructed to respond as quickly and accurately as possible by pressing the left or right button, based on the colour of the stimulus in each trial. If the location of the stimulus on the screen matches the colour of the response key (i.e., congruent condition), responses are typically faster and more accurate than when the location of the stimulus is presented on the opposite side of screen (i.e., incongruent condition). The difference in response latencies and accuracy rates between congruent and incongruent trials is called the *Simon Effect*, and it is said to measure participants' ability to overcome the interference from the location dimension of the stimulus and to focus on the colour of the stimulus. Studies which have administered the Simon task have reported that bilingual children outperform their monolingual counterparts due to the purported superior ability to suppress interfering stimuli (Bialystok, Martin, & Viswanathan, 2005; Morales, Calvo, & Bialystok, 2013). Furthermore, higher second language proficiency was associated with better performance on this task (Tse & Altarriba, 2014). It should be noted that some studies did not report any difference between monolingual and bilingual children (e.g., Carlson & Meltzoff, 2008; Duñabeitia et al., 2014). This may be due to confounding variables and an inconsistency in controlling for potentially confounding variables such as socioeconomic status and language proficiency.

2.2.3.2 Response Inhibition

While interference suppression tasks have generally reported support for a bilingual advantage with respect to IC, response inhibition tasks have mainly failed to

detect differences between monolinguals and bilinguals (Bialystok, Craik, Klein, & Viswantathan, 2004; Bialystok, Craik, & Luke, 2006; Blumenfeld & Marian, 2014; Costa et al., 2008, 2009; Hernandez, Costa, Fuentes, Vivas & Sebastian-Galles, 2010). Response inhibition tasks, which require suppressing a prepotent response, include paradigms such as go/no-go tasks, where participants must refrain from pressing a key if a specific cue appears. Other examples include day/night tasks in which participants must name a picture with its antonym, thus inhibiting the automated response of naming the picture with the 'correct' word associated with it. Studies which have administered these tasks have reported no significant differences between monolingual children in speed or accuracy (e.g., Barac, Moreno, & Bialystok, 2016; Engel de Abreu, Cruz-Santos, Tourinho, Martin, & Bialystok, 2012). One study which aimed to examine bilingual advantages by distinguishing interference suppression and response inhibition is that of Martin-Rhee and Bialystok (2008). The researchers compared monolingual and bilingual children's performance on a Simon task, to measure interference suppression, and a variation of the Stroop task (a day/night task), to measure inhibition of prepotent responses. The results of the study indicated that bilingual children were faster than their monolingual counterparts on the Simon task, yet no differences were detected on the Stroop task, thus supporting the claim that bilingualism confers advantages in interference suppression, but not response inhibition.

The discrepancy in the response patterns between the flanker and Simon tasks and response inhibition tasks, as described above, indicates that these tasks tap into the different processes underlying inhibitory control. As such, the type of task used is a methodological variation across studies which may explain some of the inconsistency in findings. More importantly, the distinction between the two subcategories of inhibitory control constitutes a counter-argument to claims that the bilingual advantage may not be

a generalisable finding. Furthermore, this distinction supports recent calls for teasing apart the different aspects of executive functions in order to gain a better understanding of the source and location of advantages that bilingualism confers.

In addition to the distinction between interference suppression and response inhibition, another distinction has been proposed to further deconstruct interference suppression: global vs. local inhibition. According to De Groot and Christoffels (2006) these two types of IC processes are involved in suppressing interference. In language production, these two processes, depending on the linguistic context, mediate the suppression of the non-target language (Guo, Liu, Misra, & Kroll, 2011). Global IC is said to be a top-down process which involves suppressing language networks holistically, especially in monolingual contexts where only one language is the target language. Local inhibition, however, is more focused in scope, as it involves bottom-up processes which manage language selection within two coactivated languages by exerting control on a restricted set of representations (e.g., specific lexical items) (2011). Both global and local IC processes are sensitive to external cues which are present in bilingual contexts where both languages can be produced. Costa et al. (2009) suggest that local inhibition processes complement the top-down conflict-monitoring processes, depending on contextual demands. This leads to the next section, *Bilingualism and Conflict Monitoring*.

2.2.3 Bilingualism and Conflict Monitoring

Conflict monitoring refers to a cognitive mechanism responsible for detecting and resolving conflicts which emerge due to competing (or conflicting) stimuli, response tendencies, or co-activated task schemata (Botvinick et al., 2001). In cognitive tasks which involve a conflict between a stimulus and the prepotent response, a conflict arises and is detected by a conflict monitoring 'device'. This device alerts pertinent executive

functions to resolve this conflict, through response inhibition, for example. Conflict monitoring, therefore, allows an adaptation to changes in behavioural goals or task requirements by engaging the necessary executive function(s) (Botvinick, Cohen & Carter, 2004). As such, this component of Miyake et al.'s (2000) model parallels Peal and Lambert's (1962) mental flexibility and the notion of attentional control through the ability to selectively focus attention, suppress irrelevant elements, and shift focus (Posner, 2012). Botvinick et al. (2001) suggested that the ability to constantly detect conflict is at the core of cognitive control, and like interference suppression, that it is achieved by the anterior cingulate cortex. Costa et al. (2009) and Hilchey and Klein (2011) propose that conflict monitoring underlies and enhances performance on executive functions. Accordingly, monitoring of attention to detect conflict is supported by behavioural and neuroimaging evidence.

Monitoring abilities, however, have received less attention in the literature, and the findings from studies which investigated this construct are conflicting. Studies examining conflict monitoring (e.g., Soveri et al., 2011) generally employ task-switching measures, where participants are required to switch their behavioural responses when presented by a cue, such as different sets of instructions. One example is that of card-sorting tasks in which participants are asked to switch from "sorting by shape" to "sorting by colour". Performance is measured by examining differences between trials preceded by the same task (i.e., *congruent* trials) and trials preceded by a different task (or *incongruent* trials). As with interference suppression tasks, bilinguals generally outperform monolinguals in task switching, especially on incongruent trials where inhibition is necessary (Prior & MacWhinney, 2010). This observation has recently been attributed to conflict-monitoring as opposed to inhibitory control (Costa et al., 2009).

Adopting the task-switching paradigm, Prior and MacWhinney (2010) reported young bilingual adults outperforming monolinguals when tasks were in mixed blocks (i.e., mixed congruent and incongruent trials), attributing this to bilinguals' efficient attention shifting between conflicting dimensions to select the correct response. Enhanced conflict monitoring abilities were attributed to young bilingual adults outperforming their monolingual counterparts in the AX Continuous Performance Task. In this task, participants were instructed to identify the occurrence of the pattern 'AX' in a continuous stream of letters. Incongruent trials included 'AY' or 'BX' patterns. Bilinguals were able to outperform monolinguals on this task, which the authors attributed to better conflict monitoring abilities involving both proactive and reactive control. These findings were further echoed by Morales and colleagues (Morales, Gomez-Ariza, & Bajo, 2013; Morales, Yudes, Gomez-Ariza, & Teresa Bajo, 2015) who reported a bilingual advantage on this task among a similar population (young adults). Morales attributed this advantage to the higher capacity of bilinguals to monitor conflict and engage inhibition. These studies, therefore, highlight the importance of conflict monitoring capacities as a bilingual advantage.

However, an advantage in conflict monitoring by bilinguals has not been consistently reported. A large-scale study by Paap, Myuz, Anders, Bockelman, Mikulinsky, and Sawi (2017), which compared the performance of 236 monolingual and bilingual participants on three cued-switching tasks, revealed no significant difference between the two groups. The researchers administered a colour-shape, a letter-number, and an animacy-size task. The three measures of switching ability had convergent validity, yet there were non-significant correlations for measures of inhibitory control. The researchers tested the bilingual advantage hypothesis by (1) contrasting bilinguals and monolinguals and (2) by placing participants on a continuum based on their degree

of bilingualism measured through self-reported language use. In all tasks, there was no significant difference between the participants irrespective of their grouping or scale assignment. These findings are similar to those of Gathercole et al. (2014) who observed no significant differences in task switching between monolingual and bilingual children, adolescents, and adults. It is possible that these studies failed to replicate the findings of previous studies due to differences in the tasks or sample populations. Costa et al. (2009) suggested a way to overcome potentially divergent findings due to methodological variations. They proposed using the same task but manipulating its conditions in order to extricate an underlying conflict monitoring demand. Their study is described next.

Costa et al. (2009) examined studies of bilingualism and executive functions which used a variety of inhibitory control tasks to measure conflict effects by comparing response latencies and accuracy rates between congruent and incongruent trials. The review revealed that inhibitory control advantages of bilinguals are observed when the two types of trials are mixed, as opposed to being presented in blocked-designs. Accordingly, Costa et al. argued that bilingual advantages may stem from increased switching between congruent and incongruent trials within the same blocks, thus indexing conflict monitoring and not just inhibitory control. To test their hypothesis, the researchers administered the flanker task to monolingual Spanish and bilingual Spanish-Catalan young adults but manipulated the amount of conflict-monitoring involved by varying the percentage of congruent to incongruent trials in five blocks. In the low-monitoring condition, Costa et al. divided trials by type as 92% congruent vs. 8% incongruent in one block and 92% incongruent vs. 8% congruent in another block. In the high-monitoring condition, the trials were equally divided between congruent and incongruent (50%-50% split), whereas in the medium-monitoring condition, the ratio was 75% congruent vs. 25% incongruent in one block and 75% incongruent vs. 25%

congruent in another block. Manipulating the distribution of trial type was therefore expected to pose variable levels of conflict, resulting in higher demand of monitoring. Results from the study revealed that the bilingual group outperformed the monolingual group only in the high-monitoring condition, as there was equal likelihood of congruent and incongruent trials. Costa et al. argued that this indicated enhanced conflict monitoring and inhibitory control, suggesting an overlap between the two processes. Furthermore, the researchers suggest that monitoring processes are trained by bilingual practices, especially codeswitching. This has propelled further research investigating the potential relationship between code-switching and executive functions (e.g., Hernandez, Martin, Barcelo & Costa, 2013).

As most studies reporting a difference between monolingual and bilingual groups have recruited young children, Costa et al.'s (2009) study is valuable as it was able to detect a difference in performance on the flanker task (rather than other tasks such as the AX Continuous Performance Task or card sorting task) between monolingual and bilingual young adults (however, see Prior, Goldwasser, Ravet-Hirsh and Schwartz (2016) for non-significant results in flanker-task conflict monitoring). As mentioned in the previous section, studies administering the flanker task to young adults did not report a bilingual advantage, and this was attributed to the young adults' being at the peak of their cognitive acuity. Following Costa et al., the manipulation of trial ratios is adopted in the presented study, as a comparable aged population is recruited.

One study which attempted to replicate the manipulation of the Flanker conditions in order to explore an advantage in conflict monitoring is that of Hofweber, Marinis, and Treffers-Daller (2016). The researchers administered the flanker task using the low-, medium-, and high-monitoring conditions by mixing congruent and incongruent trials in a 92:8, 75:25, and 50:50 ratio, respectively. Unlike previous studies which compared

monolinguals to bilinguals, the study recruited two groups of German-English bilinguals who differed in the frequency of their codeswitching. The researchers hypothesised that an advantage in conflict monitoring could be observed in participants who frequently codeswitch as suggested by Costa et al. (2009) and the theoretical framework in their study (e.g., the CPM, Green & Li, 2014). The findings of Hofweber et al. (2016) suggested this manipulation was successful at revealing differences in conflict monitoring abilities, as codeswitching frequency was positively correlated with monitoring skills. The study further supports the applicability of the trial distribution manipulation to detect differences in conflict monitoring between bilingual groups. The present study's use of this manipulation to compare trilingual groups is thus further justified.

In summary, there is a discrepancy in findings with respect to conflict monitoring, and this perhaps may be due to the overlap it shares with inhibitory control. Tasks measuring conflict monitoring involve inhibition to switch between schemas depending on the change in cues or stimuli (Koch, Grade, Schuch, & Phillip, 2010). This does not preclude, however, that a bilingual advantage in conflict monitoring exists. Creating more complex tasks through design manipulation to measure conflict monitoring has therefore been suggested and is adopted in the present study. Furthermore, many of the aforementioned studies suggest a link between codeswitching and executive functions. This is further discussed in Section 2.6.

2.2.4 Bilingualism and Working Memory

Working Memory (WM) refers to the mechanisms underlying temporary memory storage, attentional control, and the management of information in the service of complex cognition (Tsai, Au, & Jaeggi, 2016: 69). A seminal model of WM is proposed by Baddeley and Hitch (1974) and later expanded upon by Baddeley (1986, 2003). In

this multicomponent model, three specialised subcomponents—the phonological loop, the visuo-spatial sketchpad, and the central executive—represent the various processes involved in processing and storing auditory and visual information. The central executive is a control mechanism which manages functions such as directing, maintaining, and switching attention, retrieving and activating items stored in long-term memory, and managing the information exchange between the phonological loop and visuo-spatial sketchpad (Baddeley, 1986; Baddeley & Logie, 1999). The central executive, therefore, resembles Miyake et al.'s (2000) 'executive functions', with the two constructs diverging only in terms of the means of modulating cognitive functions (i.e., attention vs. inhibition, respectively) (Bialystok, 2017:249).

Since its launch, Baddeley's model of WM has sparked numerous discussions about the exact mechanisms underlying working memory, along with its capacity (and limits), amenability to training, and neurological underpinnings (Miyake & Shah, 1999; Linck, Osthus, Koeth, & Bunting, 2014). More recent theories of WM have focused on the processes rather than the components underpinning the model. One of the most influential models is that of Cowan (1999, 2008, 2016) which proposes a framework that centres on attention control and activated long-term memory. In Cowan's model, memory in the 'focus of attention' is quickly accessed and available, but it is limited in terms of capacity. On the other hand, long-term memory is not limited in capacity but may deteriorate or suffer from interference, leading to loss of memory. Another influential model is that of Engle and Kane (2004) which proposes 'executive attention' as a construct encompassing both WM and attention; the two interact to produce complex cognition. This has also been referred to as 'WM capacity' by Engle in a recent paper (Shipstead, Harrison, & Engle, 2015). Attentional control processes are said to govern the contents of WM by activating, focusing, updating, switching, and inhibiting

memory during information processing. Accordingly, some researchers argue that these processes are synonymous with 'executive functions', again highlighting the overlap and ambiguity in defining the different constructs. This overlap not only lies within theoretical constructs but also across tasks measuring executive functions. For example, Flanker and Simon tasks which are used in studies examining inhibitory control are also said to measure working memory (or executive attention). Despite the disparate theoretical perspectives on WM, there is general consensus regarding the function it has: ordering, storing, and managing information relevant to performing a current task before that information is properly incorporated into the cognitive processes. The amount of data which can be stored as well as the speed and accuracy of retrieval varies among individuals. The present study adopts Tsai et al.'s (2016) definition of 'working memory' (which is used synonymously with 'executive attention' and 'WM capacity'[2]), whereas the term 'attention' will denote the generic meaning of 'focus'.

Several measures of WM have been proposed, such as 'simple' span tasks (e.g., forward digit span, word span, or nonword span) and 'complex' span tasks (e.g., backward digit span). In simple span tasks, participants are instructed to recall a series of unrelated letters, words, digits, or objects after a brief period of presentation, whereas complex span tasks involve a level of processing (e.g., performing a simple mathematical equation) while maintaining other information presented to them, such as a string of letters, words, or objects.

Evidence for the role of WM in complex cognition stems from studies which reported high WM as a reliable predictor of performance on high-level cognitive tasks,

[2] The term 'capacity' in the construct of 'WM capacity' does not refer to storage space; rather, it refers to the degree to which resources are available to control attention, ensuring that information relevant for a current task is maintained (Engle, 2002; Engle & Kane, 2004).

including ones that involve ignoring distracting information (Piolat, Olive, & Kellogg, 2005). WM has also been central in studies examining language comprehension, where it was observed that bilinguals with higher WM capacity are better at vocabulary learning (in both languages), reading, writing, and listening comprehension (Atkins & Baddeley, 1998; Daneman & Hannon, 2007; Engle, 2002). Accordingly, individual differences in WM have been central to numerous studies in the field of applied linguistics. WM is said to underlie L2 processing (e.g., Michael & Gollan, 2005) and learning (e.g., Linck & Weiss, 2011). Although there is general agreement that WM is related to L2 outcomes, the degree to which WM and its subcomponents influence L2 outcomes is debated. Methodological issues and task type have been said to be possible explanations for inconsistencies in results (Linck et al., 2014). Complex span tasks are said to be stronger predictors of L2 development and proficiency than simple span tasks, suggesting a larger role of the inhibitory control component of WM compared to recall (short-term memory).

Evidence in support of a relation between bilingualism and WM has been observed throughout language development. In the first year of life, infants who were exposed to more than one language exhibited different patterns of attention (namely, visual attention) than monolingually raised children (Ayneto & Sebastián-Gallés, 2017). Specifically, bilingual infants would direct and focus their gaze at their caregivers' mouth, whereas monolingual infants direct their gaze at the eyes when hearing linguistic input. Older bilingual children (ages 8-11) have also been observed to have enhanced attentional control compared to monolinguals. In a study by Sorge, Toplak, and Bialystok (2017), bilingual children aged 8-11 years were observed to have enhanced performance on three executive functions tasks compared to their monolingual counterparts. In the study, participants performed a stop-signal task, a flanker task, and a

frogs-matrices task, with the latter task testing spatial working memory. Parents and teachers provided supplementary data regarding the attentional abilities of the participants. Results indicated that the bilingual children outperformed the monolingual group on all three tasks, and regression analyses revealed that higher attention ability and bilingualism were the strongest predictors of task performance.

Although studies examining WM in adults are scarce, Bialystok et al. (2012) suggest that bilingual advantages in young adults are observed when tasks or conditions are more complex and cognitively demanding. A study by Hernández, Costa, and Humphreys (2012) investigated WM, specifically visual attention, in highly proficient Spanish monolinguals and Catalan–Spanish bilingual young adults (aged 18-30 years). Results indicated that bilinguals were faster than monolinguals at directing and shifting attention and at suppressing irrelevant information in working memory. These results further suggest that bilingualism enhances general top-down shifting of attention and facilitates higher-order processes involved in complex cognition (Cockcroft, Wigdorowitz, & Liversage, 2017). Another study which examined WM in young-adult participants is that of Kerrigan, Thomas, Bright, and Filippi (2017). The researchers recruited 30 English-monolingual and 30 mixed-L1, English-L2 bilinguals to perform a series of tasks tapping into the different facets of WM. The results of the study indicated a bilingual advantage in detecting visual changes (i.e., visuo-spatial WM), with no significant group difference in verbal WM task performance.

Costa, Santesteban, and Ivanova (2006) suggest that the bilingual experience accelerates the development and maintenance of attention. As such, the bilingual experience involves constant management of language systems and is said to lead to adaptation of cognitive (and neural) systems (Bialystok, 2017). Accordingly, individual

differences in the bilingual experience are also believed to confer differential advantages on cognitive functions, and this variability is taken into account in the present study.

2.2.5 Bilingual vs. Trilingual Advantage

While there is evidence, even if specific to certain abilities, supporting the bilingual advantage, limited research has investigated whether multilingualism confers similar cognitive control advantages (Madrazo & Bernardo, 2012, 2018). It is possible that there are further advantages due to the constant need to inhibit two languages at a time when speaking in a third language, and it is conceivable, though dubious, that there are disadvantages due to the need to recruit additional cognitive resources to manage interference between more than two languages (Cockcroft et al., 2017).

Evidence in support of exploring a multilingual advantage is found in several studies which have compared IC in monolingual, bilingual, and trilingual participants. Poarch and van Hell (2012) investigated non-linguistic IC in four groups (1) German monolingual children (n=20), (2) early German-L1, English-L2 children (i.e, L2 learners) (n=19), (3) German-English bilingual children (n=18), and (4) German-English-Language X trilingual children (n=18) (ages 5;0-8;0) who had been exposed to at least two of their three languages in an immersion context (at home and/or school). Controlling for language proficiencies and socioeconomic status, the researchers administered the Simon Task and the Attentional Networks Task to assess potential cognitive differences between the groups. The results indicated that all participants responded faster in congruent than in incongruent conditions for both tasks. Further, with respect to the Simon Task, bilinguals and trilinguals exhibited less interference in the incongruent condition than their monolingual counterparts. Poarch and van Hell attributed this expected finding to the notion that language control is continuously activated by bi/trilingual speakers, leading to better general attentional control

mechanisms. However, there were no statistical differences between L2 learners and monolinguals and between L2 learners and the other bi- and trilingual participants. The researchers argued that the L2 learners' proficiency may not have reached a certain threshold similar to that of bi/trilinguals. Similarly, with respect to the Attentional Networks Task, bilinguals and trilinguals exhibited better conflict resolution control over L2 learners. The researchers concluded that bilinguals' executive control efficiency supersedes that of monolinguals, as with earlier research, especially with respect to conflict monitoring. Furthermore, it was speculated that the L3 would yield a statistical difference in this study between trilinguals and bilinguals. However, there was no difference between the highly proficient bilinguals and trilinguals, indicating that there may not be a qualitative difference between the two groups (similar to Schwieter & Sunderman, 2011). As with the argument that the L2 learners' proficiency did not reach a level that allows them to exhibit the same IC capacities as the other two groups, it can be speculated that the trilingual participants' L3 proficiency may not have reached a threshold beyond which a bilingual vs. trilingual advantage could be observed. However, a trilingual advantage is observed in the sense that the trilingual participants outperformed the monolingual group, although this may not be different from the advantage bilinguals possess.

More evidence in support of a trilingual advantage comes from Schroeder and Marian (2017) who compared IC in monolingual, bilingual, and trilingual age-matched children and young adults using the flanker task. Results from this study paralleled those of Poarch and van Hell (2012), as bilinguals and trilinguals in both groups outperformed their monolingual counterparts, yet there were no differences in performance between the bilingual and trilingual groups. These findings were further replicated in other studies such as Poarch and Bialystok (2015) and Vega-Mendoza, West, Sorace, and Bak (2015).

Other studies, however, have found no differences in IC and switching between the three groups (monolingual, bilingual, and trilingual) in young adults (e.g., Humphrey & Valian, 2012; Paap, Johnson, & Sawi, 2015). Perhaps this discrepancy in findings across studies can be ascribed to age variation among participants and subsequently the length of codeswitching experience. For instance, Poarch and van Hell (2012) recruited child participants who may not have had enough experience processing three languages, thus leading to non-significant differences between the bilingual and trilingual groups. This notion is supported by findings from studies with older adults (e.g., Kavé, Eyal, Shorek & Cohen-Mansfield, 2008) where trilinguals outperformed bilinguals, and where multilinguals outperformed bilinguals and trilinguals. Accordingly, it is conceivable that processing three languages may reveal cognitive advantages over bilingual ones in adulthood. Other possible reasons for the discrepancy include variables such as self-reports of proficiency and the different tasks being used which may influence participants' performance.

A recent study which investigated the multilingual advantage is that of Cockcroft et al. (2017). The researchers compared 39 English monolingual and 39 highly-proficient multilingual (mixed L1s and L3s, but English as L2) young adults on four components of working memory: verbal storage, visuospatial storage, verbal processing, and visuospatial processing. The multilingual group came from a low socioeconomic background unlike their monolingual peers, and they were not matched on English vocabulary size compared with the monolingual group. This was intentionally done in order to analyse results when socioeconomic status and verbal ability were statistically controlled and not controlled for. Results indicated that in both types of analyses, multilinguals outperformed their monolingual counterparts across all four components of WM, especially in visuospatial processing. This study therefore contributes to the

literature on the multilingual advantage by examining an under-researched facet of executive functions beyond the commonly investigated IC. This is in line with Hilchey and Klein's (2011) proposal that the bilingual advantage extends to working memory and supports Miyake and Friedman's (2012) 'unity and diversity' proposal which stipulates that cognitive processes are overlapping but separate abilities which share an underlying mechanism that extends to WM.

However, there is a main caveat in Cockcroft et al.'s (2017) study: the monolingual participants were indeed English native speakers, yet they learned an L2 at school. The authors state that the reason for considering them as monolingual is that they self-reported their L2 proficiency as weak and their L1 as 'good' to 'excellent'; they argue this classification had no influence on the results. However, a more accurate classification of this group would be unbalanced bilinguals (i.e., bilinguals whose proficiency in the L1 is higher than that in the L2) or L2 learners as opposed to 'monolingual'. Furthermore, the use of self-reports is subjective, and it is curious to know how participants would rate their L1 as 'good' as opposed to 'excellent' when they are considered 'native speakers' by the researchers. Furthermore, the multilingual group was varied in terms of the number of languages the participants spoke (three to five), yet there may be differences within the group based on the number of languages and proficiency in the three or five languages. Although at least three languages were acquired by all participants before the age of 6 and used on a weekly basis in the very least, proficiency levels in the languages were not controlled for or assessed by the researchers. Furthermore, the variability in the L1s and L3s of the multilingual groups, even if all were African languages, may have influenced the findings. Therefore, one could argue that Cockcroft et al. found a multilingual advantage over a bilingual one, despite within-group variability in the multilingual group. This caveat is avoided in the

present study by ensuring that participants speak an equal number of languages (and the same languages, in fact) and that proficiency levels are measured objectively.

The studies in this section offer partial support for a 'trilingual' advantage, yet very marginal (if no) support which distinguishes bilinguals and multilinguals. However, it is important to note that methodological and task-related variability may be attributed to this lack of evidence. The first caveat in drawing inferences from these studies is the way in which 'bilingualism' and 'trilingualism' are defined and measured. The aforementioned studies did not use objective measures of proficiency in the three languages, so the difference between balanced vs. unbalanced bilinguals and/or trilinguals is overlooked. However, as discussed in previous sections, language proficiency is closely tied with both linguistic and non-linguistic task performance. Perhaps the non-significant difference in performance between the groups can therefore be attributed to the comparisons between highly proficient bilinguals and unbalanced trilingual participants with low L3 proficiency. Furthermore, the studies do not distinguish between simultaneous and sequential trilinguals; however, as evidenced in the literature, factors such as age of acquisition and codeswitching experience constitute main influences to IC (Luk, De Sa, & Bialystok, 2011; Verreyt et al., 2016).

One study which overcomes this caveat is that of Madrazo and Bernardo (2012) who administered oral and written proficiency tests to both bilingual and trilingual participants and administered a go/no-go task and the sustained attention response task. Despite objectively measuring proficiency, the researchers reported no difference between bilinguals and multilinguals. However, this may be due to the type of tasks that the researchers used, which involved only response inhibition, as opposed to more complex IC tasks. In a more recent study, Madzaro and Bernardo (2018) used the Simon arrow task, a more complex IC task, in order to examine the potential trilingual

advantage between Filipino-English bilinguals and Chabacano-Filipino-English trilinguals. The researchers again used objective language proficiency tests to accurately distinguish the two groups. Results indicated that trilinguals were more accurate and had smaller Simon Effect scores than their bilingual counterparts, indicating that tasks involving both interference suppression and response inhibition can detect a trilingual advantage.

The present study similarly utilises objective proficiency measures and complex IC tasks in order to detect a possible advantage of highly-proficient (balanced) trilinguals over unbalanced trilinguals whose L3 is weaker. Given the above studies which suggest that unbalanced trilinguals perform similarly to balanced bilinguals, the present study thus predicted that the distinction made between balanced and unbalanced trilinguals would mimic the contrasts found between balanced bilinguals and balanced trilinguals in Madzaro and Bernardo's (2018) study.

2.2.6 Summary of Bilingualism and Executive Functions

Bilinguals have been observed to outperform monolinguals on a range of non-linguistic control tasks (Bialystok et al., 2012), presumably due to bilinguals' experience in managing and controlling simultaneously activated languages (Kroll et al., 2014; Morales et al., 2013). Studies which have consistently replicated the bilingual advantage have used inhibition as an underlying explanatory source. Other hypotheses suggest that specific aspects of the bilingual experience influence cognitive abilities. Specifically, bilingualism is argued to require constant conflict monitoring between the activated languages, especially in codeswitching, thus leading to enhanced domain-general non-linguistic abilities (Astheimer, Berkes, & Bialystok, 2015; Costa et al., 2009; Green, 1998; Green & Abutalebi, 2013; Verreyt et al., 2016).

While enhanced inhibitory control in bilinguals has been replicated in numerous studies, other studies did not report a bilingual advantage (e.g., Paap & Greenberg, 2013; Paap et al., 2015). This discrepancy in findings has been attributed to several factors, such as the age of L2 acquisition and L2 proficiency level. For instance, Luk et al. (2011) observed that sequential and simultaneous bilinguals differ in their performance on inhibitory control tasks, with the latter group outperforming the former. Others have suggested that how 'bilingualism' is defined and the subsequent groupings of bilinguals may influence the interpretation of the outcomes (Bialystok, 2017). Another possible factor contributing to the discrepancy in findings is not controlling for participants' codeswitching habits. Codeswitching involves specific processing experiences to manage cross-linguistic conflict (Verreyt et al., 2016), and the frequency in which participants codeswitch may influence their cognitive performance. An alternative explanation to the disparity in findings is that of methodological differences between the studies (Bak, 2016; Kroll & Bialystok, 2013). Experimental designs which incorporate IC tasks involving greater cognitive complexity have revealed a bilingual advantage, whereas studies employing less complex tasks have not consistently observed that bilinguals outperform their monolingual peers.

There is also a stronger discrepancy in the findings of studies using executive function tasks with young adults (as opposed to children) (Bialystok, 2017). Bialystok attributes the lack of replicability with young adults to task design (simplicity) as well as to the statistical analyses conducted. For instance, she argues that young adults generally have faster response latencies on both congruent and incongruent conditions (approximately 500ms) whereas the range of response latencies is much larger for children and older adults (2017). She therefore suggests the use of more complex (variations of) tasks and more sophisticated statistical analyses to uncover differences in

young adult participants. This is supported by evidence from Bialystok (2006) and Costa et al. (2009) where bilingual young adults outperformed their age-matched monolingual counterparts only on the most taxing conditions in the tasks involving IC and conflict monitoring, with comparable performance on the simple conditions in the same study.

A bilingual advantage has also been observed in other areas such as WM (e.g., Cockcroft et al., 2017). While recent calls to disentangle or compartmentalise the precise aspects of executive functions which underlie this advantage, overlapping processes make it difficult to do so (Bialystok et al., 2005; Bialystok, Craik, & Luk, 2008). However, it may be possible to explore relationships between specific aspects of the bilingual experience with precise aspects of cognitive abilities by taking into account individual differences as well as by administering different tasks which are said to tap into the different components of executive functions.

Furthermore, very few recent studies have attempted to uncover a trilingual or multilingual advantage, as opposed to a bilingual one, but have also reported inconsistent findings (e.g., Hofweber et al., 2016, Schroder & Marian, 2017). However, as with studies examining the bilingual advantage, it is important to note methodological differences as possible factors influencing the findings.

Given the above review of the literature, and as highlighted throughout the section, the present study attempts to overcome some of the main caveats of previous studies. This is achieved through using complex tasks (or manipulations of tasks) to explore group differences on cognitive task performance, using objective measures of proficiency which is argued to influence cognitive task performance, and employing advanced statistical approaches such as mixed-effects models in data analysis to account for inter-participant variability (Linck, 2016; Titone et al., 2017). This approach is similarly adopted in the task design and analysis of participants' linguistic task

performance. This is further discussed in the next section, *Language Control and Executive Functions*.

2.3 LANGUAGE CONTROL AND EXECUTIVE FUNCTIONS

Before discussing the theoretical models of language control and the overlap between language and domain-general mechanisms, it is important to clearly define and operationalise key terms. The next section offers a brief discussion of the definition and measurement of codeswitching.

2.3.1 Defining Codeswitching

2.3.1.1 Mixing, Switching, and Borrowing

Codeswitching and *codemixing* are sometimes used interchangeably, yet some studies make a distinction between the two terms. For instance, Ritchie and Bhatia (2013) claim that the two constructs differ in terms of their positions as intra-utterance or inter-utterance, where 'codemixing' appears within a clause or utterance and 'codeswitching' is inter-utterance. Meisel (1989) and Muysken (2000) designate *codeswitching* to refer to monolingually-retained utterances which are motivated by socio-pragmatic choices such as topic and *codemixing* to denote fusion of grammatical, lexical, and phonological systems. However, this is further complicated by the definition of *borrowing* or 'loan words' which is said to occur when individual lexical items from one language are assimilated morphologically or phonologically into the matrix of another language (Gardner-Chloros, 2009). As such, borrowing is synonymous to code-mixing as defined by Muysken (2000), and 'code-switching' occurs when words are not integrated into or modified by the other language. Despite the proposed distinctions between the terms, distinguishing loans from switches may be elusive, especially for

non-distant languages (Jacobson, 2011). The presence of cognates also poses a difficulty in establishing whether a switch occurred or not.

As there is no consensus on the proper distinction between these terms and unequivocally assigning language membership to all items, the present study adopts the distinction between *switches* and *loans* in assigning utterances to the three languages being examined. *Switches* are used to denote speakers' monolingually-retained utterances irrespective of the items' position or underlying motivation, whereas *loans* designate phonologically and/or morphosyntactically modified items and are assigned as belonging to the borrowing language. An example of this distinction, taken from Zantout (2014), is offered in Table 2.1:

Table 2.1 Switches vs. Loans Distinction

Example	Language Membership	Categorisation
لأ أنا كتير زهئت كيف بألولنا انو theme سخيف. La'a, ana kteer zhi'it keef bi'oloolna anno theme sakheef ('No, I am sick of how they said the **theme** is silly'). (Arabic/English)	*theme* → English	Switch (retained in its monolingual form)
وين هموركك Wein homeworkak? ('Where is your **homework'**) (Arabic/English)	*Homeworkak* → Arabic	Loan (modified by having a possessive pronoun attached and phonologically adapted by being pronounced /hom'wɔrkak/)

2.3.1.2 Codeswitching Typology

Unlike researchers whose codeswitching typologies are contingent on the universality of grammatical constraints (e.g., Myers-Scotton, 2006), Muysken (2000) offers a typology based on a wide range of empirical data from a variety of sociolinguistic contexts. He argues that two important aspects of codeswitching are generally overlooked: switch *directionality* (switching from Language A-to-B or B-to-

A), which allows researchers to distinguish the matrix (dominant) and embedded (borrowed) languages; and sociolinguistic context, which accounts for divergent codeswitching patterns. In doing so, Musyken (2000) offers a typology, summarised in Table 2.2. Although some researchers (e.g. Auer, 1984; Gafaranga, 2005) marginalise directionality, emphasising the act of codeswitching itself, this study utilises switch directionality in discussing codeswitching patterns in single-word and monologic utterances.

Table 2.2 Muysken's Typology of Codeswitching (Codemixing)

Type	Definition	Example	Sociolinguistic Context
Alternations	Stretches of words of one language alternate within a conversational turn (inter-utterance)	'maar 't hoeft niet l'i 'anna ida Šeft ana…' ('but it need not be, for when I see, I…') (Dutch/Moroccan Arabic)	Generally in stable bilingual communities which separate languages
Insertions	Words or constituents settings from one language are inserted into the utterance of another (intra-utterance)	'Chay-ta las dos de la noche-ta Chaya-ma-yku' (That-AC the two of the night-AC arrive-CIS-1pl': 'There at two in the morning we arrive') (Bolivian Quechua/Spanish)	In ex-colonial and recent migrant communities where speakers' proficiency in the two languages is unbalanced
Congruent Lexicalisations	A shared language structure is realised with words/morphemes from different languages (intra-item)	'jo moatte him ris hearer yn dat loopke yn dy mars' ('You should hear him in the melody in the march' (Frisian utterance with Dutch lexical item 'loop' and Frisian diminutive 'ke' rather than the Dutch 'je') (Frisian/Dutch)	With second-generation migrant groups and bilingual speakers of languages with quasi-equal prestige and no overt language separation

While sociolinguistic in its distinction, Muysken's (2000) typology also offers a psycholinguistic perspective on the different types of codeswitches. Muysken argues that varying degrees of language activation and integration of the linguistic systems are reflected in this typology and contends that in alternations where languages are structurally separated, activation shifts from one language to another. In insertions, however, activation in one language is said to be temporarily diminished, allowing the

language system to become pervious on a lexical level. In congruent lexicalisation, both languages are co-activated, allowing the two languages to partially share their processing systems on both the syntactic and lexical levels (Deucher, Muysken, & Wang, 2007).

Muysken's (2000) typology is partly adopted, as it is incorporated in the theoretical framework underpinning this study. However, it is adapted to match the distinction between switches and loans: alternations and insertions are combined and referred to as switches, whereas congruent lexicalisations are perceived as loans. Furthermore, criticisms of the typology are taken into account. For instance, Green and Li (2014) critique the sociolinguistic embedding of the three structural patterns and their psycholinguistic underpinnings. They argue that speakers may intentionally insert words from a lower-proficiency language because they are more precise and more efficient than a circumlocution in the current, more proficient language. The present study does not address the sociolinguistic aspect of the typology and focuses solely on the psycholinguistic underpinnings as they relate to the theories adopted.

2.3.1.3 Bilingual vs. Trilingual Codeswitching

Although codeswitching is a linguistic phenomenon common to all bi/multilingual speakers, there is a difference in the number of languages and language combinations used by bilinguals and multilinguals. Studies on trilingual codeswitching have attempted to determine whether it is idiosyncratically distinct from bilingual codeswitching. A main difference is the difficulty in identifying speakers' matrix language(s) and categorically distinguishing switches and loans. Bilinguals are able to produce three types of utterances: monolingual utterances in each language and one switching combination of the two languages. Multilingual speakers, on the other hand, and irrespective of switch directionality, can produce seven types of utterances. For example, Lebanese trilinguals can produce an utterance entirely in Arabic, English, or French, as well as four

combinations: Arabic-English, Arabic-French, English-French, and Arabic-English-French. However, trilingual codeswitching research suggests that codeswitches mostly involve two of the three languages (Edwards, 1994; Rothman & Niño-Murcia, 2008). Clyne (2003) posits that trilinguals tend to behave like bilinguals alternating between two languages, and in other cases as 'double bilinguals' where two two-language combinations are used. He ascribes this to the putative high activation of only two of three language systems in conversations. Despite their paucity, however, three-language combinations have been observed in the literature (Edwards & Dewaele, 2007; Zantout, 2014).

In a longitudinal case-study which examined trilingual codeswitching in natural conversations, Stavans (1992) recorded the interactions of two Hebrew-Spanish-English trilingual siblings from ages 2;6 to 4;2 and 5;5 to 7;1 with caretakers and friends. Analyses of the frequency of voluntary, spontaneous switches included a distinction between alternations and insertions. Results revealed that the children were able to distinguish the different interactional contexts based on their interlocutors. For instance, they only used Spanish with their Spanish-monolingual caregiver. However, they exhibited very few three-language combinations even when addressing their trilingual parents. The majority of their switches involved two-language combinations when speaking with their parents and their bilingual grandparents. These findings were later replicated in Clyne's (1997) longitudinal study which compared bilingual and trilingual codeswitching in Dutch-German-English, Hungarian-German-English, and Italian-Spanish-English trilingual families living in Australia. Results indicated that the three different groups of trilinguals used two-language switches more frequently than three-language switches, indicating low levels of 'active trilingualism'.

These studies are further supported by two additional studies which look at trilingual switches by Lebanese children. Although a different pattern was expected to emerge given the pervasiveness of both trilingual language use and three-language switches in everyday Lebanese interactions, this was not observed in either study. Edwards and Dewaele (2007) investigated the codeswitching behaviour of an Arabic-English-French trilingual mother and her trilingual 8-year-old, both of whom were residing in England since the daughter was six months old. Although the researchers stated that both interlocutors were highly proficient trilinguals, there was no mention of how proficiency was measured. In their analysis of audio-recorded parent-child interactions, the researchers noted that there were minimal three-language utterances, with the majority of switches occurring between English and Arabic. This observation lends support to Clyne's (2003) argument that trilinguals tend to behave as bilinguals.

The second study examining Lebanese trilingual codeswitching is that of Zantout (2014). Examining codeswitching patterns of two groups of children (ages 11-12 and 17-18) in their interactions with their parents and peers, the researcher analysed the frequency and directionality of switches. Results indicated an overall minimal use of French and highly infrequent trilingual switches by both groups. Although Lebanon is a context in which trilingual-language combinations are ubiquitous in social settings, Zantout noted that the participants were English-Arabic-French trilinguals whose proficiency in the three languages was not balanced. Specifically, the participants were in an English-immersion programme, with both groups being English-dominant as observed by their code-switching frequency, although no objective measure of proficiency was used. Qualitative data from interviews with the participants and their parents also shed light on the opinions of and attitudes towards French by this sample. This data revealed that French was considered a dispensable language, as the role assigned to French was

deemed an insignificant one. Indeed, participants and their parents indicated that French was not used frequently, if at all, at home or in school, and that there was no compelling need for its use. The researcher notes that there is a need to examine French-immersion speakers, as different patterns of switches are expected to emerge. This is a caveat which is also present in Edwards and Dewaele's (2007) study, since the two participants in the study spoke French as a third language, and they were residing in an English-speaking country.

Accordingly, the present study examines Lebanese trilingual adolescents from both English and French immersion programmes to overcome this shortcoming. Furthermore, this study examines the code-switching behaviour of a larger sample, as opposed to the generally adopted case-study approach in studies of spontaneous speech by trilingual speakers. This is an attempt to overcome the limited generalisability of the findings of previous to the small sample and data sizes.

2.3.1.4 Voluntary vs. Cued Codeswitching

Another distinction made by researchers examining codeswitching is that of cued vs. voluntary switching. Voluntary switches generally denote those which occur in naturalistic speech data in which a bi/multilingual is unprompted to switch to another language. These switches are typically observed in sociolinguistic studies examining dyadic conversations by interlocutors who share two (or more) languages. The switches are not only frequent but also smoothly-occurring, as speakers seamlessly alternate languages when they see fit. If an interlocutor is monolingual, however, a communicative breakdown is eluded by the bilingual speaker who would switch to the language shared by the interlocutor. Therefore, bi/multilinguals need to monitor cues, such as the language(s) spoken by their interlocutors, in order to speak in and/or switch between suitable languages when needed. Psycholinguistic studies tend to focus only on

cued switches by asking participants to speak in one language ('single-language blocks') or to alternate languages when a cue is presented either visually or auditorily ('mixed-language' condition). Accordingly, participants do not freely switch between languages, as they are required to either adhere to one language of the researcher's choice or to switch languages when cued. Code-switches in psycholinguistic tasks are therefore more prone to 'errors' and 'time lags', with participants exhibiting a longer time and/or more errors to switch when cued between languages than when speaking in a single language. Furthermore, cued language switching is said to be governed only by top-down control, as switches are prompted by external instructions, whereas voluntary codeswitching requires both top-down control and bottom-up processes (de Bruin, Samuel, & Duñabeitia, 2018:29; Gollan & Ferreira, 2009). This is explained by speakers being motivated by external cues (top-down control) as well as by lexical access or lexical priming (bottom-up control), such as when bilinguals select the items that are fastest to retrieve. A more detailed discussion of these two types of switches is offered in the following sections, in light of the language control models.

2.3.2 Measuring Codeswitching

Numerous ways of measuring codeswitching have been proposed, yet each one has its limitations. For instance, some researchers propose quantifying the number of switches from one language to another by counting the number of morphemes and then designating those to the proper language thus determining the base, or matrix language, and the borrowed, or embedded, language. This attempt to objectively and quantitatively measure the amount of language, however, may be criticised on the basis that some morphemes could belong to the different language, as in the case of 'in' and 'is' in Dutch and English (Zentella, 1999). As such, it would be difficult to determine language membership for these morphemes, therefore leading to inconsistent and inconclusive

reports on codeswitching. Switch directionality is another means of measuring codeswitching, where researchers compare the number of times a speaker switches from Language A to B and from B to A. While switching languages is very common in the production of bilingual and multilingual speakers, switch directionality may not necessarily be indicative of a speaker's codeswitching behaviour as longer utterances could be produced in one language and only one-word utterances are produced in the other. While these two measures have been quite useful in measuring naturalistic codeswitching patterns, they have been generally applied by sociolinguists. Psycholinguists, on the other hand, measure codeswitching in terms of 'switch costs'.

2.3.2.1. Switch Costs

Instead of attempting to quantify the amount of language (or number of morphemes) in an utterance, psycholinguistic research examines codeswitching within the language switching paradigm, where languages need to be constantly selected and deselected. Researchers examine the time it takes to switch from one language to another. They also examine the accuracy of a speaker in producing the target language. As almost all psycholinguistic experiments involve one-word utterances, switch costs are an efficient measure to examine codeswitching behaviour. The language switching paradigm includes two types of trials: repetition, where the target language is the same as the one in the preceding trial, and switch, where the target language is different from the one in the trial before. Responses in switch trials are typically slower and less accurate than in repetition trials. The difference between switch and repetition trials is known as 'switch costs' (Meuter & Allport, 1999; Kiesel, Steinhauer, Wendt, Falkenstein, Jost, Philipp, & Koch, 2010).

Switching between languages has been argued to be a linguistic application of task-switching. That is, switch costs entail cognitive processes which underlie updating task-

relevant representations, thus signalling both the difficulty of switching between tasks and the cognitive processes involving the selection of appropriate task responses (Festman & Schwieter, 2015). Accordingly, Rogers and Monsell (1995) suggest interpreting switch costs as an index of momentary task reconfiguration between trials. Alternatively, switch costs may be perceived as representing negative priming (or persisting activation) from the preceding task (Allport, Styles, & Hsieh, 1994).

Typically, balanced bilinguals whose proficiency in Language A and Language B is comparable exhibit symmetrical switch costs, where there is no difference in time or accuracy due to the directionality of the switch. On the other hand, unbalanced bilinguals, who are more proficient in one language, exhibit asymmetrical switch costs, where it would take longer for the participants to switch into and out of their stronger (or more dominant) language. Therefore, when examining switch costs, researchers must be aware of the language proficiency of their participants and how these may influence switch costs. This is further discussed in Section 2.4.3 within the Inhibitory Control Model.

2.3.2.2 Fluency (Pausing)

Fluency has been defined in both broad and narrow terms (Lennon, 1990). The broad definition stipulates that 'fluency' may be equated with general proficiency, whereas the narrow definition is related to the temporal aspects of speech production that 'influence the degree of fluidity' in speech, such as speech rate, pauses, repetitions, and other hesitation phenomena (Derwing, Munro, Thomson, & Rossiter, 2009:534). Segalowitz (2010) proposes a distinction between three types of fluency: (1) cognitive fluency (i.e., the efficiency of one's processing abilities); (2) utterance fluency (i.e., the temporal measures of speech production—analogous to Lennon's narrow definition); and (3) perceived fluency (i.e., listeners' judgment of a speaker's proficiency based on

specific criteria). Skehan (2003, and in Tavakoli and Skehan (2005)) proposed further segmenting 'utterance fluency' into three subcomponents: (a) *speed fluenc*y, or speech rate (measured in different units such as the number of syllables per utterance); (b) *breakdown fluency*, which examines non-lexicalised pauses (measured in the duration, location, and frequency of silent and filled pauses); and (c) *repair fluency*, which measures hesitation phenomena excluding pauses, such as repetitions, corrections, false starts, etc.

In order to measure codeswitching behaviour in naturalistic speech which exceeds one-word utterances, the present study adopts the notion of 'fluency' recommended by Green and Li (2014). While the term 'oral fluency' is often used to refer to 'oral *proficiency'*, the present study makes a clear distinction between the two constructs both in definition and in measurement. In terms of *fluency*, the present study adopts Skehan's (2003) *breakdown fluency* to define and measure *fluency* by analysing non-lexicalised pauses only and operationalising the overall construct of *fluency* as pausing. Non-lexicalised pauses include silent pauses where no speech is uttered and filled pauses which include 'meaningless hesitations;' such as *um*, *er*, *uh*, etc. (Segalowitz, 2010). This enables the quantification of fluency in a manner that is analogous to the response latencies of switch costs measured in psycholinguistic (one-word production) tasks.

In contrast, the present study utilises (oral) language *proficiency* as a more general construct which involves the ability to produce longer stretches of speech about a variety of topics and is characterised by lexical diversity and complex grammatical structures. Language proficiency is measured objectively through oral proficiency tests and is rated according to criteria[3] outlined in the methodology chapter.

[3] This partly overlaps with 'perceived fluency' as defined by Segalowitz's (2010) definition; however, to avoid ambiguity, the term 'fluency' strictly refers to pausing in this study.

It should be noted that correlation studies have reported moderate to strong relations between oral proficiency and utterance fluency (De Jong, 2018; Révész, Ekiert, & Torgersen, 2016). This has been attributed to the overlap in the measures of the two constructs, where the rubrics of oral proficiency typically contain aspects of fluency such as pausing (e.g., Kahng, 2014). Although the overlap is inevitable, the narrow scope of 'fluency' adopted in the present study compared to the more general one of 'proficiency' aids in distinguishing the two. Furthermore, the objective of using 'fluency' in the present study is to measure participants' production in a narration task, whereas proficiency is used as an independent variable which modulates performance on all productive tasks.

2.3.3 Language Control Models

Despite appearing like a simple task, producing the intended language(s) is regulated by a complex network. Researchers have previously argued that a language-specific mental process controlled which language(s) a multilingual speaker activated or deactivated to speak in the intended language (Macnamara, 1967; Grainger, 1993). Numerous models of lexical access have been put forth to explain the cognitive mechanisms governing speech production (e.g., Abutalebi & Green, 2007; Green, 1998; Schwieter & Sunderman, 2008; van Heuven, Schriefers, Dijkstra, & Hagoort, 2008). Some of these models purport that lexical selection is language-specific (Costa & Caramazza, 1999); i.e., only the activation levels of words in the intended language are considered for lexical selection. However, this assumption precludes interference from other activated words on non-intended language(s) during lexical access (Costa, 2005). An alternative perspective is that activated lexical items compete for selection to varying degrees, irrespective of their language membership (Kroll et al., 2014). In this view, language selection is thus achieved through a dual process of activating the words of the

intended language while simultaneously suppressing its equivalents in the non-intended languages. This suppression, or inhibition, is said to vary based on the level of activation of each lexical item. Furthermore, it is also believed that languages are not deactivated; rather, they are coactivated even when only one language is in use. This parallel activation has been supported by evidence from numerous studies in speech production (e.g., Colomé, 2001; Costa & Caramazza, 1999; De Groot & Starreveld, 2015; Hoshino & Kroll, 2008; Poarch & van Hell, 2012). These studies thus support the notion that there are *inhibitory control* mechanisms which govern language production and switching.

2.3.3.1 The Inhibitory Control Model

Today, most researchers subscribe to the notion that the ability to select the 'correct' language involves inhibiting the activation of the unintended language(s) through a *bilingual language control* mechanism (Runnqvist, Strijkers, & Costa, 2014). One model that has been put forth to explain how bilinguals are able to regulate control of their language systems is Green's (1998) Inhibitory Control Model (ICM) (Figure 2.1). This model postulates that bilingual speakers' *lexico-semantic system* comprises words with 'language tags', which assign the language to which each word belongs, and 'lemma representations', which include information about syntactic properties. For a bilingual to produce a word in one language without interference from the other, the *conceptualizer*, which lies outside the lexico-semantic system, devises different *language task schemas*, or action sequences, to control output. Task schemas control the mapping of a communicative action, such as speaking in Language A, to a speech act 'by coordinating a set of control processes, including those which inhibit activated but non-target language lemmas and respond to any change in the language being used' (Green 2017:726). Accordingly, task schemas regulate the level of activation of lexical representations while reactively inhibiting outputs from the unintended language. This

involves two cognitive processes: interference suppression and response inhibition. As discussed in Section 2.2.3.1, interference suppression involves ignoring any distracting language cues or input, whereas response inhibition refers to the control of a generated behavioural response. The selection of a word to be produced requires that the most highly activated target language be transmitted by the *supervisory attentional system* to the language task schema. For a bilingual to produce an utterance, the conceptualizer may use multiple words from a single language (i.e. synonyms) or from different languages (i.e. translation equivalents).

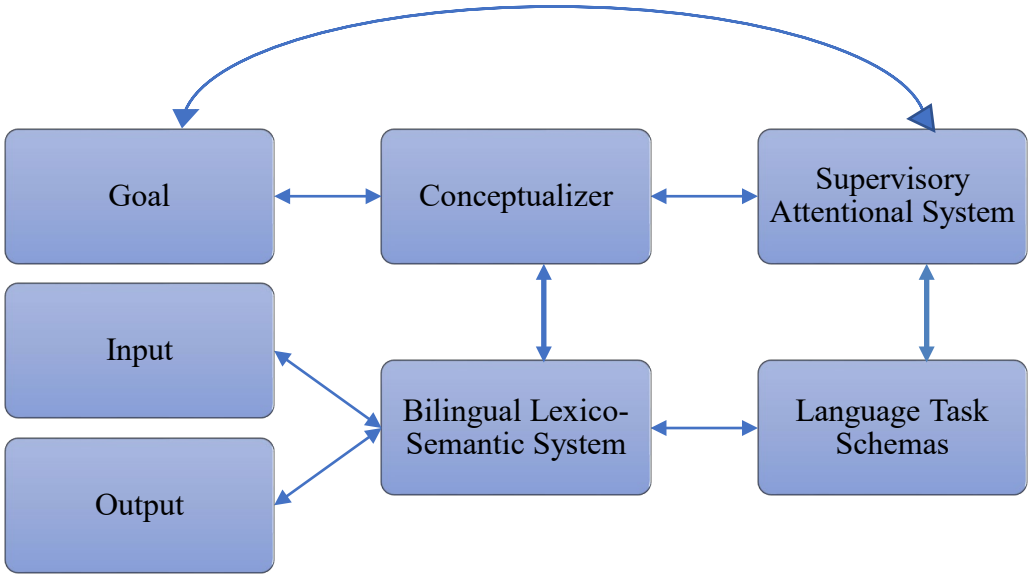


Figure 2.1 The Inhibitory Control Model (adapted from Green, 1998)

The ICM posits that the extent to which representations in the non-target language(s) are inhibited is determined by the degree of activation of those words. Accordingly, more inhibition is needed to suppress words from a more dominant language; conversely, less inhibition is required for a less dominant language. For unbalanced bilinguals, the L1 is more dominant and thus more active than the L2 at a baseline level. The L1 will therefore require more inhibition than the L2, making it more difficult to reactivate. This asymmetry in inhibition results in an asymmetry in switch

costs, or the time it takes to switch to a different language than the language used in the previous turn. As mentioned in Section 2.3.1, switch costs are typically measured by response time (in milliseconds) and accuracy (calculated by error rate) in producing the intended language(s). The ICM predicts that for unbalanced bilinguals, switching to the L1 (more dominant language) will be more difficult due to stronger prior inhibition than switching to the L2. For balanced bilinguals, a smaller (or no) asymmetry is expected.

As noted earlier, bilingual language control is usually examined through the language-switching paradigm, in which the response language of two subsequent trials can be the same (repeat/non-switch trial) or different (switch trial). A seminal study adopting this paradigm and supporting the ICM is Meuter and Allport's (1999) investigation of the effects of cross-linguistic influence and inhibition on language production. The researchers recruited 16 unbalanced English-L2 adult bilinguals with mixed L1s to complete a bilingual numeral-naming task, where participants named Arabic numerals in their L1 or L2, prompted by two background colours to indicate the target language. The participants revealed greater switch costs to the L1 than to the L2. This asymmetrical switch cost is consistent with the ICM, as the model predicts increased inhibition of the L1 in L2 production. In order to switch to the L1 after naming in the L2, participants must first change stimulus-response rules and then overcome the strong inhibition previously applied to L1 words on preceding trials. This alternation between languages results in an added cost manifested in the increased naming latencies in L2-to-L1 switching. However, in their post-hoc analyses, Meuter and Allport observed that proficiency influenced switch costs. After dividing the participants into two groups based on proficiency (measured by speed of L1 and L2 naming), they noted that one group (Group 1) was more balanced, whereas the other (Group 2) was more L1-dominant. Group 2 revealed asymmetrical switch costs, unlike Group 1, suggesting that

proficiency modulates asymmetries and that inhibition varies according to proficiency in each language. Meuter and Allport's (1999) study supports the ICM's claim that the inhibition applied to a language is proportional to the strength of the lexical activation of said language.

Other studies examining IC in language-switching tasks have yielded comparable findings to those reported by Meuter and Allport and offer further—though sometimes partial—support to the ICM. However, these studies have also highlighted other considerations that could influence switch costs, such as the importance of objectively measuring language proficiency.

Costa and Santesteban's (2004) investigation of the influence of proficiency on switch costs revealed a multifarious pattern, partly supporting the ICM. In this study, five experiments, each with twelve participants, were conducted to investigate the codeswitching patterns of three groups: (1) highly-proficient Spanish-Catalan bilingual learners of English, (2) unbalanced Spanish-Catalan bilinguals, and (3) unbalanced Korean-Spanish bilinguals. The researchers administered a picture-naming task, prompting participants to name drawings they saw on a screen in switch (L1-to-L2; L2-to-L1) and non-switch trial conditions. Switch costs were asymmetrical with low-proficiency participants, in line with Meuter and Allport's (1999) findings, yet symmetrical with high-proficiency trilinguals, even when switching between the L1 and a weaker L3. While symmetry in switch costs is predicted by the ICM for the L1 and L2, as comparable amounts of inhibition are expected for both languages, L1-L3 switching should have revealed asymmetrical patterns. Therefore, Costa and Santesteban claim that their study supports the ICM for low-proficiency bilinguals, but only partly for high-proficiency multilinguals.

However, Costa and Santesteban's (2004) unexpected findings for high-proficiency multilinguals may be attributed to the conflation of bilinguals and trilinguals in the study, especially those from different populations. Age of acquisition is another potential moderator variable which the researchers acknowledged but did not investigate. Perhaps the highly proficient trilinguals who had learned their L1 and L2 simultaneously behaved differently from their sequential-bilingual counterparts, who showed more dominance in their L1. Further, like other studies that examine proficiency as it relates to language switching, the measures used to assess proficiency are self-reported assessments (a questionnaire), a subjective measurement that likely does not accurately measure proficiency, making the unexpected finding questionable.

Other studies have aimed to extend notions of inhibitory control from bilingual to trilingual contexts. Marian, Blumenfeld, Mizrahi, Kania, and Cordes (2013), for instance, examined the effects of trilingualism, language proficiency, and language interaction on inhibitory control. The researchers recruited 26 participants (mean age=33) whose linguistic backgrounds varied considerably, and some of whom were multilingual rather than trilingual. Self-reports of proficiency were used, whereas inhibitory control was assessed by a trilingual colour-word Stroop task (measuring reaction times and error rates). Marian et al. adapted the traditional Stroop task to not only include three languages instead of two (though the stimulus was always in English), but also to include lower-frequency colours, the names of which they equated with high proficiency. Five major findings based on the researchers' five hypotheses regarding trilingualism and IC were reported: (1) trilinguals performed better in congruent trials for each of their three languages; (2) higher proficiency led to better performance; (3) within-language Stroop effect was larger than between-language effect; (4) the magnitude of between-language effect interacted with proficiency (larger interference effect when the stimulus was in L1

and when production was in L2/L3); and (5) better performance for 'focal' colour-words than non-focal ones. The researchers attributed the findings to IC and *Supervisory Attentional System* (which mediates interference when several language task schemas compete). They further claim that the magnitude of the Stroop effect is influenced by the extent of interference from the other languages. While the study lends support to the ICM, the equivalence of non-focal with low-frequency words may be disputed. Furthermore, as the stimulus was only in English, it is a confounding variable as some participants had English as an L1, whereas others had English as an L3.

Other studies (e.g., Gollan & Ferriera, 2009) have also revealed variability in switch cost patterns where low-proficiency bilinguals exhibited symmetrical switch costs, contrary to predictions of the ICM. However, this variability may also be explained by methodological differences such as predictability (i.e., when participants anticipate a switch/non-switch trial) and response preparation time.

One study which reported conflicting results is that of Philipp, Gade, and Koch (2007) who adopted the language-switching paradigm to examine IC by unbalanced trilingual young adults (mean age=22.6). The researchers asked German-L1, English-L2, and French-L3 participants to name numerals on a screen by switching between L1 and L2, L2 and L3, and L1 and L3. Participants consistently responded faster and more accurately on repetition trials than on switch trials across the three pairs of languages, as expected. Asymmetrical switch costs were also observed in each language pair, indicating that the dominant language in every pair set required increased inhibition. As such, Philipp et al. report findings which are commensurate with those of Meuter and Allport (1999) and Costa and Santesteban (2004). However, there were no significant differences in switch costs across the three pairs of languages, which was unexpected given the participants' different proficiency levels in each language (although these were

measured by 'years of formal language exposure' and self-reports). The results therefore are in contrast with the previous findings (and the ICM) which observed that asymmetries are smaller when the proficiency levels were comparable (especially in L2 and L3).

Philipp et al. (2007) suggest that this unexpected symmetry in L1-L3 and L2-L3 sets may be attributed to methodological issues (e.g. cue-stimulus interval manipulation) which were insufficient to support the dominance inhibition effect. The researchers also explored IC by measuring n-2 repetition costs, which compare n-2 language repetition sequences (e.g., L1-L2-L1) and n-2 language non-repetition sequences (e.g., L3-L2-L1). The researchers noted that words at the end of sequences were named faster in n-2 language non-repetition sequences, and that the repetition cost was significantly greater when n was the L1. However, they also noted that L3 repetition costs were larger than L2 ones, which was unexpected and attributed to persistent inhibition of the L3. In a later study, Philipp and Koch (2009) also provided evidence for a switch task where numeral naming was faster in n-2 non-repetitions than in repetition ones, which they argued to be in support of persistent inhibition of L3. While the manipulation of sequences offers support for persisting inhibition of task-schemas in language-switching tasks, it is not always observed across studies—a criticism suggested by Finkbeiner, Gollan, and Caramazza (2006) (see also, Branzi, Calabria, Boscarino, & Costa, 2016). However, the work of Philipp and colleagues is central to studies examining the relation between trilingual production and IC, and the manipulation of the language sets allows for contrasts between language pairs. This manipulation of language pairs is adopted in the present study.

Another study which offered conflicting findings in terms of switch costs is that of Verhoef, Roelofs and Chwilla (2009). The researchers conducted an event-related

potentials (ERP) study using the codeswitching paradigm to investigate IC and switch costs among 19-year-old unbalanced Dutch-English university students in a picture-naming task. Preparation time was manipulated through short (750ms) versus long (1500ms) cue-stimulus intervals. Reaction times were asymmetrical for short intervals and symmetrical for long intervals, irrespective of language and participants' proficiencies. Verhoef et al. concluded that while bilinguals may resort to inhibition in experimental tasks, they do not need inhibition to control their language use based on the time they are given to plan their speech in non-experimental settings.

Mosca and Clahsen (2016)'s investigation also highlighted the influence of preparation time. In this study, 30 German-English highly-proficient (sequential) bilinguals performed a picture-naming task using Snodgrass and Vanderwart's (1980) pictures to be named in English and/or German. The researchers manipulated the amount of time between the presentation of the stimulus and the response stimulus. Results revealed symmetrical switch costs when participants had no time to prepare for the task, which is attributed to their high proficiency in both languages. This is consistent with the predictions of the ICM and in line with previous findings (e.g., Costa et al.0, 2006). However, when participants were given a long preparation time (800 ms) before they were cued to respond, there were no switch costs. This indicates that preparation time is a variable which can influence the outcomes of language switching performance in tasks, and for highly-proficient bilinguals, preparation time may dissipate switch costs altogether. The conclusions of Verhoef et al. (2009) and Mosca and Clahsen (2016) are reiterated by Runnqvist et al. (2014) who argue that despite extensive research on codeswitching and IC, whether and to what extent bilinguals suppress their non-target languages to control their speech output in normal conversations remains a controversial, unexamined issue.

Therefore, when drawing conclusions about different groups from various experiments, it is important to examine whether methodological differences in task administration may have influenced the results. Furthermore, it is important to note that highly proficient speakers may exhibit a different (or no) pattern of switch costs.

While the aforementioned studies offer evidence in support of the ICM in terms of switch costs, some researchers have examined whether different accounts of language control may better explain language switching performance. Arguing that cross-linguistic activation does not necessarily involve competition, Finkbeiner et al. (2006), replicated Meuter and Allport's (1999) study to assess whether switch costs may be interpreted by alternative explanations and whether speaker intent to use the L1/L2 modulates differential activation. Recruiting 16 English-L1, mixed-L2 university bilinguals, the researchers conducted three experiments. Experiment 1 consisted of a numeral naming task where participants were asked to switch between their L1 and L2 based on different-colour backgrounds. The researchers included an L1-only picture-naming task. Similar to Meuter and Allport's findings, there was switch-cost asymmetry in the numeral-naming task, and switching to the L1 (reaction time) was greater than when switching to the L2. However, there were no switch costs for the picture-naming task, leading the researchers to believe that switch costs in the numeral-naming task could be attributed to a factor other than L1 inhibition. Experiment 2 (16 participants, English-L1, mixed-L2) aimed to test the *lexical suppression hypothesis*, a modified version of the inhibition control hypothesis, which claims that it is only the competing lexical nodes in the non-target language that are suppressed. Replacing the digits from Experiment 1 with dot patterns (dice), the researchers administered the task to the participants who revealed no switch costs on dot-naming patterns. The researchers then recruited 32 English-L1, mixed-L2 participants for Experiment 3 which involved only the use of L1 to examine whether ease

of processing underlies switch costs. Providing the participants with 'fast' (high-frequency, short-length, multiple-meaning) words and 'slow' words (opposite) in four different-coloured fonts, Finkbeiner et al. asked the participants to name the colour of the word when presented on a black background or to read out the word when presented on a grey background. Results indicated that switch costs were greater when switching to 'fast' words and to word naming (vs. colour naming). There was also a significant correlation in reaction time between the non-switch condition and the switch cost for an item, suggesting that 'ease of response' was an influential factor in determining switch cost.

While Finkbeiner et al. (2006) set out to offer an alternative account to IC, their research does not seem to be successful in doing so. Experiments 1 and 2 seem to be confounded by mixing task-schema switching and language-switching (Abutalebi & Green, 2007). Furthermore, Finkbeiner et al. dismiss IC in explaining Experiment 3 results and suggest response blocking based on differential activation of possible candidates, as opposed to IC's inhibition. They state that speaker intent may lead to higher activation of lexical items in the target language, which the ICM explains as the 'goal', and 'response blocking' seems synonymous with 'inhibition'.

One interesting observation in most of the aforementioned studies is that on both switch and non-switch trials, symmetrical switch costs usually involve slower naming latencies in the dominant language than in the L2 (reversed language dominance). Larger switch costs for a stronger compared to a weaker language, i.e. asymmetrical switching costs, have been taken as evidence in favour of the ICM (e.g., Costa & Santesteban, 2004; Kroll, Bobb, Misra, & Guo, 2008; Linck, Schwieter, & Sunderman, 2012; Schwieter, 2013; Wang, Xue, Chen, Xue & Dong, 2007). However, many studies with highly proficient bilinguals who have a stronger L1 and a weaker L2 have shown that,

despite the proficiency difference, switch costs for the two languages might be symmetrical. Based on these results, it has been suggested that when the difference in language proficiency is relatively small, the strength of suppression applied on the two languages is comparable, leading to a similar amount of reactivation costs (e.g., Meuter & Allport, 1999; Fink & Goldrick, 2015). For example, an unbalanced trilingual whose proficiency in the L1 is higher than in the L2 or L3 might require more inhibition to suppress the L1 when speaking in the L3 because words in the L1 are more highly activated than in the L3. On the other hand, speaking in the L1 would not require much inhibition of the L3 as the L3 baseline level of activation is assumed to be lower than that of the L1. Accordingly, when cued to produce a word in the weaker L3, the unbalanced trilingual would take less time to produce the word than when prompted to produce it in the L1.

Abutalebi and Green (2008) argue that this finding lends support to the global inhibition of the L1 throughout the switching experiments. This finding, however, also brings into question whether variability to switch costs as they relate to IC assumes different types of executive functions vis-à-vis bilinguals' proficiencies. Some researchers (e.g., Gollan & Ferreira, 2009) proposed that inhibition may be perceived as global and constant in high-proficiency bilinguals and as localised and transitory in low-proficiency bilinguals. Given the ease with which proficient bilinguals switch languages in their natural conversations, however, it is unlikely that the whole lexicon is suppressed as that would make speech a difficult endeavour for bilinguals (Runnqvist et al., 2014). Another hypothesis is offered by Kroll and Gollan (2014) who claim that the delay in L1 naming in mixed conditions is the 'genuine index of L1 inhibition'. In blocked naming conditions, the L1 is said to be faster than L2, whereas the opposite is true in mixed conditions, as observed in Costa and Santesteban (2004). Although the cued language in

blocked paradigms is forced, there is an opportunity for bilinguals to speak each language over an extended period of time, simulating the type of switching in which bilinguals engage in some contexts.

Along this line of reasoning, Misra, Guo, Bobb, and Kroll (2012) used blocked-naming contexts to examine L1 and L2 speech production. Measuring both reaction times (RTs) and the ERP waveforms in a picture-naming task with late but proficient Chinese-English bilinguals, participants who first named pictures in L1 (2 blocks) and then the same pictures in the L2 (2 blocks) differed from those who named pictures in the L2 followed by L1. Misra et al. reported shorter RTs and more positive ERPs for repeated items in the first group. For the second group, there was little evidence of repetition priming in RTs and more negative ERPs. These data were considered as reflecting sustained inhibition of L1 in contexts of blocked L2 speech. Even when L1 followed L2, there appeared to be a long-lasting global inhibition mechanism. This finding parallels that of van Assche, Duyck, and Brysbaert (2013) who tested order effects using a verbal fluency task in which Dutch-English participants were asked to list as many items as possible in each language for different semantic categories (e.g. animals). Accordingly, counterbalancing testing order conditions is crucial in assessing IC in block-naming tasks. However, it is important to note that blocked paradigms may not always be the typical interactional contexts in which bilinguals converse; individual (or group) differences in terms of habitual language switching should be considered when examining mixing and/or switch costs.

Along with proficiency and blocked- vs. mixed-naming trials, lexical robustness has also been observed as a variable affecting IC in trilinguals. Schwieter and Sunderman (2011:392) investigated the extent to which L2 and L3 *lexical robustness*, 'a particular aspect of non-native language proficiency that taps into the size and strength of the

lexicon' moderates reliance on IC by trilinguals. The researchers hypothesised that increasing lexical robustness may lead to a shift from IC to language-selective mechanisms where switch cost would be nil, yet they argued that this shift is not unidirectional and that highly-proficient trilinguals may revert to IC. Recruiting 56 English-L1, French-L2, and Spanish-L3 university students (mean age=20.6) from Ontario, they examined the participants' self-reports of proficiency (verbal fluency) in the three languages and their lexical robustness via a verbal fluency task. A trilingual picture-naming task was then administered in which the participants named pictures in the L1, L2, or L3, as determined by the colour of the background screen. L2 robustness led to faster picture naming in all three languages, whereas L3 robustness only led to faster L3 picture naming. This finding led Schwieter and Sunderman to claim that the overall benefit of L3 learning is limited with respect to the whole language system. These results support reliance on IC, which the researchers attribute to their participants' lower L2 and L3 proficiency. While the study offers promising methodological advances in testing IC among trilinguals, some of the pictures in the picture-naming task were cognates in the three languages, which may have influenced switch costs (a shortcoming overcome in the present study). Furthermore, the claim that the L3 influence is limited may be true for unbalanced trilinguals, yet it is necessary to investigate this with balanced trilinguals who may reveal differential switch patterns across the three languages. It would also be important to investigate whether and how lexical robustness correlates with IC.

As no study had reported a clear connection between IC and online speech production, Linck et al. (2012) aimed to provide evidence for a relationship between domain-general IC and trilingual speech production in a codeswitching task. The researchers hypothesised that when switching to a less dominant L2/L3, switch costs

should be greater when switching from the dominant L1 than between the L2 and L3. After administering the Simon task and a trilingual picture-naming task to 56 English-L1, French-L2, Spanish-L3 participants, the researchers used mixed-effects models to examine performance on the tasks across and between the three languages. Increased IC correlated with reduced switch costs with respect to the L1, especially when switching to the L1 or from L1 to L3, as expected. However, IC did not influence switching performance from L1 to L2, leading Linck et al. to claim that there are constraints on the condition under which general executive functions are linked to language switching. The researchers called for testing not only IC through a Simon Task but other aspects of executive functions, such as working memory, which may play a role in language processing, as suggested by Green and Li (2014) and Blumenfeld and Marian (2014). Linck et al.'s study provides an extension of the ICM to a trilingual context by proposing that switching from a dominant L1 to a less dominant L2 or L3 incurs greater switch costs than when switching between the L2 and L3. Accordingly, the amount of inhibition required on a current trial depends, in part, on the language of naming on the preceding trial. This, however, may only be the case for unbalanced trilinguals; whether it holds true for balanced trilinguals is yet to be explored.

Further support of the ICM and the overlap between cognitive and linguistic switch costs comes from Declerck, Grainger, Koch, and Philipp's (2017) study. The researchers conducted three experiments and manipulated the cues and modality of responses (verbal for both linguistic and non-linguistic tasks). Their results indicated that there was a moderate- to strong-correlation between switch costs on linguistic and non-linguistic tasks (analogous to the Simon Effect and Conflict Effect) across three experiments which used different manipulations. These findings are consistent with those of Linck et al. (2012). Another study which lends support to the findings reported by Linck et al. (2012)

is that of Liu, Rossi, Zhou, and Chen (2014) who administered a picture-naming task and Simon task to measure switch costs and inhibitory control in 47 unbalanced Chinese-English bilinguals. The results of the study revealed symmetrical switch costs for participants who had high inhibitory control (i.e., a smaller Simon Effect score) but asymmetrical switch costs for participants with low inhibitory control (i.e., larger Simon Effect scores). These findings therefore support the ICM, as lower inhibitory control is correlated with larger L1 switch costs. More importantly, the study highlights this finding as a possible explanation for the inconsistency in previous studies' findings regarding balanced and unbalanced bilinguals' performance on linguistic tasks. That is, it may not only be proficiency which influences switch costs; rather, it might be individual differences in cognitive performance that may be better predictors of switch cost (a)symmetries.

The overlap between domain-general cognitive functions, codeswitching, and language task performance has further been investigated by Jylkkä, Lehtonen, Lindholm, Kuusakoski, and Laine (2018). The study recruited 51 Finnish-English sequential bilinguals who performed the Simon task and a Flanker task (with a 50-50 congruent-to-incongruent trial distribution) to measure inhibitory control and a number-letter task to measure monitoring (general set shifting) processes. Results from the cognitive tasks were also examined as predictors of performance on a picture-naming task, to assess whether cognitive processes underlie language switching and mixing costs, as suggested by Meuter and Allport (1999). The results indicated that switch costs were asymmetrical, as L2-to-L1 switches had higher response latencies than L1-to-L2 switches, thus giving support to the ICM. However, performance on the Simon task and the flanker task did not mediate language switching costs as predicted by the ICM, thus leading the researchers to conclude that language switching (and mixing) and asymmetries in switch

costs are not due to inhibition.

This is contrary to previous findings such as those of Linck et al. (2012) who observed that a smaller Simon Effect score predicted smaller switch costs. Furthermore, the general set shifting hypothesis received partial support, where conflict monitoring, measured by the number-letter task, was said to be a key component in mixed-conditions where trials are not blocked by language. Furthermore, the researchers proposed that switching into the L2 may engage general set shifting capacity, indicating an overlap between conflict monitoring and switch costs, similar to the findings reported by Liu et al. (2014). However, the distribution of trials in the flanker task is that of a high-monitoring condition. Therefore, if conflict monitoring is indeed a predictor of switch costs, it is curious that only the number-letter task was able to detect this, but not the flanker task.

In short, the positive relationship between performance on executive functions tasks and language switching tasks in *trilingual* contexts has only been explored by very few studies (e.g., Linck et al. 2012; Mosca & Clahsen, 2016), where better performance on the Simon Task, a measure of inhibitory control, was related to better language switching capabilities when switching to and from the dominant L1. Similarly, Festman, Rodríguez-Fornells and Münte (2010) reported that better language control abilities measured by accuracy on a picture-naming task were related to enhanced executive functions task performance. Neurocognitive studies have provided further evidence for a relationship between language control and non-linguistic domain-general control (Abutalebi & Green, 2007, 2008; De Baene, Duyck, Brass, & Carreiras, 2015, Guo et al., 2011).

However, it is important to note that the inconsistency in findings of behavioural studies which examine the relationship between executive functions and linguistic

control differs in terms of the two measures of switch costs. There is evidence for a link between executive functions and language control from studies which examine accuracy rates (or more specifically, cross-linguistic errors), whereas response latencies have yielded mixed results. For accuracy, bilingual participants who had more errors on language-switching tasks were more likely to perform more poorly on a variety of executive functions tasks, including the flanker task (e.g., Festman & Münte, 2012; Soveri et al., 2011). However, in terms of response latencies, only some studies lent support to the relation between reaction times in picture- or numeral-naming tasks and cognitive tasks. Studies which examined IC through tasks such as the Simon task to measure executive functions have generally revealed correlations with language switch costs (Linck et al., 2012; Prior & Gollan, 2011). However, the consistency of this relationship depended on switch directionality and proficiency levels. These two variables are taken into account in the present study. Other studies, however, have found no correlations between switch costs in cognitive and linguistic tasks (e.g., Branzi et al., 2016; Calabria, Branzi, Marne, Hernández & Costa, 2015; Calabria, Hernández, Branzi, Marne & Costa, 2012). It is important to note, therefore, that the overlap between domain-general executive functions and language control depends on the methods used to compare the two types of control. Furthermore, when examining switch costs, exploring both accuracy and response latencies when comparing the two types of performance is necessary. Additionally, while the ICM is supported, even in part, it is important to assess how IC influences production in naturalistic discourse outside the confines of experimental settings. The Adaptive Control Hypothesis (ACH), put forth by Green and Abutalebi (2013), and the Control Process Model of Code-switching (CPM), proposed by Green and Li (2014) and Green (2018), attempt to engage with naturalistic settings and incorporate variables such as contextual and individual differences.

Discussions of the ACH and CPM are offered next.

2.3.3.2 The Adaptive Control Hypothesis

Several studies, typically adopting behavioural or neurolinguistic approaches, have explored the mechanisms which underlie cognitive processes in production, yet they have not provided a thorough account of what happens in online speech production in everyday interactions. Accordingly, Green and colleagues (Green & Abutalebi, 2013; Green & Li, 2014) have proposed the ACH (and Control Process Model) as an extension of the ICM to provide a production model for naturalistic codeswitching. This model accounts for external and internal cues which may influence codeswitching patterns. Adopting Muysken's (2000) codeswitching typology, Green and his colleagues claim that different interactional (single-language, dual-language, and dense codeswitching) contexts require task schemas to operate variably in working memory, placing different demands on eight language control processes. A description of the contexts and the processes they involve is summarised in Table 2.3.

Table 2.3 ACH: Influence of Different Contexts on Task Schemas and Control Processes

Context	Description	Task Schemas Coordination	Control Processes in Highest Demand	Explanation
Single-Language	One language used in one environment and the other in a second distinct environment.	Competition (to ensure the production of one language)	(1) goal maintenance, (2) conflict monitoring, and (3) interference suppression	This ensures that only one language is produced without interference from the other language, as the interlocutor shares only one language with the speaker.
Dual-Language	Both languages are used within one environment, but with different speakers.	Competition (to ensure the production of one language/interlocutor)	(1) goal maintenance, (2) conflict monitoring, (3) interference suppression, (4) salient cue detection, (5) selective response inhibition, (6) task disengagement, and (7) task engagement	Language A is used with one interlocutor, requiring control processes (1-3), yet switching within the same context is permissible given certain conditions, such as the introduction of a new interlocutor who speaks Language B. Accordingly, a bilingual in this context would need to be attentive to such a cue (salient cue detection process) and respond accordingly by stopping speech in Language A (selective response inhibition), thus disengaging from the original task schema and engaging with a new task that allows production in Language B. There is high demand on working memory to ensure adaptation where new goals and communicative outcomes need to be established.
Dense Codeswitching	Both languages are used within the same environment.	Cooperation (to allow utterances from both languages)	(8) opportunistic planning	The speaker and interlocutor share both languages, so the speaker must communicate given any lexical item or preset phrases that are frequently used and available to him/her, as long as it meets syntactic constraints. They may also use non-verbal gestures in order to transmit a message.

Green and Abutalebi's (2013:518) ACH postulates that in the single-language context, bilingual speakers use their two languages in distinctly separate environments (e.g., Language A is used strictly at work, whereas Language B, typically a native language, is spoken with family members at home). This context therefore places demands on global IC mechanisms (e.g., goal maintenance and conflict monitoring), to ensure the separation and appropriate selection of the languages in the two different environments. In the dual language context, both languages may be used within the same setting, but with different interlocutors; thus, codeswitching may occur within a conversation yet not within the same utterance. An example is that of a bilingual speaker using Language A at work with one colleague and Language B with another colleague. This context therefore places demands on conflict monitoring, as the bilingual speaker needs to constantly detect cues to select the appropriate language. As such, this context is argued to involve a high level of language control. The dense codeswitching context, in contrast, is believed to require the least amount of IC. In this context, bilinguals are able to converse in any language or language combination (including intra-utterance codeswitching) with their bilingual interlocutors. As such, bilingual speakers can take advantage of their linguistic repertoires without the need to inhibit either language or to place additional cognitive demands such as cue detection and goal maintenance. Green and Abutalebi flesh out the cognitive processes involved in bilingual language use which vary according to contextual demands. These processes, along with brief descriptions, are listed in Table 2.4.

Table 2.4 ACH Cognitive Processes Description

Cognitive Process	Description
1. Goal maintenance	maintaining communicative goals, such as speaking only in the appropriate (or cued) language
2. Interference control: **(a) conflict monitoring** **(b) interference suppression**	monitoring for potential conflicts with these goals, such as lexical competitors or ambient languages/noise
3. Salient cue detection	detecting cues, such as the arrival of a new interlocutor who speaks a different language, which signal the need to alter language choice
4. Selective response inhibition	inhibiting a prepotent ongoing behaviour or response, such as speaking in Language A, to allow speaking in Language B
5. Task disengagement	disengaging from the current task schema to allow for another schema to be engaged, such as when disengaging from the Language A schemas to speak in Language B
6. Task engagement	engaging a new task schema to meet a new communicative goal, such as engaging Language B task schemas to produce utterances in Language B
7. Opportunistic planning	planning utterances opportunistically to maximise communicative efficiency, specifically through adapting words from Language A to fit into the syntactic structure of Language B (i.e., intra-utterance codeswitching., which is possible only in a dense codeswitching context)

Green and Abutalebi (2013) argue that speakers who fail to produce 'appropriate' languages with respect to these contexts incur an 'interactional cost', potentially leading to a communication breakdown. Task schemas are then altered to ensure production in the shared language(s). This involves controlling the representations in the working memory to ensure the communicative outcome and to set the control process parameters. Proficiency is another variable which is said to affect success at avoiding 'interactional costs', as higher proficiency generally correlates with better interference control.

However, Green and Abutalebi emphasise the role of speakers' ecology (i.e., the speech community/ies with which bi/multilinguals typically converse), individual differences, and sensitivity to social cues and interactional cost. Accordingly, the ACH takes into account not only speakers' proficiencies and dominance, but also the sociolinguistic contexts and control mechanisms underlying codeswitching. As such, it can be argued that the ACH places codeswitching in a socio-psycholinguistic framework.

Furthermore, the ACH makes predictions given existing data on codeswitching in different paradigms. For instance, for cued codeswitching, speakers from dense-codeswitching contexts are expected to be less fluent than those in single- and dual-language contexts when prompted to use a single language because they are accustomed to using all languages available to them (relying on opportunistic planning). Speakers from dual-language contexts are believed to switch faster than those from single-language contexts because they are trained to be attentive to salient cues. Conversely, in uncued codeswitching conditions, bilingual speakers from dense-codeswitching contexts are predicted to show lower switch costs because of the cooperation between language schemas. Speakers from other contexts, however, will need to overcome the competition (their default language-schema disposition), thus exhibiting higher switch costs than dense-codeswitching speakers. As for inhibition, Green and Abutalebi (2013) and Green and Li (2014) predict that dual-language context speakers will be the most proficient in domain-general cognitive tasks, such as the Simon Task or tasks tapping into working memory (e.g., the Backward Digit Span task), because the skills needed to control codeswitching generalise to non-linguistic tasks.

Given the novelty of the ACH, there are a few studies which have tested its predictions. The present study aims to fill this gap in the literature by examining the applicability of the ACH in single-language and dense-codeswitching contexts using

tasks which involve prolonged utterances. Furthermore, the present study is the first to examine salient cue detection as a predictor of cognitive and linguistic performance.

2.3.3.3 The Control Process Model (CPM)

Green and Li's (2014) CPM may be seen as a further extension of previous models of bilingual language production put forward by Green and colleagues (Green, 1986, 1998; Green, 2011; Green & Abutalebi, 2013). This model, the researchers argue, also provides evidence for the dual activation of languages and differentiates the types of codeswitching. According to the CPM (Figure 2.2), dense-codeswitching patterns involve equal activation of the languages, whereas insertional and alternational codeswitching entail a hierarchical type of activation. This involves open-control modes for the former type and 'coupled control modes' for the other two types of switches. As with the ICM, the CPM proposes that a conceptual stage precedes a phonological output stage, and that lemmas are tagged for language membership and selected by task schemas based on sociolinguistic and pragmatic appropriateness. Words are then produced based on 'competitive queuing' where a planning stage increases the activation of a lexical item which is then entered into the phonological assembly level. The next most highly inhibited item is then the next word to be produced, and so forth. Meanwhile, non-relevant items or less-activated items are inhibited in the 'competitive choice layer'. Conceptual representations in the 'conceptual-intentional system' communicate with 'language networks' which are controlled by language schemata. These language task schemas control entry into the 'competitive queuing network' thus determining serial order. Green and Li suggest that the nature of the control processes varies according to speaker ecology (i.e., language use across different communities) and codeswitching habits.

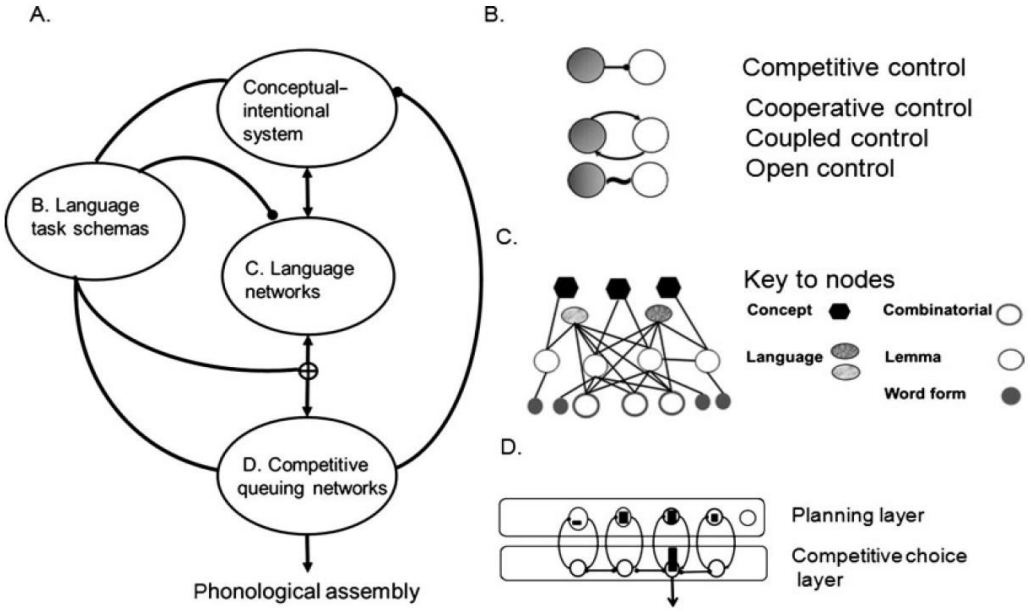


Figure 2.2 Codeswitching Control Process Model (Green & Li, 2014)

According to the ACH and CPM, inhibitory control is lower in the dense-codeswitching than in the alternation and insertion types. However, as language membership does not influence language choice, bilinguals must manage their linguistic co-activation through conflict monitoring. On the other hand, dual-language contexts are said to require increased conflict monitoring abilities than single-language contexts. The single-language context is somewhat akin to the language-blocked trials (though experiments almost always test the other languages within the same setting). The dual-language context is analogous to the language switching paradigm adopted in experimental designs, as participants are asked to alternate languages within the same context based on the introduction of a cue. Accordingly, picture-naming tasks may involve increased cognitive demands and result in switch costs. Cued language switching requires top-down control, as codeswitching is driven by explicit instructions, whereas voluntary codeswitching may require top-down control (perhaps due to an external cue)

and/or bottom-up control processes (e.g., due to lexical access in the case of opportunistic planning). Furthermore, in dense code-switching contexts, speakers may voluntarily switch languages without incurring significant (or any) switch costs. It is therefore important to compare response latencies across laboratory-based and more naturalistic contexts, yet very few studies have done so (e.g., de Bruin et al., 2018; Gardner-Chloros, McEntee-Atalianis, & Paraskeva, 2013).

In single-language contexts, speakers are believed to have global suppression of the competing non-target language(s), as they need to be inhibited at all times. In dual-language contexts, however, the cooperative control mode allows bilinguals to codeswitch for sociopragmatic reasons when appropriate. As such, bilinguals who codeswitch habitually in dual-language contexts are expected to have higher conflict monitoring skills and inhibitory control due to constantly training these capacities in their speech practices. However, participants from dense-switching contexts do not need to inhibit their languages, and as such, they may not show advantages in inhibitory control and conflict monitoring compared to other bilinguals whose speaker ecologies allow codeswitching at all times. This classification of cognitive processes involved across the different contexts offers an explanation of the inconsistencies in the literature, as bilinguals' speaker ecology is not always taken into account when comparing bilinguals to monolinguals (or to each other). However, the CPM offers explanations with the assumptions that bilinguals come from one context only, although Green and Li (2014) acknowledge that bilinguals may switch contexts. As such, the CPM does not account for bilingual speakers who, on a daily basis, may move from a single-language context to a dual-language or a dense-codeswitching one. Such bilinguals are not confined to one linguistic context, and it is difficult for the ACH and CPM to predict performance on cognitive tasks which tap into the different processes outlined by the two

models. Further, even though the CPM acknowledges that bilinguals may have several speaker ecologies, it is difficult then to categorically ascribe enhanced cognitive capacities (inhibitory and conflict monitoring or otherwise) to one context exclusively and not the other(s).

Moreover, in the dense-codeswitching context, it is unclear how the open-control mode manages the processes underlying codeswitching. Green and Li (2014) suggest that the most activated item is selected, yet it is unclear how unintended interference is managed. The CPM states that dense-codeswitching involves the selection of items which are grammatically appropriate. Any conflict in the 'fusion' of items in a morphosyntactic way must therefore be regulated, supposedly through conflict monitoring and inhibition. As such, opportunistic planning may involve a qualitatively different type of cognitive processes compared to those underpinning insertions and alternations. However, as this is speculative, it is important to examine whether bilinguals who habitually use dense-codeswitching (or congruent lexicalisations) in their speech exhibit different or enhanced cognitive performance especially in conflict monitoring, unlike the predictions of the CPM. Indeed, if this were the case, then the high-monitoring conditions which are simulated in the flanker task by Costa et al. (2009) should mimic the frequent switching which dense code-switching contexts involve.

This is a criticism of the CPM which the *extended* CPM (Green, 2018) addresses by alluding to different attentional control required in dense-codeswitching contexts compared to dual-language contexts. Green suggests that while the dual-language context requires enhanced 'narrowing of attention' due to competitive control states of the language schemas, the dense-codeswitching context is predicted to increase the 'breadth of attention' due to cooperative control. As such, Green suggests that participants from the latter context are more susceptible to interference (specifically

verbal binding competition when formulating congruent lexicalisations) and consequently become more adept at resolving it. Accordingly, speakers who frequently engage in dense-codeswitching are said to have greater facility in disengaging from interference. Green conjectures that these speakers would then exhibit reduced effects of congruency of a previous trial on the performance of a current trial in a Flanker task[4] (2018). This has indeed been evidenced in Hofweber et al.'s (2016) study which reported increased conflict monitoring abilities by bilinguals who frequently engaged in dense code-switching.

The original CPM also proposed that dense-codeswitching occurred only between typologically similar languages. However, the extended CPM (2018) suggests that elements from typologically distant languages may similarly be combined through congruent lexicalisations. The extended CPM then distinguishes the terms 'congruent lexicalisation' and 'dense-codeswitching' to refer to two different notions: 'congruent lexicalisation' refers to the construction of intra-clausal (or even intra-item) formulations whereas 'dense-codeswitching' refers to the presence of copious intra-clause switches. Of interest, however, is whether either 'congruent lexicalisations' or 'dense-codeswitches' would incur an additional cost when typologically distinct languages are combined compared to typologically similar ones. Green and Li (2014) and Green (2018) propose examining fluency (including pauses and speech rates) in order to assess the processes taking place before, during, and after switches. However, it is difficult to ascertain whether a difference in fluency reflects the type of switch (a congruent lexicalisation) or the typological distance between the languages used, especially when

[4] While the design of the present study preceded the publication of the extended CPM, the manipulation of the flanker task used in this study is in line with the extended CPM's (2018) predictions but in opposition to those of the original CPM (2014).

more than one pair of languages is spoken and switched, such as by multilingual speakers.

The ACH and CPM may offer great insights into language control by bilingual speakers, yet its applicability to multilingual contexts is unexplored. In an 'addendum' to the ICM and CPM (Green, 2017), the inclusion of the L3 in the language control models was proposed. According to Green, in language production, language task schemas must control the output by differentiating L2 and L3 schemas independently, so that interference from L2 does not inhibit L3 production. Green suggests that the different interactional contexts affect how language control is ultimately achieved. He suggests that speakers who codeswitch frequently may suffer greater language interference from their other languages when speaking in the L3, as the speakers have open control schemas which allow for this interference. On the other hand, speakers from single and dual-language contexts who are more used to separating their languages may be less likely to have overt intrusion errors when conversing in the L3 (Green, 2017:727). Green also suggests that individual differences in executive functions may predict performance errors in the L3.

While the ACH and CPM attempt to 'straddle the sociopragmatic-psycholinguistic divide', as Gardner-Chloros (2009) would describe it, the proposals have a few limitations. In short, the hypothesis and model claim to be applicable to multilingual contexts, yet they are only explained in terms of two-language codeswitching. Furthermore, they are based on the assumption that bilingual speakers are proficient in both languages, although they recognise that proficiency influences codeswitching patterns. Another assumption is that the influence of the single-language, dual-language, and dense-codeswitching contexts is static: the contexts place different demands on the control processes. However, the researchers acknowledge the possibility of bilinguals

moving across contexts, without elaborating on the consequences of such movement. While Green and colleagues criticise Muysken's (2000) typology, which is contextually-contingent, for the same reasons, the researchers appear to circumvent addressing them in their hypothesis and control model. It remains unclear which control processes are more activated in bi/multilingual speakers who experience three contexts in their everyday interactions.

In order to test the predictions of the ACH (Green & Abutalebi, 2013), Hartanto and Yang (2016) recruited a group of 133 Chinese-English bilinguals who were classified as either living in a single- language (N=58) or a dual-language (N=75) context. This classification was based on participants' responses to a questionnaire which investigated code-switching types (inter-sentential and intra-sentential) and code-switching frequency across different contexts (e.g., school, home, work, and informal situations). According to the ACH, inter-sentential switching requires greater cognitive effort, whereas intra-sentential codeswitching requires little inhibitory control as it allows simultaneous activation of both languages where task schemas are in cooperation rather than competition. It was predicted that participants who reported being in dual-language contexts and reported engaging in inter-sentential code-switching would exhibit greater cognitive control than both dual-language context participants who engage in intra-sentential switching and single-language context participants. To assess executive functions (specifically mental flexibility/monitoring), Hartanto and Yang administered a colour-shape switching task. Results indicated that dual-language context bilinguals had lower switch costs than single-language bilinguals. Furthermore, the researchers reported a positive correlation between inter-sentential codeswitching and task-switching performance, whereas intra-sentential code-switching was negatively correlated with task-switching scores. The researchers attributed this to the additional task-switching

effort required during inter-sentential codeswitching compared to intra-sentential codeswitching. This finding is in line with the predictions of the ACH and Green and Li's (2014) CPM, supporting the claim that different types of code-switching may engage and train different cognitive processes and confer different cognitive benefits.

While Hartanto and Yang's (2016) study is one of the very few studies investigating the applicability of the ACH and CPM, it is important to note that the study's measure of code-switching and exposure to contexts is based strictly on participants' self-reports. A possible caveat is participants' ability to subjectively and accurately measure their frequency of codeswitching based on the intra- vs. inter-sentential code-switching typology. Doing so involves not only knowledge of and consistent reliability in reporting differences in one's codeswitching patterns, but it also involves high levels of metalinguistic awareness and self-reflection for both voluntary and involuntary switches. Furthermore, it is unclear whether the participants were exclusively engaged in these contexts, or whether they experienced the single- and dual-language contexts at different times. This is criticism of the ACH as well, which assumes bilinguals are constantly confined to one type of context.

Furthermore, it is possible that participants who reported being engaged only in a single-language context and no code-switching may have under- or misreported their codeswitching behaviour due to possible negative attitudes towards their proficiency (e.g., inability to sustain conversations in one language), towards code-switching as a linguistic phenomenon (as highlighted by Gardner-Chloros, 2009), and/or towards misrepresentations of their identity. Negative attitudes may also be held against intra-sentential codeswitching, as opposed to inter-sentential codeswitching, also leading to underreporting. However, the use of questionnaire data remains ecologically valid. To account for subjective responses, questionnaires examining code-switching as a variable

should include items addressing participant attitudes towards the languages and towards code-switching. Attitude-based items are included in the present study's questionnaire, ensuring that this potentially confounding variable in self-reported data is taken into account. Additionally, a researcher's familiarity with the speech community and the context from which the data is collected lends further validity to the questionnaire. In the present study, participants come from a speech community which espouses multilingualism and code-switching, and individual differences in attitudes are additionally measured.

Green and Li (2014) speculate that examining fluency through pausing would permit testing the demands and adaptive nature of control in speech production. While codeswitching is said to incur 'costs' as reflected in slower response latencies when switching in certain tasks (e.g., Kutas, Moreno, & Wicha, 2009) and slower speech rates (Green, 2018), the claim that codeswitching requires increased cognitive 'effort' is debated. Codeswitching is typically effortless in naturalistic conversations, as evidenced in studies examining this phenomenon from a sociolinguistic perspective (Gardner-Chloros et al., 2013). Indeed, if it were taxing, it would not be as pervasive among bi/multilingual speakers who voluntarily switch, especially when communicating with interlocutors who share their languages. Psycholinguistic studies, even when designed to measure switch costs in artificial settings, have also suggested that codeswitching may at times be 'cost-free' (e.g., Kleinman & Gollan, 2016). However, the difference between cued and voluntary switches may not only be due to differences in response latencies (in one-word tasks) or pauses (in narration or conversation tasks). Gardner-Chloros (2009) suggests that voluntary (and naturalistic) codeswitching may involve different cognitive processes than those employed in psycholinguistic tasks.

One study which examined code-switching in a naturalistic setting through pausing is that of that Gardner-Chloros et al. (2013). In their study, the researchers analysed codeswitching between English and the Greek-Cypriot dialect by two groups of bilinguals, one residing in the United Kingdom and the other in Cyprus. The researchers operationalised fluency as 'pause length', measured in milliseconds. They tested participants in 'monolingual' and 'bilingual' modes which parallel the single- and dual-language contexts identified by Green and Abutalebi (2013). The two groups were expected to differ as the speech community in London typically switches between the languages, whereas codeswitching is less prevalent in Cyprus. Therefore, the researchers predicted that the bilingual group who resided in Cyprus would incur greater switch costs (i.e., they would pause more frequently and for longer durations) in the bilingual mode than their London-based counterparts who frequently codeswitch. Results, however, revealed no group differences in the bilingual mode, and no differences in pausing in the monolingual and bilingual modes. This suggests that codeswitching frequency may not influence bilinguals' fluency whether they are in a single- or dual-language context. These findings are opposite to the ACH predictions, as the London-based bilinguals would be expected to outperform their Cyrus-based counterparts due to more experience in switching back and forth (i.e., engaging and disengaging task schemas, etc.). Although the study offers this counter-evidence to the influence of habitual codeswitching on pauses, the use of longer utterances to measure codeswitching in naturalistic speech production is a methodological advancement, and it is adopted in the present study.

Another study which examined codeswitching in a more naturalistic way is that of Kleinman and Gollan (2016) who suggested that voluntary codeswitching may be cost-free. In their study, the researchers distinguished top-down and bottom-up processes of language selection and hypothesised that switch costs are driven by lexical access and

are therefore smaller in bottom-up switching conditions. In order to test their hypotheses of the mechanisms underlying codeswitching, the researchers recruited 171 Spanish-English bilinguals and categorised them as either Spanish-dominant or English-dominant based on scores from a multilingual naming test (although participants who performed equally in Spanish and English were classified as 'English dominant'). Kleinman and Gollan then administered a picture-naming task which asked the participants to choose the language they prefer to name words but to use the same language consistently when naming the words. For example, if a participant used 'door' when seeing a picture of a door for the first time, then he/she must consistently use 'door' and not its Spanish equivalent every time the picture of the door (stimulus) was presented. The researchers also administered a cued picture-naming tasks, in which they assigned the language to be used. As such, voluntary switches were compared to cued ones, with the aim of contrasting bottom-up and top-down processing mechanisms, respectively. Results indicated that bilinguals could maintain and use two languages as efficiently as a single language, switching between them frequently without incurring a switch cost when items in the target language had higher lexical access.

These findings resonate with those of a similar study by De Bruin et al. (2018). The researchers administered cued and voluntary language switching tasks to 54 Basque-Spanish young adults (mean age=23). Similar to Kleinman and Gollan (2016), de Bruin et al. noted that language choice was based on ease of lexical access. Surprisingly, however, switch costs were observed in both cued and voluntary task switches, yet cued switching showed a mixing cost. This cost was interpreted as the cost of using two languages instead of one, whereas a mixing benefit was observed for the voluntary switching task. This suggested that using two languages within naturalistic speech voluntarily may be less costly than speaking in one language only. While the present

study does not administer a picture-naming task which manipulates voluntary and cued switches, it does so in a narration task in which participants are asked to use only one language or codeswitch between their three languages. Accordingly, whether voluntary codeswitching is cost-free in blocked and mixing conditions is examined within a more naturalistic task.

Another study which observed no switch costs in voluntary codeswitching is that of Blanco-Elorrieta and Pylkkänen (2017). In their study, Blanco-Elorrieta and Pylkkänen administered a picture-naming task to 19 Arabic-English bilinguals. The cues in the task included artificial cues (colours), faces of monolingual speakers ('monolingual cues' indicating one appropriate language to use), and faces of bilingual speakers ('bilingual voluntary cues'). Results indicated that switch costs were incurred in trials involving artificial cues, but not natural cues. This highlights a possible experimental caveat, as the language-switching paradigm tasks involve colours or flags to cue the specific language to be used. Furthermore, it elucidates the importance of creating tasks which mimic naturalistic conversational contexts where cue detection may indeed be highly sensitive to the nature of the cues and how closely these are to naturally occurring cues in daily conversations. The difference between the 'artificial' and 'naturalistic' cues was further supported by imaging data from the study. When bilinguals switched in response to artificial cues, increased activation in two areas of the brain associated with language switching and executive control was observed (specifically, the anterior cingulate cortex and dorsolateral prefrontal cortex). However, in the monolingual cues, increased switch activation was observed in the earliest time window of analysis only, and in the bilingual voluntary cues, there was no increased activation at all. Blanco-Elorrieta and Pylkkänen therefore concluded that switching may

be cost-free in naturalistic settings at both the behavioural and neural levels, whereas it is most effortful when cued by artificial cues.

The notion that codeswitching may be 'cost free' is further supported by evidence form Hlavac (2011) who examined highly proficient Croatian-English speakers who codeswitched in naturalistic speech. The researcher noted that participants produced few hesitation and monitoring phenomena in codeswitched speech, and that when these phenomena occurred, they preceded a codeswitch which did not contain the phonological or morphological attributes of the base language. Kroll, Gullifer, McClain, Rossi, and Cruz Martin (2015) interpret Hlavac's findings as evidence that disfluencies do not stem from lexical gaps; rather, they state that hesitation phenomena signal to their interlocutors of their intention to codeswitch. Furthermore, Kroll et al. suggest that these findings support the ACH as speakers opportunistically exploited the coactivation of their two languages especially since the hesitation phenomena amounted to only 2% of the total speech samples.

Unlike the findings presented above which suggest codeswitching is 'cost-free', other studies have indicated that switch costs can be observed even in voluntary conditions. Gollan and Ferreira (2009) examined voluntary codeswitching in Spanish-English balanced and unbalanced bilinguals, in order to explore factors influencing switch costs. The researchers noted that codeswitching was frequent in both groups. Unbalanced bilinguals, however, codeswitched fewer times (24% of trials) than their balanced counterparts (35% of trials), indicating that proficiency in the L2 is a predictor of codeswitching frequency. Additionally, the researchers noted a switch cost, although the switching was performed voluntarily. When speaking in single-language blocks, however, balanced bilinguals incurred mixing costs, as opposed to unbalanced bilinguals in the L2. The researchers attributed this 'mixing benefit' to the monitoring of lexical

access in the L2; i.e., using the L2 only when it is easily accessible. This may reflect a strategy by the unbalanced group to avoid naming items which are less accessible. However, in further analyses, Gollan and Ferreira reported a mixing benefit even for items which did not show strong language preference. Therefore, lexical access may only partly explain the mixing benefit. Gross and Kaushanskaya (2015) examined voluntary codeswitching among children and provided further evidence that language choice is not fully explained by lexical access. The researchers noted three main factors which influenced the production of items in the less-dominant language: (a) when the L2 item was acquired early on in life; (b) when the L2 item had a higher frequency of use than its L1 equivalent; and (c) when the L2 item did not have an L1 equivalent. Similar to Gollan and Ferreira's (2009) study, mixing costs differed across the L1 and L2; a mixing cost was found for the L1, but not the L2, in reaction times and accuracy rates, whereas a mixing benefit was found for the L2. However, the presence of switch costs indicated that voluntary switching was costly, unlike the findings of studies presented earlier (e.g., Kleinman and Gollan, 2016).

2.3.3.4 Summary of Language Control Models

In summary, the ICM has generally received substantial support from behavioural studies examining the relationship between language control and executive functions, particularly inhibitory control. Overall, analyses of switch costs have revealed symmetrical switch costs between languages of similar proficiency levels and asymmetrical switch costs when one language has a higher level of proficiency, with the dominant language typically requiring more inhibition than the weaker language(s). However, there has been an inconsistency in findings due to numerous factors which include methodological variations, such as the manipulation of preparation time and differences in operationalising 'proficiency', as well as a lack of controlling for other

factors such as individual differences. The ACH and CPM were therefore proposed as extensions of the ICM by taking into account different external and participant-internal factors which may better explain the inconsistency in findings. The two recent proposals suggest that linguistic contexts vary in terms of the cognitive processes they involve, as bilinguals may either alternate languages freely between their languages or separate them to varying degrees. They also take into account individual differences such as participants' frequency of codeswitching which may influence switch costs.

Green and his colleagues (Green, 2017, 2018; Green & Abutalebi, 2013; and Green & Li, 2014) further suggest examining language switching in naturalistic contexts through fluency (such as pausing). This is based on the premise that codeswitching is cost-free *in situ*, i.e., in voluntary conditions compared to cued, experimental settings. However, like the findings of switch costs, findings from studies examining pauses in codeswitching are inconsistent. Methodological considerations which take into account cued and voluntary conditions have attempted to investigate the mechanisms underlying codeswitching and whether top-down processes which inhibit items from a language as opposed to bottom-up processes which are said to involve lexical access. Although many of the studies aim to investigate whether codeswitching is costly compared to the cost-free single-language production, it is conceivable that speakers who are more accustomed to switching between languages would incur costs when speaking in a monolingual mode. This is investigated in the present study, where individual differences in language experience and in speech behaviour are considered as modulators of language production and codeswitching across single-language and trilingual conditions. This leads to the next section of the chapter, which highlights some of the factors which influence language (and cognitive) control.

2.4 FACTORS INFLUENCING COGNITIVE AND LINGUISTIC CONTROL

While the notion of the bi/multilingual advantage and the exact mechanisms underlying language production and executive functions are still debated, deconstructing 'bilingualism' as a unitary concept has been supported by several researchers (Luk & Bialystok, 2013; Treffers-Daller, 2009). This involves examining correlations between variables within the bilingual experience and executive functions in order to take into consideration inter-participant variation (e.g., Abutalebi & Green, 2016; Bak, 2016). Such an approach, therefore, allows researchers to identify the specific aspects of bilingualism that would confer a bilingual advantage. As executive functions are modulated by experience, sociolinguistic practices which involve frequent switching between languages must therefore be examined as influencing executive functions (de Bruin & Della Sala, 2016; Valian, 2015; Yang et al., 2016). One of the main objectives of this study is to contribute to the present understanding of the interaction between language proficiency, language experience, and executive functions, taking into account the different experiences individuals and groups have in their daily interactions.

2.4.1 Language Proficiency

The most widely examined aspect of bilingualism and cognition research is language proficiency. Measures of proficiency generally include self-reports, performance on language tests, and third-party metrics, such as school grades or scores from standardised tests. Based on these measures, *language balance* is typically derived, where 'balance' refers to the difference in proficiency levels between two languages (Kupisch & De Weijer, 2016). To complicate matters further, 'balance' has also been used interchangeably with *language dominance*, and the measures of both have been sometimes used to measure proficiency. Some studies use 'balance' and 'dominance' to refer to the difference in frequency of language use, typically measured through self-

reports. For speech communities in which multilinguals constantly use their languages, however, it is difficult to establish which language is the strongest or weakest in an objective way. Further, balance and dominance have also been measured through the use of *lexical robustness* which sometimes denotes proficiency, as it refers to lexical access (Linck et al., 2012; Schwieter & Sunderman, 2008, 2009, 2011). This conflation in both definition and measurement has been attributed to much of the inconsistencies in the findings of studies that examine the same construct but define and/or measure it differently (Silva-Corvalán & Treffers-Daller, 2016). This study distinguishes 'language proficiency' from the two other terms by defining and measuring this construct by how well a speaker performs on an oral proficiency test and is able to perform all the linguistic tasks involved in the test. This includes being able to discuss a variety of topics using complex syntactic structures. The study measures 'lexical robustness' through the use of a verbal fluency task, in order to draw comparisons with previous studies which have used this construct. However, the present study avoids references to 'dominance' or 'balance' based on lexical robustness scores. As lexical robustness is said to tap into a specific area of proficiency (ease of lexical access) (Schwieter & Sunderman, 2011), it is discussed within the confines 'lexical access' and is analysed separately from proficiency scores as a predictor of linguistic and cognitive task performance.

The notion that language proficiency and/or lexical robustness influence switch costs has been highly supported in the literature (e.g., Meuter & Allport, 1999), as discussed at length in the previous section of this chapter (Section 2.3). Similarly, the two constructs have been said to influence cognitive performance (e.g., Rosselli, Ardila, Lalwani, & Vélez-Uribe, 2016; Tse & Altarriba, 2012, 2014). However, very few studies have examined the influence of the two factors on both linguistic and non-linguistic tasks. Studies which have examined individual differences in language proficiency and

executive functions have generally reported that higher proficiency levels (especially in the L2) tend to be related to higher accuracy rates, faster reaction times, smaller Conflict effects (Flanker Task) and/or smaller Simon effects (Simon Task). It is important, however, to stress that the relationship between language proficiency and cognitive task performance is not a causal one, and that whether proficiency influences executive function or is influenced by it (or mutually influences and is influenced by it) remains to be a source of controversy.

Some studies (e.g., Festman & Schwieter, 2015; Koch et al., 2010), suggest that language proficiency and its relation to switch costs and cognitive task performance may not offer a complete account of the interactions between the constructs. As such, some researchers (e.g., Bak, 2016; Green & Li, 2014; Paap et al., 2015) emphasise the importance of examining other variables such as speakers' codeswitching behaviour and idiosyncrasies in the bi/trilingual experience. This supports the notion that language control and executive functions are related and that they may rely on similar or overlapping processes depending on individual differences in the participants' bilingual experiences. Some of these individual differences are explored in the present study and are elaborated upon next.

2.4.2 Language Experience

2.4.2.1 Codeswitching Frequency

As purported by the ACH and CPM, individual differences in the frequency of codeswitching and language use may involve different cognitive processes. Accordingly, it is necessary to examine the influence of language experience on linguistic and cognitive task performance. While some studies have recently explored codeswitching frequency as a predictor, there has been an inconsistency in measuring this variable.

In order to operationalise and measure code-switching in a systematic way, Rodriguez-Fornells, Kraemer, Lorenzo-Selva, Festman, and Muente (2012) developed and administered the Bilingual Switching Questionnaire to 566 Spanish-Catalan bilingual participants. The questionnaire included items on a 7-point Likert scale, tapping into the frequency and intentionality of code-switching. Furthermore, the researchers administered a Flanker-Stroop task where trials were equally divided between congruent- and incongruent conditions, and included stop-signal trials. This manipulation of trials required constant conflict monitoring, making the task a high-monitoring one. The results of the study revealed a positive correlation between the questionnaire scores and measures of both interference suppression (Conflict Effect scores) and response inhibition (task switching from stop-signal trials). However, the researchers only administered the task in a high-monitoring context, so the conflict effect may have been due to either inhibition or conflict monitoring. This study therefore lends support to the role which code-switching plays in terms of enhancing executive control, yet it does not specify or tease apart the elements of inhibitory control, specifically conflict monitoring (see also Liu, Jiao, Wang, Wang, Wang, & Wu, 2018 for similar results).

Furthermore, Prior and Gollan (2011) examined two groups of bilinguals who reported disparate code-switching habits. A group of 41 Spanish-English frequent code-switchers and a group of 43 English-Mandarin non-code-switchers performed a colour-shape task-switching task. A monolingual English group (n=47) also performed the task. A fluency task was administered to assess language-switching performance between the two bilingual groups. The results indicated that the code-switching bilinguals outperformed the monolingual group on the task-switching, but there was no significant difference in performance between the non-codeswitching bilingual and monolingual groups. Furthermore, frequent code-switchers outperformed the non-codeswitching

bilinguals in language switching by having smaller switch costs, suggesting a correlation between cognitive task switching and language switching. This study therefore highlights the importance of examining the influence of code-switching habits when looking at both linguistic and non-linguistic performance. However, as noted by Verreyt et al. (2016), a caveat in Prior and Gollan's findings is not taking into account the typological distance between the two sets of languages spoken by the two bilingual groups. As English and Mandarin are typologically more distant than English and Spanish, it is possible that the English-Mandarin bilinguals would be able to separate their languages more easily due to the lack of shared structures, lexemes, and orthography. As such, the English-Mandarin non-codeswitchers would not experience bilingual language control which is as taxing as that of their Spanish-English counterparts, leading to differences in task performance. While this is indeed an important variable to take into account, Prior and Gollan's findings may still be valid even if conducted on more comparable bilingual groups.

In order to overcome the confound of typological differences between the languages spoken by participants, Verreyet et al. (2016) examined the influence of codeswitching frequency and language proficiency on cognitive performance in Dutch-French bilinguals in Brussels. In their study, the researchers classified bilinguals as balanced or unbalanced using self-reported language proficiency scores. They also asked participants to report the frequency of their code-switching on a scale of 0 (never) to 7 (very often). Results from this classification indicated that unbalanced bilinguals reported minimal to no frequent code-switching (almost never), whereas balanced bilinguals reported a range of code-switching frequency. Accordingly, Verreyet et al. subdivided the balanced bilingual group into 'frequent' and 'infrequent' code-switchers. The researchers then administered a Simon Task and a Flanker task to assess inhibitory

control. Results of the study indicated that balanced bilinguals who frequently code-switched outperformed both unbalanced bilinguals and balanced, infrequent-codeswitching bilinguals. Furthermore, self-reported code-switching frequency was positively correlated with Flanker and Simon Task performance on both congruent and incongruent trials, suggesting that code-switching correlates positively with conflict-monitoring, as suggested by Costa et al. (2009). However, differences between balanced and unbalanced bilinguals were not observed, suggesting that proficiency, unlike code-switching frequency, is not a significant predictor of cognitive control. The researchers conclude that code-switching modulates both inhibitory control and conflict-monitoring.

While the studies described so far have reported a correlation between codeswitching frequency and at least one aspect of executive functions, other studies have found no significant correlations between the two. Rather, they reported other variables which predict enhanced cognitive performance, such as the age of L2 acquisition (Soveri et al., 2011) and language proficiency (Kang & Lust, 2018). Despite their lack of support for codeswitching frequency as a significant predictor of cognitive performance, these studies highlight the importance of looking at other variables which are of interest to the present study as well. The studies are described next.

Soveri et al. (2011) investigated self-reported code-switching behavioural patterns and executive functions by examining individual differences among 38 Swedish-Finnish bilinguals. Rodriguez-Fornells et al.'s (2012) questionnaire was used to assess code-switching habits, the Simon and Flanker tasks were used to measure inhibition, and a number-letter task was used to measure task-switching. Unlike Rodriguez-Fornells et al.'s findings, however, there was no correlation between questionnaire scores and cognitive task performance. The results were also in contrast with earlier Verreyet et al.'s (2016) findings which reported a positive correlation between codeswitching frequency

and performance on the Simon and Flanker tasks. However, Soveri et al. reported a strong correlation between cognitive performance and other background variables such as the age of L2 acquisition. Furthermore, the researchers proposed that the lack of correlation between code-switching frequency scores and switch costs suggests that code-switching does not train continuous switching; rather, it trains high-level monitoring processes which manage simultaneously co-activated languages. These processes, as explained by the authors, may not be observed in non-verbal switch costs which are trial-dependent and involve transient, short-term control.

Another study which examined the influence of codeswitching frequency, as well as other individual differences in language experience, on cognitive task performance is that of Kang and Lust (2018). The study recruited 43 English-Chinese bilingual children (aged 8) residing in a codeswitching-rich context, Singapore. Questionnaires exploring the amount of exposure to each language, codeswitching frequency, and language proficiencies indicated that codeswitching was frequent in this sample. End-of-year school examination scores for English and Chinese were also examined as an (additional) objective measure of language proficiency and as a means of categorising participants as balanced or unbalanced bilinguals. Kang and Lust then administered an elicited imitation task, a codeswitching task, and three executive functions tasks: a Stroop task, a verbal fluency task, and a nonverbal Hearts and Flowers/Dots task. The elicited imitation task was used to further assess participants' proficiency in the two languages by being able to produce a small variety of sentence structures using comparable vocabulary levels. The code-switching task, a novel task which the researchers developed, required participants to produce voluntary switches in response to prompts, with no language being cued by the researchers. However, children typically switched due to the language switch of their interlocutor in the preceding trials, as

expected and observed in numerous studies (see Myers-Scotton, 2006, for an overview). The results of the study reported a non-significant correlation between codeswitching performance and cognitive task performance. However, the researchers noted that the 'degree of bilingualism' was a significant predictor. The degree of bilingualism was measured by questionnaire proficiency scores, the end-of-year school scores, and reported language use and language exposure. While the researchers suggest that the role of codeswitching may be indirect and perhaps unessential in cognitive task performance, methodological limitations in the study may explain these findings, such as the operationalisation of codeswitching and the conversational task which involves a confounding variable—the interlocutor effect. This does not preclude, however, the influence of the other variables which were reported to be significant predictors.

A similar approach to analysing the role of codeswitching in cognitive task performance is adopted by Yim and Bialystok (2012). Similar to the results of Kang and Lust (2018), the findings of the study provided counter-evidence to the influence of codeswitching frequency (and type) on cognitive tasks, yet they indicated a correlation between codeswitching and enhanced language task performance (switch costs). Yim and Bialystok explored whether codeswitching frequency and type predicted better cognitive and linguistic performance by measuring codeswitching patterns through elicited conversation tasks. This objective measure of codeswitching frequency was said to overcome the self-reporting measures which other studies adopted. Distinguishing intra- and inter-sentential codeswitching by 78 English-Cantonese bilinguals, the researchers analysed the frequency of switches, although no further analyses of the typology were reported in the article. The researchers also administered a fluency task which involved monolingual and bilingual switching blocks, a switching task which required participants to alternate sorting by colour and sorting by category, and a version

of the Simon task. Yim and Bialystok reported a negative correlation between codeswitching frequency and language switch costs from the fluency tasks, yet they did not observe any significant correlations between codeswitching frequency and cognitive task performance. This led the researchers to conclude that there is a dissociation between linguistic and cognitive control, and that the processes underlying codeswitching frequency are language-specific, as they do not seem to correlate with or draw upon cognitive mechanisms.

While Yim and Bialystok's (2012) claims may be true for their particular sample, it is difficult to generalise these conclusions as there are methodological considerations which may have influenced the findings. As discussed in the critique of Prior and Gollan's (2011) study, the dissociation between the two types of control may be ascribed to the typologically distant languages spoken by the participants, as cross-linguistic interference may not be significant enough to influence or be detected by cognitive task performance. Furthermore, the objective measure used for code-switching frequency may be influenced by other factors such as topic, which is a manipulation the researchers introduced in the task. Accordingly, the code-switching patterns elicited by the conversation task may not have been representative of the naturalistic codeswitching behaviour in which the participants typically engaged.

Despite the inconsistency in the findings of the studies described so far, the correlation between linguistic practices and cognitive and linguistic task performance is still generally supported. Costa et al. (2009) suggested that code-switching and mental flexibility are correlated, and that enhanced bilingual performance on the high-monitoring conditions of the flanker task can be attributed to code-switching. It has also been suggested that training in code-switching can generalise to cognitive domains (Kutas et al., 2009; Timmer, Calabria, & Costa, 2019). Further support comes from

neural evidence which supports the hypothesis that codeswitching trains conflict monitoring and interference suppression (Abutalebi & Green, 2007). In high-monitoring conditions, such as those in which bilinguals may need to codeswitch, speakers must constantly manage resources from both languages and monitor external and internal cues. This enables speakers to accommodate their interlocutors by responding in the appropriate language(s) through selecting or deselecting the language(s), while inhibiting the non-intended language(s). This is analogous to the monitoring and inhibitory control processes which participants are asked to engage in tasks such as the Flanker and Simon tasks. Costa et al. (2009) hypothesised that the bilingual advantage may stem from codeswitching patterns of participants, yet the researchers did not measure the codeswitching frequency of the participants in their study. However, as the participants were Catalan-Spanish bilinguals, it is likely that they were dense codeswitchers. Categorising the participants as such would further explain the finding in terms of their performance on the different conditions of the flanker task.

The notion of the bilingual experience 'training' cognitive capacities and thus resulting in a bilingual advantage has been investigated by Timmer et al. (2019). This study investigated the relationship between language control and domain-general cognitive processes and explored whether codeswitching training enhances non-linguistic task switching performance. Sixty-four Catalan-Spanish young adults (mean age=22) were randomly assigned to two groups: a group who was trained in language switching and a group who was trained in blocked language naming. For the language switching task, participants performed a picture-naming task, and for the cognitive switching task, participants performed a colour-shape judgment task. The researchers administered a pre-test and a post-test for both tasks one week apart, with two training sessions on language switching in between. Results indicated that both linguistic and

non-linguistic switch costs decreased significantly more for the first group than the latter group, but there was no difference between the groups in mixing costs for the single-language blocks which were similarly reduced for both groups. Timmer et al. attributed the group difference in switch costs to the transfer from the linguistic to the non-linguistic domain for specific sub-processes, thus suggesting at least a partial overlap of the control mechanisms across the two domains. While this study offers a new approach to the topic of the overlap between these two domains, it should be noted that individual differences may not have been properly accounted for given the random allocation to groups and the variability among these participants in terms of their codeswitching practices. Furthermore, the researchers used self-reported proficiency ratings to establish that the groups were matched, yet more detailed and objective background information about the language experiences and proficiencies of the participants would have perhaps allowed the researchers to control for potentially confounding variables.

In summary, findings from studies on bilingualism and executive functions have propelled researchers to explore the relationship between codeswitching, language switching tasks, and domain-general control. Some studies have reported a correlation between codeswitching frequency and high-level monitoring abilities but not switch costs (e.g., Soveri et al., 2011; Costa et al., 2009), whereas other studies stated that codeswitching modulates both monitoring and switch costs (Prior & Gollan, 2011; Hartanto & Yang, 2016) and inhibition (Verreyt et al., 2016).

2.4.2.2 Language Use

Few studies (e.g., Hartanto & Yang, 2016; Hofweber et al., 2016; Soveri et al., 2011; Verreyt et al., 2016) have examined the influence of individual differences in language use on executive functions, and while many of the studies conflate codeswitching frequency with language use, the present study distinguishes the two. The

present study operationalises 'language use' as the self-reported percentage of time each language is used across different contexts and with different interlocutors, whereas codeswitching frequency refers specifically to self-reported frequency of switching between languages within the same utterances or conversations in different contexts and with various interlocutors. In general, the findings of previous studies suggest that increased use of the different languages is linked to faster reaction times on cognitive tasks (Kang & Lust, 2018; Dong & Xie, 2014). In addition, it is linked to smaller conflict effects, especially in high-monitoring conditions (Costa et al., 2009). Similarly, as discussed in the previous section, increased codeswitching frequency has been observed to confer cognitive advantages reflected in smaller Conflict/Simon effects and enhanced monitoring and control capacities (Hartanto & Yang, 2016; Hofweber et al., 2016). Overall, studies indicate that increased—or perhaps more balanced use—of both languages is related to enhanced cognitive performance. However, the specific aspects of executive functions which are improved depend on the nature of language use given the different communicative contexts, as suggested by the ACH and CPM. Accordingly, the present study examines the percentage of self-reported language use and frequency of codeswitching as separate predictors of individual differences across different tasks measuring executive functions.

2.4.2.3 Language Exposure

Bilinguals and multilinguals vary considerably in terms of their exposure to more than one language. This exposure differs based on individuals' contexts, and the ways in which speakers interact with their communicative contexts is reflected in their speech behaviour (Green & Abutalebi, 2013). Typically, studies operationalise language exposure as the percentage of time bilinguals are exposed to each language or as bilinguals' Age of Acquisition (AoA). Individual differences in language exposure in

terms of percentage of exposure (e.g. Gullifer, Chai, Whitford, Pinvena, Baum, Klein, & Titone, 2018) may play a crucial role in bilinguals' proficiency and may also tie in with their language use (Titone et al., 2017). However, studies do not typically examine the potential overlap between language exposure and language use—a contribution made by the current study in understanding the influence of language experience.

As for AoA, several studies have examined the influence of L2 and L3 age of acquisition on executive functions (e.g., Luk et al., 2011; Tse & Altarriba, 2012). These studies generally compare monolingual groups, simultaneous bilinguals, and sequential bi/multilinguals, with the assumption that the earlier L2 AoA is, the higher the demands on language control. Sequential bilinguals have also been subdivided as 'early' and 'late' L2 learners, in order to assess the influence of bilingualism on executive functions across different stages of development. However, there has been an inconsistency in specifying the age range of 'early' and 'late' AoA (Bonfieni, Branigan, Pickering, & Sorace, 2019). For instance, some studies refer to early bilinguals as individuals who were exposed to the L2 before formal schooling (i.e., age range of 3-5), whereas other studies refer to 'late' bilinguals as those who learn their L2 after age 3 (any age between 3 and 12). Moreover, other studies use 'late' bilingualism to refer to languages learned after the age of 12. Given the overlap of age range in categorising speakers as 'early' and 'late bilinguals', the role of AoA in executive functions has been difficult to establish when comparing studies. Additionally, some researchers operationalise AoA as the first instantiation of L2 exposure.

The present study distinguishes language exposure and AoA by using self-reports of language exposure to each language in different contexts (in percentage). An objective index of AoA is used to classify the instantiation of formal and informal exposure. As the multilinguals in the study are exposed to the three languages, even if in varying

degrees, to the three languages, informal AoA is set as 'from birth'. Formal instruction, on the other hand, is indexed by the year in which formal instruction in each language is given. As such, formal instruction in Arabic and the immersion language is set at age 3 and the third language is set at age 11.

2.4.2.4 Language Typological Distance

It is suggested that typological distance between languages may influence the means by which language control operates. When languages are similar, they are said to interfere with one another, leading to larger switch costs, compared to more distant languages. However, few studies have explored language typology as a modulating factor influencing cognitive task performance (see influence of script in Coderre & van Heuven, 2014 and Paap, Darrow, Dalibar, & Johnson, 2015). This may be attributed to the complexity of and difficulty in identifying language distance across domains, such as phonology, orthography, syntax, and vocabulary. In studies examining trilingualism, a language may share two domains with one language and two other domains with the third language, making it difficult to establish typological distance between the three languages and to quantify typology as a measurable construct. Many studies examining this typological distance in studies of cognitive functions typically contrast language pairs which share few features with those that are similar (e.g., Korean and English vs. German and English). However, results from these findings are mixed, as some studies reported generally slower response latencies by bilinguals who spoke typologically similar languages (e.g., Bialystok et al., 2005; Paap et al., 2015), whereas other studies reported smaller conflict effects (Coderre & van Heuven, 2014) and smaller switch costs (Prior & Gollan, 2011) for bilinguals speaking non-distant languages.

To investigate this potential influence, Costa et al. (2006) compared two groups of highly proficient bilinguals, with one group switching between two typologically similar

languages (Spanish and Catalan) and the other group switching between two distant languages (Spanish and Basque). Results indicated symmetrical switch costs between the two groups, suggesting that language similarity does not significantly influence switch costs, and more generally, the way in which bilinguals control their languages. However, recent studies have suggested that typological distance between languages may influence how *multilinguals* process their third language. This is attributed to the dual—yet varying—influence of the other two language on the third language. Indeed, one of the most important determinants of which language influences the third language more is that of typological distance. It is hypothesised that the language closer to the weaker third language will constitute the main source of cross-linguistic influence during L3 production. If both languages are equally similar to the L3, then the weaker of the two is more likely to be the source of influence (Cenoz, 2003). This is based on the assumption that unbalanced multilinguals are able to inhibit their L1 more easily than their weaker L2 when speaking in the L3 (de Bot, 2004; Hammarberg, 2001). This is further supported by the notion that L2 strategies are typically adopted when a third (or more) language is learnt, thus rendering the L2 as the main source of influence during L3 processing.

Numerous studies examining switch costs and language control contrast language pairs to determine which language takes more time to suppress and reactivate in speech production tasks such as picture-naming tasks. However, analyses are typically more centred around participants' proficiency in the three languages, without taking into account the potential influence of the languages' typological distance. In his multilingual production model, de Bot (2004) suggests that the multilingual lexicon comprises subsets of language-specific items based on linguistic closeness. Based on typological distance, the subsets may overlap with one another to varying degrees. When the level of

activation of a language increases, the associated subsets it shares with other languages are also activated. The greater the overlap between the shared subsets, the higher the activation and subsequent influence of the two languages. In this model, increased inhibition of the more dominant language allows for more influence from the less dominant language, and the degree of this influence is amplified by the typological closeness of the L2 and L3.

The present study examines three relatively 'equally' distant languages—Arabic, English, and French, and the multilingual participants are subdivided as highly proficient in two or three languages. Unlike English and French, which are closely related to Indo-European languages, Lebanese Arabic is a Semitic dialect that has no standardised script (Comrie, 1991). It also differs from Modern Standard Arabic (MSA), which is rarely spoken and used mainly in writing or highly formal contexts, although the two varieties share some lexical and grammatical features. However, Lebanese Arabic shares several phonological and grammatical features with French, and as it is infused with lexical items from English and French, it may be argued that it is equally similar to both languages. As typological distance is beyond the remits of the present study, and given the above discussion, it would of interest to assess whether switch costs vary depending on the language pairs, while taking into account the proficiency levels of the balanced vs. unbalanced trilingual groups. As the three languages are shared by all participants, typological distance is therefore not a confounding variable, and cross-linguistic influences are not investigated further.

2.4.2.5 Age

Age has been argued to be a highly influential factor in bilingualism and cognition research, as developmental factors are tied to executive functions (Anderson, Chung-Fat-Yim, Bellana, Luk, & Bialystok, 2018). As indicated in the discussion on the bilingual

advantage, bilingual children and older adults generally outperform their age-matched monolingual counterparts on tasks tapping into cognitive functions. Studies on the bilingual advantages in children have argued that bilingualism influences their executive functions which in turn enhances their metalinguistic and cognitive skills, culminating in higher educational attainment (Blain-Briere, Bouchard & Bigras, 2014). For older adults, a surge of studies examining the influence of bilingualism on cognitive functions has reported a slower and delayed decline in mental capacities due to old age in bilinguals compared to monolinguals (Bialystok et al., 2008; Bialystok et al., 2004). Furthermore, researchers have found age-related advantages in both late as well as early (and/or balanced) bilinguals. Bak, Nissan, Allerhand and Dreary (2014), for example, reported that late bilingualism enhances general intelligence and reading skills in older adults. Studies on young adults, however, have revealed multifarious results, as this age group is said to be at peak performance. However, as discussed earlier, the use of more complex tasks has been suggested as a means of circumventing this issue to reveal differences between monolingual and bilingual young adults (Costa et al., 2008). The present study which examines the cognitive and linguistic performance two groups of young adults (ages 17-18 years) uses complex versions of the tasks.

2.4.2.6 Socioeconomic Status

Another factor which is argued to influence cognitive and linguistic task performance is that of Socioeconomic Status (SES). Studies of SES which have examined children from different SES backgrounds have reported lower levels of language skills (both receptive and productive) for children of low SES (Locke, Ginsborg & Peers, 2002). SES has also been observed to influence cognitive development and executive-function components including inhibition, planning, monitoring, and switching (Morton & Harper, 2007; Noble, Norman, & Farah, 2005).

Most studies which examine bilingualism and cognition therefore control for this variable. One study which examined the bilingual advantage among low SES and middle-class SES children is that of Calvo and Bialystok (2014). The study recruited 175 six-year-old children which were divided into four groups: low-SES monolingual, low-SES bilingual, middle-SES monolingual, and middle-SES bilingual. The participants performed a series of tests which included the Peabody picture vocabulary task, a picture-naming task, and a Flanker task, and basic intelligence tests, among others. The children performed similarly on intelligence tests, but executive function tasks were influenced by both SES and bilingualism. Specifically, middle-class children outperformed low-SES children on all tasks, and bilingual children outperformed monolinguals on cognitive tasks. However, monolingual children outperformed bilingual children on language tests, which was expected. Although there were no significant interactions between bilingualism and SES on task performance, the study highlights the importance of examining each variable, even if separately, as it may influence outcomes on different tasks. As such, the present study controls for SES by recruiting participants from the same (high) socioeconomic background.

2.4.3 Summary of Factors Influencing Task Performance

To summarise this section, it is important to note important variables which may modulate linguistic and non-linguistic task performance in order to either explore their influence or to control for it. Language proficiency, codeswitching frequency, language use, and language exposure have been highlighted in the literature as highly influential factors affecting performance, and they are all examined as predictors in the current study. In addition to these factors, salient cue detection, which has not been investigated, yet which is hypothesised by the ACH as a cognitive process underlying speech and non-linguistic processing, is also explored as a predictor. Other factors which are said to

influence linguistic and cognitive processing include typological distance, age, and socioeconomic status. These three variables are controlled for in the present study.

2.5 RATIONALE AND OVERARCHING HYPOTHESES

This chapter highlighted recent research undertaken to explore the relationship between language control, executive functions, and the effects of bi/multilingualism on language production as well as on non-linguistic processing. In reviewing the literature, there appear to be several factors which influence performance on language and cognitive tasks, such as language proficiency, patterns of language use, frequency of codeswitching, the typological distance between languages, age of acquisition, socioeconomic status, and age. Furthermore, theoretical and methodological differences in defining and measuring constructs have led to an inconsistency in the findings (Paap et al., 2016). The conflation of terms and the variation in tasks which ostensibly measure bilingualism and executive functions pose a challenge in drawing conclusions from studies on both language control and bilingual advantages. Recent calls for investigating the interplay between language and executive functions have proposed doing so by taking into account individual differences in proficiency and language experience (Green & Abutalebi, 2013; Green & Li, 2014). Additionally, while multilingualism is becoming increasingly pervasive, very few studies have investigated multilingual advantages. Further, language control models do not seem to offer a complete account of the multilingual experience.

Given the above, the study hypothesises that speakers who have a higher proficiency in the three languages will exhibit a multilingual advantage, at least in one aspect of executive functions (inhibitory control, conflict monitoring, or working memory). As there have been inconsistent findings, administering different tasks of

complex designs aim at elucidating differences in performance between balanced and unbalanced trilinguals. The present study also predicts that participants who frequently engage in codeswitching would exhibit better conflict monitoring skills, reflected in their performance on the high-monitoring condition of the flanker task. Participants who report using their languages more separately are expected to possess better salient-cue detection and exhibit enhanced inhibitory control skills as measured by the Simon task. In addition, speakers who are highly proficient in the three languages and constantly codeswitch are expected to have smaller and symmetrical switch costs on the trilingual picture-naming task and narration task in the trilingual condition. However, it is expected that unbalanced trilinguals would exhibit larger, asymmetrical switch costs in the picture-naming task and narration task. Multilinguals who do not typically engage in codeswitching are expected to incur lower switch costs (i.e., pause less frequently and for shorter durations) in the single-language conditions of the narration task, yet they are expected to incur switch costs in the trilingual condition of the task. Accordingly, the present study adopts a multifaceted design which deconstructs 'multilingualism' and 'executive functions' while taking into account individual differences. It hypothesises that specific components of multilingualism and language experience map onto specific components of language control. By teasing apart both constructs, it aims to fill gaps in the literature about the individual processes involved in multilingual language production. The next chapter further details the study design and hypotheses.

2.6 CHAPTER SUMMARY

This chapter presented a review of research carried out in bilingualism and cognition research, defined key terms of constructs including 'executive functions' and 'codeswitching', and offered critical discussions of theoretical models which have been

put forth to explain the relationship between language control and executive functions. It also discussed factors which may influence performance on both cognitive and linguistic tasks, linking the above to the rationale and overarching hypotheses which concluded the chapter.

CHAPTER 3: METHODOLOGY

3.1 INTRODUCTION

Following the literature review and the rationale for the present study, this chapter presents the overall research design which includes the variables and research questions. The chapter continues with a description of the participants and the materials, providing a thorough discussion of each instrument developed, adapted, and/or used in the study. The chapter then discusses the overall data collection procedure, presents a summary of the pilot study, and lists the ethical considerations of this project.

3.2 RESEARCH DESIGN

This is an exploratory study which examines whether two 'mirror-groups' of Arabic-English-French multilingual speakers of varying language proficiencies, yet from the same sociolinguistic context, perform differently on cognitive and linguistic tasks. It also examines whether differences in language experience, such as self-reported codeswitching habits, modulate performance on both types of tasks. Combining measures of executive functions and switch costs (operationalised by response time/pausing and accuracy), the study adopts a purely quantitative approach. In so doing, it offers a platform through which it is possible to draw comparisons between the two multilingual groups, as well as within each group, on both linguistic and non-linguistic measures. Furthermore, the project uses a novel instrument – the trilingual narration task – whose quantitative analysis of pauses allows for parallels to be drawn with participants' response latencies on a single-word utterance task.

3.2.1 Variables

Table 3.1 lists the dependent, independent, and control variables in this study and provides a description of how each variable is measured.

Table 3.1 Description of Variables

Variable Type	Name	Operationalised As	Measured By
Dependent	Switch Costs	Response Latency	Reaction time, measured in milliseconds, in *one-word* utterances
		Fluency	Operationalised as (a) *pause frequency*, measured as number of pauses, and (b) *pause length*, measured in milliseconds, in *utterances exceeding one word*
		Accuracy	Producing the cued language, measured in number of errors
	Codeswitching Patterns	Switch Directionality	The direction of switches (Arabic-English, Arabic-French, English-Arabic, English-French, French-Arabic, and French-English)
	Domain-general Executive Performance	Inhibitory Control (Simon Effect)	Performance on the Simon Task (accuracy and reaction time)
		Monitoring Cost	Performance on the flanker task (accuracy and reaction times in low and high-monitoring conditions)
		Conflict Effect	Performance on the flanker task (accuracy and reaction times in all conditions)
		Working Memory	Performance on the Backward Digit-Span task
Independent	Immersion Type	Medium of Instruction	English (EMI) or French (FMI)
	Proficiency	Oral Proficiency Score	Performance on Oral Proficiency Tests in each language
		High-Proficiency Mono/Bi/Tri-lingualism	A proficiency score of 8 or above in one, two, or three languages
		Lexical Robustness[5]	Performance in each language on the Verbal Fluency Task (number of exemplars)
	Language Experience	Self-reported Language Use, Language Exposure,	Response from relevant Language Background Questionnaire items

[5] As indicated in Section 2.5.1.2, lexical robustness is included under the variable 'proficiency' due to the theoretical assumptions of previous studies, yet analyses performed on this construct and subsequent discussions make a clear distinction between oral proficiency scores and verbal fluency scores.

		Codeswitching Behaviour, and Salient Cue Detection	
Control	Data collection context	All data collection was carried out at the participants' homes	
	Socio-Economic Status	High (determined by school tuition and admission criteria)	
	Education	All participants are from the same school which they have attended since nursery	
	Language Exposure at School	All participants are enrolled in an immersion program	

3.2.2 Research Questions

Based on the research aims and hypotheses outlined in Section 2.5, five research questions (RQ) were formulated. Two results chapters address these questions: RQ1-2 are addressed in Chapter 4, and RQ3-5 are addressed in Chapter 5. The questions have not only been divided according to the underlying variables they measure, but also with respect to an overarching question for each set. The principal question in the first results chapter, Chapter 4, explores whether EMI and FMI students perform differently in domain-general executive tasks and whether their performance on each task is predicted by individual differences, namely language proficiency and language experience. The second results chapter, Chapter 5, focuses on whether the two groups differ in their linguistic performance and explores the predictors of language-task performance. Furthermore, it explores the relationship between cognitive and linguistic task performance. The following lists each RQ, along with the variables, measures, instruments, and analytical strategy, in tabular form.

The first research question investigates whether two multilingual groups who come from the same background—linguistic and otherwise, behave differently on executive

control tasks. As discussed in the literature review, studies which compare monolingual and bilingual groups have observed differences in performance on such tasks, reporting a bilingual advantage. This study hypothesises that the two groups of multilinguals may not behave similarly given the differences in their L3 proficiency whereby FMI are identified as 'balanced trilinguals' whereas EMI are 'unbalanced trilinguals'. The prediction is that the FMI group will outperform the EMI group, demonstrating a potential for a *trilingual* advantage. The first RQ is formulated as follows:

RQ1: Do EMI and FMI participants differ in performance across the domain-general cognitive tasks: (A) the flanker task, (B) the Simon task, and (C) the Backward Digit-Span task?

The first sub-question investigates performance on the flanker task and is presented as follows:

RQ1A: Do the two groups exhibit different accuracy rates and response latencies on the flanker task, including different monitoring costs and conflict effects?

Variable Type	Variable Name	Operationalised As	Analysis
Dependent	Accuracy	Number of Errors	Generalised Linear Mixed-Effects Model
Dependent	Reaction Time	Milliseconds (for correct responses only)	
Dependent	Monitoring Cost	(A) Difference in accuracy (high and low monitoring conditions) (B) Difference in RTs (high and low monitoring conditions)	
Dependent	Conflict Effect	(A) Difference in accuracy (congruent and incongruent trials) (B) Difference in RTs (congruent and incongruent trials)	
Independent	Monitoring Condition	low, medium, and high-monitoring conditions	
Independent	Congruence Condition	Congruent vs. incongruent trials	
Independent	Language Group	EMI or FMI	

The second part of RQ1 examines whether the two groups differ in performance on accuracy, response latencies, and Simon Effect scores on the Simon task. It is stated as follows:

RQ1B: Do the two groups exhibit different inhibitory control in the Simon Task?

Variable Type	Variable Name	Operationalised As	Analysis
Dependent	Accuracy	Number of Errors	Generalised Linear Mixed-Effects Model
Dependent	Reaction Time	Milliseconds (for correct responses only)	
Dependent	Simon Effect	Difference in RTs (congruent and incongruent trials)	
Independent	Congruence Condition	Congruent, Incongruent, or Control trials	
Independent	Language Group	EMI or FMI	

The third part of RQ1 examines whether the two groups differ in performance on the working memory task. It is stated as:

RQ1C: Do the two groups exhibit different working memory capacities as measured by the Backward Digit Span Task?

Variable Type	Variable Name	Operationalised As	Analysis
Dependent	Working Memory	BDST Score	Generalised Linear Model
Independent	Language Group	EMI or FMI	

Following on from RQ1A-C, the last part of the first research question explores whether the two groups' performance across the three cognitive tasks is correlated. In other words, it explores whether participants exhibiting a high working memory capacity would exhibit better inhibitory control by having a low conflict effect score on the flanker task and/or a low Simon effect score on the Simon Task. As many studies employ one or two of the cognitive measures used in this study, it would be of interest to

explore whether the different cognitive tasks tap into different control mechanisms which overlap or which are dissociated from one another. This sub-question is therefore formulated as follows:

RQ1D: Are participants' working memory scores, Simon Effect scores, and/or Conflict Effect scores correlated?

Variable Name	Operationalised As	Analysis
Working Memory	BDST Score	Correlations
Conflict Effect	Difference in RTs (congruent and incongruent trials)	
Simon Effect	Difference in RTs (congruent and incongruent trials)	

The second research question examines whether participants' proficiency in each language influences their cognitive performance and whether being highly proficient in one, two, or three languages predicts this performance. Another purported index of proficiency, the Verbal Fluency Task score for each language, is also explored as a potential predictor of cognitive task performance. The second part of RQ2 explores whether individual differences in participants' self-reported codeswitching behaviour, language use, and language exposure, i.e., 'language experience', influence performance on cognitive tasks. The second research question is stated as follows:

RQ2: Is EMI and FMI participants' performance on non-linguistic, domain general tasks modulated by their language proficiency and/or language experience?

The first part of RQ2 addresses whether performance on the flanker task, Simon task, and BDST is influenced by language proficiency. It is stated as:

RQ2A: Is EMI and FMI participants' performance on non-linguistic, domain general tasks modulated by their language proficiency?

Variable Type	Variable Name	Instrument	Operationalised As	Analysis
Dependent	Working Memory	BDST	BDST Score	Mixed Effects Models
Dependent	Conflict Effect	Flanker Task	Difference in RTs (congruent and incongruent trials)	
Dependent	Simon Effect	Simon Task	Difference in RTs (congruent and incongruent trials)	
Independent	Language Proficiency	OPT	Score (1-10) on OPT in each language	
		High-Proficiency Mono/Bi/Triling ualism (HPBT)	Being highly proficient in 2 vs. 3 languages (having scores over 8 in 1, 2, or 3 languages)	
		Lexical Robustness	VFT scores in each language	
Independent	Language Group		EMI or FMI	

The second part of RQ2 examines whether performance on the three cognitive tasks is influenced by the constructs underlying 'language experience'.

RQ2B: Is EMI and FMI participants' performance on non-linguistic, domain general tasks modulated by language experience?

Variable Type	Variable Name	Instrument	Operationalised As	Analysis
Dependent	Working Memory	BDST	BDST Score	Mixed Effects Models
Dependent	Conflict Effect	Flanker Task	Difference in RTs (congruent and incongruent trials)	
Dependent	Simon Effect	Simon Task	Difference in RTs (congruent and incongruent trials)	
Independent	Language Experience	Self-reported Language Use, Language Exposure, Codeswitching Frequency and Salient Cue Detection	Response from relevant Language Background Questionnaire items	
Independent	Language Group		EMI or FMI	

Following cognitive task analyses in Chapter 4, the second results chapter focuses on answering the research questions pertaining to switch costs and fluency which are measured by the trilingual picture-naming task and the trilingual narration task, respectively. The research questions that were formulated for this aspect of the study are presented next.

The literature suggests that the L1 is more strongly inhibited than the L2 or L3. However, this study's participants are simultaneous multilinguals, which makes language assignment rather elusive, especially for the L1 and L2. The two groups are believed to differ in their L3 proficiency, which is the basis of dividing the groups as 'balanced' and 'unbalanced' trilinguals. In order not to make false assumptions about which language/s is most dominant, however, and taking into account individual differences in proficiency levels and language use, the languages are examined without a pre-assignment of dominance or balance. The third question explores whether EMI and FMI participants exhibit variable inhibitory control levels of their different languages. In other words, it examines whether EMI or FMI participants exhibit symmetrical or asymmetrical switch costs during the picture-naming task. This analysis also includes an examination of preceding trials; in other words, it examines both the languages switched *into* and the languages switched *from*. Moreover, it examines whether EMI and FMI participants differ in fluency on the narration task and whether switch costs and pauses are correlated. The third research question is stated as follows:

RQ3: Do EMI and FMI participants differ in performance on the linguistic tasks: (A) the trilingual picture-naming and (B) the narration task?

The first part of RQ3 examines participants' performance on the picture-naming task, and is formulated as follows:

RQ3A. Do EMI and FMI participants differ in their performance on the picture-naming task in terms switch costs (accuracy and reaction times) for each language?

Variable Type	Variable Name	Operationalised As	Analysis
Dependent	Switch Costs	Accuracy, measured as a percentage of correct responses (Total – Number of Errors)	Generalised Linear Mixed Effects Models
		Reaction Time, measured in milliseconds (for correct responses only)	
Independent	Trial Type	Repetition vs. Switch	
Independent	Language	Target Language: Arabic, English, or French	
Independent	Prior Language	Preceding Language: Arabic, English, or French (i.e., indicating switch directionality)	
Independent	Language Group	EMI vs. FMI	

The second part of RQ3 examines fluency in the narration task which incorporated two linguistic contexts: (1) a single-language context in which participants were cued to produce narrations in Arabic, English, and French separately; and (2) a trilingual context in which participants were instructed to produce narrations using a mixture of the three languages. For the single-language condition, it was hypothesised that FMI participants would have fewer and shorter pauses than EMI participants when speaking in the weakest language. Pauses from the single-language condition were also contrasted with pauses from the trilingual condition, which is a more naturalistic linguistic context in which the participants of this study typically converse in social, though not academic, settings.

It was hypothesised that pauses in the trilingual condition would be fewer and shorter in duration for both groups compared to the single-language condition, as the trilingual condition is a context which allows opportunistic planning and language use, as

stipulated by the ACH and CPM. Accordingly, the second part of the third research question is formulated as follows:

RQ3B. Do EMI and FMI participants exhibit the same frequency and duration of pauses in the single-language and trilingual conditions of the narration task?

Variable Type	Variable Name	Operationalised As	Analysis
Dependent	Fluency	(a) Pause frequency (number of pauses, converted into ratios) (b) Pause duration, measured in milliseconds	Mixed Effects Models
Independent	Language Condition	Single-language condition (Arabic-only, English-only, and French-only) vs. Trilingual condition (Arabic-English-French)	
Independent	Language Group	EMI vs. FMI	

As a main contribution of the present project to this topic, the third part of this research question attempts to draw parallels between performance on the picture-naming task and on the narration task in the trilingual conditions. That is, this sub-question explores whether switch costs in single-word utterances map onto those in prolonged (monologic) utterances with respect to pausing and response latencies. Specifically, switch costs from the picture-naming task were contrasted with pauses from the narration task from both linguistic contexts both within the same language (analogous to repetition trials) and between languages from the trilingual context (analogous to switch trials). The third part of this question is formulated as follows:

RQ3C. Do EMI and FMI participants' pause durations in the narration task correlate with their RTs in the picture-naming task?

Variable Type	Variable Name	Instrument	Operationalised As	Analysis
Dependent	Reaction Time	TPNT	Milliseconds	Correlations
Dependent	Fluency	Narration Task	Milliseconds	
Independent	Language Condition	TPNT and Narration Task	(1) Repetition trials vs. Single-language condition (2) Switch trials vs. Trilingual-language condition	
Independent	Language Group		EMI or FMI	

Similar to cognitive performance, linguistic performance, measured as switch costs and fluency, on the trilingual picture-naming and narration tasks was predicted to be modulated by language proficiency and language experience. It was hypothesised that participants who typically codeswitch on a regular basis, have high lexical robustness scores, and/or are highly proficient in the three languages would exhibit symmetrical switch costs. The fourth research question, therefore, is formulated as follows:

RQ4. Do language proficiency and language experience predict performance within and across EMI and FMI participants in the picture-naming task and narration task?

The first part of the RQ4 explores language proficiency as a predictor variable of TPNT and NT task performance, and it is formulated as follows:

RQ4A: Does language proficiency influence switch costs in the TPNT and pause frequency and duration in the NT for the two groups?

Variable Type	Variable Name	Instrument	Operationalised As	Analysis
Dependent	Switch Costs	TPNT	Response latencies measured in milliseconds	Generalised Multilevel Model
Dependent	Pause Frequency	NT	Number of Pauses (converted to ratios)	
Dependent	Pause Duration	NT	Milliseconds	
Independent	Language	TPNT	Arabic, English, or French	
Independent	Language Condition	NT	Single-language vs. Trilingual	
Independent	Language Proficiency	OPT	Score (1-10) on OPT in each language	
		HPBT	High Proficiency in two vs. three languages	
		Lexical Robustness	VFT scores in each language	
Independent	Language Group		EMI or FMI	

The second part of RQ4 explored whether the constructs underlying self-reported language experience influence performance on the TPNT and NT. The question is formulated as follows:

RQ4B: Does language experience influence switch costs in the TPNT and pause frequency and duration in the NT for the two groups?

Variable Type	Variable Name	Instrument	Operationalised As	Analysis
Dependent	Switch Costs	TPNT	Response latencies measured in milliseconds	Mixed-Effects Models
Dependent	Pause Frequency	NT	Number of Pauses (converted to ratios)	
Dependent	Pause Duration	NT	Milliseconds	
Independent	Language	TPNT	Arabic, English, or French	
Independent	Language Condition	NT	Single-language vs. Trilingual	
Independent	Language Experience	Self-reported Language Use, Language Exposure, Codeswitching Frequency and Salient Cue Detection	Response from relevant Language Background Questionnaire items	
Independent	Language Group		EMI or FMI	

The final research question was designed to test some of the assumptions presented by the ICM, ACH, and CPM and to assess whether codeswitching in simulated naturalistic tasks (in the picture-naming and narration) can be examined in a comparable way to psycholinguistic tasks in experimental settings. It asks whether the participants' cognitive performance and codeswitching behaviour overlap across the two production tasks, in order to explore whether the underlying processes of executive and language control are shared. Specifically, this research question examined whether better inhibitory control capacities, conflict monitoring skills, and working memory capacity predict smaller switch costs in the picture-naming and narration tasks. The fifth research question in its two parts is therefore presented as:

RQ5. What is the relationship between cognitive mechanisms and temporal switching costs in the linguistic tasks?

RQ5A. Do inhibitory control capacities, conflict monitoring skills, and working memory capacity predict RT switch costs in the picture-naming task?

Variable Type	Variable Name	Instrument	Operationalised As	Analysis
Dependent	Switch Costs	TPNT	Response latencies (in milliseconds)	Linear Mixed-Effects Models
Independent	Language	TPNT	Arabic, English, or French	
Independent	Conflict Effect	Flanker Task	Difference in RTs (congruent and incongruent trials)	
Independent	Simon Effect	Simon Task	Difference in RTs (congruent and incongruent trials)	
Independent	Working Memory	BDST	BDST score	
Independent	Language Group	-	EMI or FMI	

RQ5B. Do inhibitory control capacities, conflict monitoring skills, and working memory capacity predict pause duration in the trilingual condition of the narration task?

Variable Type	Variable Name	Instrument	Operationalised As	Analysis
Dependent	Pause Duration	NT	Length of pauses (measured in milliseconds) in between- and within-language switches	General Mixed-Effects Models
Independent	Switch Directionality	NT	Six types of switches: Arabic-English, Arabic-French, English-Arabic, English-French, French-Arabic, and French-English	
Independent	Conflict Effect	Flanker Task	Difference in RTs (congruent and incongruent trials)	
Independent	Simon Effect	Simon Task	Difference in RTs (congruent and incongruent trials)	
Independent	Working Memory	BDST	BDST score	
Independent	Language Group	-	EMI or FMI	

3.3 SAMPLING FRAME

In addition to the implementation of a thorough research design and the use of appropriate instruments, rigorous research must employ suitable sampling procedures and ensure the adequacy of sample size. This section describes the sampling frame adopted in this project and outlines sample size calculations employed both before and after data collection.

3.3.1 Sampling Procedure

Sampling refers to the process of setting specific inclusion criteria to select a subset (i.e., a sample) of a population of interest. Depending on the aims of a study, sampling may be based on the premise of generalisation to the population or of an in-depth examination of a specific subgroup within the population. The sample, therefore, may not necessarily be representative of the characteristics of the entire population. While this

may preclude making generalisations, it allows for specific insights into minority or unrepresented groups in research. The present study focuses on a specific segment of Lebanese adolescent students which is not representative of the general population. Rather, the sample selected for this study is constrained to high socioeconomic class students who are privately educated at a selective school. As such, it is necessary to highlight that the conclusions drawn from the findings of this project may not be generalisable to students enrolled at public (or state) schools or at lower-ranking private schools in Lebanon. Nevertheless, the findings of the study may be applicable to other contexts (e.g., Singapore, India, United Arab Emirates, and Saudi Arabia), where scholastically high-achieving students are enrolled in immersion programmes at elite schools.

Due to safety precautions and accessibility restrictions, convenience sampling was adopted in the selection of the school from which all participants were recruited. Ranked as one of the top three schools in Lebanon, the target school is an independent school with two campuses in Lebanon. All communication and recruitment were done through the Beirut campus, thus ensuring that all participants were also residents of the same city. The school, which is highly selective in student admission and retention, is renowned for the exceptional performance of its students on Lebanese official exams (Lebanese Baccalaureate) and international standardised exams such as the Scholastic Aptitude Tests (SATs) and Le Diplôme Approfondi de Langue Française (DALF). Students in this school typically belong to upper-class families whose patriarchs and matriarchs include members of parliament, government officials, bank managers, and business leaders. High tuition fees and strict academic and admission/retention criteria render the student body rather uniform in itself yet easily distinguishable from the majority of those of other schools. That is, students from the present study's target school are exemplary students

who not only outperform their peers from other Lebanese schools on almost all official and standardised exams, but they also excel in their future academic studies. In 2018, the school's website reported that every student in its senior class was accepted to their first and/or second choice universities, both local and abroad. The universities listed were all top-ranking, prestigious institutions which are also known for their strict admission criteria.

Having recruited participants from this school for a previous study, the researcher had established ties with the administration, facilitating the data collection procedure. Furthermore, as an alumna of this school, the researcher was highly familiar with the characteristics of the student body and possessed specific knowledge of the immersion programmes, adding to the development of more precise hypotheses about the sample selected for the study. Specifically, the researcher was confident that the immersion languages were strictly adopted during class time with very minimal to no code-switching to other languages by the teachers during class time. Furthermore, students in this school also adhere to speaking in the immersion language except in classes where another language must be used (e.g., Arabic during Arabic lessons). This may be attributed to the school's language policy which is implemented since nursery, where children are required to use the appropriate language with their teachers, many of whom are monolingual teachers, mainly recruited from the United States and France. While objective measures of assessing language choice during class-time would have further consolidated these claims, the high familiarity of the researcher with both the school and its students offer credibility to the assumptions and hypotheses made in this project.

While the selection of the school was based on convenience, a criterion-sampling procedure was used to recruit participants. Students who had not been enrolled in the school throughout their lives, who were not born to Lebanese parents, or who spoke

additional languages were excluded from the study. All participants were 17 or 18 years of age and were right-handed with no reported visual, auditory, or mental impairments. They had not resided in any foreign country or travelled abroad longer than three weeks or taken extra-curricular language courses. A language-background questionnaire (Appendix A) enabled screening of the participants to ensure that the above selection criteria were met before the data collection procedure commenced.

3.3.2 Sample Size

In order to make inferences based on the sampled data, sample size must be adequate to conduct statistical analyses. The two strategies typically employed in determining sample size are to conduct statistical power analyses (1) *a priori* (i.e., before collecting the data, and/or (2) *a posteriori* (i.e., after collecting the data) (Mertler & Reinhart, 2016). The adequacy of the sample size in the present study was established by adopting both strategies. *A priori* power analyses were run through the G*Power software (version 3.1) (Faul, Erdfelder, Lang, & Buchner, 2007). Several analyses were run to ensure that the sample size for each group was large enough to conduct within- and between-group analyses. For instance, a total sample size of 36 (18 per group) was determined using an *a priori* power calculation for F-tests obtained from repeated measures ANOVA whose assumptions are analogous to those of linear mixed-effects models (Linck & Cunnings, 2015). The input parameters were set at (a) a large effect size ($\eta^2 = 0.25$), (b) an alpha level set at 0.05, (c) a power set at 0.95, (d) a group number of two, (e) a set of four measurements, (f) a correlation between repeated measures at 0.5, and (g) a non-sphericity correction at 1, and the output is presented in Table 3.2.

Table 3.2. A Priori Sample Size Calculations for Repeated Measures ANOVA (F-Tests)

Test	**F tests - ANOVA: Repeated measures, within-between interaction**				
Analysis	**A priori: Compute required sample size**				
Input	Effect size f	= 0.25	**Output**	Non-centrality parameter λ	=18.00
	α err prob	= 0.05		Critical F	=2.69
	Power (1-β err prob)	= 0.95		Numerator df	=3.00
	Number of groups	= 2		Denominator df	=102
	Number of measurements	= 4		Total sample size	=36
	Correlations among repeated measures	= 0.5		Actual power	=0.95
	Non-sphericity correction ε	= 1			

Another type of analysis for correlations was also run, such as correlations between Simon Effect scores and Conflict Effect scores. The power analysis output indicated that 21 participants were needed per group for an estimated effect size of 0.5 (large effect size) for a correlation analysis set at the 0.05 level, as depicted in Table 3.3. G*Power analyses were also conducted by entering estimated means and standard deviations of variables of interest, such as the Simon Effect scores, which were based on results from the pilot study (section 3.7). The effect sizes were computed to further confirm the sample sizes necessary to achieve sufficient statistical power. All analyses indicated that the sample size chosen, which was 60 participants (30 participants per group), was sufficient to perform statistical analyses with enough power. This was also later confirmed through *a posteriori* analyses (discussed in Chapters 4 and 5 when relevant) which further confirmed that the sample met the sample-size assumptions for every statistical test that was run.

Table 3.3 A Priori Sample Size Calculations for Correlations (T-Tests)

Test	**t tests - Correlation: Point biserial model**				
Analysis	A priori: Compute required sample size				
Input	Tail(s)	= 2	**Output**	Non-centrality parameter δ	= 3.74
	Effect size |ρ|	= 0.5		Critical t	= 2.02
	α err prob	= 0.05		Df	= 40
	Power (1-β err prob)	= 0.95		Total sample size	= 42
				Actual power	= 0.95

3.4 PARTICIPANTS

This study recruited 60 simultaneous multilinguals whose proficiency in their first two languages, Lebanese Arabic and English/French, is high, with intermediate- to high-proficiency levels in the L3 (French/English) (see Section 3.5.2 on Proficiency). The participants were recruited based on their enrolment at the same school from nursery till Year 12 in either an EMI or an FMI programme. A summary of the participants' profiles and language exposure is provided in Table 3.4. All participants were simultaneous multilinguals who were also simultaneously multiliterate.

Table 3.4 Participant Profile Summaries by Immersion Type

Means, (SDs), and % of participant characteristics and self-ratings[6] of language proficiency, language exposure, and language use

	EMI (N=30)						**FMI (N=30)**					
	Arabic		**English**		**French**		**Arabic**		**English**		**French**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Formal instruction age (years)	3	-	3	-	11	-	3	-	11	-	3	-
Formal Instruction (% of hours/week) in language and language-taught subjects	12%	-	82%	-	6%	-	12%	-	6%	-	82%	-
Exposure in school (outside class)	34.6	(14.02)	51.28	(12.68)	14.02	(10.79)	31.78	(13.64)	21.11	(9.99)	47.11	(15.03)
Exposure at home	58.92	(14.41)	37.28	(14.56)	3.79	(5.34)	62.87	(11.56)	17.30	(9.06)	19.83	(7.72)
Exposure through media	22.29	(14.21)	71.99	(17.75)	5.72	(7.15)	25.86	(16.17)	56.13	(18.06)	18.02	(11.69)
Language use at school (outside class)	41.71	(14.19)	51.40	(17.21)	6.89	(10.03)	43.88	(12.75)	32.01	(13.90)	24.10	(16.12)
Language use at home	53.77	(12.03)	42.86	(13.91)	3.37	(6.47)	70.46	(11.63)	12.41	(8.77)	17.12	(8.43)
Language use online	22.15	(12.38)	70.78	(15.91)	7.08	(8.10)	22.90	(12.83)	50.56	(16.92)	26.54	(11.59)
Language use in social contexts (excluding home, school, and online)	36.40	(19.39)	58.98	(21.36)	4.62	(10.94)	53.22	(25.24)	24.83	(20.59)	21.94	(10.27)
Self-reported proficiency (out of 10)	7.43	(1.07)	9.07	(0.64)	5.53	(1.61)	7.43	(1.22)	8.20	(1.09)	9.10	(0.712)

[6] All data from this table is obtained from the language background questionnaire which is discussed at length in Section 3.5.1.1 of this document.

The first group, consisting of 30 EMI participants, was taught all subjects, aside from Modern Standard Arabic and Social Studies, in English since the age of three. They received formal instruction in French, their L3, since age 11 (Year 6). This group was expected to codeswitch mainly between Arabic and English, given their extended exposure to and instruction in the two languages, with marginal use of French, as observed in Zantout (2014). Despite their exposure to French through society and schooling, it was expected that this group would only use French when cued, while English would be used most. Furthermore, it was hypothesised that participants in this group would struggle to produce long utterances in French, and perhaps even Arabic, without relying on the use of English, given that their strongest language was self-reported to be English, followed very closely by Arabic, and with French being a distant third language.

The second group comprised 30 FMI participants who were taught all subjects in French, except for Modern Standard Arabic and Social Studies, and had received formal instruction in English since age 11 (Year 6). This group was expected to exhibit more L3 use than their EMI counterparts given the ubiquity of English in the media and social settings. FMI participants were also believed to be able to sustain prolonged speech utterances in the three languages and any combination of the three when prompted, with minimal struggle when speaking only in the L3 (English). As no study had examined Lebanese FMI participants, it was hypothesised, given the researcher's familiarity with this sample and context, that this group was relatively balanced in the three languages and that their L3 proficiency level would be high, unlike EMI participants whose L3 proficiency may range between intermediate to high. This is further corroborated by participants' self-reported proficiency (and by the proficiency tests discussed in Chapter 4).

As mentioned earlier, given the ways in which the participants use their languages across different contexts, designating specific language labels as L1, L2, and L3 for the languages spoken by this sample—and perhaps this population—is both tenuous and variable. The profile summary of each group (Table 3.4) reveals that it is difficult to determine whether Arabic or the first 'foreign' language learned (i.e., English for EMI and French for FMI) is the first (or native) language. Despite the argument that language use and/or exposure (or as the literature at times refers to it as 'dominance') can determine the status of a language as a first or second language, it is not clear-cut in this sample whose social context is a trilingual one, yet whose academic context is an immersion one. Furthermore, whether the *third* language is indeed the last language formally taught is equivocal. For some EMI and FMI students, for instance, having equal proficiency across the three languages or perhaps having higher proficiency and self-reported use of the language which is taught in Year 6 officially makes the language assignment of 'L3' to the language taught at Year 6 inaccurate. Accordingly, individual differences in language proficiency, language use, and language exposure are also taken into consideration in the analysis. That is, analyses first examine the overall group distinctions and then look at individual differences irrespective of group membership when necessary.

3.5 MATERIALS

The following sections describe each task and its administration in detail. All tasks, which were piloted and re-piloted in some cases, were administered by the researcher, except the Oral Proficiency Tests, which were administered by three university-level language teachers. In terms of data coding and some analyses, the researcher and two of

the language teachers discussed and agreed on ratings, and intraclass correlations were used to determine inter-rater reliability.

3.5.1 Language Background Questionnaire (LBQ)

Participants first completed a language background questionnaire (Appendix A), allowing the researcher to screen the participants and collect information on language use and self-reported proficiency levels for all three languages. Data from the questionnaire, which included 106 items (147 including sub-items) and was based on a 5-point Likert scale, allowed for a more in-depth examination of self-reported language use, language exposure, and proficiency, and whether self-reported data matched the observed codeswitching patterns. The data from the LBQ thus provided information about *language experience*, offering further insights into individual variation in language use and code-switching, as well as other related variables such as salient cue detection.

Partly based on the questionnaire developed by McKendry and Murphy (2011), the present LBQ was divided into six sections: Section I allowed for screening and asked demographic questions, such as the gender and age of the participants, the languages they spoke, etc. Section II investigated participants' interpersonal language use and code-switching behaviour at home with their parents and siblings, with their peers inside and outside the school context, and in everyday interactions. Section III explored the participants' media language exposure and the frequency in which they received input in each language. Section IV explored participants' use of language and code-switching online, such as the language(s) in which doing homework and communicating with friends and family is used. Section V addressed participants' language choice, such as whether they preferred using a certain language in a specific context and whether they intentionally choose to speak in one language or code-switch. Section VI asked

participants to report on their opinions about language, especially with regards to the languages in the Lebanese context.

Participants were given the opportunity to choose between the English and the French versions of the questionnaire which was translated and back-translated by two simultaneous French-English bilingual doctoral students. Participants completed the questionnaires by hand and at their own pace. The researcher was present during this session but was seated at a distance from the participants. When asked, she provided clarifications on the rare occasions when participants were uncertain about a word or a question. However, the researcher did not provide any leading information that might have influenced participants' responses. This is one advantage to disseminating a paper-based questionnaire where the researcher is present to address questions, which is not possible with an online, remotely-filled questionnaire.

Although the first section of the LBQ provided the screening data, the researcher requested that the entire questionnaire be completed as she could not see the responses and screen the participants until after the questionnaire was handed in. As the participants were informed through the leaflet and word-of-mouth about the requirements, only one participant was excluded from the study as he had lived in Canada for the first five years of his life and was not enrolled in the school since nursery. The participant was informed at the end of the LBQ session that he did not meet the criteria and was thanked for his time and for volunteering to take part in the study. The remaining 60 participants who had met all screening criteria carried on with the remaining data-collection sessions.

3.5.2 Oral Proficiency Tests (OPTs)

Very few studies that examine codeswitching and use psycholinguistic tasks employ separate objective measures to test language proficiency. Generally, researchers

rely on self-reported language proficiency, yet this qualitative measure is highly dependent on participants' perception of their overall linguistic abilities. Some studies break down proficiency into the four skills (speaking, writing, reading, and listening), yet the scores remain variable and non-objective. This is a limitation which several studies report, and which the present study overcomes.

Oral proficiency in the three languages was assessed using OPTs for Lebanese Arabic, English, and French. These tests provided information about the individual participants and the groups they belonged to. As Lebanese Arabic is a spoken dialect with no standardised script, and as there are no speaking (or written) tests in this variety, a researcher-designed OPT (Appendix B) was developed, based primarily on the American Council on the Teaching of Foreign Languages Oral Proficiency Interviews (OPI). The test comprised two sections: (1) a set of questions which began with easy items and became more complex (e.g., Where did you grow up? How is it similar/different from 'City X'?); and (2) a set of items which provided speakers with an opportunity to produce more difficult constructions typical of educated native speakers (e.g., Some parents believe that teenagers should not be allowed to use social media till they are 18. Do you agree? How would you try to convince someone that your view is the right one?). Given the nature of this test, the three examiners were trained to adapt to the participants' answers and select from a variety of questions, depending on the observed performance of each participant, making the test semi-structured. Each test was audio-recorded and took approximately 7-18 minutes to complete.

As with all the tasks, the OPTs took place in the participants' homes, and the researcher was seated in a different room while each examiner and participant sat in a room to complete the task. The participants were not informed that this task would rate their proficiency, but they were informed that they would be asked a series of questions

to which they had to respond solely in the target language. They were also told that while the task set-up was similar to that of an interview, it was a very relaxed setting and that they did not need to provide personal information or discuss topics they felt were sensitive or private. The piloting of the OPT questions ensured that no questions enquired about such information or tackled such topics, and the OPT included hypothetical questions with a role-playing section which the participants in both the pilot and main studies found engaging.

Three versions of the OPT were developed and translated and back-translated to English and French by native speakers of each language for all three versions. The versions were counterbalanced to ensure that there was no overlap of tasks and questions between the languages to mitigate any test-related confounds. The examiner, the researcher and a second (external) examiner provided individual scores out of 10 based on the OPT scoring rubric (Appendix C). The scoring criteria allowed the raters to distinguish between sublevels (low-intermediate-high) of three categories (novice, intermediate, advanced) and 'superior'. In case of discrepancy between ratings, a consensus for a final score was reached through discussions. Inter-rater reliability was established through intra-class correlations; a discussion of the analysis is offered in Chapter 4 (Section 4.4.1). Scores from the OPTs were analysed for each group, further confirming the distinction between the two groups and the underlying assumption regarding the language proficiencies of EMI and FMI participants. OPT scores were also used as the measures of proficiency for each language, and they were fitted as potential predictors of both cognitive and linguistic task performance. The OPT scores also offered the opportunity to assign participants as 'highly proficient' in one, two, or three languages, having obtained scores of at least 8 in each OPT. Accordingly, whether being

'balanced' trilinguals or not conferred linguistic and cognitive benefits was investigated based on this categorisation.

3.5.3 Verbal Fluency Task

As a measure of lexical robustness, verbal fluency tasks are said to reveal cognitive executive function in bi/multilinguals through the speed of lexical retrieval (Costa et al., 2006; Linck et al., 2012). They are also said to correlate with participants' self-ratings of proficiency (Schwieter & Sunderman, 2008, 2009, 2011). The verbal fluency task involves accessing the mental lexicon for word retrieval in the cued language(s), and suppressing irrelevant words, while avoiding repetition, which involves cognitive processing. As such, this task taps into both verbal ability and executive control (Fisk & Sharp, 2004). However, the dual aim of the task implies that verbal fluency scores do not provide a 'pure' measure of either proficiency or processing performance (Shao, Janse, Visser, & Meyer, 2014). While many studies on codeswitching have examined VFT scores without the use of other tasks to control for both proficiency and processing, this study employed the VFT in conjunction with other tasks that assess proficiency and processing separately.

The task has variants based on the cue categories employed, such as *semantic categories* (e.g. vegetables), *formal single-letter categories* (e.g. words beginning with 'b'), and *double-letter categories* (e.g. words beginning with 'bi'). As formal categories have been reported as being slightly more difficult than semantic categories (Crowe, 1998), this study included an equal number of semantic and formal category cues (15 each) (adapted from Gollan, Montoya, & Werner, 2002, and Runnqvist, Strijkers, Alario, & Costa, 2012). The instructions were altered from 'words beginning with 'b'' to 'words *starting with the sound* 'b'' to account for Lebanese Arabic (LA) not having a standardised script. The task categories are depicted in Table 3.5. The categories were

piloted and repiloted to ensure that the formal categories elicited comparable numbers of items and that each sound was clearly distinguished from the others.

Table 3.5 Verbal Fluency Task Categories

Formal Categories	Semantic Categories
Words starting with the sound /m/	Things with wheels
Words starting with the sound /k/	Countries
Words starting with the sound /s/	Furniture
Words starting with the sound /z/	Sports
Words starting with the sound /æ/	Cities
Words starting with the sound /f/	Musical Instruments
Words starting with the sound /l/	Occupations
Words starting with the sound /d/	Vegetables
Words starting with the sound /j/	Electrical Appliances
Words starting with the sound /t/	Animals
Words starting with the sound /w/	Fruits
Words starting with the sound /ʤ/	Things at School
Words starting with the sound /b/	Clothes
Words starting with the sound /n/	Colours
Words starting with the sound /r/	Things in the Kitchen

This task was conducted in Lebanese Arabic, English, and French separately for each participant and administered as follows: for each language, five semantic categories and five first-sound categories were presented to participants verbally. Participants had to produce as many exemplars as possible within 60 seconds. To avoid language switching, categories were counterbalanced and blocked by language and then randomized for each participant. The total number of unique category exemplars produced by the participant in each language served as an indicator of lexical robustness.

Participants were randomly assigned to six groups with 5 EMI participants and 5 FMI participants per group. Before the task was administered, participants were given oral instructions about the task. The sound ‘o’ was used and the participants were asked to give exemplars as practice and were additionally given the exemplars *oval*, *orchestra*, and *orange* in English, *ovale*, *orchestre*, and *orange*, in French, and *orchestra*, *orange*, and *ossa* in Arabic. Participants were given samples of incorrect exemplars in each language which included proper names, Modern Standard Arabic words, repetitions (e.g.,

open, *opening*, and *opened*), different pronunciations such as 'Oliver', and incorrect pronunciation of words that would assign a word to another language rather than the target language. The researcher used a digital timer and participants could look at the clock if they wished. A beep announced the start of the recording, and a long beep indicated the end of the recording for each item. A 30-second break was given between the categories.

The responses were audio-recorded and rated by the researcher and two external raters as correct or incorrect. A lexical robustness score for each language was then calculated by adding all correct responses from all ten categories for each of the three languages. All repeated words and words produced in other languages (including Modern Standard Arabic) were considered incorrect and excluded from the score calculation. Data from this task were analysed using negative binomial regression models to examine both within- and between-group performance. The data from the VFT were also correlated with OPT scores to examine whether the two instruments overlapped in terms of assessing language proficiency. Finally, data from the VFT were used to predict group and/or participant performance on cognitive and linguistic task performance, allowing comparisons with previous studies.

3.5.4 The Flanker Task

Tasks which present conflicting stimuli such as the flanker task (Eriksen & Eriksen, 1974) are typically used to measure interference suppression. Each stimulus consists of a row of five chevron arrows (see Table 3.6), and participants are asked to use the right and left arrow buttons to identify the (left or right) orientation of the central arrow and ignore the distractor stimuli (flanking arrows). Flanking arrows on either side of the central arrow could be congruent (pointing in the same direction as the target, hence posing no conflict) or incongruent (pointing to the opposite direction of the target,

creating conflict). In incongruent trials, participants need to inhibit the distractor stimuli in order to select the central arrow's orientation, resulting in increased reaction times and reduced accuracy compared to congruent trials. In order to measure the conflict effect (i.e., inhibitory cost), reaction times and errors from congruent trials are subtracted from those of incongruent trials. The smaller the conflict effect, the greater the inhibitory control. In neutral trials, where the central arrow was flanked by diamonds, the flanking distractors provide no facilitation or interference to response selection, thus data from these trials is not usually analysed.

Table 3.6 Flanker Task Stimuli Conditions

◊ ◊ › ◊ ◊	Neutral Right
◊ ◊ ‹ ◊ ◊	Neutral Left
› › › › ›	Congruent Right
‹ ‹ ‹ ‹ ‹	Congruent Left
‹ ‹ › ‹ ‹	Incongruent Right
› › ‹ › ›	Incongruent Left

As discussed in Chapter 2 (Section 2.2.3.1), the flanker task is used to measure interference suppression, which constitutes part of inhibitory control. One advantage to this task is that it is a non-linguistic measure which allows the comparison of different groups of bi/multilinguals without conflating language proficiency (language processing demands) with accuracy and reaction time (nonverbal cognitive demands), such as with the Stroop Task. Furthermore, as it incurs a minimal working memory load due to the relative simplicity of its instructions and performance demands, the flanker task arguably measures inhibition more 'purely' than other measures such as the Stroop and Simon tasks which confound working memory and inhibitory efforts (Costa et al., 2008; Hofweber et al. 2016). Nevertheless, as highlighted earlier, it may be difficult to completely disentangle the different, yet overlapping, elements of cognitive control such

as working memory, conflict monitoring, and interference suppression. Due to its strong psychometric properties and the associated generally high replicability of findings from numerous studies employing it, however, the flanker task was adopted in the present project. Participants generally perform at ceiling in terms of accuracy due to the task's simplicity and large number of trials; thus, reaction times are what typically distinguish participant performance. This is especially true for blocks where congruent and incongruent trials are mixed at relatively equal proportions.

While the flanker task has typically adopted a fixed ratio of congruent to incongruent trials (generally 75:25), recent studies have advocated modifying the task to reflect the different monitoring contexts which bilinguals encounter. Pioneering work in looking at how different ratios of (in)congruency incur low- to high-conflict demands is that of Costa et al. (2009). As detailed in Chapter 2, in a low-monitoring context (92% congruent vs. 8% incongruent trials), participants are said to require less conflict-monitoring thus providing faster responses, whereas in a high-monitoring context (50% congruent vs. 50% incongruent trials), it is observed that responses are slower due to the increased demands on executive control mechanisms and mental flexibility (Table 3.7). The monitoring cost is then computed by subtracting overall RTs in the low-monitoring block from the high-monitoring block. A smaller monitoring cost thus indicates better conflict-monitoring skills. Contrary to the high-monitoring context where trial switching is high and almost balanced, however, low-monitoring contexts present the most difficulty in terms of inhibitory control, as the number of incongruent trials is minimal, reducing the need for distractor monitoring and inhibition. As such, it is more difficult and time-consuming for participants to re-activate their schemata on these trials. Thus, smaller RTs in the low-monitoring block indicate better ability to activate inhibitory skills.

Table 3.7 Flanker Task Block Congruency Summary

Block	Congruent Trials (%)	Incongruent Trials (%)	Monitoring Cost	Inhibitory Load
1	92	8	Low	High
2	75	25	Medium	Medium
3	50	50	High	Low

Unlike Costa et al. (2009), yet similarly to Hofweber et al. (2016), the three blocks were presented to the participants in the same session. However, unlike either of these two studies, participants consisted of two groups of multilinguals as opposed to bilingual and monolingual groups in the former study and two bilingual groups in the latter. Furthermore, the blocks were randomised and counter-balanced. While the other studies did not present the blocks in order of increasing demands to monitoring skills, it is difficult to ascertain whether the order in which they were presented (low-, high-, and then medium-monitoring) may have had an effect on the overall performance in these tasks. Moreover, Costa et al. (2009) noted that the monitoring cost was actually highest in the 75-25 trials rather than the 50-50 trials, unlike their predictions. Thus, the block sequence which they had selected may have yielded the unexpected results given their allocation of the 'medium-monitoring' block as the last block.

Given the above, in this study, participants were randomly allocated to groups as described in Table 3.8:

Table 3.8 Flanker Task Block Sequence Counterbalancing Summary

	First Block	Second Block	Third Block	Block Sequence
Group 1	75-25	50-50	92-8	2, 3, 1
Group 2	75-25	92-8	50-50	2, 1, 3
Group 3	50-50	75-25	92-8	3, 2, 1
Group 4	50-50	92-8	75-25	3, 1, 2
Group 5	92-8	75-25	50-50	1, 2, 3
Group 6	92-8	50-50	75-25	1, 3, 3

Participants were seated approximately 60 cm from the researcher's computer screen and were given both verbal and written instruction prior to the task. Participants completed the flanker task, which was created using SuperLab (software version 4.5), on a Sony Vaio laptop, where stimuli were presented via a 15.5-inch monitor and response pads of Cedrus were used. The screen was white, and the chevron arrows were black, 20mm in size each. The response pads were covered by a black sticker with one white right-pointing and one white left-pointing arrow (→ and ←).

Following the experimental design adopted by Costa et al. (2009) (see Figure 3.1), each trial began with a fixation point at the centre of the screen for 400 milliseconds, followed by a single row of five arrows which disappeared after 1000ms or when a response was given. After the disappearance of the stimulus, a blank screen for an additional 500ms (or until a key was pressed) was presented, thus allowing participants to provide a response from the onset of the stimulus presentation till 1500ms. After the 1500 milliseconds, an inter-trial interval was triggered for 200-1500 milliseconds. Inter-trial intervals were jittered to avoid a monotonous response rhythm. Participants were instructed to respond as quickly and accurately as possible. The task began with 12 practice trials, followed by three blocks of 96 trials each.

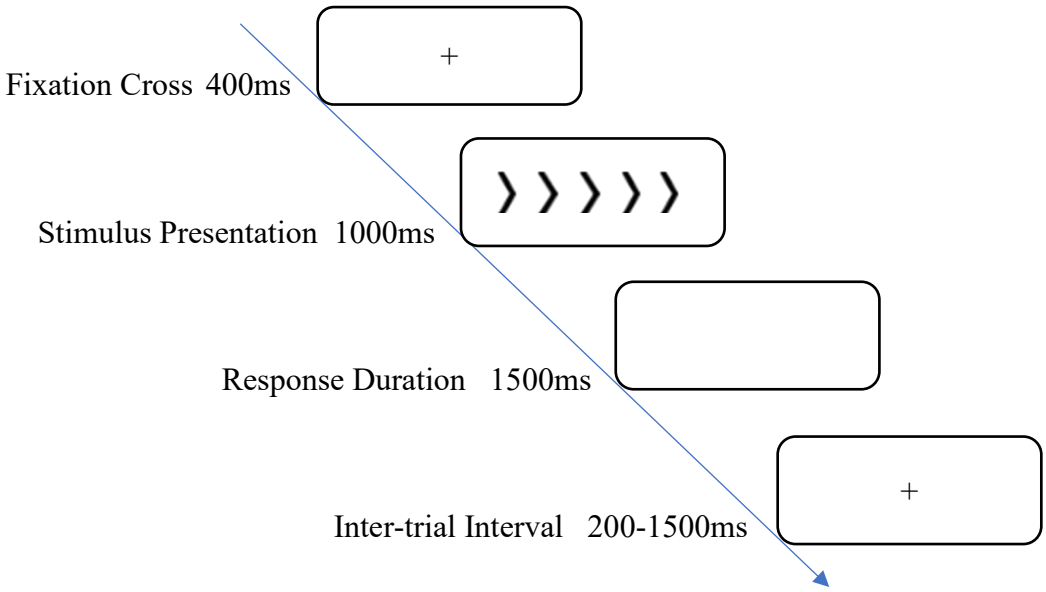


Figure 3.1 Schematic representation of the events in an experimental trial of the flanker task

3.5.5 The Arrow Simon Task

Commonly used in bilingual studies and more recently trilingual studies, such as Blumenfeld and Marian (2014) and Tse and Altarriba (2014), the arrow Simon Task, a version of Simon and Ruddell's (1967) Simon Task, is an experimental paradigm based on stimulus-response compatibility that allows the measurement of domain-general inhibitory control, In this version, the stimuli consist of left- or right-oriented arrows, whose position on the screen either matches or mismatches the position of the response keys (left or right). When there is a mismatch, a conflict is created, resulting in lower accuracy rates and slower reaction times. For example, a right-pointing arrow appearing on the left side of a screen creates conflict between two overlapping perceptual dimensions (right direction vs. left location). This conflict, which requires inhibitory control to be resolved, is analogous to the conflict elicited by the Stroop Task (reading 'red' written in a green font colour) (Bialystok, 2006). However, and as mentioned in the flanker task section, with other inhibitory control tasks, e.g. a multilingual Stroop task, it is difficult to disentangle linguistic interference and overcome the different-script

confound. Hence, using a task which taps into the same attentional constructs as the Stroop task without conflating language and executive performance makes it possible to compare the findings of the present study to those of previous studies in which the Stroop task is used.

Another advantage to using the arrow Simon Task as opposed to the traditional (standard) colour-square Simon Task is that the original version may not be as sensitive in detecting inhibitory control variability among proficient multilingual participants. In a study comparing the traditional and arrow versions of the Simon task, Bialystok (2006) reported that the arrow Simon task elicits greater perceptual conflict and longer response latencies. As this study involved two groups of multilingual speakers, the selection of the more complex version of the Simon Task increased the likelihood of detecting the potential effects of proficiency across the three languages on inhibitory control, as demonstrated by Blumenfeld and Marian (2014).

The Arrow Simon Task, also termed the Spatial Arrow Task and the Spatial Simon Task (Hilchey & Klein, 2011), includes two stimulus dimensions (direction and position) and a response dimension which overlap (right and left). In *congruent* trials, the stimulus direction and position (e.g., a right-pointing arrow is placed on the right of the screen) overlap with the response, where participants press a button on the righthand side. In *incongruent* trials, the stimulus, e.g. a right-pointing arrow, is placed on the left-hand side of the screen, and participants had to press the red (right-hand) button to indicate directionality while ignoring the position of the arrow on the screen. When an arrow is placed in the centre of the screen, it is a *control* or neutral trial and serves as a baseline. This is depicted in Figure 3.2.

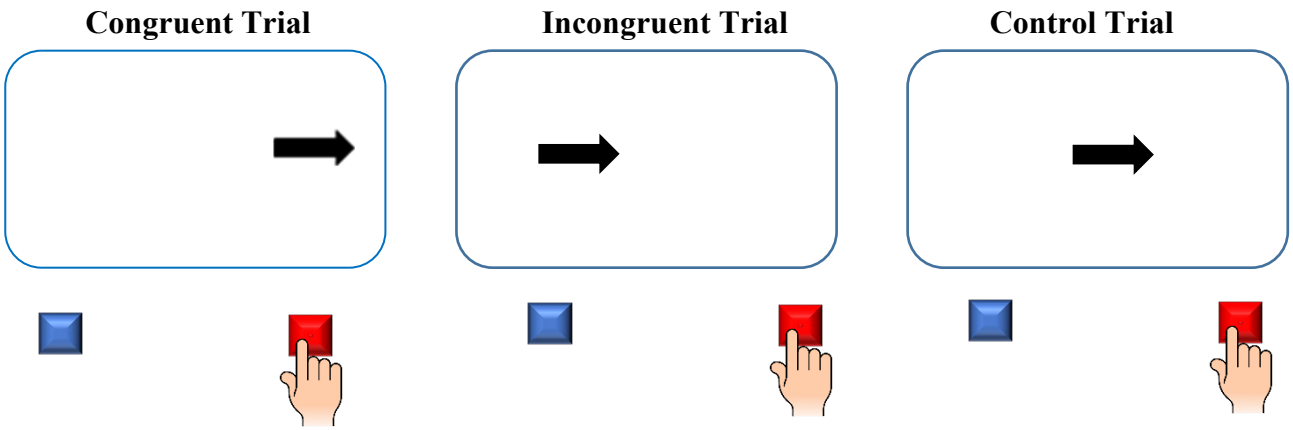


Figure 3.2 Arrow Simon Task Trials

The arrow Simon task taps into participants' ability to (1) maintain the instructions throughout the task in their working memory, (2) monitor the immediate demand for a switch from a congruent or incongruent trial, (3) suppress irrelevant stimulus dimensions, (4) activate task schemata that fulfil the response requirement, (5) suppress irrelevant task schemata that were activated on previous trials if the following trial was a switch, (6) reactivate the relevant task schemata, and (7) reconfigure the stimulus-response mapping by providing the correct response (Hilchey & Klein, 2011). These processes therefore indicate that the Simon task, like the flanker task, involves WM and conflict monitoring while measuring interference suppression. As discussed in Chapter 2, Miyake et al.'s notion of the 'unity and diversity' of executive functions reveals that the processes may not be entirely separable. However, tasks such as the Simon task which measure interference suppression differ from tasks measuring other elements, such as WM, despite partially sharing an underlying mechanism. Admittedly, it is precarious to claim that tasks measuring one element of executive functions are 'pure' measures of that element alone. Nevertheless, task design, validity, and replicability lend support to the use of these tasks in measuring the different elements of executive functions.

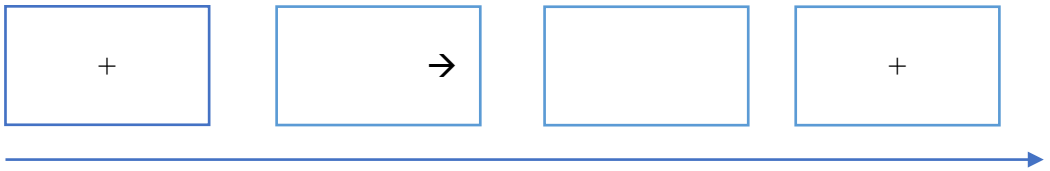
In the Simon task, the degree of the difference in response times on congruent and incongruent trials is labelled 'the Simon effect' and is seen as an indicator of one's ability 'to inhibit the prepotent tendency to respond based on the (task-irrelevant) location of the stimulus' (Linck et al., 2012: 655). Typically, a response on incongruent trials is 20-30ms longer than on congruent trials in the standard Simon task (Lu & Proctor, 1995). To compute the Simon effect, the difference between mean reaction times (RTs) on incongruent trials and the mean RTs on congruent trials is calculated. A smaller Simon effect indicates greater interference suppression abilities. As the arrow version of the Simon Task is used, the Simon effect may be attributed to the (1) stimulus-stimulus conflict (i.e., conflict between two dimensions of the same stimulus); (2) stimulus-response conflict (i.e., conflict between stimulus dimension and response dimension); or both (Tse & Altarriba, 2014). Studies such as Bialystok et al. (2008) argue that bilingual advantages are more likely to occur in tasks that involve both types of conflict. Although the control trials, where the arrow appears at the centre of the screen, are not included in computing the Simon effect, they allow the researcher to tease apart facilitation due to location-direction congruency and interference due to location-direction incongruency that may contribute to the Simon effect. This is because the central position of the arrow does not require additional processing and can be used as a baseline to establish response speed. This is a shortcoming which other studies such as Tse and Altarriba (2014) failed to overcome, yet which is addressed in the current project.

In the present study, congruent and incongruent trials were divided in equal proportion (50%-50%) and were randomly mixed; this was advocated by Costa et al. (2009) and Bialystok (2006) as a means of increasing task sensitivity to the detect the role of monitoring. Accordingly, the congruency was parametrically manipulated such

that there were 104 congruent trials with arrows pointing left, 104 trials with arrows pointing right, 104 incongruent trials with arrows pointing left, and 104 incongruent trials with arrows pointing right. Control trials were divided into 104 central arrows pointing left and 104 pointing right. The total number of trials was therefore 624 in total.

Participants were seated in front of the researcher's Sony Vaio computer's 15.5-inch screen (with SuperLab software version 4.5) and given verbal instructions about the task. They were asked to respond as quickly and as accurately as possible when they viewed the stimulus. Participants were instructed to ignore the arrow's location on the screen and respond according to the arrow's direction, irrespective of its position, by pressing either the red or blue key in the response box. Following the experimental design adopted by Bialystok and her colleagues (2004; 2005), each trial began with a white fixation cross (+) in the centre of the screen for 150ms (see Figure 3.3). Participants were then presented with the arrow stimuli whose peripheral locations (left and right) were situated 5º on the horizontal plane from the central fixation. The white arrow stimuli measured 40mm in length and 20mm in height, against a black background. The stimuli appeared for 400ms on the left or right side of the screen for the experimental trials and in the centre for the control trials. Participants were given a 1200-ms blank interval (from the onset of the visual stimulus) during which they could provide a response. After the 1200 milliseconds, an inter-trial interval was triggered for 200-1500 milliseconds. Each recording block comprised 52 trials, with the first four excluded to avoid primacy effects. Eight blocks of the experimental trials (Simon Task) and four blocks of the control condition (central stimulus) were alternately tested using pairs of experimental blocks interspersed between the controls and a short rest (6800ms) between each block. The mapping of the stimulus (direction) to response key was counterbalanced across participants within and across each group. Before the trials

began, participants were given eight practice trials. The task took approximately 15-20 minutes to complete.



Fixation Cross	*Stimulus Presentation*	*Response Duration*	*Inter-Trial Interval*
150ms	400ms	1200ms	200-1500ms

Figure 3.3 Schematic representation of components of a congruent trial in the Arrow Simon Task

3.5.6 Backward Digit-Span Task

As the cognitive processes underlying codeswitching have been observed to correlate with switch costs (e.g., Linck et al., 2012), it was important to examine the relationship between Working Memory (WM) and codeswitching. Participants with varying levels of proficiency and trilingual competences may differ in their codeswitching performance in relation to WM. The backward digit-span task provides a non-linguistic, domain-general processing measure which taps into an aspect of executive functions, updating, that differs from IC. However, as indicated earlier, an overlap of some of the mechanisms underlying IC and WM has been observed and are therefore acknowledged.

As WM is a complex construct which involves several components, it is measured by a variety of tasks as discussed in Chapter 2 (Section 2.2.4). The present project adopted the use of a complex span task – the backward digit span task (BDST), which is a widely used measure of WM. Researchers such as Olsthoorn, Andringa and Hulstijn (2014) argue that while the simple forward span task may reflect recall (storage), the BDST reflects processing in WM. Furthermore, they advocate the use of

numerical stimuli which have been demonstrated to be reliable items when measuring verbal WM by multilingual speakers. This is due to overcoming the language confound which may influence performance on this task. That is, multilingual speakers of varying language proficiencies may perform differently (partly) due to the target language of the WM task, as opposed to their WM capacity only.

Participants in this study were visually presented with a series of digits (stimuli) of increasing length from 3-9 digits, which they were asked to verbally repeat in their immersion language while reversing the order of the digits. Each length was presented twice as two trials, with each trial only appearing once. A participant's span is defined as the last length at which one or both lists from a series of a particular length were recalled correctly. The numbers (0→9) were presented in 75-point Arial font in blue on the centre of a white computer screen for a duration of 1000ms each, with no blank screen in between the items of a list. Using slides from a PowerPoint presentation, the participants moved to the next length unless stopped by the researcher to signal the end of the task either due to the completion of the entire task or due to making errors on both trials of the same span length. The researcher was present during the task but sat behind each participant at a distance. She marked the series that were correctly given and asked the participants to stop after failing to produce a length correctly after two trials. Before starting the testing session, the participants were familiarised with the procedure, and testing commenced only after the participants demonstrated understanding on four practice trials using two three-digit lengths and two four-digit lengths. The task took between 5-8 minutes to complete, depending on the total length that each participant reached. For each length span, the participants receive a 1 (Pass) or 0 (Fail) (see Table 3.9). A total raw score of the longest span score was therefore computed. The participants received a score of 0 for failing to produce any trial on the first block length

correctly to 14 for correctly producing the right order on both trials for every span length.

Table 3.9 Backward Digit Span Task

Span Length	Trial 1	Pass=1, Fail=0	Trial 2	Pass=1, Fail=0
3	172		849	
4	6287		3524	
5	41372		27058	
6	735149		589421	
7	1427365		9165382	
8	29034715		41729568	
9	351930748		729856413	

3.5.7 Trilingual Picture-Naming Task (TPNT)

Commonly used in bilingual studies (e.g., Costa & Santesteban, 2004; Verhoef, et al., 2009), the picture-naming task measures switch costs - the time it takes to switch to a different language than the language used in the previous turn. Switch costs are typically measured by response time (in milliseconds) and accuracy (calculated as error rate) in producing the target language(s). The task presents participants with visual stimuli (typically drawings of common objects) which they are asked to name in the language set by the researcher. The stimuli are either in switch (L1-to-L2; L2-to-L1) and non-switch trial conditions (L1-L1; L2-L2). As a variant of the picture-naming task, the trilingual picture-naming task in the current project measured response time and accuracy to look at switch costs for the three languages and whether/how they varied according to language proficiency and language experience. However, unlike other studies that have added a third response language to the bilingual picture-naming task (e.g., Schwieter & Sunderman, 2011), the present study avoided using some of the pictures typically adopted from Snodgrass and Vanderwart (1980) in order to avoid the use of cognates (e.g. English ‘table’ and French ‘table’) and phonologically similar-sounding items (e.g. English ‘cat’ and French ‘chat/chatte’). The items used in this study

were therefore black and white line drawings of a pencil, house, car, dog, book, bear, and heart (taken from the standardised list) and belt, star, and sun replacing *cat*, *chair*, and *table* from the original list. The modified list was piloted to verify whether the linguistic repertoires of the main-study participants comprise the selected words (which are also high-frequency words in English and French; no frequency lists are available for Lebanese Arabic). There was equal production of all three languages in the experiment (based on a 100% accuracy rate).

Participants performed this task on the researcher's computer and in front of a microphone which recorded the verbal responses. Participants were verbally instructed to name the pictures in Lebanese Arabic, English, or French, when presented with yellow, red, and blue background colours, respectively. An additional cue, as used by Festman and Mosca (2016), was an icon of the Lebanese, American, or French flag which was presented in the top right corner of the screen. The use of the double cue is recommended by researchers such as Logan and Bundesen (2003) who observed that single cues may incur a processing advantage on non-switch trials, leading to priming effects, and resulting in longer RTs on switch trials.

In this task, participants were presented with individual black-and-white pencil drawings (size 198 x 283 pixels) of the 10 selected pictures on the computer screen in 60 lists ranging from 5-14 pictures in length. Each list began with a fixation point (+) which was presented in the centre of the screen on a white background for 500ms (see Figure 3.4). Then a picture appeared on a red, blue, or yellow background for 2000ms or until the participant spoke into the microphone. Another fixation point (+) then appeared on a white background for 500ms, followed by another picture on a coloured background in either a switch (where the trial is named in language different from the previous trial) or a non-switch trial (where the previous trial is named in the same language). This was

repeated until the list ended, indicated by an asterisk for 1000ms, which offered a brief break before commencing the next list.

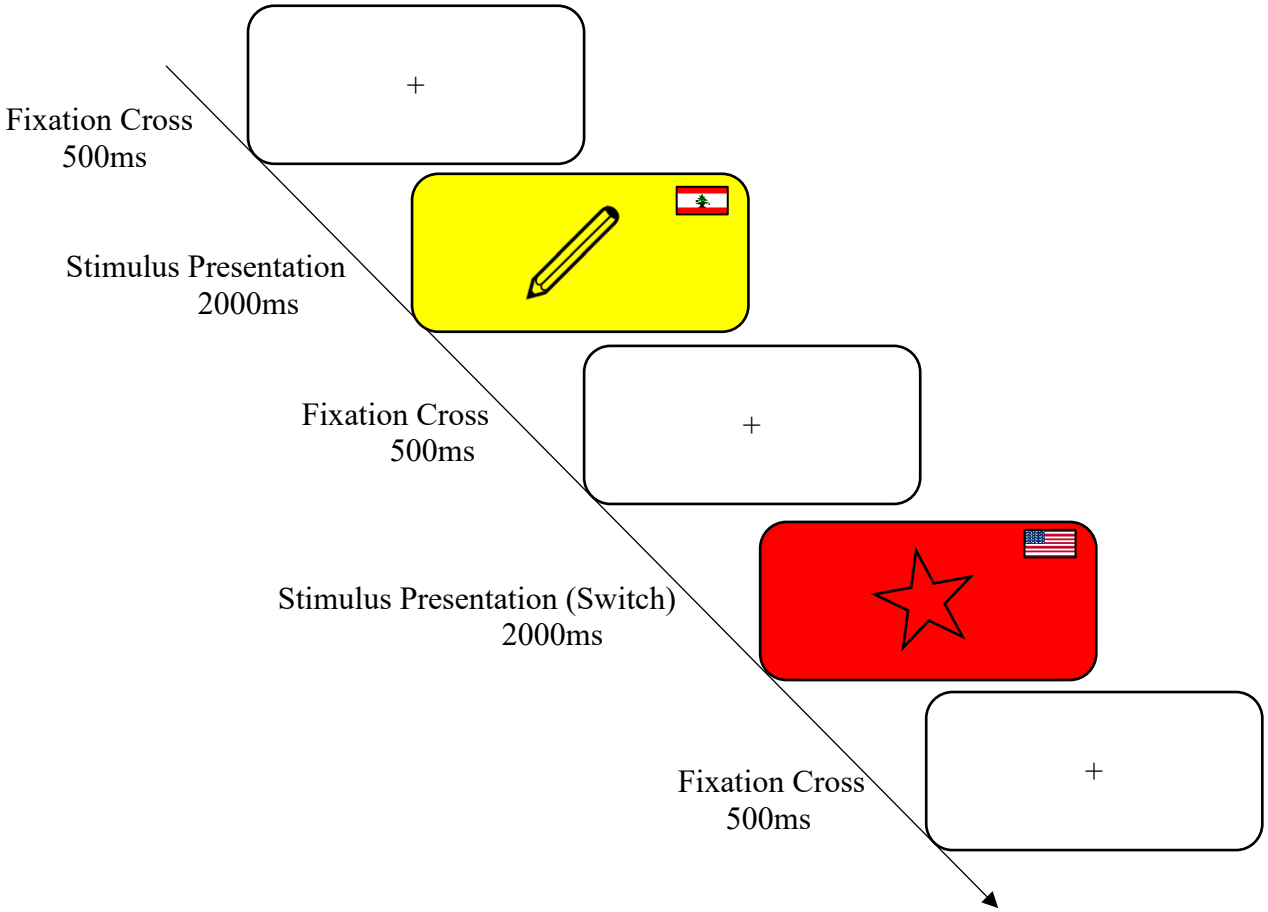


Figure 3.4 Schematic representation of a switch trial event from Lebanese Arabic to English in the trilingual picture-naming task

Within the lists, each trial was classified as a non-switch (or repetition) or a switch trial. Following Linck et al. (2012), each of the 10 pictures appeared 60 times, where 42 (70%) were non-switch trials and 18 (30%) were switch trials. Switch trials were equally distributed across the six possible language combinations (i.e., English-Arabic; English-French; French-Arabic; French-English; Arabic-English; and Arabic-French), resulting in 30 switches per type. Each list included 0-4 switch trials, and for lists with series of 5-10 pictures, no picture appeared twice. For lists with 11-14 pictures,

the repeated pictures were placed at least three trials away from their first presentation. Each picture appeared once in each of the possible languages. Instructions were given verbally in a combination of Arabic, English, and French, in order to avoid priming in one language, and in order to match the demands of the task.

All responses were timed by the computer using SuperLab (version 4.5) and recorded and coded for accuracy by the researcher and two external raters. Inaccurate trials and trials with RTs faster than 200ms or slower than 2000ms were excluded from the analysis. This was due to the assumption that the minimum reaction time required to process the cue and stimulus and trigger a verbal response exceeds 200ms. Therefore, a response provided in less than 200ms is considered a case of microphone mis-triggering, such as in the case of a cough, and was excluded from the analysis. Responses exceeding 2000ms were also excluded as they were considered a lapse in a participant's concentration. Other excluded responses included cases such as hesitations, mispronunciations, microphone mis-triggering, providing wrong words, or speaking in the non-target language. The participants practiced on six lists, each presenting the 10 pictures, to familiarise the participants with the pictures and cues (colour codes and flags) prior to the experimental lists. The participants received a one-minute break after the practice list and then started the experiment. Upon the completion of every 10 lists, a five-minute break was offered to avoid fatigue effects.

3.5.8 Narration Task

As mentioned in Section 2.3.2.2, fluency can be used to measure codeswitching in utterances exceeding one word. One way of measuring fluency is through administering speech elicitation tasks. However, these tasks have not been standardised, and due to the specific nature of the tasks influencing speakers' L2 fluency, it is difficult to reconcile discrepant findings (Segalowitz, 2010:42). One of tasks which is

recommended to elicit speech is picture stories or wordless cartoons with one or several frames which require participants to describe the storylines (e.g., Tavakoli & Foster, 2008; Tavakoli & Skehan, 2005). The advantage of using such narration tasks is that they constrain the topics about which the participants may talk, thus reducing the confounds of topic familiarity and variation. That is, such tasks require all participants to discuss the same topic, which is typically highly common and almost equally familiar. If topic selection were participant-based, speakers may speak only about topics they find easy which would then mask any potential dysfluency. Furthermore, unlike conversations, these tasks are not influenced by the researcher's choice of words in previous turns or even sentence constructions which participants might mimic. Narration tasks also do not incur demands on memory such as the case with tasks when participants are instructed to recall and describe previous events.

Although there are no clear criteria in terms of creating or selecting pictures and storylines, Rossiter, Derwing, and Jones (2008) recommend numerous suggestions to ensure there is no bias when narration tasks are administered to different participants and samples. Such criteria include ensuring the universality of the themes, avoiding culture-specific visuals such as body language images, and ensuring clarity of the storyline by avoiding flashbacks and flashforwards. It is imperative that such tasks are pilot-tested on a sample from the target population before administering them to the participants of the main study. The present study took the aforementioned into consideration.

While the trilingual picture-naming task elicits one-word utterances, the narration task was developed to elicit utterances (narrations) ranging between five to 12 minutes. The task was administered in single-language contexts which specified the language to be used or a trilingual context which required participants to codeswitch using the three

languages in a voluntary manner. Pauses in the single-language and trilingual linguistic contexts were used to measure the temporal latencies incurred when speaking in one language and/or when switching from one language to another. In line with psycholinguistic task analyses, therefore, pauses in the narration task may be perceived as being equivalent to reaction times in the picture-naming task.

Adopting Gardner-Chloros et al.'s (2013) measurement of pausing, and as suggested by the ACH and CPM, pauses were expected to provide a macrolinguistic indicator of codeswitching. Pauses were measured as silences greater than 250ms, as recommended by Huensch and Tracy-Ventura (2017) and Wright and Tavakoli (personal communication, 16/10/2015) and were analysed as intra-language pauses or inter-language pauses (in this case, switches) to determine a 'conversational' cost. Silent and filled pauses were included in the analysis, and performance on single-language and trilingual language contexts was compared. The details of the analyses are further elaborated upon in Chapter 5.

In the narration task, participants were shown a series of pictures (storyboard and narration task transcription samples are offered in Appendix D) lined vertically to avoid unintentional language cuing, given the opposite direction of the Arabic and English/French scripts. Participants were verbally instructed to tell the stories in as much detail as they could in the language assigned by the researcher. Each participant produced three monolingual, or single-language narrations (i.e., Arabic-only, English-only, and French-only) and a trilingual narration which combines the three languages. Each storyboard, printed in A0 size and consisted of 18 pictures. In the trilingual condition, participants were cued to switch between three languages, yet the frequency of switches and amount of languages used was not specified and was expected to vary among participants given their language proficiency and use.

Similar to the picture-naming task, the pictures appeared in front of coloured backgrounds indicating monolingual language choice, i.e. yellow for Arabic, red for English, and blue for French. The language was also written on top of the posters in the respective script, as a reminder of which language to use. In the language-combination trials, the languages that were assigned were written next to each other with the background colours split vertically on the posters. The tasks were counterbalanced across conditions and participants. After each storyboard, a three-minute break was offered. The researcher transcribed the audio-recorded data from the narration task to measure switch directionality, and this was checked by an external rater. Pause frequency and duration were measured and analysed using the Praat software.

3.6 OVERALL DATA COLLECTION PROCEDURE

Participants were recruited from one school in Beirut, Lebanon. The researcher's familiarity with the school and its student body, as well as with the school's high enrolment rate, allowed access to a large number of Year-12 students. Following ethical procedures outlined in Sections 3.8, the researcher provided an information sheet and consent forms (Appendix E) to all students enrolled in Year 12. Additionally, the researcher provided a brief explanation of the study at the school and answered questions raised by potential participants or their parents by e-mail or phone. The researcher asked the school to collect the consent forms after one week from the date of distribution. The experiment was conducted over a period of 2.5 months (nine sessions) at each participant's home. A breakdown of the data collection schedule is provided in Appendix F. Each testing session lasted approximately one hour, and breaks were provided within that hour, to minimise fatigue.

Before commencing any task, participants received verbal instructions in a Lebanese Arabic-English-French combination, as well as practice sessions to familiarise them with the task and reduce potential anxiety. Some tasks also provided written instructions on the computer screen. All participants completed the LBQ first, to verify that the participants met the recruitment criteria. The other tasks were administered in different orders to avoid task-order bias. Accordingly, six counterbalancing schemes (Appendix F) were created to systematically vary the order of task administration. Each scheme was used with five participants per immersion type.

3.7 PILOT STUDY

A *pilot study*, an 'abridged version' of the main study, is typically conducted in behavioural studies to assess the quality of the research design and to test and amend the instruments developed and used (Cohen, Manion, & Morrison, 2007). To pilot the study's instruments, the researcher recruited 21 (11 EMI and 10 FMI) participants from the same target school. The participants had just completed their academic year (Year 12) and were therefore of the same age group and grade as the target sample. Upon receiving ethical clearance and the assent and consent of the adolescents and their parents, the data collection procedure designed for the main study was followed. Data were collected over the course of three weeks (July-August 2015) as an initial phase, followed by a second phase for re-piloting three tasks with four new additional participants (2 EMI, 2 FMI) in September 2015.

Based on the outcome of the initial phase of the pilot study, two amendments were made to the Verbal Fluency Task, whereas one modification was made to the Narration Task. The following table (Table 3.10) summarises the issues that arose in the first pilot data collection phases and how these were addressed through modifications.

Table 3.10 Summary of Pilot Study Modifications

Task Name	Problematic Items	Modification
Verbal Fluency Task	Standardised categories, such as 'words starting with 'a' and 'e'', were problematic, as these two vowels represent the same letter and sound in Arabic, but they are different in English/French. Phrasing of 'words starting with' (a letter) elicited MSA words, as LA does not have a standardised script.	A non-standardised letter using the consonants 'm' and 'k' replaced the vowels. The category phrasing was modified to 'words *sounding* with' along with the phonetic sound of a letter.
Narration Task	Storyboards originally consisted of eight detailed pictures/story, yet the task duration ranged between 1.5-6 minutes with the 17 participants.	A storyboard with 18 pictures was piloted with the four additional participants in September. Their production ranged between 7 and 12 minutes, thus overcoming the challenge in comparing performance across tasks.

By testing the tasks' ability to elicit speech from the participants and reveal differences within and between the groups, the second phase of the pilot study indicated that the instruments were suitable for use in the main study. Partial descriptive analyses revealed preliminary, yet promising, indications regarding the possible findings of the main study. A brief summary of the descriptive analyses of the pilot study in both phases is offered in Appendix G.

3.8 ETHICAL CONSIDERATIONS

Upon receiving ethical clearance from the Central University Research Ethics Committee of the University of Oxford (Appendix E), the researcher contacted the target school's president, informed him of the study's aims, and asked for permission, subject to parental consent, for access to potential participants (Appendix E). At the school, the

researcher disseminated pamphlets introducing the study and consent forms requesting parents' consent and the students' assent. The documents (Appendix E) included an introduction to the study and an explanation of the objectives and procedures involved. They also highlighted the study's anonymous, confidential nature and the rights that participants had in terms of withdrawing at any point without providing a reason for doing so. Additional disclosure was offered to some parents who contacted the researcher by phone to further discuss the procedure of the research study. As mentioned in the previous section, all ethical procedures followed in the study were adopted in the pilot study (Appendix G). Upon collecting the data, all ethical considerations were abided by, and to ensure anonymity, each participant was assigned a unique identification code in all computer files. All data were kept in a secure folder in a secure location to which only the researcher had access throughout the course of the project.

3.9 CHAPTER SUMMARY

This chapter began with a discussion of the research design, the variables, and the five research questions that were formulated based on the literature and the hypotheses of the study. The research questions were divided thematically, with the first two questions addressing whether the two multilingual groups exhibit differences in executive performance and whether performance differences are predicted by language proficiency and language experience. The second set of questions focused on the participants' language production and codeswitching, addressing the overarching question of whether the two groups behave similarly in single-word and monologue speech production with respect to switch costs and codeswitching patterns. The final question in the second set attempted to bring together performance on non-linguistic and linguistic tasks and

examine group differences across these tasks. The subsequent sections of this chapter offered detailed descriptions of the participants and an in-depth discussion of the instruments used in the study. Each instrument was described, the rationale for using, adapting, or developing it was provided, and the procedure in which it was implemented was offered. A description of the overall procedures used to collect the data was then presented, along with brief sections on the study's pilot study and ethical considerations.

CHAPTER 4. RESULTS AND DISCUSSION CHAPTER I

4.1 INTRODUCTION

This chapter addresses the first two research questions which examine cognitive task performance. The overall analytical strategy is first presented, followed by in-depth descriptions of the analyses and results of Research Questions 1 and 2 and the independent variables (language proficiency and language experience). Summaries and brief discussions of the findings are offered after each section. The chapter concludes with an overall summary of cognitive task performance.

4.2 GENERAL ANALYTICAL STRATEGY

All data from the cognitive tasks were compiled and entered into the statistical package R (version 3.5.2) (R Development Core Team, 2018). Analyses were carried out using the MASS, nlme, and lme4 packages. Mixed-effects models were used to analyse RQ1A-C and RQ2, as well as the independent variables, whereas correlations were used to analyse RQ1D.

Mixed-effects models (MEMs) have become prominent in the analysis of psycholinguistic data and are for this project. MEMs are able to analyse both *fixed effects* (i.e., independent or predictor variables) and *random effects* (i.e., participant-level and item-level variance), producing more accurate analyses than ANOVAs and regression models which only account for participant variance (Baayen, Davidson, & Bates, 2008; Linck & Cunnings 2015). This advantage of MEMs is especially important when multiple observations are nested within both participants and items in hierarchical models such as the ones examined in the present study. In addition to controlling for several types of variance within the same model, MEMs overcome some of the assumption violations which would render ANOVAs and regressions either less robust or

invalid (Linck et al., 2012). Specifically, MEMs can be fitted to continuous, ordinal, and categorical data even when data are not normally distributed, as is the case with many task results in this study. MEMs are also argued to offer more accurate models than other analytical tools with respect to missing data and different group sizes (Field, Miles, & Field, 2012). Despite their advantages, the models must be first suitably fitted to ensure that their output is accurate.

To construct and select the best-fit models for the present study, several steps were followed based on recommendations of common practice and the researcher's discretion. Upon fitting the dependent variables for each research question, the next step in creating mixed-effects models was selecting the explanatory, or predictor, variables based on the hypotheses in the study. For each research question, these variables, as well as with interactions between them, were entered as fixed effects when necessary and possible. The second step was defining random effects and fitting them into the model. Several options are suggested for determining the random effects structure. One is starting with a *maximal random effects structure* which includes intercepts and slopes for all fixed effects (Barr, Levy, Scheepers, & Tily, 2013; Judd, Westfall, & Kenny, 2017). However, a complex random effects structure may lead to convergence errors if there are not enough observations to calculate the variance-covariance matrices of each random slope and intercept (Eager & Roy, 2017). This would also result in a decrease in statistical power by over-fitting the model (Matuschek, Kliegl, Vasishth, Baayen & Bates, 2017). Accordingly, if the model does not converge or is over-fitted, then a minimal random structure which includes intercepts only for participants and items is recommended, as these random intercepts account for repeated measures (Bates, Kliegl, Vasishth, & Baayen, 2015). It should be noted that although this may improve the fit of the model, this may lead to false positives through subject- or item-specific error.

In order to ensure that the MEMs in this study avoid the pitfalls of fitting random effects and potentially obtaining false positives or convergence errors, models were first fitted with a maximal random structure. They were examined to see if they converged through the distribution of residuals. If a model failed to converge, (multi)collinearity[7] between predictors was checked through the VIF.mer package (Frank, 2011). In all models, VIFs which were below 2.5 with *t* and *z* values > 2 were interpreted as significant (Gelman & Hill, 2007). If collinearity was significant, then the correlated predictors were identified and analysed in separate models. Singularity was also checked to reassess the structure by removing either intercepts and/or slopes nested within participants and items. Checking for influential values was then conducted using the influence.ME package (Nieuwenhuis, Pelzer, & Grotenhuis, 2017) by rerunning the model several times. Each time the model was rerun, either one participant or one item was removed. Beta-estimates and influential values were then compared between the tested model and the different simulations, in order to determine whether a participant or an item had a significant influence on the model outcome. If singularity was normal, only the models which did not fail to converge were selected for comparisons.

In comparing the models, the parameters were estimated for fixed effects through Maximum Likelihood and for random effects using Restricted Maximum Likelihood. The models' *goodness-of-fit* or *quality-of-fit* was checked through visual inspection of the distribution of residuals and determined by Akaike's Information Criterion (AIC; Akaike, 1998), and their significance by the likelihood ratio chi-square tests. The best-fit models were then selected in a stepwise approach by removing non-significant fixed effects as recommended by Baayen (2008). For generalised MEMs, the glmulti package

[7] Multicollinearity denotes high correlations between predictor variables (typically r=+/-0.9). Singularity refers to a perfect correlation between two explanatory variables (r=+/-1) (Field et al., 2012).

(Calcagno & de Mazancourt, 2010) was also used for confirmation of the best-fit model. Although the algorithms of the automated selection of models using this package allow for thousands of simulations to be run on the data for the best fit, there is a caveat of dropping variables which are of interest to the hypotheses under investigation. Accordingly, the glmulti package was used after 'manual' stepwise selections were made, for verification purposes only. In case of a discrepancy between the selected model and the recommended glmulti model, both models were compared and the final best-fit model was selected based on the researcher's discretion, as recommended by the authors of this package. The following offers a more detailed account of how the research questions addressing cognitive task performance were analysed with respect to the type of data at hand.

4.3 Cognitive Task Performance

> *RQ1: Do EMI and FMI participants differ in performance across the domain-general cognitive tasks: (A) the flanker task, (B) the Simon task, and (C) the Backward Digit-Span task?*

In order to address the first research question which compares EMI and FMI participants in terms of their cognitive task performance, quantitative data from the flanker task, the Simon Task, and the Backward Digit-Span Task were entered into R. Before statistical analyses were conducted, each dataset was cleaned based on the inclusion/exclusion criteria set for each task. In brief, both correct and incorrect responses in the Flanker and Simon tasks were included in the accuracy rates' analyses, whereas only correct responses were included to analyse reaction times in both tasks. Missing responses were classified as 'incorrect' in both tasks. Reaction times below 250ms and above 1500ms were excluded from the data as indicated in Chapter 3. Upon cleaning the data, Shapiro-Wilks' tests, along with visual inspections of the datasets and

examinations of skewness and kurtosis for each task were conducted to examine normality. An *a priori* screening for outliers and a visual model inspection are said to be two critical steps within the mixed-effects modelling approach to prevent model distortion (Baayen & Milin, 2010). To approximate residuals' normality and to ensure homoscedasticity (i.e., stabilise residual variance), data transformations were explored. Two types of transformations were conducted on the data in this chapter: log-transformations and box-cox power transformations, both of which ensure that residuals are approximately normal, given the lambda parameter of the transformations. Lambda values indicated that for the Flanker RTs and Simon Task RTs, log transformations were appropriate. Visual inspections of the data, along with the skewness and kurtosis tests, confirmed that these transformations were appropriate to approximate normality in the two tasks (data skewness <0.15).

4.3.1 RQ1A - The flanker task

> *RQ1A: Do the two groups exhibit different accuracy rates and response latencies on the flanker task, including different monitoring costs and conflict effects?*

The first part of the first research question explored whether the participants from both groups perform similarly on the flanker task. To address this, the data examined the accuracy and RTs in each of the three conditions (low, medium, and high) in congruent and incongruent trials. It then examined the monitoring costs and conflict effects.

4.3.1.1 Accuracy

To address the first research question, accuracy binary data (correct= 0; incorrect= 1) were fitted as the outcome variable in a generalised linear MEM with a binary link function (Bates, 2010). Congruency, Condition, and Group, along with their interactions, were fitted as predictors. The overall design of this analysis was therefore a 3x2x2 factorial design: 3 (Monitoring Condition: low vs. medium vs. high) x 2

(Congruence: congruent vs. incongruent) x 2 (Group: EMI vs. FMI). As indicated in Table 4.1 and Figure 4.1, participants performed very close to ceiling, with an error rate of 1.22% (or a total of 211 errors of 17,280 trials for all participants). The Shapiro-Wilk test ($p<0.0001$), normal Q-Q plots, as well as histograms revealed that accuracy rates were not normally distributed. Non-parametric tests were therefore used to assess this dependent variable using the MASS and nlme packages. Specifically, the generalised linear model (GLMM) with multivariate normal random effects with Penalized Quasi-Likelihood (PQL) was used. This model is regarded as a mixed-effects implementation of a logistic regression, or a mixed logit model (Venables & Ripley, 2002). The advantage of using the GLMM is that it offers the same analyses as the ANOVA without the assumption of a normally distributed data set or a continuous dependent variable (Jaeger, 2008; Linck, 2016). Furthermore, this model allows comparisons to be made across and within participants and conditions without the need to first calculate the average scores for each condition by subjects (in this case, congruency) and then analysing these scores using ANOVA. As such, it is argued that this model is superior for binary data analysis to repeated measures ANOVA (Linck & Cunnings, 2015).

Table 4.1 Mean and Standard Deviations of Flanker Task Accuracy Rates

Condition/Congruence	**Accuracy Mean Rate in % (Std. Deviation)**			
	EMI		FMI	
	N=30		N=30	
	M	**SD**	**M**	**SD**
Low Monitoring				
Congruent	99. 5	(0.66)	99.37	(0.69)
Incongruent	89.44	(8.17)	90.02	(8.16)
Medium Monitoring				
Congruent	99	(0.83)	98.89	(0.91)
Incongruent	96.67	(3.03)	95.16	(2.78)
High Monitoring				
Congruent	98.5	(1.25)	98.58	(1.26)
Incongruent	98.83	(1.27)	99	(1.25)

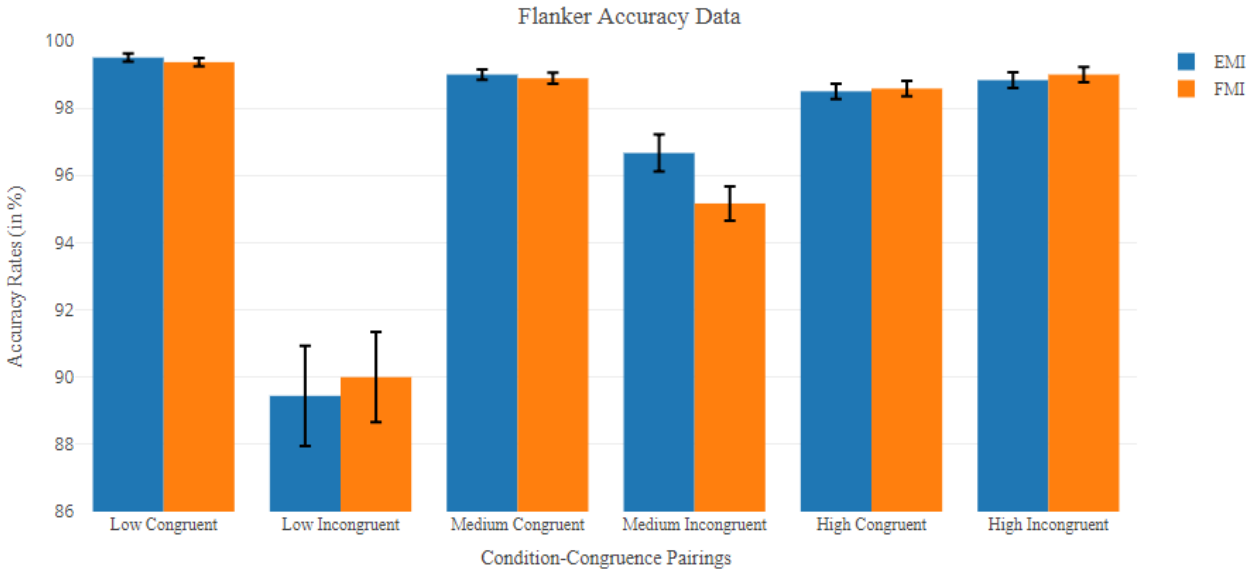


Figure 4.1 Flanker Accuracy Data (EMI vs. FMI)

To analyse the accuracy data, the model was fitted with a dummy independent coded variable, where one level of that factor serves as the baseline condition against which all other levels of the factor are compared. The model was fitted and refitted in order to change the baseline and obtain planned pairwise comparisons. A sample of the model's output is presented in Table 4.2 which lists maximum likelihood estimates, where the baseline is the low-congruent condition for the EMI group against which all other conditions as well as the FMI group (on this particular condition) are compared.

Table 4.2 GLMM Output for Flanker Accuracy Data (Estimated coefficients, standard errors (SE), z-values, and p-values)

Asterisks (*) indicate: p< .05 (*), p< .01 (**), p<. 001(***) and p< .0001 (****)

Predictor	Coefficient	SE	z-value	p-value
Fixed Effects				
Intercept	-5.30	0.21	-25.09	0.00****
FMI	0.24	0.28	0.86	0.39
Low-Incongruent	3.17	0.27	11.75	0.00****
Medium-Congruent	0.71	0.26	2.65	0.00**
Medium-Incongruent	1.94	0.26	7.36	0.00****
High-Congruent	1.12	0.27	4.18	0.00****
High-Incongruent	0.86	0.28	3.06	0.00**
FMI: Low-Incongruent	-0.24	0.36	-0.65	0.51
FMI: Medium-Congruent	-0.14	0.36	-0.37	0.71
FMI: Medium-Incongruent	0.14	0.34	0.41	0.68
FMI: High-Congruent	-0.30	0.36	-0.81	0.41
FMI: High-Incongruent	-0.40	0.39	-1.01	0.31
Random Effects				
Participants	0.12			
Items	0.08			
Residuals	0.68			

*Formula: Accuracy ~ Group*Monitoring*Congruence + (1 + Congruence*Monitoring| subject) + (1| item)*; t-value = Coefficient/SE, with t-ratio values over 2.0 indicating the coefficient is significantly different from zero (Gelman & Hill, 2007).

A summary of the model's ANOVA detailing all main effects and interactions is offered in Table 4.3. Overall, the model revealed that there were significant main effects for congruence ($F(1,290)=123.79$, $p<0.0001$) and monitoring conditions ($F(2, 290) =44.93$, $p<0.0001$), yet there was no significant Group main effect ($p>0.05$). There was also a significant interaction between congruence and monitoring conditions ($F(2,290)=64.83$, $p<0.0001$) (see Figure 4.2). These results indicated that irrespective of the manipulation of the monitoring conditions, both groups responded similarly by having significantly lower error rates on congruent than on incongruent trials. However, difference in the *distribution* of congruent-to-incongruent trials affected the accuracy rates.

Table 4.3 GLMM ANOVA (Type III) Summary for Flanker Accuracy Data

	F-Statistic	DF1	DF2	p-value
Intercept	225.28	1	290	<0.0001
Group (EMI vs. FMI)	0.37	1	58	0.54
Congruence (congruent vs. incongruent trials)	123.79	1	290	<.0001
Monitoring Condition (low vs. medium vs. high)	44.93	2	290	<0.0001
Group*Congruence	0.27	1	290	0.61
Group*Monitoring	0.56	2	290	0.57
Congruence*Monitoring Condition	64.86	2	290	<0.0001
Group*Congruence*Monitoring	0.43	2	290	0.68

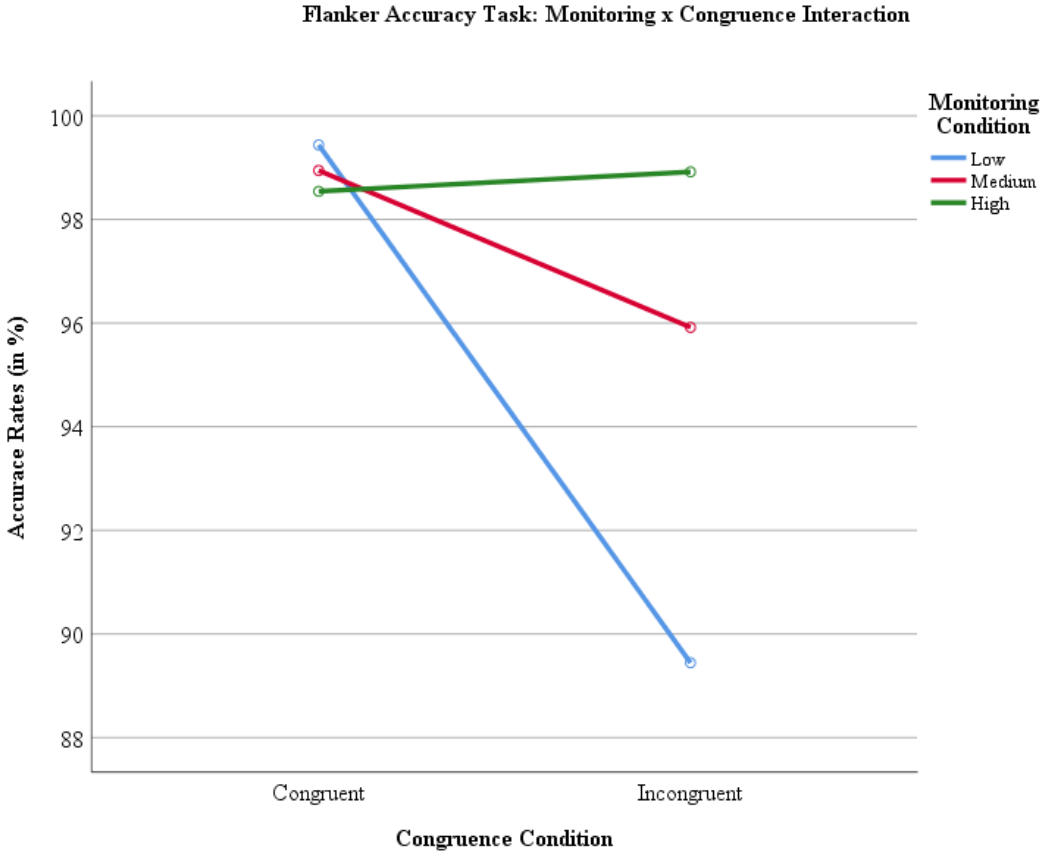


Figure 4.2 Flanker Accuracy Interaction (Monitoring x Congruence)

In order to further investigate the interaction between congruence and monitoring conditions, *post-hoc* analyses were performed. Simple main effects were examined through a generalised mixed effects model which was fitted with accuracy rates as the outcome variable and with Congruence (2 levels), Monitoring (3 levels), and their

interaction, as the predictor variables. As there was no significant main effect for Group or significant interaction between Group and the other predictor variables, data from all participants were combined in the analyses.

The results of the *post-hoc* analyses revealed that accuracy rates were significantly influenced by the Congruence predictor. Overall, congruent trials were significantly more accurate than incongruent trials (β=1.76, SE=0.27, z=6.52, $p<0.0001$). However, accuracy rates were modulated by monitoring conditions, as depicted in Table 4.4. Specifically, accuracy rates were significantly higher in congruent trials compared to incongruent trials in the low-monitoring (β= 3.15, SE= 0.27, z= 11.67, $p< 0.0001$) and medium-monitoring (β=1.33, SE=0.21, z= 6.37, $p< 0.0001$) conditions, but not in the high-monitoring condition ($p>0.05$). An unexpected observation was that in the high-monitoring condition, participants responded more accurately on incongruent trials than on congruent trials. This may be explained by the increased cognitive demands, especially conflict monitoring, required by the 50-50 distribution of trials in this condition. A heightened attention to the frequency of incongruent trials may have thus led to anticipatory action, exhibited by the slightly more accurate responses on these trials.

Table 4.4 Flanker Accuracy Monitoring x Congruence Contrasts

β=Coefficient estimate, SE=Standard Error, z= z-value, p=p-value; boldface= significant contrasts

		Low								**Medium**								**High**							
		Congruent				Incongruent				Congruent				Incongruent				Congruent				Incongruent			
		β	**SE**	**z**	**p**	**β**	**SE**	**z**	**p**	**B**	**SE**	**z**	**p**	**β**	**SE**	**z**	**p**	**β**	**SE**	**z**	**p**	**β**	**SE**	**z**	**p**
Low																									
	Congruent	-				**3.15**	**0.27**	**11.67**	**0.00**	**0.66**	**0.26**	**2.52**	**0.02**	**1.89**	**0.26**	**8.23**	**0.00**	**0.96**	**0.26**	**3.71**	**0.00**	0.53	0.27	1.93	0.05
	Incongruent	**3.15**	**0.27**	**11.67**	**0.00**	-				**2.36**	**0.23**	**10.26**	**0.00**	**1.01**	**0.22**	**4.61**	**0.00**	**2.11**	**0.24**	**8.75**	**0.00**	**2.40**	**0.26**	**9.25**	**0.00**
Medium																									
	Congruent	**0.66**	**0.26**	**2.52**	**0.02**	**2.36**	**0.23**	**10.26**	**0.00**	-				**1.33**	**0.21**	**6.37**	**0.00**	0.32	0.23	1.41	0.17	0.13	0.25	0.53	0.56
	Incongruent	**1.89**	**0.26**	**8.23**	**0.00**	**1.01**	**0.22**	**4.61**	**0.00**	**1.33**	**0.21**	**6.37**	**0.00**	-				**1.06**	**0.22**	**4.81**	**0.00**	**1.33**	**0.24**	**5.54**	**0.00**
High																									
	Congruent	**0.96**	**0.26**	**3.71**	**0.00**	**2.11**	**0.24**	**8.75**	**0.00**	0.32	0.23	1.41	0.17	**1.06**	**0.22**	**4.81**	**0.00**	-				0.31	0.25	1.18	0.24
	Incongruent	0.53	0.27	1.93	0.05	**2.40**	**0.26**	**9.25**	**0.00**	0.13	0.25	0.53	0.56	**1.33**	**0.24**	**5.54**	**0.00**	0.31	0.25	1.18	0.24	-			

Monitoring

Regarding accuracy rates across the monitoring conditions, participants' accuracy rates in both congruent and incongruent conditions were influenced by the monitoring variable. The low-monitoring condition had the highest accuracy rates for congruent trials and the lowest accuracy rates for incongruent trials across all monitoring conditions. Specifically, accuracy on the low-congruent trials was significantly higher than all other trials ($p<0.001$) except for high-incongruent trials, where accuracy rates were comparable ($p>0.05$). As for the low-incongruent condition, accuracy rates were significantly lower than all other trials ($p<0.0001$). This is indicative that the 92-8% distribution of the congruent-to-incongruent trials reduced the conflict monitoring requirements of this task by facilitating responses on the congruent trials, whereas more inhibitory control and interference suppression were needed to produce accurate responses on incongruent trials. The 92% congruent trials in the low monitoring condition were expected to be the highest in accuracy, yet it is interesting that participants performed almost similarly on the low-congruent and high-incongruent trials. Participants were predicted to incur a cost in accuracy due to the constant switching between congruent and incongruent trials. It is this constant switching, however, which seems to have triggered a shift in attention and resulted in more accurate responses.

For the medium-monitoring condition, where the distribution was 75%-25%, accuracy results were relatively more varied than those in the low-monitoring condition. Participants responded significantly more accurately on the medium-congruent trials than medium-incongruent trials ($\beta=1.33$, $SE=0.21$, $z=6.37$, $p< 0.0001$), as expected, as well as on the low-incongruent trials ($\beta=2.36$, $SE=0.24$, $z=10.26$, $p< 0.0001$) which had the lowest accuracy rates across all conditions. Conversely, participants were less accurate in

the medium-congruent trials than on the low-congruent trials (β=0.66, SE=0.26, z=2.52, p<0.01), as the low-congruent trials had the highest accuracy rates across all conditions. There were no significant differences between accuracy on the medium-congruent trials and high-congruent condition or high-incongruent condition (p>0.05). As such, the 75-25 distribution was similar to the 50-50 distribution of congruent to incongruent trials for the medium-monitoring condition with *congruent* trials. However, the accuracy rates on incongruent trials in this condition were significantly lower than those in the low-congruent, medium-congruent, high-congruent, and high-incongruent conditions (p<0.0001 for all contrasts, see Table 4.4). Medium-incongruent accuracy rates were significantly higher than those in the low-incongruent condition, as the latter condition was the least accurate (β=1.01, SE=0.22, z=4.61, p< 0.0001).

For the high-monitoring condition, there were no significant differences between accuracy on congruent trials and incongruent trials (p>0.05) or the medium-congruent trials (p>0.05), as indicated earlier. However, participants were significantly more accurate on the high-congruent trials than on the low-incongruent (β=2.11, SE=0.24, z=8.75, p<0.0001) and medium-incongruent trials (β=1.06, SE= 0.22, z=4.81, p<0.0001). High-congruent trials were significantly less accurate than on the low-congruent condition (β=0.96, SE=0.26, z=3.71, p<0.001). For the high-incongruent condition, there were no significant differences between accuracy on these trials and on the low-congruent, medium-congruent, and high-congruent trials (all p>0.05). Accuracy on high-incongruent trials was significantly higher than that on low-incongruent trials (β=2.40, SE=0.26, z=9.25, p<0.0001) and medium-incongruent trials (β=1.33, SE=0.24, z=5.54, p<0.0001). This finding indicates that the 50-50 condition which had a higher number of incongruent trials resulted in more accurate rates than the other conditions which had fewer incongruent trials compared to congruent ones. This result therefore indicates that

accuracy rates are susceptible to a subtle change in task design, such as the change of proportion of incongruent trials relative to congruent ones.

A schematic representation of accuracy across the monitoring-congruence conditions is offered in Figure 4.3:

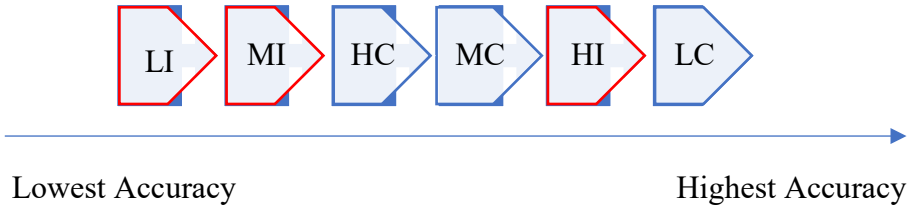


Figure 4.3 Flanker accuracy rates (L= Low, M= Medium, H=High, C=Congruent, IC=Incongruent)

To summarise, participants appear to have exhibited a reverse pattern of accuracy across the monitoring conditions for congruent vs. incongruent trials. For congruent trials, the low-monitoring condition had the highest accuracy rates, followed by the medium-condition, and then the high-monitoring condition. There was no significant difference in accuracy rates between the low- and high-monitoring conditions or between the medium- and high-monitoring condition for these congruent trials. Accordingly, the congruent-to-incongruent ratio manipulations of 92-8 vs. 50-50 and 75-25 vs. 50-50 did not influence response accuracy on congruent trials. For incongruent trials, the low-monitoring condition yielded the lowest accuracy rates, followed by the medium-monitoring condition and then the high-monitoring condition, with significant differences across these three monitoring conditions. This pattern reveals that the manipulation of the ratios led to higher accuracy when trials were equally distributed between congruent and incongruent trials. As the proportion of incongruent trials decreased compared to congruent trials, however, accuracy rates decreased as well.

These findings are in line with Costa et al.'s (2009) predictions, as the low-monitoring condition, especially the incongruent trials, had a higher error rate relative to

the other conditions. This supports the underlying assumption that when incongruent trials are in much lower proportion to congruent trials, more inhibitory effort is needed to suppress distracting stimuli. That is, more recruitment of conflict resolution mechanisms is required, in order to solve the conflict created by each of the incongruent trials, leading to a larger conflict effect (2009). Furthermore, the high-monitoring condition revealed comparable performance between congruent and incongruent trials, indicating that the constant shifting between the two types of trials called for an increase in monitoring and attention, leading to more accurate responses between the two types of trials. As such, the high-monitoring condition may be perceived as mirroring 'cognitive processes involved in the micro-management of languages' (Hofweber et al., 2016). Accordingly, it is possible to infer that as participants engage in frequent codeswitching, their conflict monitoring skills are heightened, and they are more attentive to cues which demand constant monitoring (i.e., salient cue detection). These findings are in opposition to the predictions of the CPM (yet in line with those of the extended CPM), but they parallel those of Hofweber et al. (2016). The findings are further discussed and corroborated in the analyses of predictors (especially codeswitching frequency and salient cue detection) in Section 4.4.

4.3.1.2 Reaction Time (RT)

Reaction time data were not normally distributed for one of the conditions as indicated by the Shapiro-Wilk Test ($p<0.001$). To approximate normality, a reciprocal transformation for the data was conducted through the Box-Cox function of the MASS package in R (Venables & Ripley, 2002). Upon applying this transformation, visual inspection of the distribution of the entire dataset, as well the distribution of reaction times for each participant, indicated that the reciprocal transformation was indeed the most appropriate transformation of the data. A multilevel (hierarchical) linear model

from the nlme package was then selected to analyse the data, as it is able to fit different measures with continuous and categorical data with different levels. The model was fitted with RTs as the response variable and with Condition (low, medium, high), Congruency (congruent and incongruent), and Group (EMI and FMI), along with interactions between these variables, as predictors. A maximal random-effects structure was adopted, as indicated in the General Analytical Strategy section. Table 4.5 offers descriptive statistics of the raw RT scores before the data transformation.

Table 4.5 Mean and Standard Deviations of Flanker Task Reaction Times (RTs)

Condition/Congruence	**RTs in ms (Std. Deviation)**			
	EMI		**FMI**	
	N=30		N=30	
	M	**SD**	**M**	**SD**
Low Monitoring				
Congruent	520.40	(15.28)	523.17	(16.54)
Incongruent	550.00	(27.08)	556.57	(28.57)
Medium Monitoring				
Congruent	522.63	(22.77)	509.17	(33.94)
Incongruent	554.90	(47.65)	557.57	(30.16)
High Monitoring				
Congruent	555.37	(22.24)	552.27	(29.06)
Incongruent	574.03	(29.51)	581.90	(43.36)

Similar to the generalised linear model used for the accuracy data, the present model is a linear mixed-effects model (MEM) but fit by a restricted maximum likelihood given the data were transformed. As with the accuracy data, the RTs from the flanker task indicated a main effect for congruence ($F(1, 290)=121.76$, $p<0.0001$) and a main effect for monitoring ($F(1, 290)=61.25$, $p<0.0001$). There was no significant main effect for Group ($F(1, 58)=0.00$, $p>0.05$). There were also no significant interactions between any of the predictor variables (all $p>0.05$). A summary of the ANOVA (Type III) obtained from the MEM is presented in Table 4.6.

Table 4.6 MEM ANOVA (Type III) Summary for Flanker RTs

	F-Statistic	DF1	DF2	p-value
Intercept	57515.55	1	290	<0.0001
Group (EMI vs. FMI)	0.00	1	58	0.98
Congruence (congruent vs. incongruent)	121.76	1	290	<0.0001
Monitoring (low, medium, high)	61.25	2	290	<0.0001
Group*Congruence	1.80	1	290	0.11
Group*Monitoring	0.26	2	290	0.61
Congruence*Monitoring	2.10	2	290	0.15
Group*Congruence*Monitoring	0.18	2	290	0.67

*Formula: Reaction Times ~ Group*Monitoring*Congruence + (1 + Congruence*Monitoring| subject) + (1| item)*

Congruence

The results of the model revealed that participants were significantly faster on congruent trials than incongruent trials across the three monitoring conditions. Specifically, congruent trials were significantly faster than incongruent trials in the low-monitoring condition, (β= 0.10, SE= 0.02, t= 4.85, p< 0.0001), the medium-monitoring condition (β= 0.13, SE= 0.02, t= 6.30, p< 0.0001), and the high-monitoring condition (β= 0.07, SE= 0.02, t= 3.35. p< 0.001). These results were expected, yet it should be noted that unlike accuracy rates, the high-monitoring condition revealed a difference between congruent and incongruent trials.

Monitoring

The results of the model indicated that participants exhibited the shortest reaction times in the low-monitoring condition of the flanker task. Specifically, responses on low-monitoring trials were significantly faster than on medium-monitoring trials (t=4.28, p<0.0001) and on high-monitoring trials (t=6.97, p<0.0001). Furthermore, RTs on the medium-monitoring condition were significantly faster than in the high-monitoring condition (t=3.87, p<0.01). This indicates that the 75-25 distribution was less demanding in terms of conflict monitoring compared to the 50-50 distribution.

Overall, results from the RT data indicate that the distribution of trials in the high-monitoring conditions required more conflict resolution, as depicted by longer response latencies (Figure 4.4[8]). This is true for both congruent and incongruent trials, and is in line with the predictions of Costa et al. (2009). Taken together with the accuracy rates, it appears that participants look a longer time to respond on this condition and this may have resulted in more accurate responses. As such, it can be inferred that the conflict effect is more evident in this measure of the flanker task (RTs) rather than accuracy rates.

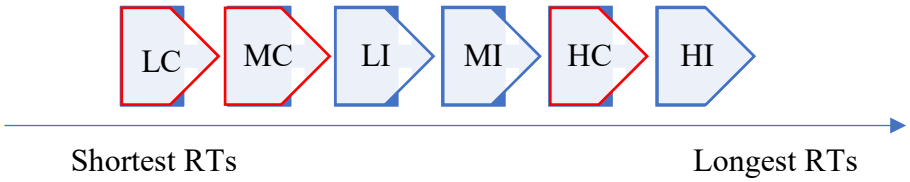


Figure 4.4 Flanker RTs (L= Low, M= Medium, H=High, C=Congruent, IC=Incongruent)

To summarise, both groups had smaller RTs on congruent trials than on incongruent trials within each condition. Furthermore, RTs were the smallest in low-monitoring condition, followed by the medium-monitoring condition, and finally the high-monitoring condition, with significant differences across each condition. This supports the task design which is based on the premise that the high-monitoring condition was taxing on participants' performance, though this is observed only in the RT (not accuracy) data. This could be attributed to the assumption that in the high-monitoring condition, participants incur a high monitoring cost due to the high frequency of congruent-to-incongruent trial switching.

[8]Although the analyses did not distinguish congruence within each monitoring condition due to the non-significant Congruence x Monitoring interaction, this schematic representation of response latencies across the monitoring-congruence conditions is done purposefully to highlight the difference between Flanker accuracy and RT measures. Please refer to Appendix H for Monitoring x Congruence contrasts for each group as supplementary *post-hoc* analyses.

Post-hoc Analyses: Block Order

As noted in Chapter 3, different monitoring blocks were counterbalanced to ensure that the sequence of blocks was not presented in order of increasing demands to monitoring skills. Despite controlling for this potential moderating variable, exploratory *post-hoc* analyses were conducted to investigate whether block sequence had an influence on performance. The six groups to which participants were randomly allocated were contrasted; each group comprising 5 EMI and 5 FMI participants, as stipulated in Section 3.5.4. A generalised linear model was fitted with Group (2 levels: EMI and FMI) and Block Order (6 levels), along with Group x Block order, as predictors with accuracy rates as a dependent variable. Another linear model was also fitted with these predictors yet with RTs as the outcome variable. There were no significant main effects for block order for Flanker accuracy rates ($F(1, 58)=0.97$, $p>0.05$) nor for Flanker RTs ($F(1, 58)=1.25$, $p>0.05$). Furthermore, there was no significant main effect for Group and no interaction between Group x Block Order ($p>0.05$). These results indicate, therefore, that the sequence in which the blocks were introduced did not significantly influence participants' performance for either group throughout the experiment. The results of the models' ANOVA output are summarised in Table 4.7 and Table 4.8.

Table 4.7 GLM ANOVA (Type III) Summary of Block Order (Flanker Accuracy)

	F-Statistic	DF1	DF2	p-value
Intercept	57515.55	1	290	<0.0001
Group (EMI vs. FMI)	0.16	1	58	0.69
Block Order (6 levels)	0.97	5	290	0.38
Group*Block Order	0.22	5	290	0.77

Table 4.8 GLM ANOVA (Type III) Summary of Block Order (Flanker RTs)

	F-Statistic	DF1	DF2	p-value
Intercept	5253.71	1	290	<0.0001
Group (EMI vs. FMI)	0.09	1	58	0.96
Block Order (6 levels)	1.25	5	290	0.15
Group*Block Order	1.12	5	290	0.14

4.3.1.3 Monitoring Cost

4.3.1.3.1 Accuracy

As discussed in Section 3.5.2, *monitoring cost* (as termed by Hofweber et al., 2016) is measured by subtracting the number of errors in the low-monitoring block from that of the high-monitoring block. The monitoring cost for each participant was calculated and then data from both groups was assessed for normality. Visual inspection revealed that the data were positively skewed. The Shapiro-Wilk test indicated a non-normal distribution of the accuracy monitoring cost ($p<0.001$). Furthermore, the raw data revealed (mathematically) negative outcomes where the number of errors in the low-monitoring condition exceeded that in the high-monitoring condition. One potential explanation for the results in this study is that the high-monitoring condition entailed participants' increased attentiveness throughout the experiment, resulting in higher accuracy rates.

As the monitoring-cost data were not normally distributed, and they ranged between (-2) and (+2), a non-parametric test was used for analysis. The Mann-Whitney U test could not be selected to compare the two groups, as its assumptions were violated due to the presence of numerous ties. Accordingly, a Fisher (exact) permutation test was selected to analyse the data. The advantage of this test is that it does not assume a normal distribution; rather, it builds a distribution by resampling the observations (Fay & Shaw, 2010). Furthermore, it is not sensitive to the presence of ties. Results revealed that the EMI and FMI groups were not significantly different ($z=0.22$, $p=0.83$) in terms of their monitoring cost accuracy.

4.3.1.3.2 Reaction Time

For reaction time, the monitoring cost was computed by subtracting overall RTs in the low-monitoring block from those in the high-monitoring block. After the

computation was completed for each participant, the data were checked for normality and were revealed to be normally distributed by the Shapiro-Wilk test ($p>0.05$ for each group). Visual inspection of the normal Q-Q plots confirmed the normality of the distribution. Descriptive statistics are offered in Table 4.9, including the minimum and maximum scores obtained by participants in each group. These scores are indicative of the range, where it is evident that some participants exhibited very low monitoring costs whereas others exhibited much higher ones with respect to RTs. In other words, some participants performed almost identically on the low-monitoring and high-monitoring conditions.

Table 4.9 Descriptive Statistics of Monitoring Costs

Group	N	Mean	SD	Minimum	Maximum
EMI	30	43	(24.75)	4	120
FMI	30	41.33	(26.20)	3	114

In order to compare the EMI and FMI groups, an independent-samples t-test was conducted. There was no significant difference in the monitoring costs of the EMI ($M=43$, $SD=24.75$) and FMI ($M=41.33$, $SD=26.20$) groups; $t(58)=0.25$, $p=0.80$. These results indicate that the participants in both groups had comparable monitoring costs; that is, they exhibited similar conflict-monitoring skills.

While 'monitoring cost', as a novel construct put forth by Costa et al. (2009), is indicative of an important manipulation of congruent-to-incongruent ratios in the blocks, the means by which it is assessed should be approached with caution. The simple subtraction of low-monitoring errors and reaction times from those in the high-monitoring block may be a sensible calculation when comparing highly dissimilar groups, such as monolinguals and bilinguals, as in Costa et al.'s study. However, for the two groups in the present study, this computation was not sensitive enough to detect

significant differences between the two groups' performance on the low- and high-monitoring conditions, especially in terms of accuracy. Furthermore, the outcomes of this measure seem to highly overlap with the contrasts obtained in the analysis of accuracy and reaction times across all the monitoring conditions. However, the contrasts obtained from the generalised linear model are more sensitive to detecting variation in performance across the conditions and within the two groups.

4.3.1.4 Conflict Effect

The *Conflict Effect* measures the potential increase in inhibitory effort from congruent to incongruent trials and is believed to parallel inhibitory control measured in the Simon Task (Hofweber, 2017, personal communication). The score obtained from the conflict effect is indicative of inhibitory mechanisms, where a small conflict effect is associated with better inhibitory control abilities, specifically interference suppression.

4.3.1.4.1 Accuracy

To compute the conflict effect in terms of accuracy, the number of errors in congruent trials across all conditions was subtracted from the number of errors in incongruent trials. Upon calculating the score for each participant, data were checked for normality, and it was revealed through the Shapiro-Wilk test that the distribution was not normal ($p<0.01$). As there were several ties in the scores, where the conflict effect for accuracy ranged between -3 and 3, a non-parametric Fisher permutation test was used. The test revealed that the two groups were not significantly different in terms of the accuracy aspect of *conflict effect* ($z= -0.56$; $p=0.58$). However, as mentioned earlier, the number of errors was extremely low and the computation of the conflict effect resulted in negative scores as some participants made more errors on congruent trials than incongruent ones given the distribution of the two trial types. As such, this calculation of

the conflict effect may not be the most suitable way of assessing the construct despite its pervasive use in the literature. Perhaps it is a sound computation when the sample does not perform at ceiling and/or when the two groups are distinct in other ways, as in the case of bilinguals and monolinguals.

4.3.1.4.2 Reaction Times

As with accuracy rates, for RTs, the conflict effect was calculated as the difference between RTs in incongruent and RTs in congruent trials (i.e., RT incongruent - RT congruent). After computing the conflict effect for each participant, the data were assessed for normality. The scores were non-normally distributed, as indicated by the Shapiro-Wilk test ($p<0.01$), and a Mann-Whitney U test could not be used due to the presence of several ties. Accordingly, a Fisher permutation test was used to compare the EMI and FMI groups. Contrary to the prediction that FMI participants would exhibit lower conflict effects than their EMI counterparts, the results indicated that the two groups were not significantly different in terms of the response latency aspect of their conflict effect ($z= -0.88$, $p= 0.38$). It is important to note that the range of conflict effect was between 0 and 97, meaning that there was much variation among participants. The Conflict Effect (RT) is later examined along with other factors, such as the Simon Effect and working memory, as it may shed more light on executive performance than when examined alone.

4.3.2 RQ1B – The Simon Task

RQ1B: Do the two groups exhibit different inhibitory control in the Simon Task?

Performance on the Simon task, a measure of inhibitory control, was assessed through accuracy rates and reaction times. The Simon task also provided a score, the Simon Effect, where the lower the Simon Effect score, the better the inhibitory control that a multilingual possesses. This section addresses the second part of the RQ1 by

comparing EMI and FMI participants in terms of their accuracy, reaction times, and Simon Effect scores on this task.

4.3.2.1 Accuracy

Participants in both EMI and FMI had high accuracy rates on the Simon task, with an error rate of 2.45% (or a total of 919 errors of 37,440 trials for all 60 participants). Tests for normality (Shapiro-Wilk $p<0.0001$) indicated that the accuracy data were non-normally distributed. Therefore, a generalised linear mixed-effects model with a logistic link function (Bates, 2010) from the MASS and nlme packages was selected for analysis. Accuracy binary data (correct= 0; incorrect= 1) were fitted into the model as the outcome, whereas Congruence (3 levels: congruent, incongruent, and control) and Group (2 levels: EMI and FMI), as well as the Congruence x Group interaction, were fitted as predictors. A summary of the descriptive results of accuracy means for the different congruence trials for each group is summarised in Table 4.10.

Table 4.10 Mean and Standard Deviations of Simon Task Accuracy Rates

Congruence	**Accuracy Mean Rate in % (Std. Deviation)**			
	EMI		FMI	
	N=30		N=30	
	M	**SD**	**M**	**SD**
Congruent	97.88	(1.88)	98.41	(1.58)
Incongruent	96.29	(1.84)	95.77	(2.11)
Control	97.98	(1.76)	98.6	(1.33)

A sample of the model's output is offered in Table 4.11. The results of the model revealed a significant main effect for Congruence ($F(2,319)=23.21$, $p<0.001$), with no other significant results (all $p>0.05$). That is, for participants in both groups, differences in accuracy were observed across the congruent, control, and incongruent conditions. The model was releveled to obtain contrasts by congruence level. The results indicated that responses were significantly more accurate on congruent trials than on incongruent

trials (β=0.81, SE=0.20, z=4.13, p<0.0001) and on control trials than on incongruent trials (β=0.90, SE=0.21, z=4.43, p<0.0001). There were no significant differences between the congruent and control trials (p>0.05) (see Figure 4.5).

Table 4.11 GLMM Output for Simon Task Accuracy Data (Estimated coefficients, standard errors (SE), z-values, and p values). Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	Simon Accuracy Rates			
Predictor	**Coefficient**	**SE**	**z-value**	**p-value**
Fixed Effects				
Intercept	-4.99	0.15	-33.35	0.00****
FMI	-0.37	0.23	-1.59	0.12
Congruent	0.05	0.21	0.22	0.82
Incongruent	0.61	0.19	3.28	0.00**
FMI: Congruent	0.08	0.32	0.26	0.80
FMI: Incongruent	0.51	0.28	1.82	0.07
Random Effects				
Participants	2.85			
Items	0.68			
Residuals	0.95			

*Formula: Accuracy ~ Group*Congruence + (1 + Congruence|Participant)*

The accuracy results of the Simon task are expected, as participants were predicted to exhibit the least accuracy on incongruent trials. This is because of the underlying premise of this task's design which purports that inhibitory control is needed to resolve the conflict between two spatial dimensions in the incongruent trials and to suppress any prepotent tendency to respond based on the location of the stimulus, as opposed to the directionality of the stimulus (Hilchey & Klein, 2011; Linck et al., 2012). Although it was predicted that responses on incongruent trials would be the least accurate, the two groups were expected to behave differently: FMI participants were expected to produce significantly more accurate responses given the assumption of a trilingual advantage for FMI over EMI. However, as with the flanker task, accuracy rates on the Simon task were almost at ceiling, and no group differences were observed.

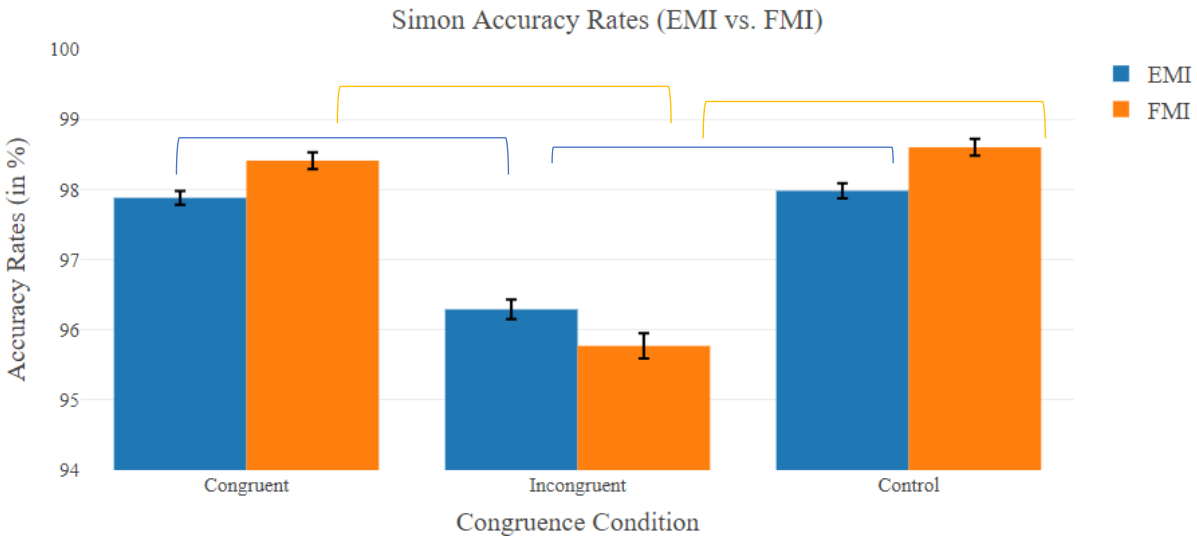


Figure 4.5 Simon Accuracy Rates -- Congruence Contrasts (brackets = significant differences)

4.3.2.2 Reaction Time

RT data for the Simon task were not normally distributed, so a transformation to approximate normality was conducted through the Box-Cox function of the MASS package in R. Similar to the RT transformation in the flanker task, a reciprocal transformation was adopted and checked for normality through visual inspection of the QQ-plots. A generalised linear model was then selected for analysis, and its fit was assessed as detailed in the analytical strategy presented in the first section of this chapter. While Table 4.12 summarises the raw scores of the two groups across the Simon Task conditions, the analysis reports on the transformed data.

Table 4.12 Mean and Standard Deviations of Simon Task RTs

Congruence	**Reaction Time in milliseconds (Std. Deviation)**			
	EMI		FMI	
	N=30		N=30	
	M	**SD**	**M**	**SD**
Congruent	385.072	(23.78)	381.401	(31.89)
Incongruent	473.463	(29.44)	468.504	(34.78)
Control	401.961	(26.87)	400.307	(36.63)

The generalised linear model was fitted with reaction time as the response variable and with Group (2 levels: EMI vs. FMI) and Condition (3 levels: Congruent,

Incongruent, and Control), along with the Group x Condition interaction, as predictor variables. Output from the model is presented in Table 4.13 where the baselines are EMI for the Group condition and Congruent trials for the Congruence condition.

Table 4.13 GLM Output for Simon RTs (Estimated coefficients, standard errors (SE), t-values, and p-values). Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	Simon RTs			
Predictor	**Coefficient**	**SE**	**t-value**	**p-value**
Fixed Effects				
Intercept	385.07	5.64	68.26	0.00****
FMI	-3.67	7.98	-0.46	0.65
Control	16.90	6.55	2.58	0.01*
Incongruent	88.40	6.56	13.50	0.00****
FMI: Control	1.73	9.26	0.19	0.85
FMI: Incongruent	1.30	9.27	-0.14	0.89
Random Effects				
Participants	17.36			
Items	18.92			
Residuals	24.93			

*Formula: RT~Group*Congruence + (1+Congruence|Participant) + (1|item)*

The model revealed a significant main effect for Congruence ($F(1, 319)=23.47$, $p<0.001$). Group and the Group x Congruence predictors did not reveal significant results ($p>0.05$, for both predictors). The model was then releveled to obtain contrasts across the different congruence levels. The results indicated that for both EMI and FMI participants, response latencies, unlike accuracy rates, were significantly different across the three conditions: Congruent-trial reaction times were significantly smaller than those on the control ($\beta= 0.12$, SE=0.04, $t= 2.98$, $p<0.001$); and incongruent trials ($\beta= 0.51$, SE=0.04, $t=12.75$, $p<0.0001$). Furthermore, control-trial RTs were significantly smaller than incongruent-trial RTs ($\beta= 0.39$, SE=0.04, $t=9.77$, $p<0.0001$) (see Figure 4.6).

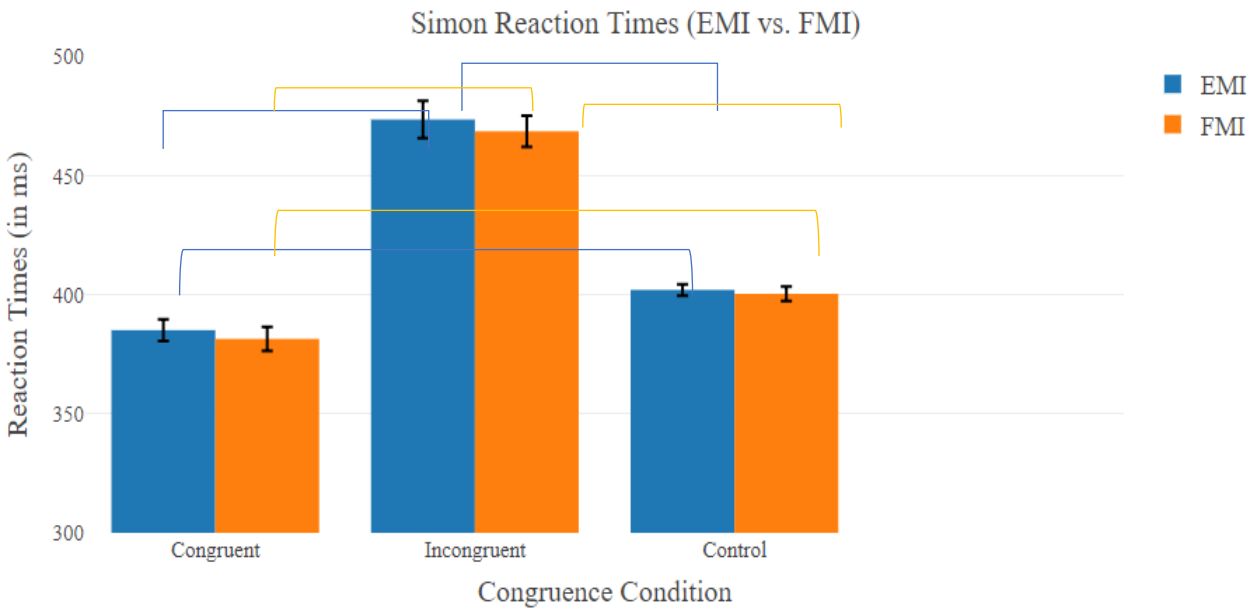


Figure 4.6 Simon RTs – Congruence contrasts (brackets = significant differences)

As one of its contributions to the literature, the present study set out to examine facilitation on congruent trials and interference on incongruent trials in the Simon Task by using control-trial performance as the baseline. For both groups, congruent trials had significantly smaller RTs than the control trials, and incongruent trials had significantly larger RTs than the control trials. Accordingly, a facilitation effect for congruent trials and an interference effect in the incongruent trials are evident in the task in terms of response latencies. However, it is difficult to establish whether the source of facilitation or interference was the stimulus location, stimulus direction, both location and direction, or the overlap/mismatch between one of the stimulus dimensions and the response dimension. It is proposed, therefore, that examining accuracy may potentially shed light on the dimension(s) influencing performance.

Responses on congruent and control trials were not significantly different in terms of accuracy. In other words, whether an arrow was positioned left, right, or centre did not result in higher accuracy as long as the direction of the arrow and response

position were congruent. Accordingly, the influence of facilitation due to the overlap between stimulus location and stimulus direction could not be established, and the source of facilitation appears to be that of stimulus direction. It may therefore be possible to claim that interference in incongruent trials, which were significantly less accurate than control trials, could be attributed to the conflict between stimulus direction—not location—and the response dimension. Whether it is possible to extend the claims from accuracy rates to response latencies is uncertain, yet this study proposes teasing apart the possible sources of conflict incurred by the mismatch of stimuli dimensions and response dimension in this version of the Simon Task.

4.3.2.3 Simon Effect

4.3.2.3.1 Accuracy

To compute the Simon Effect in terms of accuracy, the number of errors in congruent trials was subtracted from the number of errors in incongruent trials. Upon calculating the score for each participant, data were checked for normality, and it was revealed through the Shapiro-Wilk test that the distribution was not normal ($p<0.0001$). As there were several ties in the scores, where the Simon Effect for accuracy ranged between -2 and 6, a non-parametric Fisher permutation test was used. The test revealed that the two groups were not significantly different in terms of the accuracy aspect of Simon Effect ($z= 0.193$; $p=0.156$). However, as with the Conflict Effect, the total number of errors was generally low, with some participants (N=9 in both groups) making more errors on congruent than incongruent trials, leading to negative results. Furthermore, a large number of participants performed similarly (same number of errors) on both congruent and incongruent trials (N=16 for both groups). Accordingly, the accuracy element of the Simon Effect was not investigated in further analyses, as the negative numbers and 'zero' differences make this construct's computation fallacious.

As indicated in the Conflict Effect accuracy section, a more accurate measure to investigate differences between congruent and incongruent trials would be through the use of generalised linear models, as opposed to mathematically subtracting the number of errors between the trials and then comparing these differences.

4.3.2.3.2 Reaction Time

For reaction times, the Simon Effect is computed as the difference between mean RTs on incongruent trials and the mean RTs on congruent trials. Upon calculating the Simon Effect for each participant, an inspection of the data indicated a non-normal distribution (Shapiro-Wilk Test, $p<0.05$). In order to transform the data to normality, the data was squared and then analysed using the non-parametric Fisher permutation test from the 'Perm' package (Fay & Shaw, 2010). As there were several ties between the participants, thus precluding the use of the Mann-Whitney-Wilcoxon test, the Fisher permutation test was used to assess whether EMI and FMI groups (independent variable) differ in their Simon Effect scores (dependent variable). Results indicated that the two groups were not significantly different in terms of their inhibitory control ($z=0.276$, p-value=0.78).

To summarise, on the Simon Task, EMI and FMI participants performed similarly in terms of accuracy, reaction times, and Simon Effect scores. The two groups were significantly more accurate and faster on congruent and control trials than on incongruent trials. The findings are contrary to the prediction that FMI participants would outperform EMI participants, given the assumption that FMI participants' frequent use of and codeswitching between *three* languages might confer cognitive advantages, especially in terms of inhibitory control skills. However, as noted in the Conflict Effect section, the range of scores on the Simon Effect was big (56 – 114), so participants within each group may exhibit different inhibitory control capacities due to

other variables (e.g., differences in language experience) which are investigated in Section 4.7. Contrasts between the Simon Effect and Conflict Effect are also further elaborated upon in the RQ1D.

4.3.3 RQ1C – BDST

In bilingualism research, approaches to BDST score analysis are highly varied (Leeser & Sunderman, 2016). To compare participants, raw scores are generally converted to z-scores and then analysed through t-tests or ANOVAs. However, the assumptions of these tests are sometimes overlooked as the non-normal distribution often aids in categorising participants. Many studies (e.g., Sunderman & Kroll, 2006) group participants based on their WM scores as having high vs. low WM capacities. However, the construct being measured is not dichotomous in nature, and categorising participants based on their performance in this task is both subjective and problematic. For instance, it may be difficult to establish the score or range of scores which could be interpreted as 'high' or 'low'. This division strategy is also contingent on a specific sample's performance; the mean for each sample could be different, and therefore categorising a group with high WM capacity (e.g., a raw score of 11) could differ from another sample where the mean is 11 and the 'high-capacity' group has a range of 13-14.

Furthermore, BDST scores are an individual-level score—not a group measure—according to MacCallum, Zhang, Preacher, and Rucker (2002) whose methodological paper critiqued the dichotomisation of WM measures on a conceptual and a statistical level. The researchers argue that this dichotomisation may result in a misinterpretation of the construct underlying the measure, a loss of effect size and power, and a loss of measurement reliability. Other studies conceptualise the BDST score, which usually ranges between 0 and 14, as a continuous variable and analyse scores using correlations and/or regressions. However, as it is a count variable with a ceiling, it

should not be treated as a continuous variable or analysed as one. Leeser and Sunderman (2016) attribute the common use of either approaches to BDST (and WM) scores to ease of use and/or to underestimating the implications of dichotomising variables such as WM.

This study avoids the aforementioned conceptual and statistical pitfalls of analysing BDST scores as described. Upon checking the distribution of the raw scores, it was determined that it was non-normal, so a generalised linear model with a binomial link function (Guo & Zhao, 2000) was employed. The GLMM model was also chosen because the BDST score is a discrete count variable with a ceiling, and there were numerous ties between participants given the small range; all participants scored between 8 and 14 (see Figure 4.7). Participants' raw scores were fitted as the response, and the two groups (EMI or FMI) were fitted as the independent factor. As the goodness-of-fit chi-squared test of residual deviance was not appropriate due to the saturated model, the validity of the model's goodness-of-fit was determined with a half-normal Q-Q plot, as recommended by Collet (2002:129). Visual inspection of the half-norm Q-Q plot revealed that the distribution of the scores were within the envelope, indicating the model's goodness-of-fit. The results of the generalised linear model analysis, summarised in Table 4.14, indicated that there was no significant difference between the EMI and FMI participants on this task ($p=0.252$); that is, the two groups exhibited similar working memory capacities.

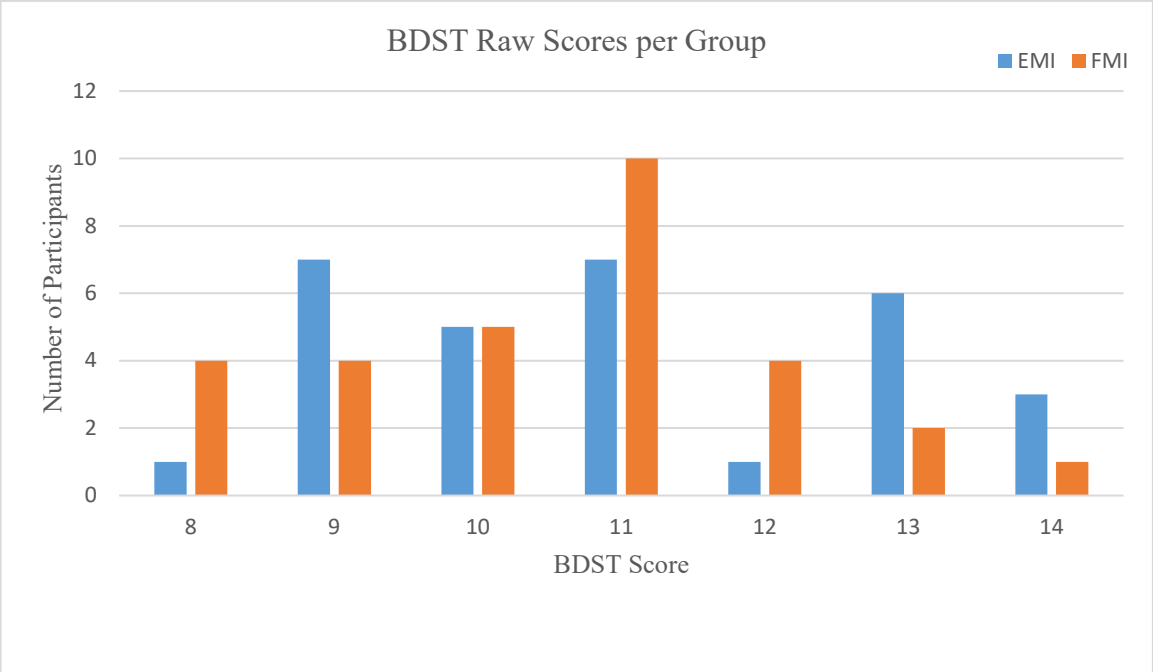


Figure 4.7 BDST Raw Scores

Table 4.14 GLMM output for BDST scores (Estimated coefficients, standard errors (SE), z-values, and p-values). Asterisks (*) indicate: p< .05 (*), p< .01 (**),p< .001(***) and p< .0001 (****)

	BDST Score			
	B	SE	z-value	Pr(>|z|)
Intercept	1.2993	0.1189	10.926	<2e-16***
Group (EMI vs. FMI)	-0.1879	0.16141	-1.145	0.252

Formula: BDST ~ Group + (1|Participant)

4.3.4 RQ1D – Correlations between Working Memory, Conflict Effect, and Simon Effect

RQ1D: Are participants' working memory scores, Simon Effect scores, and/or Conflict Effect scores correlated?

The last part of the first research question investigated the possible correlations in performance on the cognitive tasks. Specifically, it addressed whether participants[9] who have better working memory capacity also have lower Simon Effect scores and/or lower

[9] This sub-question was not concerned with group differences. Further, as the two groups did not significantly differ in performance on any of the cognitive tasks, the overall correlations include all 60 participants. Exploratory between-group differences in correlations revealed similar results, and these are offered schematically in Figure 4.8.

Conflict Effect scores. As discussed in Chapters 2 and 3, working memory involves an element of inhibition (Miyake et al., 2000), and it is in less demand in the flanker task, as maintaining instructions to perform this task is simple. However, in the Simon task, working memory is in higher demand given the need to remember to dismiss the stimulus location and respond based on the stimulus direction. Furthermore, this RQ investigates whether the Simon Effect scores and the Conflict Effect scores are correlated, as both the Simon and Flanker tasks are said to measure inhibitory control, specifically interference suppression. However, the manipulation of trial distribution on the flanker task added to the complexity of the task by increasing demands for conflict monitoring. As for the Simon task, the inclusion of a bivalent stimulus was the source of complexity for the task. While correlations were expected on a theoretical basis, there has been a lack convergent validity between different tasks which attempt to measure executive functions (Paap & Sawi, 2014). Accordingly, as non-traditional versions of the tasks were employed in the present study, it was worthwhile investigating whether there was a correlation in performance on the three tasks.

Working memory for each participant was measured on a scale of 0-14, whereas the Simon Effect and Conflict Effect scores were computed as a difference between incongruent RTs and congruent RTs and were therefore continuous variables. The distribution for each of the three variables was not normal, yet the data analysed and reported in this research question were not transformed. This is because analyses were run on both transformed and non-transformed data, and the results were approximately identical. Although the data was not normally distributed and had tied ranks, the use of non-parametric tests such as Spearman's Rho and Kendall's Tau could not be performed as the data was continuous and not ordinal (Fields et al., 2012). The best way to investigate possible correlations between the three variables was therefore to use

Pearson's R. The R statistical 'stats' package was then used to perform correlations, and a correlation matrix is offered in Table 4.15. A scatterplot of the correlation matrix is offered in Figure 4.8.

Table 4.15 Correlation Coefficient Matrix of Cognitive Tasks

	Working Memory	**Conflict Effect**	**Simon Effect**
Working Memory	1.0000	-0.4572	-0.3165
Conflict Effect	-0.4572	1.0000	0.04574
Simon Effect	-0.3165	0.0457	1.0000

Results revealed that for both EMI and FMI participants, working memory and Conflict Effect were negatively correlated ($r= -0.46$, $p<0.001$), as were working memory and the Simon Effect ($r= -0.32$, $p<0.05$). In other words, the better working memory capacity a participant had (i.e., the higher their BDST score), the lower their Conflict Effect and Simon Effect scores. That is, increased working memory capacity was related to better inhibitory control skills. This is in line with the predictions made by this study. However, the Simon effect and Conflict Effect did not seem to be correlated ($r=0.05$, $p=0.73$), which was contrary to expectations. This finding indicates that although the two tasks measure an underlying inhibitory mechanism, the type of inhibitory control processes may be different. Claims about the correlations between the Simon Effect and Conflict Effect come from studies that have used the traditional Simon task, however. As the arrow version of the Simon task was used, it was expected to involve higher complexity than the traditional Simon task. This may explain the lack of correlation between the arrow Simon and the flanker tasks.

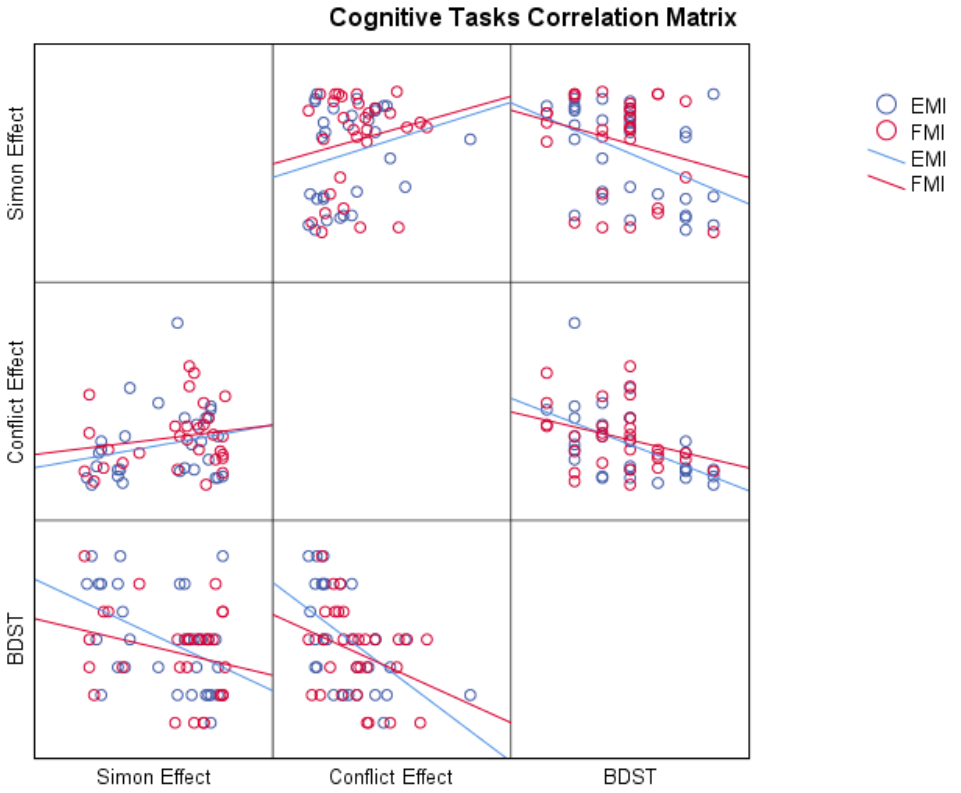


Figure 4.8 Scatterplot of Cognitive Task Correlations

An important observation regarding the Simon Effect scores is the difference in the mean and range from those of the Conflict Effect scores: Simon Effect range=56-114; Conflict Effect range= 0-97. The higher Simon Effect scores may be attributed to the more complex nature of the Simon Task, especially the arrow version, which arguably entails more inhibitory mechanisms than the flanker task which is perceived as being less demanding. Another interesting and related finding is the difference in performance on the Flanker and Simon tasks as a result of the manipulation of congruent-to-incongruent trials. Trials in the Simon task were equally distributed between congruent and incongruent trials (50-50 distribution), in order to increase the complexity of the task. This manipulation is similar to the distribution of trials in the high-monitoring condition of the flanker task. However, the different tasks yielded disparate results: while the 50-50

distribution of congruent to incongruent Simon trials resulted in significantly lower accuracy rates, the same distribution resulted in similar accuracy rates in the flanker task, with incongruent trials being marginally—though not statistically—more accurate than congruent trials. However, for RTs, both tasks incurred longer RTs for the 50-50 distribution.

One possible explanation for the different effect of the 50-50 distribution for accuracy rates is that the Flanker and Simon tasks tap into different cognitive processes, which is further confirmed by the lack of correlation between the two tasks. Specifically, while both tasks are said to measure interference suppression, the arrow Simon task seems to requires increased response inhibition, whereas the flanker task requires more conflict monitoring. Alternatively, it is also possible that the additional dimension (bivalent stimulus) in the Simon task resulted in increased errors. As such, it can be argued that the distribution of trials may not elicit a significant difference in accuracy rates in tasks of a less complex nature, such as the Flanker. Conversely, one could argue that the 50-50 manipulation was successful in increasing conflict monitoring capacities on the Flanker. Another possible explanation is that the 50-50 distribution added a third 'dimension' to the Simon task—that is, it increased the cognitive load of the task which already had a bivalent stimulus to add to the complexity. Accordingly, the lower accuracy rates of the Simon task could be a result of this additional load. As the Simon task was only administered in the most complex distribution, it is difficult to categorically ascribe performance to the distribution of the trials as opposed to (1) the nature of the task, (2) its complexity with respect to including two dimensions, or (3) the cognitive processes it requires.

4.3.5 Summary of Cognitive Task Performance

This section presented the results of the first RQ in its four subdivisions. It investigated whether EMI and FMI participants exhibited comparable performance on domain general tasks, namely the flanker task, the Simon Task, and the BDST. It also investigated whether the scores obtained from these tasks were correlated. Most predictions regarding the overall behaviour of participants, such as having higher accuracy rates and smaller RTs on congruent tasks, were confirmed. However, the hypothesis that the two groups would exhibit different performance on the tasks was not validated. This is therefore in contrast to the hypothesis that balanced trilinguals outperform unbalanced trilinguals on cognitive tasks.

This section also discussed shortcomings of the computation and analysis of certain constructs, such as 'monitoring costs', as they failed to paint an accurate picture of the data. However, upon correlating the different measures, significant correlations were observed between BDST and Conflict Effect scores and between BDST scores and Simon Effect scores. However, scores on the Simon and Flanker tasks were not significantly correlated, which was attributed to task design and manipulation, leading to marginally overlapping cognitive processes underlying the two tasks. As mentioned throughout the chapter, the measures, while seemingly different, may tap into different cognitive mechanisms and correlate with other variables (e.g., codeswitching frequency) which are investigated in RQ2. Accordingly, variation within each group is further explored, and the analytical strategy adopted will allow for examining the data at the participant-level.

4.4 ANALYSIS OF THE INDEPENDENT VARIABLES

This section offers an analysis of the instruments measuring the predictor variables in this study, namely the Oral Proficiency Tests (OPT), Verbal Fluency Task (VFT), and Language Background Questionnaires (LBQ). This allows addressing the second research which examines the influence of language proficiency and language experience on cognitive and linguistic task performance.

4.4.1 Oral Proficiency

4.4.1.1 Oral Proficiency Tests (OPT)

As discussed in Chapter 3, oral proficiency test scores were obtained for Arabic, English, and French from three raters. For each language, the raters included (a) the researcher, (b) the examiner who conducted the interview, and (c) an external rater who holds a master's degree in linguistics and teaches each respective language at university level. Accordingly, each participant was rated by seven raters for the three proficiency tests. Despite the variation in raters across the languages, all participants were rated by the same raters for each language. In order to assess the inter-rater reliability of the scoring, an alternative to Cohen's kappa, the Fleiss' kappa (Fleiss & Cohen, 1973) was initially selected. This generalised reliability measure was chosen as there were more than two raters—a drawback of Cohen's kappa (McHugh, 2012)—and because the scores were ordinal (numerical). However, as there was very high agreement among the raters, the Fleiss' kappa could not be used as it proved inadequate in interpreting the results. This is termed as the paradoxical behaviour of the Fleiss' Kappa, as it reports a lower coefficient or even a negative coefficient based on its assumptions, rather than indexing a higher, positive coefficient when there is very high agreement among raters (Falotico & Quatto, 2015). Accordingly, as recommended by Hallgren (2012), intraclass correlations were used to assess agreement, where the coefficient is calculated as the

ratio of between-subjects variation to total variation[10]. There was very high agreement across all proficiency ratings with intra-class coefficients of 0.960 for Arabic, 0.966 for English, and 0.989 for French.

In order to confirm the hypothesis that the two groups differed in terms of their language proficiency, sliding contrasts on the mean scores of the three languages were conducted using the 'contr.slide' function of the lmer package in R. Results revealed that for the EMI group, Arabic and English were not significantly different (β=0.13, SE=0.17, t=0.78, p>0.05), Arabic was significantly different from French (β= 2.30, SE=0.24, t=9.76, p<0.0001), and English was significantly different from French (β= 2.433, SE=0.26, t=9.32, p<0.0001). The hypothesis that this group was more balanced in two languages instead of three was therefore supported.

For the FMI group, sliding contrasts revealed no significant differences across the three sets of language pairs. That is, proficiency scores in Arabic were not significantly different from English (β= 0.10, SE= 0.15, t=0.65, p>0.05) or French (β=0.267, SE=0.17, t=1.610, p>0.05). English proficiency scores were also not significantly different from French scores (β= 0.17, SE= 0.19, t=0.90, p>0.05). These contrasts therefore confirm that FMI participants are balanced trilinguals given their similar proficiency scores on all three languages.

The scores for each language were then analysed across the two groups. Table 4.16 offers descriptive statistics of the means in each language. The data were first checked for normality. Results from the Shapiro-Wilks test indicated that the data was not normally distributed for any language for either group (p< 0.05 in all conditions).

[10] Based on Shrout and Fleiss (1979), the following formula was developed:
Individual Variance (between subjects) / Individual Variance + Rater Variance (Between raters) + Residual variance (Error). In SPSS, this was considered as Type A, given that all participants had the same raters for each language. The computation was therefore two-way mixed effects/absolute agreement.

Accordingly, non-parametric tests were conducted to explore between-group differences. As the assumptions of the Mann-Whitney U test were violated due to the presence of numerous ties, a Fisher (exact) permutation test was selected to analyse the data. Results revealed that the EMI and FMI participants did not significantly differ in the distribution of their Arabic proficiency scores ($p=0.75$) and English proficiency scores ($p=0.29$), but they were significantly different in their French scores ($p<0.0001$) where FMI participants outperformed their EMI counterparts. These analyses therefore support the assumption that the EMI and FMI groups differ in their *third*-language proficiency.

Table 4.16 Means and Standard Deviations (in parentheses) of OPT Scores

	EMI		**FMI**	
	M	*SD*	*M*	*SD*
OPT Arabic	8.60	(0.68)	8.53	(0.68)
OPT English	8.73	(0.74)	8.56	(0.67)
OPT French	6.30	(1.42)	8.81	(0.76)

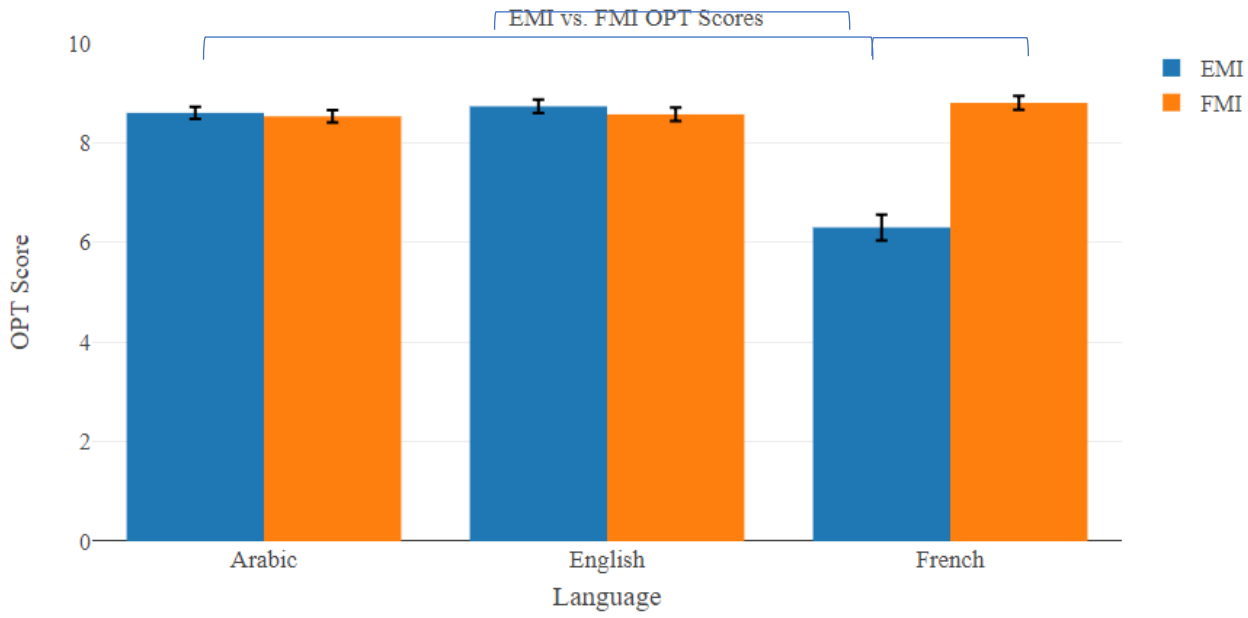


Figure 4.9 OPT scores by group, bars= Standard Errors, brackets= significant difference

4.4.1.1.3 High Proficiency in One, Two, or Three Languages

It should be noted that while the analysis of between-group differences offers insights into the general sample and confirms the hypotheses given the different

immersion programmes, individual differences were considered in the analyses, even when no significant group differences were observed. This is achieved by including participant-level analyses which, in mixed effects models, constitute part of the Random Effects, to ensure that no over-generalisations are made. This is an advantage of using mixed-effects models, as some participants in the EMI may be considered balanced trilinguals, whereas some FMI participants are considered balanced bilinguals. In addition to Random Effects, participant variation is also explored through the 'high-proficiency mono/bi/trilingualism' variable (hereafter referred to as HPBT) which categorised each participant as highly proficient in one, two, or three languages based on the OPT scores. Proficiency scores in any language of 8 and higher indicated that participants were highly proficient in said language. As all participants achieved a score of 8 in at least two languages, the variable was then transformed into a dichotomous one (i.e., a binary variable) in which participants were categorised as highly proficient in either two or three languages. The descriptive results for each group are summarised in Table 4.17. The HPBT variable was especially extracted from the OPT task scores to explore whether being highly proficient in three languages, as opposed to two, would significantly influence cognitive and linguistic task performance.

Table 4.17 High Proficiency in Two vs. Three Languages (HPBT)

	Two languages	**Three Languages**
EMI	25	5
FMI	8	22

4.4.1.2 Verbal Fluency Task (VFT)

In addition to the scores obtained from the OPTs, scores from the Verbal Fluency Task (VFT) were also analysed for each group. It was hypothesised that FMI participants would outperform their EMI counterparts on all linguistic tasks, as well as on the VFT,

especially in the weakest language. Although this project assesses performance on this task in order to compare current findings with previous ones, it does not use the VFT scores as the sole indices of language proficiency and/or dominance. Rather, it employs the task, which for literature-motivated reasons is subsumed under the overarching 'proficiency' variable, as a hybrid cognitive-linguistic measure. This is because it involves language production as well as an element of executive functions, namely working memory and (self) monitoring, as instructions and produced words need to be maintained throughout every category and each language. As discussed in Chapter 3, results from the VFT were therefore examined in terms of whether they correlate with proficiency scores from the OPT.

Furthermore, VFT scores for each language were examined as predictor variables of both cognitive and linguistic task performance. As a corollary investigation, exploring differences between of OPT and VFT scores in predicting performance was conducted to assess whether the findings are commensurate. This allowed for a more detailed discussion of the use—and usefulness—of the VFT as a pure linguistic measure, as has been used in some of the literature.

The VFT data were not normally distributed between the languages within and across the groups, especially in the French-language (Shapiro-Wilk's test, $p<0.0001$), and this was further verified by examining the normal Q-Q plots and histograms. Accordingly, non-parametric tests were used to analyse within- and between-group differences across the three language conditions. Specifically, a negative binomial regression model was used as the VFT scores were over-dispersed count data; that is, the variances exceeded the means of the scores across the three languages. While a Poisson regression model was initially tested to analyse this data, the dispersion parameter indicated that a negative binomial generalised linear model was more appropriate (Long

& Freese, 2006). The goodness-of-fit of the negative binomial model was assessed by chi-squared testing of the deviance residuals. Results indicated that the model was to be accepted, and the repeated measures multilevel model was fitted using the 'nb' and the Laplace Approximation (glmerMod) functions in the MASS library. Table 4.18 summarises the descriptive data from the VFT. The model revealed that there was a significant main effect of Group as a predictor variable ($p<0.0001$), as well as a significant interaction between Group x Language (all levels having $p<0.001$). These results are further analysed through post-hoc analyses discussed next.

Table 4.18 Means and Standard Deviations (in parentheses) of the VFT Scores

	EMI		**FMI**	
	M	*SD*	*M*	*SD*
VFT Arabic	101.93	(15.385)	113.73	(14.814)
VFT English	159.67	(19.810)	121.00	(16.973)
VFT French	60.03	(14.712)	152.90	(19.787)

In order to explore the significant interaction, planned contrasts were performed by changing the reference level (i.e., the intercept) for each level of the VFT and Group variables. The distribution of the scores across the languages for each group is presented in Figure 4.10. The results of the analyses revealed that the VFT scores of the EMI and FMI groups were significantly different in each language condition: Arabic ($\beta= 0.11$, $SE= 0.04$, $z=2.76$, $p<0.01$), English ($\beta= 0.28$, $SE= 0.4$, $z=7.3$, $p<0.0001$), and French ($\beta= 0.28$, $SE= 0.4$, $t=22.46$, $p<0.0001$). Specifically, EMI participants produced significantly more exemplars in English than FMI pariticpants (by approximately 12 words per category), whereas FMI participants produced significantly more exemplars than the EMI group in both Arabic (by 10 words per category) and French (by 15 words per category). These results do not align with those obtained from the OPTs, as the two groups were only significantly different in their French proficiency scores.

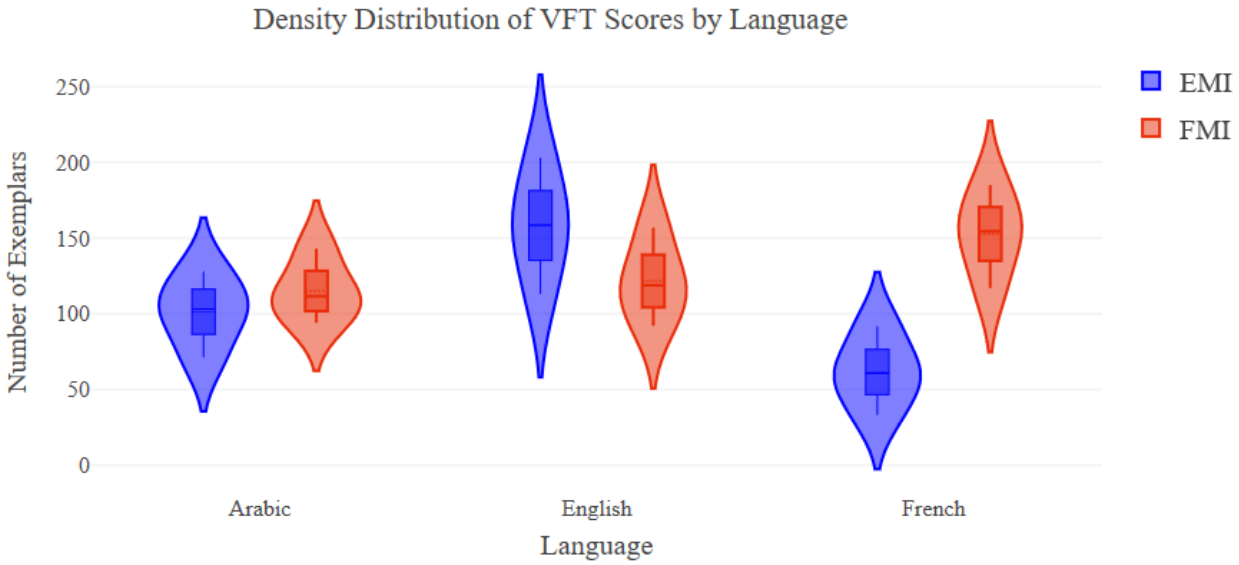


Figure 4.10 Violin plots showing the distribution of the correct exemplars for the VFT for each language and group. The plot outline shows the density of data points at the different scores. The boxplot shows the interquartile range, and the thin vertical line indicates minimum and maximum scores for each language. The solid horizontal line represents the mean.

Refitting the negative binomial regression model offered contrasts for each group's scores in the three language pairs: Arabic vs. English, Arabic vs. French, and English vs. French. As summarised in Table 4.19, EMI participants produced significantly more exemplars in English, followed by Arabic, and then French. In order to further interpret the results, exponentiation of the β-estimates allowed for the calculation of incidence rate ratios. Accordingly, for the overall task, EMI participants were more likely to produce 60 words more in English than in Arabic, 159 more words in English than in French, and 102 more words in Arabic than in French. Put differently, for each category in English, EMI participants were able to produce (in 60 seconds) six more exemplars than in Arabic and 16 more exemplars than in French, as well as 10 more exemplars in Arabic than in French.

Table 4.19 VFT Language-Pair Contrasts (EMI)

	β	SE	z-value	p-value
Arabic vs. English	0.45	0.04	12.06	0.00****
Arabic vs. French	0.53	0.04	12.73	0.00****
English vs. French	0.98	0.04	24.46	0.00****

Data from the FMI group were analysed in the same manner, with the negative binomial model refitted to obtain contrasts for each language pair. The results, summarised in Table 4.20, indicated that FMI participants produced significantly more exemplars in French than in Arabic and English ($p<0.000$ for each contrast), yet they produced a similar number of exemplars in English and Arabic ($p>0.05$). The incidence rate ratios for these contrasts revealed that FMI participants were able to produce 12 more exemplars in French than in English and 11 more than in Arabic for each category in 60 seconds. The difference between English and Arabic exemplars for each category was approximately one exemplar, which explains the non-significant difference between the two language overall scores.

Table 4.20 VFT Language-Pair Contrasts (FMI)

	β	SE	z-value	p-value
Arabic vs. English	0.06	0.04	1.65	0.10
Arabic vs. French	0.29	0.04	8.03	0.00
English vs. French	0.23	0.04	6.39	0.00

As with the between-group analyses, the within-group analyses of the VFT scores depict the two groups rather differently from the results offered by the OPT scores. The VFT results would lead to the identification of the EMI group as an unbalanced trilingual group; i.e., that EMI participants are English-dominant with a significantly higher proficiency in Arabic (L2) than in French (L3). Furthermore, the FMI group would also be classified as an unbalanced trilingual group which is French-dominant with a non-significantly different L2-L3 proficiencies (in English and Arabic). Given previous interpretations of VFT scores in the literature, this would lead to the (mis)classification of the two groups, as these results are inconsistent with the scores obtained from the OPTs. Furthermore, it is difficult to determine which language is considered the L2 or L3 in the FMI group. Taken together, these findings led to the necessary exploratory

examination of whether the two groups' performance on the VFT and OPT scores are statistically correlated, offered in the next section.

4.4.1.3 OPT vs. VFT scores

In order to test the validity of using the VFT as an index of proficiency, this section aimed to assess whether VFT scores and OPT scores statistically align. A Pearson product-moment correlation coefficient was computed to explore the relationship between participants' performance on the VFT and the OPT in each language condition. The results revealed that there were significant correlations between the VFT and OPT scores when data from both groups was aggregated: Arabic, $r(58)= 0.32$, $p<0.05$; English, $r(58)=0.28$, $p<0.05$; and French, $r(29)=0.79$, $p<0.0001$. However, when conducting within-group analyses, there were conflicting findings in terms of the VFT and OPT scores matching. Sample size, effect size, and statistical power analyses which were determined both *a priori* and *a posteriori* were checked for potential explanations of these results, yet as these criteria were met (effect sizes ranged between 0.7 and 0.9), it was necessary to probe into the within-group analyses, as data from the VFT is typically used to determine differences in language proficiency within each group. Therefore, it is counter-intuitive to adopt an instrument which provides a potentially inaccurate basis of classifying languages as L1, L2, and L3.

For the EMI group, correlation analyses between the VFT scores and OPT scores revealed disparate results: Arabic, $r(29)= 0.349$, $p=0.058$; English, $r(29)=0.323$, $p=0.082$; and French, $r(29)=0.646$, $p<0.0001$. Accordingly, the VFT scores only correlated with the significantly lower-proficiency language for this group. An overlay scatter plot for the three languages is presented in Figure 4.11.

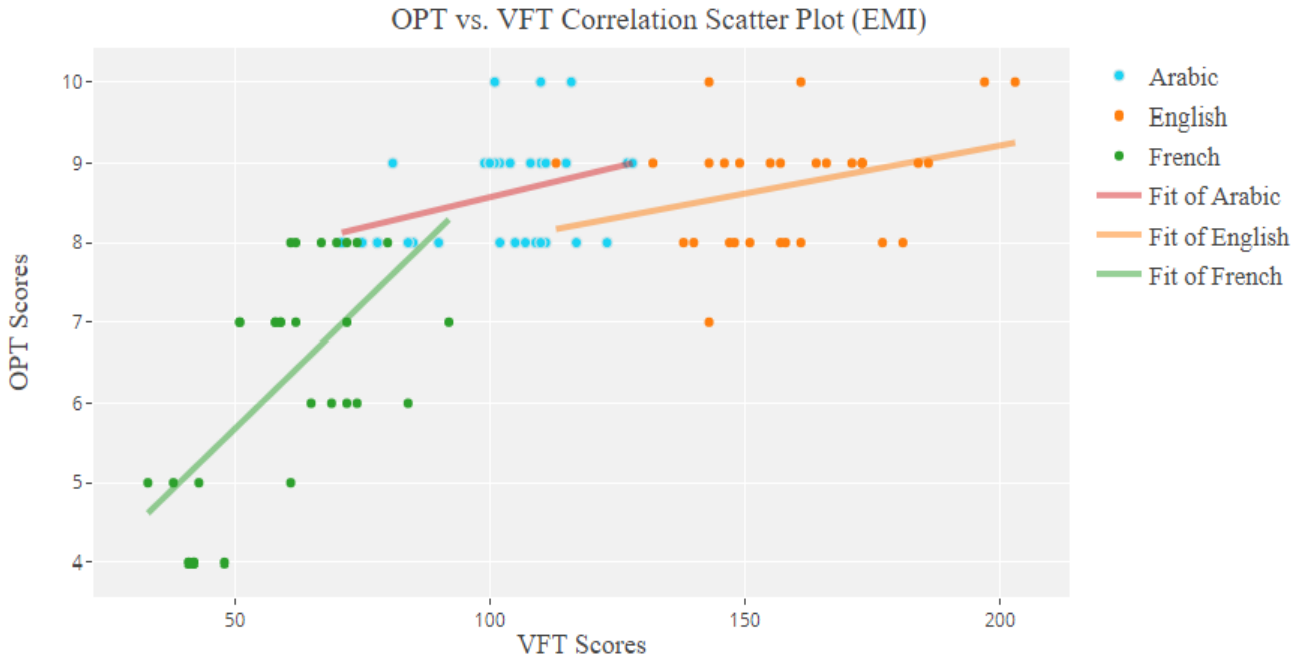


Figure 4.11 Correlations between OPT and VFT scores (EMI)

As for the FMI group, correlation analyses between the VFT scores and the OPT scores revealed further conflicting results: Arabic, $r(29)= 0.431$, $p<0.05$; English, $r(29)=0.212$, $p=0.260$; and French, $r(29)=0.354$, $p=0.055$. That is, only the Arabic scores on the VFT and OPT correlated with each other for this group. Again, with the FMI group, the weakest language as indicated by the OPT was correlated with lowest VFT scores. A scatterplot for each language is presented in Figure 4.12.

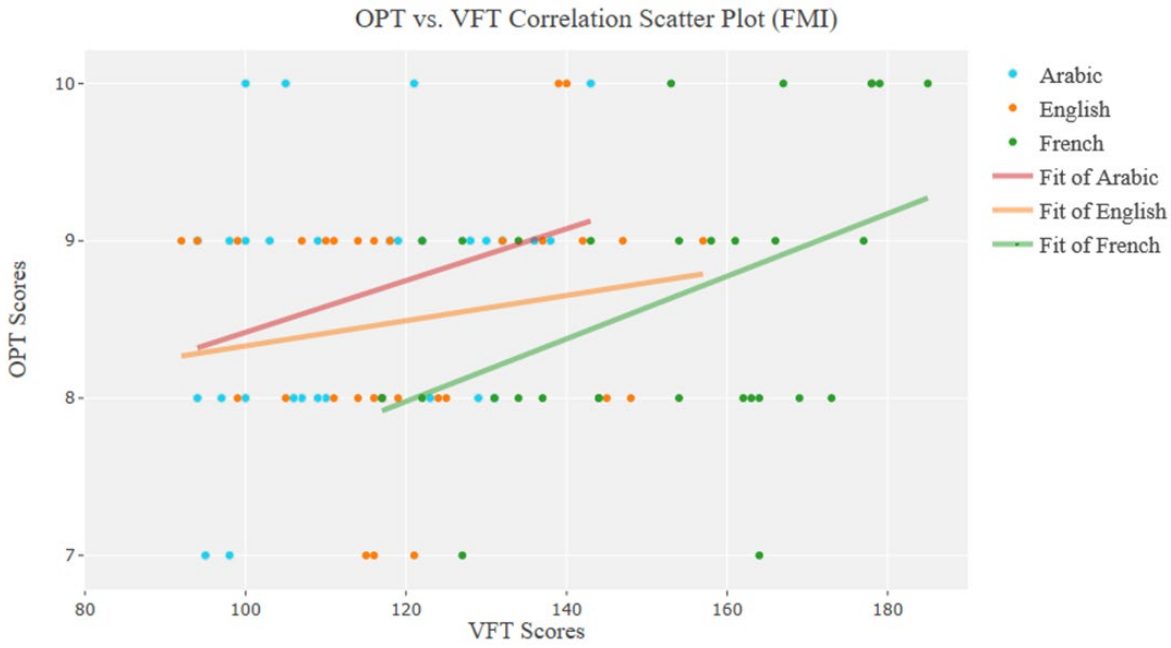


Figure 4.12 Correlations between OPT and VFT Scores (FMI)

These findings might therefore indicate that the VFT is only sensitive in detecting the weakest language among the three, although this weakest language may not be significantly different from the others when measured by the OPT. Despite the correlation for the weakest language of each group, the lack of correlation for the other languages must be addressed. It may be attributed to several factors, such as the validity of either (or both) the OPT or (and) the VFT as measures of proficiency. It could also be due to the nature of the tasks; while both are productive linguistic tasks, they require different processes with the OPT tapping into general proficiency and the VFT specifically tapping into only specific areas of proficiency (lexical access and retrieval). Furthermore, the lack of correlation could possibly stem from the confounding element of working memory which the VFT possesses. While it is difficult to ascertain which reason, or perhaps set of reasons, underlie the lack of correlation, it is imperative that all analyses, as has been done in this study, differentiate the two instruments in their predictive ability as factors in the mixed-effects models. Moreover, unlike previous

studies, the present project does not use the VFT to determine the status of each language as the most/more dominant language. This would have ramifications on the interpretation of the linguistic tasks; thus, OPT, VFT, and linguistic tasks (picture-naming and narration) performance are retained as 'per language' as opposed to assigning L1, L2, and L3 for each language for either group. This is further exemplified and discussed in Chapter 5.

4.4.2 Language Experience

Language experience was operationalised as participants' self-reported language use, language exposure, codeswitching behaviour, and salient cue detection and measured by the language background questionnaire (LBQ) (Appendix A). The data were entered into the SPSS software (version 25) for descriptive analyses which compared group means, frequencies, and percentages. Inferential statistics were conducted through R (MASS package) to examine group differences. In all analyses, individual Likert items were treated as ordinal data, but when combined under the main constructs they were measuring, multi-item scales were treated as continuous variables. It should be noted that the constructs were not considered as *latent* variables as they could be directly measured through self-reported frequency-of-use. Therefore no inferential statistics were conducted in the analysis of this variable in itself.

Although the LBQ was set out initially to capture as much information as possible about each participant's language experience and use, only four relevant constructs were included in the analyses as predictors of cognitive and linguistic behaviour. Excluded constructs involved opinions about language status and about language use (including code-switching). The rationale behind not including these constructs is two-fold: (1) they may not have a direct influence on linguistic behaviour

and are not relevant to cognitive behaviour; and (2) they were explored as possible covariates in initial analyses and were shown to have no significant predictive power. Accordingly, the main constructs that were theoretically motivated and directly measured were: salient cue detection, code-switching frequency, language use (for each language), and language exposure (for each language). Table 4.21 offers a description of each construct and lists the relevant LBQ items. Each question investigating this construct was rated on a five-point Likert scale ranging from 'never' to 'all the time'.

Table 4.21 Language Experience -- Construct Description and Measurement

Construct	Description	LBQ Items
Code-switching Frequency	The frequency of switching between languages at home, school, with peers, and online	3E, 4E, 5E, 8E, and 9E, and 53
Salient-cue Detection	Noticing codeswitching by self and others and switching to another language due to an interlocutor switch, a topic change which requires a switch, or a (potential) breakdown in communication.	12, 13, 14, 15, 16, 17, 18, 19, 24, 25, 63, and 65
Language use	the frequency of using a language interpersonally across different contexts with different interlocutors. That is, this construct measures the overall use of each language, irrespective of how/when it is used[11].	Arabic: 3A, 4A, 5A, 6A, 8A, 9A, 10A, 36, 40, 44,47, and 50
		English: 3B, 4B, 6B, 8B, 9B, 10B, 37, 41, 45, 48, and 51
		French: 3C, 4C, 6C, 8C, 9C, 10C, 38, 42, 46, 49, and 52.
Language Exposure	The amount of input received through various mediums and contexts (excludes formal instruction)	Arabic: 1A, 2A, 7A, 11A, 26, 29, and 32
		English: 1B, 2B, 7B, 11B, 27, 30, and 33
		French: 1C, 2C, 7C, 11C, 28, 31, and 34

[11] The contexts included in this analysis exclude in-classroom language exposure and use. In the description of participant profiles (Section 3.4, Table 3.4), language use and language exposure across different contexts are offered in percentiles. The constructs used from the LBQ as predictor variables, however, examine the overall use of each language.

The data from the LBQ were analysed to establish the normality of the distribution of variables. For all variables, skewness and kurtosis levels were between -1.72 and +1.39, which is within the recommended limits of normality which are +/-2 (Field, 2013). A visual inspection of the generated histograms and Q-Q plots further suggested normality. In comparing EMI and FMI participants, homogeneity was checked with the Levene's test of equality of variance. There were no significant differences between the two groups ($p>0.5$), indicating that the variance was homogenous. Table 4.22 offers descriptive data for each construct for the EMI and FMI participants. As the number of items for each construct differs, data is presented in percentiles for ease of comparison, yet raw means were used in the analyses.

Table 4.22 Language Experience Constructs -- Descriptive Data

	EMI		FMI	
	Mean (%)	SD	Mean (%)	SD
Codeswitching Frequency	82.53	(14.61)	77.07	(14.44)
Salient Cue Detection	83.39	(12.19)	85.39	(10.42)
Arabic Language Use	36.55	(10.42)	43.61	(8.14)
English Language Use	57.63	(13.99)	33.59	(9.81)
French Language Use	5.82	(6.52)	22.79	(6.77)
Arabic Language Exposure	38.82	(10.34)	42.52	(9.21)
English Language Exposure	52.41	(12.01)	36.74	(10.18)
French Language Exposure	8.77	(4.63)	20.74	(6.28)

Although *language experience* is not a task-based variable, it was important to explore differences in the means of the constructs it comprises between the two groups. Doing so not only elucidated results obtained from further statistical analyses performed later, but it also shed light on basic contrasts in self-reported data by the EMI and FMI groups.

Codeswitching Frequency and Salient Cue Detection

As codeswitching frequency and salient cue detection were normally distributed, and as both were one-level constructs, they were treated as continuous variables which were analysed using independent-samples t-tests. The results of the analyses revealed that EMI and FMI participants were comparable in terms of their self-reported code-switching frequency ($t(58)=1.46$, $p=0.15$) and salient cue detection ($t(58)=-0.68$, $p=0.49$). That is, both groups were similar in terms of how frequently they switched between their languages, irrespective of which languages they used, and how capable they were in detecting a switch or a need for a switch due to a variety of reasons such as an interlocutor's switch in a prior turn.

Language Use

In terms of how frequently each language was used, i.e. 'language use', a general linear model was fitted with Group (2 levels: EMI and FMI) and Language (3 levels: Arabic, English, and French) as predictors, along with the interaction between the two variables, and with the self-reported frequency of language use as the outcome variable. According to this model, there were significant differences between the two groups in all three languages, as indicated by the significant interaction of Group x Language ($t=15.98$, $p<0.0001$). Although the two groups were expected to differ in their use of English and French, their significant difference in Arabic was not expected. The results of the between-group analyses are summarised in Table 4.23 and illustrated in Figure 4.13.

Table 4.23 Language Use EMI vs. FMI

	β	SE	t-value	p-value
Arabic	7.07	2.48	2.85	0.01*
English	24.05	2.48	9.82	0.00****
French	16.98	2.49	6.83	0.00****

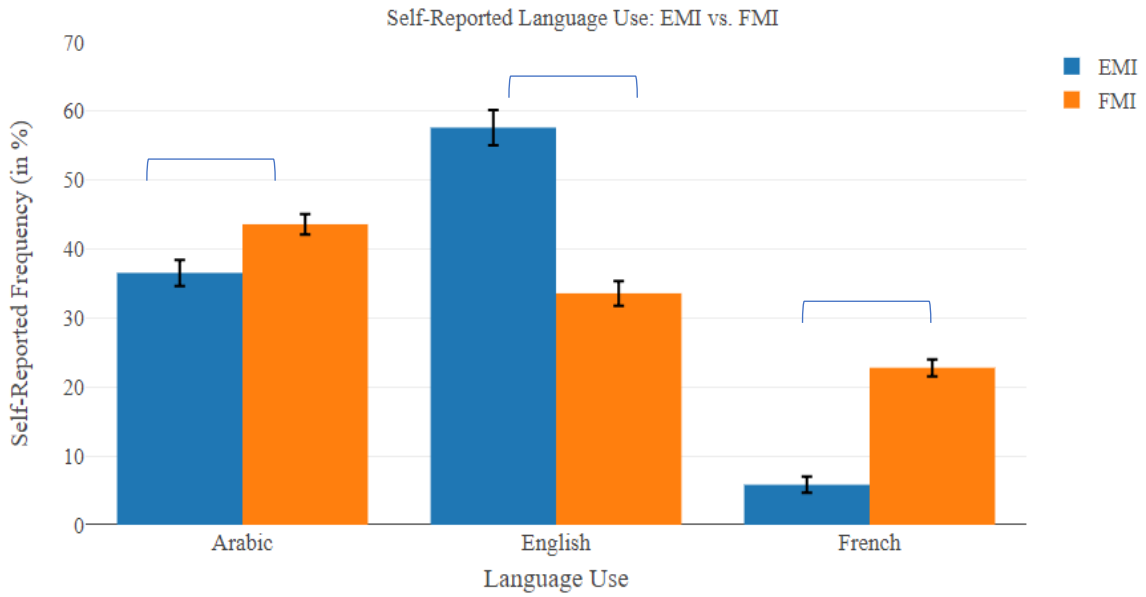


Figure 4.13 Language Use by group, error bars = standard error, brackets= significant difference

In order to examine language use for each group, the model was releveled by Group and Language by changing the reference (intercept) for each level of the variables. The results of the analysis indicated that EMI participants used English significantly more than Arabic and French, and Arabic significantly more than French. The results which show the language contrasts are summarised in Table 4.24 and are illustrated in Figure 4.14 which depicts the density of data points for each language.

Table 4.24 Language Use Contrasts (EMI)

	β	SE	t-value	p-value
Arabic vs. English	21.08	2.48	8.49	0.00***
Arabic vs. French	30.73	2.48	12.37	0.00****
English vs. French	51.81	2.48	20.86	0.00****

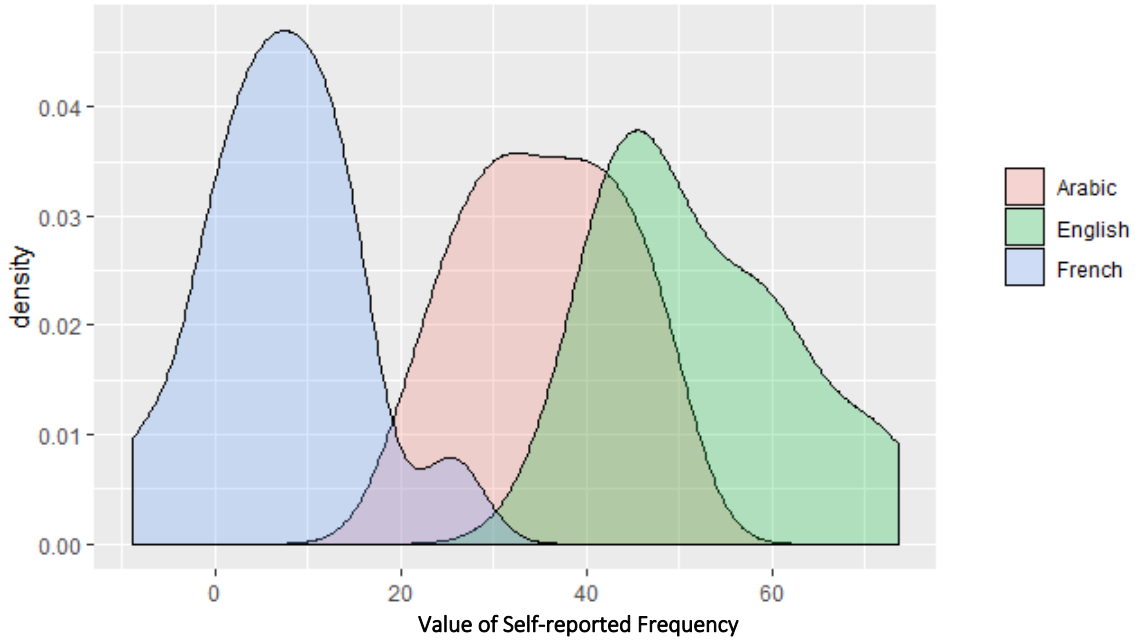


Figure 4.14 Density distribution plot of language use (EMI)

For the FMI group, results also indicated that the distribution of language use was significantly different for each language. Specifically, Arabic was used significantly more frequently used than English, and English was significantly more frequently used than French. Table 4.25 summarises the results, showing contrasts between languages in terms of use. Furthermore, Figure 4.15 depicts the density of distribution points based on self-reported language use data by the FMI group.

Table 4.25 Language Use Contrasts (FMI)

	β	SE	t-value	p-value
Arabic vs. English	10.03	2.48	4.04	0.00***
Arabic vs. French	20.81	2.48	8.38	0.00****
English vs. French	10.79	2.48	4.43	0.00****

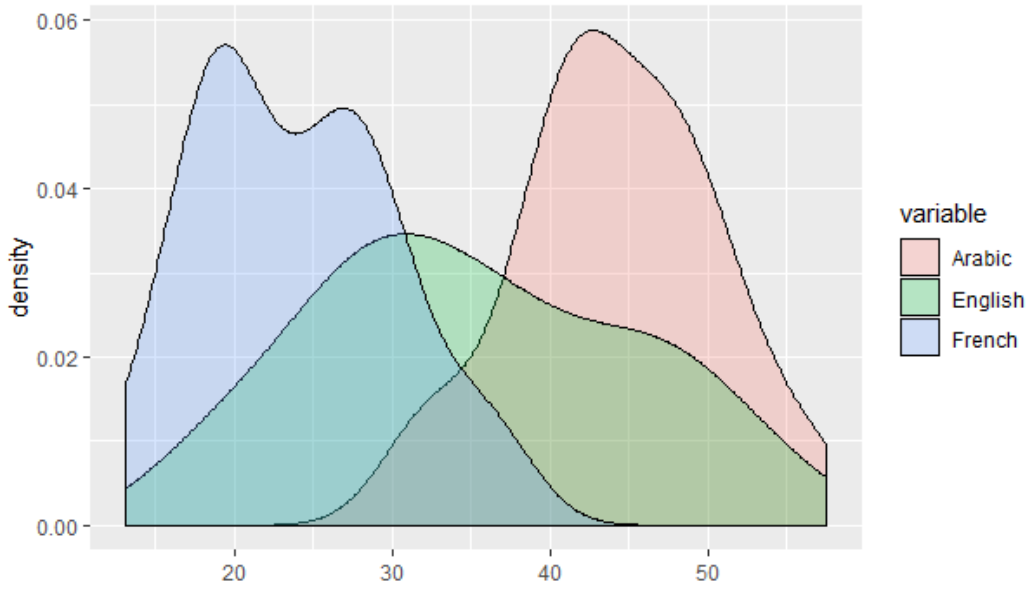


Figure 4.15 Density distribution plot of language use (FMI)

Language Exposure

As for the groups' exposure to the three languages, a general linear model was fitted with the same predictor variables (Group, Language, and Group x Language) yet with self-reported 'language exposure' data (continuous variable) as the outcome variable. The results revealed a significant Group x Language interaction ($t=9.33$, $p<0.0001$) indicating that the groups differed across the languages. Simple main effects and post-hoc analyses were therefore performed. The results, summarised in Table 4.26, indicated that EMI and FMI participants had similar exposure to Arabic, yet their exposure to English and French was significantly different.

Table 4.26 Language Exposure EMI vs. FMI

	β	SE	t-value	p-value
Arabic	3.70	2.36	1.57	0.12
English	15.67	2.36	6.65	0.00****
French	11.97	2.36	5.08	0.00****

To explore the significant interaction and differences within each group, the intercept was releveled for each level of the two variables, Group and Language. The

results indicated that EMI participants reported receiving significantly more input in English than in Arabic and French, and significantly more input in Arabic than in French. The linear model analysis results are presented in Table 4.27.

Table 4.27 Language Exposure Contrasts (EMI)

	β	SE	t-value	p-value
Arabic vs. English	13.59	2.38	5.76	0.00***
Arabic vs. French	30.05	2.38	12.79	0.00****
English vs. French	43.64	2.38	18.51	0.00****

FMI participants also reported significantly different levels of input in each language. The linear model analyses, summarised in Table 4.28, revealed that this group received significantly more Arabic input than English or French, and significantly more English input than French. As with language use, these results are unexpected, especially in terms of the minimal exposure to French by the FMI group in social contexts. These results also highlight the utility of English superseding that of French in non-academic linguistic contexts for this group.

Table 4.28 Language Exposure Contrasts (FMI)

	β	SE	t-value	p-value
Arabic vs. English	5.78	2.36	2.45	0.02*
Arabic vs. French	21.78	2.36	9.24	0.00****
English vs. French	16.00	2.36	6.79	0.00****

Summary of Language Experience Analyses

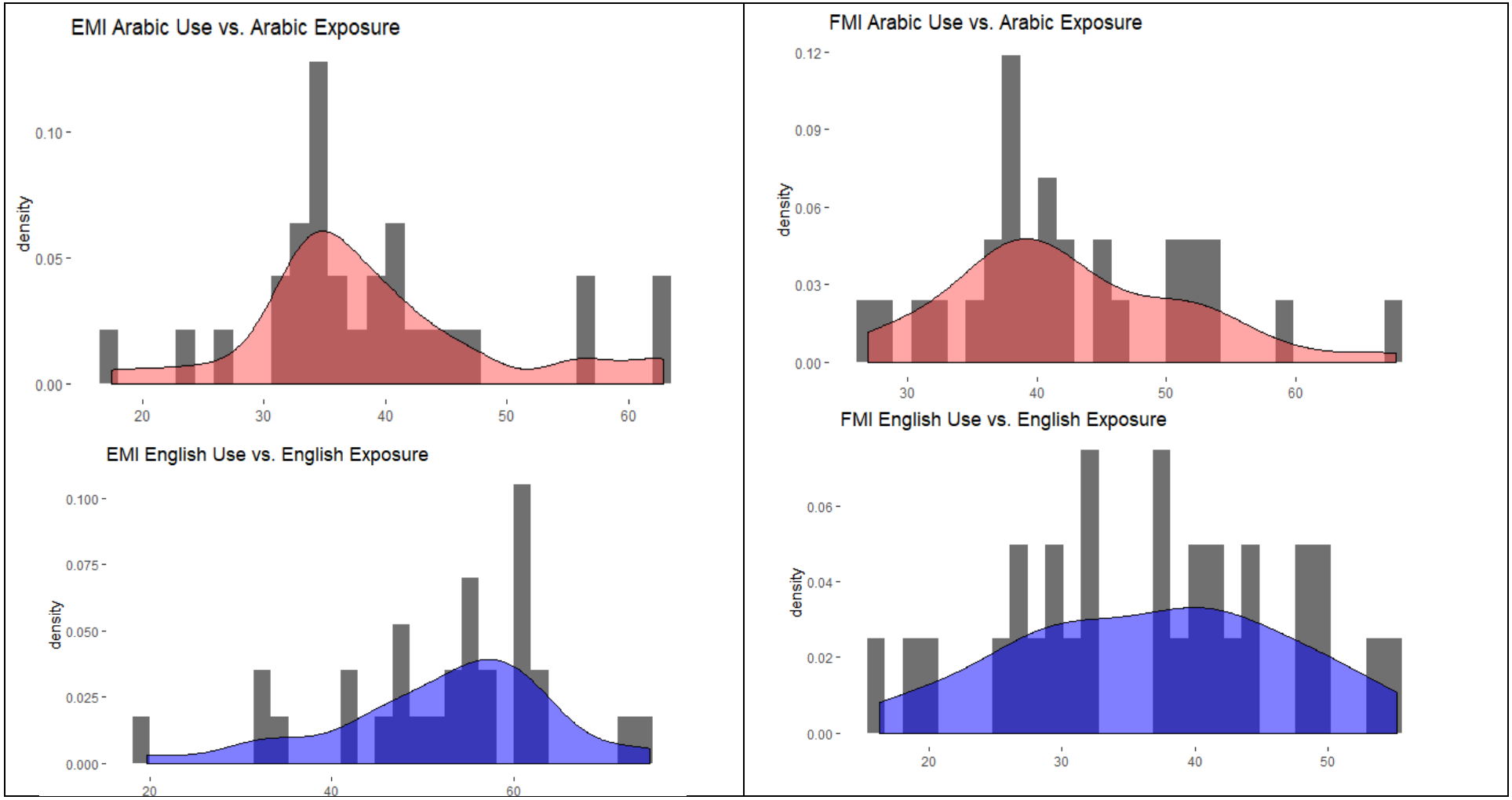
The overall analyses of language experience revealed several interesting findings in the self-reported data of the participants. In terms of language use, EMI participants used Arabic significantly less frequently than their French counterparts despite having similar Arabic language exposure to FMI participants and despite using French minimally in their speech. For the EMI group, English was the predominant language in terms of self-reported use and exposure. On the other hand, FMI participants reported

using their languages in a slightly more 'balanced' way with Arabic as the most frequently used language, followed by English, and then French. The relatively low frequency of exposure to and use of French for the FMI group in non-classroom settings compared to the other two languages is striking. This is indicative of a more general representation of language use by this specific segment of the Lebanese community where Arabic and English are more ubiquitous than French. This further highlights the contrast between the two groups in terms of their 'third' language exposure and use; EMI participants have very minimal exposure to and use of French, whereas FMI participants are highly exposed to and frequently use English[12].

Another important observation is that overall, for each group, the use of each language highly matched the exposure to that language. The overlap between language use and language exposure for each group is presented in Figure 4.16. This finding is in line with the notion that language exposure is a key individual difference which influences language use (Titone et al., 2017). As indicated in the literature review, examining the relationship between these two variables is one of the contributions of the present study to current research. This overlap also supports sociopragmatic theories underpinning speech production, where the language of the speaker tends to match that of their interlocuter (Gardner-Chloros, 2009; Green & Abutalebi, 2013). That is, participants typically reported using the same language(s) in which they are spoken to by family members and friends. However, not all input comes from speech; items in the LBQ related to the amount of language exposure from books, television, and computer-use also matched the language use reported by the participants. This is also a significant observation which emphasises the reciprocity between language use and language

[12] Post-hoc analyses on language use and language exposure are presented in Appendix H. While these analyses are of relevance to the discussion on language experience, they are not directly addressed in the research questions, and so they have been included in the appendix.

exposure, whether it comes from face-to-face interactions or other media. This diversity of both language exposure and language use is ‘permitted’ by the speech community where frequency of use of each language can vary by the individual with potentially minimal to no risk of communicative breakdown.


EMI Arabic Use vs. Arabic Exposure
density
0.10
0.05
0.00
20
30
40
50
60
EMI English Use vs. English Exposure
density
0.100
0.075
0.050
0.025
0.000
20
40
60
FMI Arabic Use vs. Arabic Exposure
density
0.12
0.09
0.06
0.03
0.00
30
40
50
60
FMI English Use vs. English Exposure
density
0.06
0.04
0.02
0.00
20
30
40
50

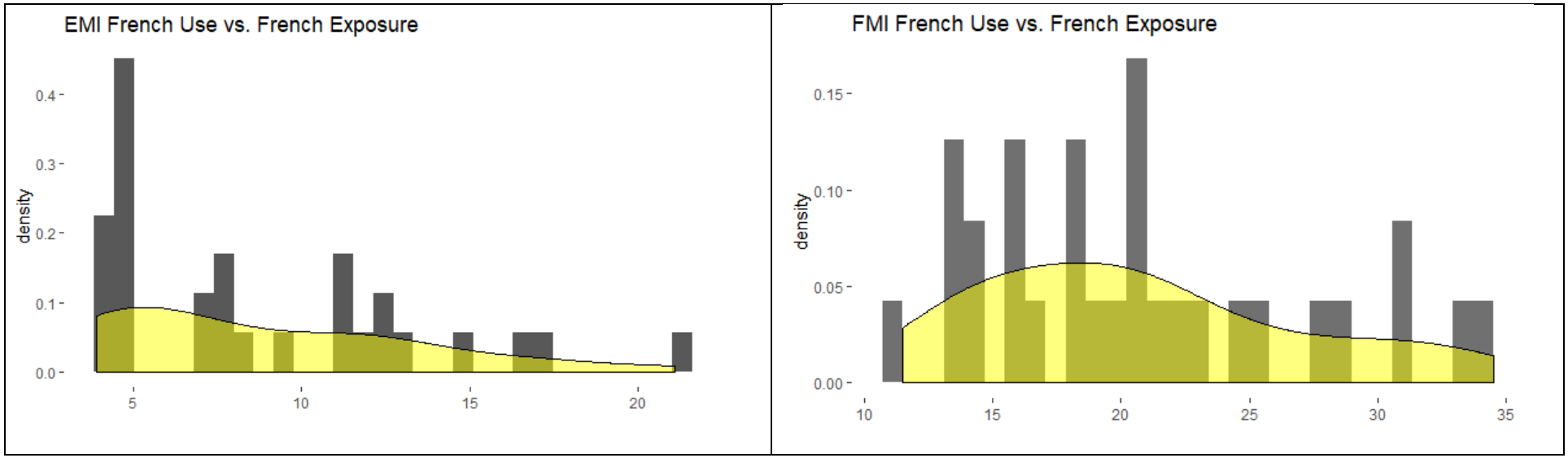


Figure 4.16 Overlapping Distributions of Language Use vs. Language Exposure (EMI vs. FMI); grey histogram- language use, coloured distribution plot = language exposure

4.5 RQ2-- Predictors of Cognitive Task performance

> *RQ2. Is EMI and FMI participants' performance on non-linguistic, domain general tasks modulated by their language proficiency and/or language experience?*

The second research question explores the predictors of cognitive task performance. A series of linear mixed-effect models was used to analyse the data in R using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). The Simon Effect scores, Conflict Effect scores, and Backward Digit Span Task scores were entered as dependent variables in separate models. The log-transformed data from these tasks were used to satisfy the normality assumption. To reduce collinearity, all continuous fixed effects were z-scored whereas the two-level categorical predictors (Group and HPBT) were coded as -0.5 and +0.5. Participants were included in the models as random effects.

An initial step was creating a full model including all fixed effects of interest, random intercepts for participants, and slopes for all predictors (i.e., a maximal structure, Barr et al., 2013). Keeping the random effects structure constant, as analyses at the participant-level were of interest to this project, the best fitting model was then selected through a stepwise procedure. First, non-significant predictors were removed, and a comparison between the simpler model and the full model was conducted through Akaike's Information Criterion (AIC) and likelihood-ratio chi-square tests using the ANOVA function in R (Baayen, 2008). If the simpler model did not significantly differ from the full model, then the simpler model which had fewer predictors was selected. After that, each significant predictor was removed from the selected simple model to examine whether the model improved. Comparisons of the models with different fixed effects being removed were conducted with parameters estimated with Maximum Likelihood for fixed effects and Restricted Maximum Likelihood for random effects comparisons (Pinheiro & Bates, 2006). The models were then checked again for

collinearity between the significant predictors through the VIF.mer function in R (Frank, 2011). All models whose VIFs were below 2.5 and whose t- and z-values were >2 were considered significant (Gelman & Hill, 2007). For each cognitive task, a discussion of the overall best-fit model is presented, followed by further (post-hoc) analyses conducted when necessary.

4.5.1 Language Proficiency and Cognitive Performance

RQ2A: Is EMI and FMI participants' performance on non-linguistic, domain general tasks modulated by their language proficiency?

4.5.1.1. Language Proficiency and the flanker task

In order to explore whether having higher language proficiency predicted better cognitive performance on the flanker task, a generalised linear model was fitted with a maximal structure. The predictors were OPT Scores (3 levels: Arabic, English, and French), VFT scores (3 levels: Arabic, English, and French), HPBT (2 levels: Bilingualism and Trilingualism), and Group (2 levels: EMI and FMI). All interactions between these variables were also included as predictors. The dependent variable was the z-scored Conflict Effect scores[13]. The full model did not converge; accordingly, four separate minimal-structure models were used to analyse the data: (1) Group x OPT scores, (2) Group x VFT Scores, (3) HPBT x OPT Scores, and (4) HPBT x VFT Scores. The structuring of these models falls in line with the theoretical assumptions about (and prior analyses of) OPT and VFT scores' lack of correlation. It also fits with the post-hoc nature of the HPBT variable analyses, further circumventing the singularity confound in

[13] Accuracy data from the Conflict Effect calculation was excluded from further analyses given the very low number of errors, with some participants having 100% accuracy rates. A generalised linear model was fitted with the binary data of responses as outcome variables and with the language proficiency and language use as predictors. The residuals did not meet the prerequisite assumption of homoscedasticity and revealed a negative dispersion with fitted values approaching 1. This indicated that the ceiling effect created a bias in the model, so no further analyses were performed on accuracy rates of the Conflict Effect.

the maximal structure model. The models were releveled by group when fitted in order to obtain within- and between-group contrasts.

Model 1: Group x OPT Scores

This 2x3 model investigated whether having a higher proficiency in any specific language or set of languages would enhance performance on the flanker task. The analyses revealed no significant main effect for any predictors or interactions for either group (all $p>0.05$). A summary of the mixed-effects model's ANOVA results is presented in Table 4.29 (for full-model estimates, see Appendix H). That is, for participants in both groups, having high proficiency scores in any language or set of languages did not predict enhanced conflict monitoring capacities. A possible explanation for this is that language proficiencies were generally high, so a lack of significant results of OPT scores as an explanatory variable was not observed. Another explanation is that the measure itself was not sensitive enough to detect differences in performance on this particular task.

Table 4.29 MEM ANOVA (Type III) Summary for Conflict Effect x OPT Scores

	F-Statistic	DF1	DF2	p-value
Intercept	97.36	1	57	<0.0001
Group (EMI vs. FMI)	0.74	1	57	0.28
OPT Scores (Arabic vs. English vs. French)	1.37	2	57	0.25
Group*OPT Scores	0.36	2	57	0.55

Formula: Conflict Effect ~ Group OPT Scores +(1|Participant)*

Although the model converged, the lack of significant findings prompted exploratory analyses of even simpler models. That is, the present model was further divided into two models by (1) removing Group as a predictor and analysing the influence of OPT scores on Conflict Effect scores for all participants; and (2) removing all interactions but keeping main effects for all predictors including Group. Both exploratory models resulted in non-significant results despite their relatively simple structures (see Appendix H).

Model 2: Group x VFT Scores

The second model was fitted with the Conflict Effect scores as the dependent variable and with Group (2 levels) and VFT scores (3 levels) as the predictors, along with all interactions between the two variables. The best-fit model included main effects for Group and VFT scores, with no interactions between the two variables. As with the OPT scores, the results of this model revealed no significant main effects for either predictor (both $p>0.05$), as summarised in the model's ANOVA summary in Table 4.30. That is, having higher lexical robustness scores in any language or combination of languages did not predict a smaller Conflict Effect score. A summary of the model's output is presented in Table 4.31.

Table 4.30 MEM ANOVA (Type III) Summary for Conflict Effect by VFT Scores

	F-Statistic	DF1	DF2	p-value
Intercept	97.49	1	58	<0.0001
Group (EMI vs. FMI)	0.89	1	58	0.35
VFT Scores (Arabic vs. English vs. French)	0.63	2	58	0.43

Table 4.31. MEM Output Summary for Conflict Effect by VFT Scores (Estimated coefficients, standard errors (SE), t-values, and p-values)

	Conflict Effect Scores			
	Estimate	**SE**	**t-value**	**Pr(>|t|)**
Intercept	47.81	2.84	1.68	0.10
Group (EMI vs. FMI)	19.62	1.62	2.21	0.23
VFT Arabic	0.12	0.18	0.62	0.54
VFT English	0.02	0.15	0.12	0.91
VFT French	0.15	0.16	0.94	0.35

Formula: Conflict Effect ~ Group + VFT Scores + (1|Participant)

As with the OPT scores, exploratory models were created to examine the potential influence of VFT scores. The models included the removal of Group and splitting the VFT variable into three sub-variables. No significant results were obtained, as presented in Appendix H.

Model 3: HPBT x OPT

The third model was structured in order to explore whether being highly proficient in three languages as opposed to two significantly influenced Conflict Effect scores. Although the HPBT variable is binary and is derived from the OPT scores, it was important to explore whether having high proficiency in a *specific* language set(s) enhances cognitive performance. That is, not only did this model examine whether having a higher proficiency in a third language reduces Conflict Effect scores, but it also examined which language or language combination(s) is able to predict a lower Conflict effect. Furthermore, in this model, there was no distinction between participants based on the EMI vs. FMI grouping. Accordingly, the use of the mixed-effects model allowed for the unequal distribution of participants as highly proficient in two vs. three languages, irrespective of which two (at least) languages they are highly proficient in. Model 3 was therefore fitted with the Conflict Effect scores as the outcome variable, and with HPBT (2 levels) and OPT Scores (3 levels) as predictors.

The results of the model, summarised in Table 4.32, revealed no significant main effects for any of the predictors or interactions between them (all $p>0.05$); that is having high proficiency scores in two or three languages did not influence participants' cognitive performance. The 'additive' cognitive advantage which is typically observed in low vs. high-proficiency (or balanced vs. unbalanced) bilinguals and trilinguals was not replicated in this model. As this variable was central to the main aims of the study, and given the collinearity between predictors (which was not high enough to influence the results as verified by the VIF values), different simulations specifically for HPBT were conducted in both simpler and more complex models. In all simulations, HPBT did not significantly predict better cognitive performance on the flanker task. This may be attributed to the fact that participants in both groups had reached a minimum level of

proficiency in the L3 which did not confer a cognitive benefit that can be significantly detected by the flanker task. Another possible explanation is that the Flanker task, or the Conflict Effect score itself, was not a sensitive measure of differences in cognitive performance for this specific simultaneous trilingual sample.

Table 4.32 MEM ANOVA (Type III) Output Summary for Conflict Effect by HPBT and OPT

	F-Statistic	DF1	DF2	p-value
Intercept	92.77	1	57	<0.0001
HPBT	1.89	1	57	0.22
OPT Scores	1.68	2	57	0.19
HPBT*OPT Scores	0.01	2	57	0.99

Formula: Conflict Effect ~ HPBT OPT Scores + (1|Participant) + (1|Language)*

Model 4: HPBT x VFT Scores

With the same assumptions as those discussed in Model 3, the fourth model was fitted with HPBT as a predictor variable, along with the VFT scores (instead of OPT) of each language, and with Conflict Effect scores as the outcome variable. No interactions were included in this model between the variables. As with the results obtained from Model 3, there were no significant main effects (all $p>0.05$), as summarised by the model's output in Table 4.33. Exploratory models, offered in Appendix H, also indicated no significant main effects (or interactions). Accordingly, being highly proficient in two vs. three languages did not predict a smaller Conflict Effect score. Furthermore, in this model, having a higher lexical robustness score in any language or language combination did not predict better cognitive performance on the flanker task. As with the conclusions drawn from the results from Model 3, despite the variation in performance on the VFT across participants and despite participants' categorisation as balanced or unbalanced trilinguals, these differences could not predict better cognitive performance on the flanker task.

Table 4.33. MEM Output Summary for Conflict Effect by HPBT and VFT scores (Estimated coefficients, standard errors (SE), t-values, and p-values)

	Conflict Effect Scores			
	Estimate	**SE**	**t-value**	**Pr(>|t|)**
Intercept	48.51	2.81	1.72	0.09
HPBT	10.70	6.35	1.68	0.10
VFT Arabic	0.04	0.19	0.18	0.85
VFT English	0.07	0.14	0.47	0.64
VFT French	0.05	0.09	0.61	0.54

Formula: Conflict Effect ~ HPBT + VFT Scores + (1|Participant)

In summary, despite the use of different measures to operationalise 'language proficiency' and the use of different exploratory models, there were no significant main effects of this predictor on Flanker performance. Furthermore, there were no significant interactions between Group x OPT Scores, Group x VFT scores, HPBT x OPT Scores, and HPBT x VFT scores (all $p>0.05$). For both EMI and FMI participants, being highly proficient in two or three languages did not significantly predict smaller Conflict Effect scores. Furthermore, having a higher lexical robustness score in any of the three languages did not predict enhanced interference suppression and conflict monitoring capacities as measured by the flanker task.

4.5.1.2. Language Proficiency and the Simon Task

In order to explore whether having higher language proficiency predicts better cognitive performance on the Simon Task, a mixed-effects model was fitted with a maximal structure. The predictors were OPT Scores (3 levels: Arabic, English, and French), VFT scores (3 levels: Arabic, English, and French), HPBT (2 levels: Bilingualism and Trilingualism), and Group (2 levels: EMI and FMI). All interactions between these variables were also included as predictors. The dependent variable was the transformed Simon Effect scores (squared). The full model did not converge; accordingly, four minimal-structure separate models which eliminated the confound of

collinearity were used to analyse the data: (1) Group x OPT scores, (2) Group x VFT Scores, (3) HPBT x OPT Scores, and (4) HPBT x VFT Scores. As mentioned in the model structuring of the flanker task, the structuring of the present models falls in line with the theoretical assumptions about and prior analyses of the lack of correlation between the OPT and VFT scores. It also fits with the post-hoc nature of the HPBT variable analyses, further circumventing the singularity confound in the maximal structure model.

Model 1: Group x OPT Scores

In this 2x3 model, the hypothesis of whether having a higher proficiency in any specific language or set of languages would enhance performance on the Simon task was addressed. The best-fitting model included Group, OPT Scores, and interactions between the two variables, whereas the log-transformed Simon Effect was fitted as the outcome variable. The results of the model's ANOVA are presented in Table 4.34.

Table 4.34 MEM ANOVA (Type III) Table of Main Effects and Interactions for Simon Effect scores by OPT

	F-Statistic	DF1	DF2	p-value
Intercept	389.91	1	57	<0.0001
Group (EMI vs. FMI)	2.19	1	57	0.18
OPT Scores	3.17	2	57	0.05*
Group*OPT Scores	2.56	2	57	0.22

The results of the analysis indicated a significant main effect for OPT scores ($F(2,57)=3.17$, $p<0.05$). The model's output, offered in Table 4.35, indexed OPT scores in English as significantly predicting smaller Simon Effect scores, yet there were no significant main effects of OPT scores for Arabic or French or any significant interaction between the predictors (all $p>0.05$). For both groups, therefore, the only predictor of enhanced performance on the Simon task was a higher proficiency in English. That is, irrespective of group membership, participants with high proficiency in English exhibited greater interference suppression abilities.

Table 4.35 MEM Output Summary for Simon Effect by OPT scores (Estimated coefficients, standard errors (SE), t-values, and p-values) Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	Simon Effect Scores			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	143.84	23.91	6.02	0.00****
Group (EMI vs. FMI)	3.18	7.62	0.42	0.68
OPT Arabic	2.68	2.14	1.25	0.22
OPT English	5.95	2.79	2.13	0.04*
OPT French	3.88	2.18	1.78	0.08
Group*OPT Arabic	0.05	0.05	0.99	0.33
Group*OPT English	0.07	0.04	1.68	0.10
Group*OPT French	0.07	0.05	1.47	0.15
OPT Arabic*OPT English	0.01	0.01	1.51	0.14
OPT Arabic*OPT French	0.02	0.01	1.41	0.17
OPT English*OPT French	0.02	0.01	1.86	0.07
OPT Arabic*OPT English*OPT French	0.02	0.01	1.42	0.15
Group*OPT Arabic*OPT English	0.01	0.01	1.22	0.23
Group*OPT Arabic*OPT French	0.01	0.01	1.11	0.28
Group*OPT English*OPT French	0.00	0.00	1.56	0.13
Group*OPT Arabic*OPT English*OPT French	0.00	0.00	1.31	0.20

Formula: Simon Effect ~ Group OPT Scores + (1+OPT Score|Participant)*

Model 2: Group x VFT Scores

The second model was fitted with the Simon Effect scores as the dependent variable and with Group (2 levels) and VFT scores (3 levels), along with their interactions, as the predictors. The results of the model's output (Table 4.36) revealed no significant main effects or interactions between any of the predictors. That is, having higher lexical robustness scores in any language or combination of languages did not predict a smaller Simon Effect score. The model was releveled by group to examine potential within-group differences, yet no significant results were obtained, and no post-hoc analyses were therefore conducted.

Table 4.36 MEM Output Summary for Simon Effect by VFT scores (Estimated coefficients, standard errors (SE), t-values, and p-values) Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	Simon Effect Scores			
	Estimate	SE	t-value	p-value
Intercept	157.26	46.41	0.34	0.74
Group	241.74	39.49	0.61	0.54
VFT Arabic	0.58	0.42	0.14	0.89
VFT English	2.28	0.36	0.63	0.53
VFT French	1.10	0.62	0.18	0.86
Group*VFT Arabic	1.55	0.35	0.45	0.66
Group*VFT English	2.59	0.33	0.79	0.43
Group*VFT French	0.46	0.37	0.12	0.90
VFT Arabic*VFT English	0.01	0.00	0.41	0.68
VFT Arabic*VFT French	0.02	0.01	0.38	0.71
VFT English*VFT French	0.00	0.00	0.09	0.93
VFT Arabic*VFT English*VFT French	0.00	0.00	0.13	0.90
Group*VFT Arabic*VFT English	0.02	0.00	0.63	0.53
Group*VFT Arabic*VFT French	0.00	0.00	0.08	0.94
Group*VFT English*VFT French	0.01	0.00	0.41	0.68
Group*VFT Arabic*VFT English*VFT French	0.00	0.00	0.19	0.85

Formula: Simon Effect ~ Group Scores + (1|Participant) + (1|Language)*

Model 3: HPBT x OPT

This model included the HPBT variable which explores whether being highly proficient in three languages as opposed to two significantly influences performance on the Simon task. Having observed that OPT Scores in English predicted better Simon task performance, it was expected that this 'post-hoc' analysis would reveal a similar finding. Furthermore, as this model does not distinguish participants based on group membership, an interaction between HPBT and OPT English scores was expected to reveal whether being highly proficient in two languages (one of which being English) or three languages would predict lower Simon Effect scores. The model was fitted with the Simon Effect

scores as the outcome variable, and with HPBT (2 levels) and OPT Scores (3 levels) as predictors.

The model (Table 4.37) revealed only one significant main effect, which is that of OPT English. There were no other significant main effects or interactions. Accordingly, having high proficiency scores in two or three languages did not influence a participant's cognitive performance (by having lower Simon Effect scores).

Table 4.37 MEM Output Summary of Simon Effect by HPBT and OPT (Estimated coefficients, standard errors (SE), t-values, and p-values) Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	Simon Effect Scores			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	-1.70	1.56	1.09	-0.28
HPBT (Bilingual vs. Trilingual)	1.40	1.51	0.93	0.36
OPT Arabic	0.20	0.18	1.10	0.28
OPT English	0.37	0.14	2.64	0.01*
OPT French	0.21	0.17	1.22	0.23
HPBT*OPT Arabic	-0.16	0.17	0.92	0.36
HPBT*OPT English	0.16	0.18	0.89	0.38
HPBT*OPT French	0.17	0.16	1.02	0.31
OPT Arabic*OPT English	-0.02	0.02	1.07	0.29
OPT Arabic*OPT French	-0.02	0.02	-1.21	0.23
OPT English*OPT French	0.02	0.02	1.20	0.24
OPT Arabic*OPT English*OPT French	0.00	0.00	1.19	0.24
HPBT*OPT Arabic*OPT English	0.02	0.02	0.89	0.38
HPBT*OPT Arabic*OPT French	0.02	0.02	1.01	0.32
HPBT*OPT English*OPT French	0.02	0.02	1.01	0.32
HPBT*OPT Arabic*OPT English*OPT French	0.00	0.00	0.99	0.33

Formula: Simon Effect ~ HPBT OPT Scores + (1|Participant)*

Model 4: HPBT x VFT Scores

The fourth model, summarised in Table 4.38, was also fitted with HPBT as a predictor variable and VFT scores as predictors (along with interactions between them) and with Simon Effect scores as the outcome variable. The results revealed no significant

main effects or interactions (all p>0.05); i.e., being highly proficient in two vs. three languages did not predict a smaller Simon Effect score. Furthermore, having a higher lexical robustness score in any language or language combination did not predict better cognitive performance on the Simon Task. Despite the difference in performance on the VFT across the different participants and the differences between highly proficient bilinguals vs. trilinguals, these differences could not predict better cognitive performance. These results are similar to those obtained from the Conflict Effects model (Model 4), which also revealed that VFT scores did not predict cognitive task performance (on the flanker task).

Table 4.38 MEM Output Summary for Simon Effect by HPBT and VFT (Estimated coefficients, standard errors (SE), t-values, and p-values)

	Simon Effect Scores			
	Estimate	SE	t-value	p-value
Intercept	0.52	0.48	1.07	0.29
HPBT (Bilingual vs. Trilingual)	0.52	0.47	1.10	0.28
VFT Arabic	0.00	0.00	1.08	0.29
VFT English	0.00	0.00	0.97	0.34
VFT French	0.00	0.00	0.78	0.44
HPBT*VFT Arabic	0.00	0.00	1.11	0.27
HPBT*VFT English	0.00	0.00	1.04	0.30
HPBT*VFT French	0.00	0.00	0.92	0.36
VFT Arabic*VFT English	0.00	0.00	1.00	0.32
VFT Arabic*VFT French	0.00	0.00	0.83	0.41
VFT English*VFT French	0.00	0.00	0.63	0.53
VFT Arabic*VFT English*VFT French	0.00	0.00	0.82	0.42
HPBT*VFT Arabic*VFT English	0.00	0.00	1.05	0.30
HPBT*VFT Arabic*VFT French	0.00	0.00	0.94	0.35
HPBT*VFT English*VFT French	0.00	0.00	0.68	0.50
HPBT*VFT Arabic*VFT English*VFT French	0.00	0.00	0.84	0.41

Formula: Simon Effect ~ HPBT Scores + (1|Participant) +(1|Language)*

In summary, for Language Proficiency as a predictor of cognitive performance on the Simon Task, there were no significant interactions between Group x OPT Scores, Group x VFT scores, HPBT x OPT Scores, and HPBT x VFT scores (all $p>0.05$). Furthermore, there was no significant main effect for any predictor variable for Simon Effect scores except for OPT English scores. For both EMI and FMI participants, being highly proficient in two or three languages did not significantly predict smaller Simon Effect scores. Furthermore, having a higher lexical robustness score in any of the three languages did not predict better performance on the Simon Task.

4.5.1.3 Language Proficiency and Working Memory Task (BDST)

In order to explore the influence of language proficiency on BDST scores, a generalised linear model with a binomial link function was selected using the MASS library in R. This is due to the BDST scores constituting a discrete count outcome variable with a ceiling. The model was fitted with Group, HPBT, OPT scores, and VFT scores, along with interactions between them, as predictors. As with the Simon Effect and Conflict Effect scores, the model did not converge due to being overfitted and due to singularity. Several model simulations were compared to select the best-fit model, and a similar breakdown of simpler models was selected.

Model 1: Group x OPT Scores

The first model, summarised in Table 4.39, included main effects for the predictors Group and OPT scores and two-level interactions between the levels of these two explanatory variables yet excluding interactions within the OPT score levels as determined by AICs and likelihood tests and further verified by the glmulti package. Results from this model indicated a significant main effect for OPT scores in English ($p<0.05$), similar to the result obtained from the Simon Task. As both BDST and Simon

Effects scores were correlated (as indicated by RQ1D), sharing a significant predictor was not unexpected.

Table 4.39 GLMM Output Summary for BDST by OPT (Estimated coefficients, standard errors (SE), z-values, and p-values) Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	BDST Scores			
	Estimate	**SE**	**z-value**	**Pr(>|z|)**
Intercept	-2.96	3.66	-0.81	0.42
Group (EMI vs. FMI)	-0.41	2.13	-0.19	0.85
OPT Arabic	-0.08	0.27	-0.28	0.78
OPT English	0.94	0.40	2.33	0.02*
OPT French	-0.34	0.28	-1.25	0.21
Group*OPT Arabic	0.14	0.18	0.79	0.43
Group*OPT English	0.49	-0.26	-1.85	0.07
Group*OPT French	0.38	0.22	1.75	0.08

*Formula: BDST ~ Group + OPT Arabic + OPT English + OPT French + Group*OPT Arabic +Group*OPT English + Group*OPT French + (1|Participant) + (1|Language)*

Model 2: Group x VFT Scores

A second generalised linear model with a binomial link was fitted with Group and VFT scores as predictors, along with interactions between them. As the VFT entails an element of working memory, it was expected that scores of the VFT might predict performance on the BDST. However, the model indicated that there were no significant main effects for VFT scores or Group and no significant interactions between them. The results of the model's output are summarized in Table 4.40 (see Appendix H for exploratory analyses which also revealed no significant effects of VFT scores as predictors of working memory).

Table 4.40 GLMM Output Summary for BDST and VFT (Estimated coefficients, standard errors (SE), z-values, and p-values)

	BDST Scores			
	Estimate	**SE**	**z-value**	**Pr(>|z|)**
Intercept	107.44	280.52	0.38	0.70
Group	201.83	238.69	0.85	0.40
VFT Arabic	0.53	2.52	0.21	0.84
VFT English	1.56	2.19	0.71	0.48
VFT French	0.76	3.76	0.20	0.84
Group*VFT Arabic	1.47	2.09	-0.71	0.48
Group*VFT English	2.11	1.97	1.07	0.29
Group*VFT French	0.58	2.24	-0.26	0.80
VFT Arabic*VFT English	0.01	0.02	0.52	0.61
VFT Arabic*VFT French	0.01	0.03	0.39	0.70
VFT English*VFT French	0.00	0.03	0.00	1.00
VFT Arabic*VFT English*VFT French	0.00	0.00	0.19	0.85
Group*VFT Arabic*VFT English	0.02	0.02	0.95	0.34
Group*VFT Arabic*VFT French	0.00	0.02	0.07	0.94
Group*VFT English*VFT French	0.01	0.02	0.54	0.59
Group*VFT Arabic*VFT English*VFT French	0.00	0.00	0.36	0.72

*Formula: BDST~Group*VFT Scores + (1|Participant)*

Model 3: HPBT x OPT Scores

The third model was fitted with WM scores as the outcome variable and with HPBT and OPT scores as the predictor variables. As OPT English scores significantly predicted BDST performance, it was of interest to explore whether this variable, along with others, interacted with the categorisation of participants as being highly proficient in two or three languages. The results of the model revealed that only OPT English scores were a main predictor of enhanced working memory capacity, with no difference between the groups ($p>0.05$). There were no other significant main effects or interactions, as shown in Table 4.41. An exploratory model with only the HPBT variable as a predictor was performed, and the results were also non-significant ($F(1,58)=0.46$,

p=0.50), indicating that whether a participant was highly proficient in two or three languages did not influence their working memory performance.

Table 4.41 GLMM Output Summary for BDST by HPBT and OPT (Estimated coefficients, standard errors (SE), z-values, and p-values) Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	BDST Scores			
	Estimate	**SE**	**z-value**	**Pr(>|z|)**
Intercept	320.47	852.94	0.38	0.71
HPBT (Bilingual vs. Trilingual)	-253.16	826.77	-0.31	0.76
OPT Arabic	-45.29	98.65	-0.46	0.65
OPT English	79.54	25.73	3.07	0.03*
OPT French	57.14	94.17	0.61	0.55
HPBT*OPT Arabic	34.54	94.13	0.37	0.72
HPBT*OPT English	21.51	98.51	0.22	0.83
HPBT*OPT French	44.89	89.42	0.50	0.62
OPT Arabic*OPT English	4.97	11. 7487	0.42	0.67
OPT Arabic*OPT French	7.79	11.04	0.71	0.48
OPT English*OPT French	6.51	10.96	0.59	0.56
OPT Arabic*OPT English*OPT French	-0.92	1.29	-0.72	0.48
HPBT*OPT Arabic*OPT English	3.23	11.18	0.29	0.77
HPBT*OPT Arabic*OPT French	5.83	10.22	0.57	0.57
HPBT*OPT English*OPT French	4.54	10.35	0.44	0.66
HPBT*OPT Arabic*OPT English*OPT French	0.62	1.18	0.52	0.60

Formula: BDST ~ HPBT OPT Scores + (1+OPT Scores|Participant)*

Model 4: HPBT x VFT Scores

The fourth model was also fitted with HPBT as a predictor variable, along with the VFT scores of each language, and with BDST scores as the outcome variable. There were no significant main effects or interactions, as indicated in Table 4.42. This reiterates the findings of Models 2 and 3 that participants whose proficiency is high in three languages do not outperform participants who have high proficiency in two

languages on the BDST. Furthermore, the lexical robustness scores in any language or combination of languages did not predict better working memory capacity.

Table 4.42 GLMM Output Summary for BDST by HPBT and VFT (Estimated coefficients, standard errors (SE), z-values, and p-values)

	BDST Scores			
	Estimate	**SE**	**z-value**	**Pr(>|z|)**
Intercept	-88.25	266.01	-0.33	0.74
HPBT (Bilingual vs. Trilingual)	153.51	258.57	0.59	0.56
VFT Arabic	0.68	2.34	0.29	0.77
VFT English	0.57	1.96	0.29	0.77
VFT French	-0.11	1.70	-0.07	0.95
HPBT*VFT Arabic	-1.25	2.26	-0.55	0.58
HPBT*VFT English	-0.97	1.91	-0.51	0.62
HPBT*VFT French	0.61	1.59	-0.38	0.70
VFT Arabic*VFT English	0.00	0.02	0.22	0.82
VFT Arabic*VFT French	0.00	0.01	0.11	0.91
VFT English*VFT French	0.00	0.01	0.15	0.88
VFT Arabic*VFT English*VFT French	0.00	0.00	0.19	0.85
HPBT*VFT Arabic*VFT English	0.01	0.02	0.47	0.64
HPBT*VFT Arabic*VFT French	0.00	0.01	0.35	0.73
HPBT*VFT English*VFT French	0.00	0.01	0.29	0.77
HPBT*VFT Arabic*VFT English*VFT French	0.00	0.00	-0.26	0.79

Formula: BDST ~ HPBT VFT Scores + (1|Participant)*

To summarise, the BDST score was only predicted by English OPT scores, but there were no other significant main effects or interactions in any of the models that were constructed. These results are analogous to those obtained from Simon Effect scores.

4.5.1.4 Summary of Language Proficiency as a Predictor of Cognitive Performance

Several models were put forth to examine whether language proficiency, which was operationalised using three sub-factors, influenced cognitive task performance.

Specifically, OPT scores, VFT scores, and the HPBT variable were examined as predictors of the Conflict Effect, Simon Effect, and BDST scores. Group (EMI vs. FMI) was included as a predictor in models with OPT and VFT scores, but not with HPBT due to collinearity. The overall results from the models revealed that the Conflict Effect was not predicted by any variable; that is, whether participants were EMI or FMI, whether they had high proficiency scores in any language(s), whether they had high lexical robustness scores, or whether they were highly proficient in two or three languages did not influence their interference suppression and conflict monitoring abilities. However, for the Simon Task and the BDST, OPT scores in English (for both EMI and FMI) were significant predictors of better inhibitory control skills (lower Simon Effect scores) and better working memory capacity (higher BDST scores). There were no other significant main effects or interactions.

4.5.2 Language Experience as a Predictor of Cognitive Performance

RQ2B: Is EMI and FMI participants' performance on non-linguistic, domain general tasks modulated by language experience?

4.5.2.1. Language Experience and the flanker task

In order to investigate the influence of language experience on the flanker task, generalised mixed-effects models were fitted with the transformed Conflict Effect as the outcome variable. The predictors included Group (2 levels: EMI vs. FMI), Language Exposure (3 levels: Arabic, English, and French), Language Use (3 levels: Arabic, English, and French), self-reported code-switching frequency (CSF), and self-reported salient-cue detection (SalC). However, due to multicollinearity between language exposure and language use, separate models were used to investigate the influence of the factors on Conflict Effect. After removing the correlations, the models converged

with all subject and item slopes. The fit of each model was assessed according to the procedure outlined at the beginning of Section 4.5.

Model 1: Group x Language Use x CSF x SalC

The first model was constructed to examine the influence of Group (2 levels: EMI and FMI), Language Use (3 levels: Arabic, English, and French), Codeswitching Frequency (CSF), and Salient-Cue Detection (SalC) on Conflict Effect scores. The best-fit model included interactions between Group and each of the variables, as well as between CSF and SalC. All other interactions were removed as they resulted in a poorer fit, as examined through the residuals and confirmed by the glmulti package.

The results of the model, summarised in Table 4.43 indicated that there was a significant main effect of Salient Cue Detection. All other variables did not significantly predict Conflict Effect scores (see Appendix H for further analyses). Accordingly, the results indicate that having a higher ability to monitor cues during speech predicts lower Conflict Effect scores.

Table 4.43 GLM Output Summary for Conflict Effect and Language Experience (Model 1) (Estimated coefficients, standard errors (SE), t-values, and p-values)
Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	Conflict Effect Scores			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	38003.79	22944.95	1.66	0.10
Group	29490.41	14742.13	1.94	0.06
Arabic Use	380.19	229.45	1.66	0.10
English Use	362.52	235.50	1.54	0.13
French Use	270.31	210.44	1.27	0.21
CSF	0.99	0.55	1.81	0.08
SalC	0.67	0.31	2.18	0.03*
CSF*SalC	0.43	0.66	0.65	0.52
Group*Arabic Use	294.96	147.43	1.97	0.06
Group*English Use	264.25	203.07	1.30	0.19
Group*French Use	203	165.04	1.23	0.20
Group*CSF	0.62	0.31	1.97	0.06
Group*SalC	0.37	0.19	1.94	0.07

Formula: Conflict Effect ~ Group(Language Use+ CSF + SalC) + CSF*SalC* + (1|Participant) + (1|Language)

These results support the notion that performance on the flanker task as a measure of conflict monitoring capacity is influenced by participants' ability to monitor cues both external and internal, such as their interlocutor's language choice. This finding further supports the claim that the mechanisms underlying conflict monitoring and salient cue detection may be overlapping. Furthermore, it gives credence to the overall notion that individual experiences in different linguistic contexts which require constant monitoring of cues and interference suppression may enhance performance on a cognitive task which involves similar non-linguistic demands. Although Salient Cue Detection has not been investigated in the literature, and as there are no empirical studies to which this study's results may be compared, this finding is supported by the theoretical assumptions of the ACH and CPM. This is further discussed vis-à-vis the

theoretical framework in the general discussion chapter (Chapter 6).

An unexpected finding, however, is that codeswitching frequency did not predict performance on the flanker task. While studies such as Rodriguez-Fornells et al. (2012) and Hofweber et al. (2016) reported a positive relation between the two, this was not replicated in the present study. As discussed in the review of the literature, the lack of replicability in findings may be ascribed to methodological variances between studies. The present study differs in task type from that of Rodriguez-Fornells et al. (who used a Flanker-Stroop task) and in the measurement of codeswitching frequency between the present study and that of Hofweber et al.

Another possible explanation is that the flanker task measures conflict monitoring capacities and interference suppression, but not response inhibition. While codeswitching involves processes of conflict monitoring, interference suppression, and response inhibition, the participants in the present study do not need to constantly inhibit prepotent responses in social settings. This is due to their linguistic context(s) which allows them to freely switch between languages, especially as the majority of their interlocutors share their three languages. On the other hand, their ability to attend to cues during conversations, such as topic change, in order to select appropriate language(s) is in line with the conflict monitoring demands which underlie the flanker task. Thus, the finding that salient cue detection is a significant predictor of Conflict Effect scores is further attested.

Model 2: Group x Language Exposure x CSF x SalC

The second model examining language experience as a predictor of Conflict Effect scores included Group (2 levels), Language Exposure (3 levels), as well as CSF and SalC. Although Group, CSF, and SalC were examined in the previous model, their interaction with language exposure and their influence on the overall structure of the

model necessitated their inclusion. As with model 1, interactions between Group, CSF, and SalC were included, as well as between Group and Language Exposure.

The results, summarised in Table 4.44, indicated that there were no significant main effects or interactions between any of the predictors, with SalC being marginally, though not statistically, significant in this model. These findings suggest that like Language Use, Language Exposure was not a predictor of Flanker task performance. Although the groups differed significantly in the amount of input they receive from each language, this did not influence their conflict monitoring performance. This was further verified by creating models with only Group x Language Exposure as predictors of Conflict Effect scores with no significant main effects for either variable.

Table 4.44 GLM Output Summary for Conflict Effect scores by Language Experience (Model 2) (Estimated coefficients, standard errors (SE), t-values, and p-values) Asterisks (*) indicate: p< 0.05 (*), p< 0.01 (**), p< 0.001(***) and p< 0.0001 (****).

	Conflict Effect Scores			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	904.36	201.37	4.49	<0.001***
Group	400.92	228.53	1.75	0.08
Arabic Exposure	19.17	21.33	0.89	0.37
English Exposure	6.31	20.35	0.31	0.76
French Exposure	1.94	19.63	0.10	0.92
CSF	1.29	0.86	1.51	0.14
SalC	0.66	0.34	1.98	0.06
CSF*SalC	1.25	0.88	1.43	0.16
Group*Arabic Exposure	4.06	118.54	0.03	0.97
Group*English Exposure	0.59	0.60	0.98	0.33
Group*French Exposure	0.22	0.19	1.16	0.23
Group*CSF	0.14	0.08	1.75	0.08
Group*SalC	0.38	0.20	1.84	0.07

Formula: Conflict Effect ~ Group(Language Exposure + SalC + CSF) + CSF*SalC* + (1|Participant)

4.5.2.2. Language Experience and the Simon Task

In order to investigate the influence of language experience on performance on

the Simon task, and specifically Simon Effect scores, a generalised mixed-effects model was fitted with the transformed Simon Effect scores as the outcome variable. The predictors included Group (2 levels: EMI vs. FMI), Language Exposure (3 levels: Arabic, English, and French), Language Use (3 levels: Arabic, English, and French), self-reported code-switching frequency (CSF), and self-reported salient-cue detection (SalC). As with the Conflict Effect models, the variables were examined in separate models due to the multicollinearity between Language Use and Language Exposure. The selected models' comparisons were determined by AICs and likelihood tests and further verified by the glmulti function and through the distribution of residuals.

Model 1: Group x Language Use x CSF x SalC

The initial model included main effects for Group, Language Use, CSF, and SalC, as well as interactions Group x CSF, Group x SalC, CSF x SalC, and Group x CSF x SalC. The results indicated that Group and Language Use were not significant predictors (both $p>0.05$, see Appendix H), yet they were retained in the model as they were theoretically-motivated. However, the removal of within-variable interactions in Language Use as well as all interactions involving Group and Language Use improved the fit of the model. The glmulti function further confirmed the goodness-of-fit of the model. Results of the analyses, summarised in Table 4.56, revealed significant main effects for CSF and SalC, as well as a significant interaction in CSF x SalC ($F(1,56)=7.12$, $p<0.01$). That is, the self-reported codeswitching frequency and salient-cue detection were strong predictors of Simon Effect scores. In other words, the more frequently participants engaged in codeswitching, the better their inhibitory control (i.e., the smaller their Simon Effect scores). Furthermore, the more attuned participants were to cues in their environment, the lower their Simon Effect score. The significant interaction between CSF and SalC indicated that the higher the CSF and SalC scores, the

lower the Simon Effect score, but having the CSF score was a stronger predictor of Simon task performance than SalC. As there were no group differences as observed in the non-significant Group interactions and as revealed by the releveling of the model for each group, data from both groups is presented in Figure 4.17.

Table 4.45 GLM Output Summary for Simon Effect by Language Experience (Model 1) (Estimated coefficients, standard errors (SE), t-values, and p-values)
Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗)

	Simon Effect Scores			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	5289.96	1874.65	2.83	0.00**
Group (EMI vs. FMI)	0.41	0.24	1.67	0.10
Arabic Use	0.27	0.24	0.92	0.36
English Use	0.20	0.24	0.84	0.41
French Use	0.24	0.24	0.98	0.32
Code-switching Frequency	5.29	1.35	3.89	0.00***
Salient-cue Detection	4.67	1.58	2.96	0.00**
CSF*SalC	2.73	0.89	3.11	0.00**

*Formula: Simon Effect ~Group+ Language Use + CSF*SalC +(1+SalC*CSF|Participant)*

These results are in line with both empirical and theoretical studies and with the predictions made in the present study, specifically the notion that an increased codeswitching frequency predicts better cognitive performance on the Simon task, indicating better inhibitory skills. These findings parallel those of Verreyet et al. (2016), for instance, who reported a positive correlation between codeswitching frequency and Simon task performance. They are also in line with the predictions made by Costa et al. (2009) and Green and Abutalebi (2013). Given the complexity of the arrow Simon task, it could be argued that participants recruited processes similar to those involved in codeswitching and salient cue detection. Specifically, the need to monitor and resolve the high-levels of conflict created by the bivalent stimuli and 50-50 trial distribution, as well as to suppress prepotent responses, is similar to the processes involved in SalC and CSF.

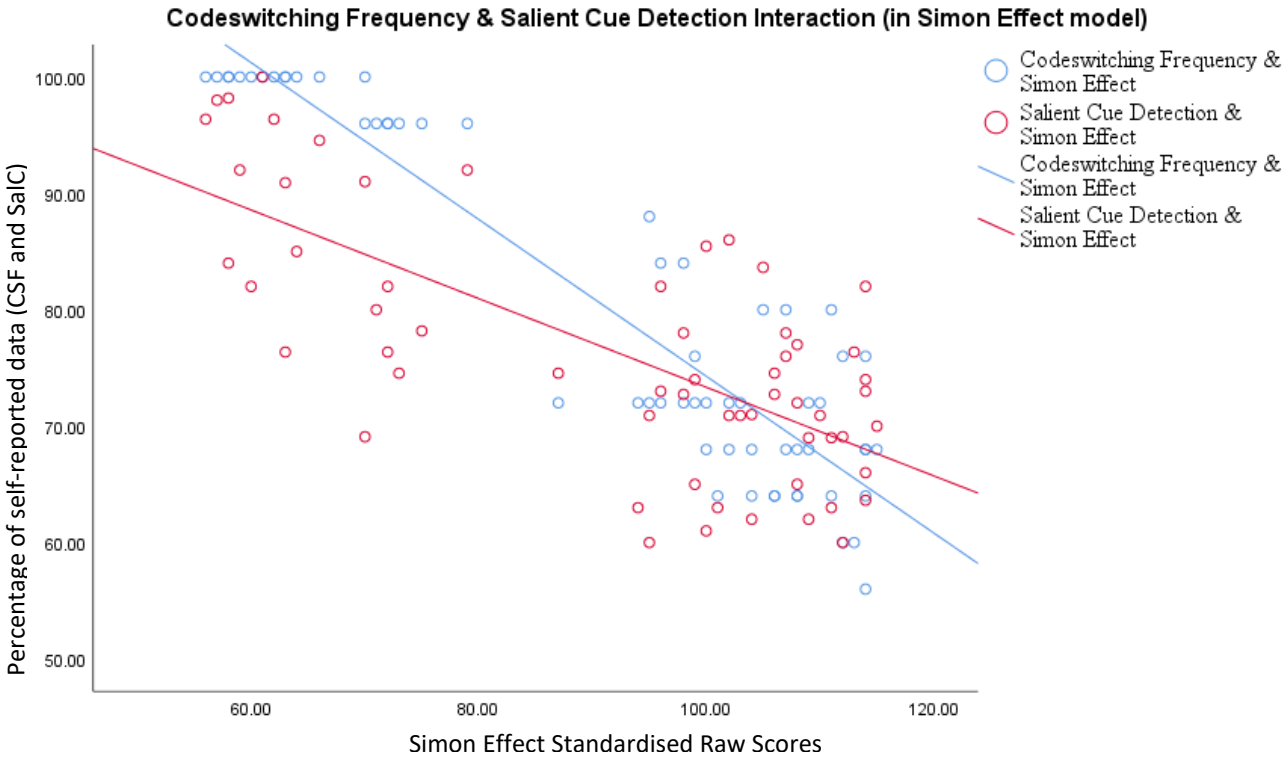


Figure 4.17 Interaction between CSF and SalC in the Simon Task

Model 2: Group x Language Exposure x CSF x SalC

A general linear model was then created to examine the influence of Language Exposure, as opposed to Language Use in Model 1, on Simon Effect scores. The log-transformed Simon Effect scores were fitted as the outcome variable, whereas Group, Language Exposure, CSF, and SalC were fitted as predictor variables, along with interactions between SalC and CSF. As with Model 1, Group and Language Exposure were not significant predictors, yet they informed about the variables which were at the crux of the model and were therefore retained (for simpler models, see Appendix H). The results of the analyses are offered in Table 4.46.

Table 4.46 GLM Output Summary for Simon Effect by Language Experience (Model 2) (Estimated coefficients, standard errors (SE), t-values, and p-values
Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗)

	Simon Effect Scores			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	2035.51	1402.84	1.45	0.15
Group (EMI vs. FMI)	429.00	411.06	0.41	0.68
Arabic Exposure	0.62	3.90	0.16	0.87
English Exposure	23.65	12.52	1.89	0.07
French Exposure	0.04	0.16	0.25	0.80
Code-switching Frequency	0.52	0.16	3.26	0.00***
Salient-cue Detection	0.45	0.17	2.67	0.01*
CSF*SalC	0.07	0.02	3.99	0.00***

*Formula: Simon Effect ~ Group+Language Exposure + CSF*SalC +(1+CSF*SalC|Participant)*

The results therefore revealed that Language Exposure did not serve as a significant predictor of Simon task performance; that is, the differential amounts of input from each language did not influence the inhibitory control abilities of the participants in this study. This dissociation may be due to the participants' overall high exposure to at least two languages, thus potentially attenuating the influence of this variable on cognitive task performance. However, as in Model 1, CSF and SalC, as well as the interaction between these two variables, significantly influenced Simon Effect scores. These results point to the importance of examining language experience when interpreting findings from cognitive task performance. This also highlights the potential overlap between the processes underlying linguistic and non-linguistic behaviour.

4.5.2.3 Language Experience and Working Memory (BDST)

In order to explore the influence of language experience on BDST scores, a generalised linear model with a binomial link function was selected using the MASS library in R. This is due to the BDST scores constituting a discrete count outcome variable with a ceiling. The model was fitted with Group, Language Use, Language

Exposure, CSF, and SalC. As with the Simon Effect and Conflict Effect scores, the multicollinearity between Language Use and Language Exposure required that the separate models be used to analyse the data. These models were structured in the same way as those adopted in the analysis of the Flanker and Simon tasks.

Model 1: Group x Language Use x CSF x SalC

For Model 1, the best-fit structure included main effects for the predictors Group (2 Levels), Language Use (3 Levels), CSF, and SalC with no interactions between them. The results of this model, summarised in Table 4.47, revealed no significant main effects for any predictors (all $p>0.05$). The exclusion of Group did not influence the outcome of the model. The non-significant findings may be attributed to the nature of the BDST task, as it does not involve the same processes underlying frequent use of each language, frequent codeswitching and constant conflict monitoring.

As Simon Effect scores were predicted by CSF and SalC, and given the correlation between Simon Effect scores and BDST scores (RQ1), it was of interest to examine whether these two variables influence BDST scores when examined on their own, with the exclusion of Group, Language Use, and Language Exposure. The results of the analysis, offered in Appendix H, revealed non-significant main effects and a non-significant interaction ($p>0.05$). Despite the correlation of the two cognitive measures, language experience differed in its influence on the two tasks.

Table 4.47 GLMM Output Summary for BDST by Language Experience (Model 1) (Estimated coefficients, standard errors (SE), z-values, and p-values)

	BDST Scores			
	Estimate	**SE**	**z-value**	**Pr(>|z|)**
Intercept	7786.02	7960.89	0.98	0.33
Group (EMI vs. FMI)	4741.14	5003.84	0.95	0.35
Arabic Use	77.88	79.61	0.98	0.33
English Use	43.41	97.18	0.45	0.66
French Use	23.26	38.55	0.61	0.54
Code-switching Frequency	0.05	0.18	0.26	0.80
Salient-cue Detection	0.03	0.10	0.29	0.78

Formula: BDST ~ Group +Language Use + SalC +CSF+(1|Participant)

Model 2: Group x Language Exposure x CSF x SalC

This model was similarly structured to the preceding model, with Language Exposure replacing Language Use as a predictor. Initially, a full model (with interactions) was included (see Appendix H), but as an inspection of the model's ANOVA results indicated no significant main effects or interactions ($p>0.05$), only the main predictors which were theoretically driven were assessed. The model still revealed no significant results, as indicated in Table 4.48.

Table 4.48 GLMM Output Summary for BDST by Language Experience (Model 2) (Estimated coefficients, standard errors (SE), z-values, and p-values)

	BDST Scores			
	Estimate	**SE**	**z-value**	**Pr(>|z|)**
Intercept	7786.02	7960.89	0.98	0.33
Group (EMI vs. FMI)	4741.14	5003.84	0.95	0.35
Arabic Exposure	77.88	79.61	0.98	0.33
English Exposure	43.41	97.18	0.45	0.66
French Exposure	23.26	38.55	0.61	0.54
Code-switching Frequency	0.05	0.18	0.26	0.80
Salient-cue Detection	0.03	0.10	0.29	0.78

Formula: BDST ~ Group +Language Exposure + SalC +CSF+(1|Participant)

4.5.2.4 Summary of Language Experience as a Predictor of Cognitive Performances

To summarise, the results of the models investigating the influence of self-reported language use, language exposure, codeswitching frequency, and salient cue detection indicated that Conflict Effect scores were significantly predicted by SalC for both the groups ($\beta=0.67$, $SE=0.31$, $t=2.18$, $p<0.05$). That is, the more attuned participants reported being to cues in their linguistic contexts, the better their conflict monitoring abilities (i.e., the lower the Conflict Effect scores) as measured by the flanker task. Simon Effect scores were significantly predicted by SalC ($\beta=0.45$, $SE=0.17$, $t=2.67$, $p<0.05$) and CSF ($\beta=0.52$, $SE=0.16$, $t=3.26$, $p<0.001$), and there was a significant interaction between these two variables ($\beta=0.07$, $SE=0.02$, $t=3.99$, $p<0.001$). These findings indicate that participants who frequently codeswitched and constantly monitored external cues possessed enhanced inhibitory control skills (i.e., lower Simon Effect scores), and that CSF was a stronger predictor of Simon task performance than SalC. On the other hand, BDST scores were not predicted by any variable, as the models examining working memory revealed no significant main effects or interactions.

4.6 CHAPTER SUMMARY

This chapter first offered a detailed discussion of the general analytical strategy adopted in the analyses of the cognitive tasks. It then presented the results of the tasks by comparing EMI and FMI participants' performance on the flanker task, Simon task, and BDST (RQ1A, B, & C). Correlations between the three tasks were then performed (RQ1D), thus concluding the analysis of RQ1. The independent variables, Language Proficiency and Language Experience, were then analysed, highlighting group differences as well as variations in and/or overlaps between the different instruments measuring these variables. The chapter then presented the analyses of the RQ2 which

addressed the influence of the independent variables on cognitive task performance. Given the complexity of the independent variables and multicollinearity between them, separate mixed-effects models were used to analyse the data. The results of each model were presented and discussed separately, followed by a general summary of the results of each independent variable's influence on the cognitive tasks.

CHAPTER 5: RESULTS AND DISCUSSION CHAPTER II

5.1 INTRODUCTION

Following the analysis of the cognitive tasks in Chapter 4, the present chapter answers the research questions pertaining to the linguistic tasks, namely the trilingual picture-naming task (TPNT) and the narration task (NT). For each task, the general analytical strategy is described, followed by detailed analyses and brief discussions of the results. Predictors of linguistic task performance are then analysed and discussed, followed by the analysis of the final research question of this thesis which attempts to combine both cognitive and linguistic task performance under one analytical paradigm.

5.2 RQ3—LINGUISTIC TASK PERFORMANCE

RQ3. Do EMI and FMI participants differ in performance on the linguistic tasks: (A) the trilingual picture-naming and (B) the narration task?

The third research question explores whether EMI and FMI participants differ in their performance on the two linguistic tasks: the TPNT and the NT. The first sub-section offers analyses of the TPNT, the second sub-section analyses data from the NT, and the final sub-section correlates data from both tasks.

5.2.1 Trilingual Picture-Naming Task

RQ3A. Do EMI and FMI participants differ in their performance on the trilingual picture-naming task in terms switch costs (accuracy and reaction times) for each language?

The first part of RQ3 explores whether EMI and FMI participants differ in their accuracy, reaction times (RTs), and switch costs (for both accuracy and RTs) on the picture-naming task. To investigate the hypotheses outlined in the Rationale (Section 2.5), switch costs were measured for the three languages for each group without

language assignments as L1, L2, and L3. As mentioned earlier, this is due to the difficulty in (1) finding a clear definition of language dominance; and (2) establishing a distinction in the three languages as to being the participants' first, second, or third languages given that the participants were simultaneous trilinguals and given the differences between the OPT and VFT scores. Accordingly, the results are presented with the languages 'intact'; i.e., without an *a priori* assignment of the status for each language as the most dominant or weakest language.

5.2.1.1 General Analytical Strategy

As with the analyses conducted on the cognitive tasks, analyses of the TPNT were carried out in the statistical package R (version 3.5.2) (R Development Core Team, 2018) using the MASS, nlme, and lme4 packages (Bates et al., 2015). For the TPNT, the dependent variables were accuracy (number of errors) and response latencies (reaction times). As accuracy was high, no item or participant was excluded from analysis. Both correct and incorrect responses were used to analyse accuracy data, whereas only correct responses were used in RT analyses. Accuracy binary data (correct= 0; incorrect= 1) were fitted in a generalised linear mixed-effects model (GLMM) with multivariate normal random effects using the Penalized Quasi-Likelihood (PQL) method with a binary link function (Bates, 2010). Reaction times, which are continuous, were fitted in linear mixed effects models.

The best-fit models were selected using a stepwise procedure, where only significant predictors were used to compare nested models of increasing complexity. Models which failed to converge were excluded from further comparisons. The best-fit model was then determined by the Akaike's Information Criterion (AIC; Akaike, 1998) and its significance by the likelihood ratio test (see also Matuschek et al., 2015). For the TPNT, the best-fit models for accuracy and response latencies included the

experimental factors of Group (EMI vs. FMI), Language (Arabic, English, or French), and Trial Type (Repetition vs. Switch). All models had fully random slopes and intercepts by participants and items, i.e., they had maximal random-effects structures (Barr et al., 2013; Bates et al., 2015). Before accepting the models, the goodness-of-fit of each model was checked through visual inspections of the distribution of the residuals (Baayen, 2008). For RTs, the residuals revealed a non-normal distribution. To approximate residuals' normality and to stabilise variance, meeting the homoscedasticity condition (Baayen, 2008), response latencies were log-transformed.

After the selection of the models and data transformation, the variables were centred and coded. Group and Trial Type, which are two-level factors, were coded as 0.5 and - 0.5, whereas the three-level Language factor was dummy coded, where one level of this factor served as a baseline reference against which the other two levels were compared. Using the relevel function in R, the order of the levels was changed to obtain the contrasts for the groups, languages, and trial types. A Holm-Bonferroni adjustment was applied for multiple comparisons. Details for each model and the transformations made are offered in the subsequent sections.

5.2.1.2 Accuracy

Accuracy rates on the TPNT were calculated as the total number of errors in each language subtracted from the total number of responses. The errors included hesitations, mispronunciations, microphone mis-triggering, providing wrong words, or speaking in the non-target language. Accuracy data indicated that participants performed close to ceiling in terms of response accuracy, with an overall accuracy rate of 95.86% (or a total of 1,491 errors of 36,000 trials for all participants). The high level of accuracy may be attributed to participants' frequent codeswitching in their linguistic contexts. The descriptive results of accuracy rates for each group and each language in both repetition

and switch trials (along with switch costs) are summarised in Table 5.1.

Table 5.1 TPNT Accuracy Means (in %) and Standard Deviations (EMI and FMI)

	Arabic		English		French	
	M	**SD**	**M**	**SD**	**M**	**SD**
EMI						
Repetition	94.78	(2.08)	97.31	(1.77)	95.02	(2.21)
Switch[14]	93.86	(3.41)	96.50	(2.64)	93.99	(4.16)
Switch Cost	0.92		0.81		1.03	
FMI						
Repetition	96.02	(1.62)	96.81	(1.64)	97.98	(1.17)
Switch	95.09	(2.76)	95.83	(2.73)	97.12	(1.96)
Switch Cost	0.93		0.98		0.86	

The Shapiro-Wilk's test ($p<0.05$), along with visual inspections of normal Q-Q plots and histograms, revealed that accuracy rates were not normally distributed. A generalised mixed effects model was therefore used to assess this dependent variable using the MASS and nlme packages. Accuracy binary data were fitted as the outcome variable, whereas Group, Language, and Trial Type were fitted as predictors. The model also included all interactions between the predictors. A summary of the model's ANOVA results is presented in Table 5.2.

Table 5.2 GLMM ANOVA (Type III) Summary for TPNT Accuracy Data

	F-Statistic	DF1	DF2	p-value
Intercept	12493.06	1	530	<0.0001
Group (EMI vs. FMI)	4.98	1	58	0.03
Trial Type (Repetition vs. Switch)	21.40	1	530	<0.0001
Language (Arabic vs. English vs. French)	27.71	2	530	<0.0001
Group*Trial Type	0.24	1	530	0.78
Group*Language	32.41	2	530	<0.0001
Trial Type *Language	0.06	2	530	0.93
Group*Trial Type*Language	0.11	2	530	0.89

[14] Following research conventions, switch trials include switches from both other languages. However, in post-hoc analyses of the data, switches are separated by language to examine switch costs for each language pair.

As indicated in Table 5.2, the model revealed that there were significant main effects for Group, Trial Type, and Language. The analysis also revealed a significant interaction between Group x Language ($F(2,530)=32.41$, $p<0.0001$), illustrated in Figure 5.1. No other significant main effects or interactions were obtained from the model.

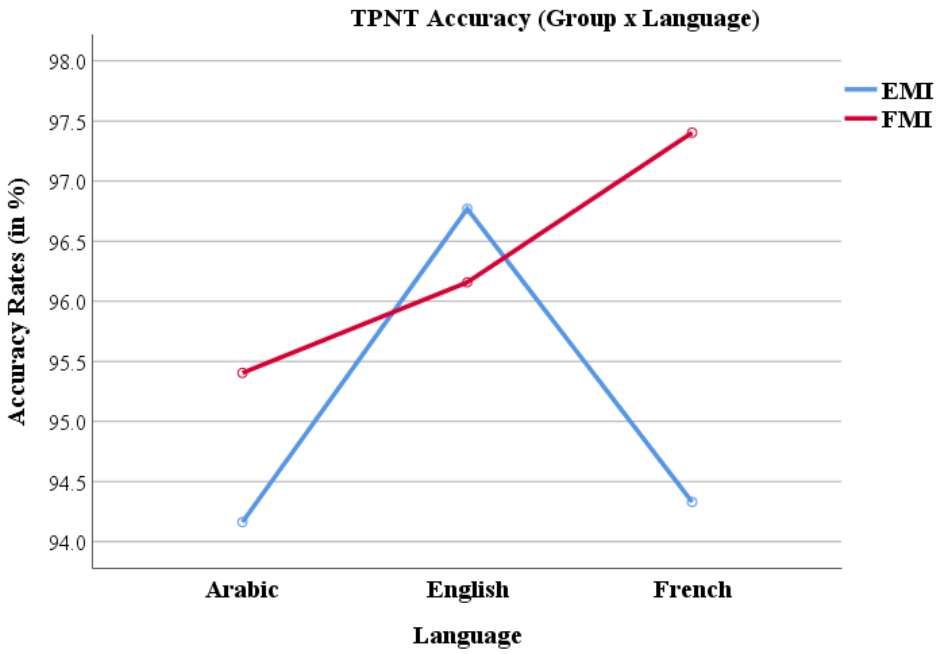


Figure 5.1 TPNT Accuracy Interaction Plot (Group x Language)

Overall, the results of the analysis indicate that for both EMI and FMI, repetition trials were significantly more accurate than switch trials (EMI: $\beta=0.91$, $SE=0.31$, $z=2.93$, $p<0.01$; FMI: $\beta=0.92$, $SE=0.31$, $z=2.96$, $p<0.01$). This was independent of Language and Group, as suggested by the non-significant Language x Trial Type interaction and the Group x Trial Type interaction ($p>0.05$ for both interactions).

In order to investigate the significant Group x Language interaction, simple main effects were analysed by changing the factor contrast coding for the mixed model, and *post hoc* contrasts were analysed by changing the reference (intercept) for each level of Group and Language. The GLMM output is offered in Table 5.3, where the baseline group is EMI, the baseline language is Arabic, and the baseline trial type is Repetition.

Table 5.3 GLMM Output Summary for TPNT Accuracy Data (Estimated coefficients, standard errors (SE), z-values, and p-values) Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗)

Predictor	Coefficient	SE	z-value	p-value
Fixed Effects				
Intercept	94.78	0.44	211.76	0.00****
Group	1.25	0.69	1.83	0.07
Switch	0.92	0.38	2.42	0.02*
English	2.53	0.38	6.65	0.00****
French	0.22	0.47	0.48	0.61
Group*Switch	0.01	0.60	0.01	0.99
Group*English	1.74	0.66	2.65	0.01**
Group*French	1.21	0.71	2.42	0.02*
Switch*English	1.73	0.54	0.21	0.83
Switch*French	0.11	0.60	0.14	0.88
Group*Switch*English	0.08	0.85	0.19	0.85
Group*Switch*French	0.16	0.89	0.17	0.865
Random Effects				
Participants	1.93			
Items	0.01			
Residuals	2.39			

*Formula: Accuracy ~ Group*Language*Trial Type + (1 + Language*Trial Type| Participant) + (1| item)*

Post-hoc analyses revealed that FMI participants responded significantly more accurately in French (β= 3.07, SE=0.61, z= 5.01, p<0.0001) and Arabic (β= 1.24, SE=0.61, z= 2.02, p<0.05) than EMI participants, whereas there was no significant difference in accuracy between the two groups in English (p>0.05). Furthermore, the analyses revealed that EMI participants responded significantly more accurately in English than they did in Arabic (β=2.53, SE=0.38, z=6.65, p<0.0001) and French (β=2.31, SE=0.38, z=6.08, p<0.0001), with no difference in error rates between Arabic and French responses (p>0.05). FMI participants, on the other hand, were significantly more accurate in French than in English (β=1.17, SE= 0.54, z= 2.16, p<0.05) and Arabic (β=1.95, SE=0.54, z=3.62, p<0.001), with no difference in accuracy between the latter two languages (p>0.05).

These results indicate that TPNT responses were significantly more accurate for the immersion language of EMI and FMI participants; English and French,

respectively. While these results were predicted, differences between the immersion languages and Arabic were not expected for either group, given both groups' high proficiency in Arabic. Furthermore, FMI participants were expected to exhibit similar accuracy rates across the three languages, especially as they were believed to be balanced trilinguals and given their high proficiency scores in the three languages, as measured by the OPT. Another unexpected finding is that EMI participants had a comparable number of errors in both Arabic and French, despite having a lower proficiency score in French. Accordingly, it is perhaps not proficiency, but language use, which modulates accuracy rates on this task.

5.2.1.3 Prior Language Effects

In the preceding analysis, for each language, switch costs were aggregated for the two non-target languages. For instance, Arabic switch costs reflect the mean cost to switch from English *and* French (i.e., English-Arabic and French-Arabic trials). However, as mentioned in Chapter 3, a more detailed examination was adopted to disentangle the effects of previous trial language on switch costs. Planned-contrast analyses were performed on the data which was split into three language pairs (Arabic-English, Arabic-French, and English-French) for each group. Each language pair included repetition and switch trials from the two relevant languages. For example, for the Arabic-English language pair, Arabic repetition trials, English repetition trials, Arabic-English switch trials, and English-Arabic switch trials were included. To explore the influence of the preceding trial on switch costs, a variable with two levels (Arabic vs. English, Arabic vs. French, or English vs. French) was coded for repetition trials and switch trials as +0.5 and -0.5. For repetition trials, this variable does not signify a change; accordingly, a significant interaction between Trial Type (repetition vs. switch) and Preceding Language (the two languages of each language pair) indicates an

influence of preceding trials on switch trials (i.e., on switch costs).

Although no language assignments were given, switch costs were predicted to differ by group. For EMI, switch costs were expected to be larger for Arabic and English compared to the French, yet they were expected to be symmetrical between Arabic and English, as stipulated by the ICM. For FMI, it was predicted that switch costs would be symmetrical for the three languages as this group was expected to be have similar proficiencies in the three languages. Descriptive data of the language pairs are summarised in Table 5.4.

Table 5.4 TPNT mean accuracy rates in percentages and Standard Deviations for Language Pairs

	EMI				**FMI**			
Arabic & English	**Arabic**		**English**		**Arabic**		**English**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Repetition	94.78	(2.08)	97.31	(1.77)	96.02	(1.62)	96.21	(1.64)
Switch	93.56	(4.49)	96.00	(2.82)	95.16	(3.28)	95.32	(3.42)
Switch Cost	1.22		1.31		0.86		0.89	
Arabic & French	**Arabic**		**French**		**Arabic**		**French**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Repetition	94.78	(2.08)	95.02	(2.21)	96.02	(1.62)	97.98	(1.17)
Switch	94.16	(3.00)	95.00	(4.44)	95.03	(3.41)	97.40	(1.90)
Switch Cost	0.62		0.02		0.99		0.58	
English & French	**English**		**French**		**English**		**French**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Repetition	97.31	(1.77)	95.02	(2.21)	96.21	(1.64)	97.98	(1.17)
Switch	97.01	(2.82)	92.99	(4.39)	95.12	(2.94)	96.83	(2.54)
Switch Cost	0.31		2.03		1.09		1.15	

For EMI, analyses of the planned contrasts models, illustrated in Figure 5.2 (top row), revealed that for the Arabic-English language pair, repetition trials were significantly more accurate than switch trails ($\beta=0.71$, $SE=0.29$, $z=2.44$, $p<0.05$), and Arabic-English switch costs were symmetrical due to the non-significant interaction between Preceding Language and Trial Type ($p>0.05$). For the Arabic-French language pair, switch trials were less accurate than repetition trials, but this did not reach

significance ($p>0.05$). There was a significant interaction between Preceding Language and Trial Type given the very small switch costs in French; that is, switch costs in Arabic were significantly larger than French switch costs ($z=2.35$, $p<0.05$). For the English-French language pair, responses were significantly more accurate in English than in French ($z=4.14$, $p<0.0001$), and switch costs were significantly smaller for English than in French as evidenced by the significant interaction between Preceding Language x Trial Type ($z=7.36$, $p<0.0001$). This indicates that switch costs were asymmetrical for the English-French language pair with English switch costs being smaller than French switch costs. Furthermore, comparing the language pairs revealed that for Arabic, switch costs are larger when switching from English. Switch costs in English were larger when switching from Arabic. Finally, switch costs in French were larger when switching from English.

Taken together, the analyses of the language pairs revealed that switch costs were symmetrical for the Arabic-English, but they were asymmetrical for English-French and Arabic-French language pairs. These results are in line with the predictions of the ICM; however, the 'reversal' of switch costs in the English-French language pair is an unexpected finding.

The ICM suggests that switch costs are larger for the 'stronger' language, leading to asymmetrical switch costs between the 'stronger' and 'weaker' languages. However, the asymmetry in the English-French language pair stems from the larger switch costs for the 'weaker' language, French. One possible explanation is participants' high proficiency in English and very high performance on this task may have led to a ceiling effect which obfuscates switch costs for English trials.

Alternatively, it is possible that the typological distance (or proximity in this case) between English and French modulated switch costs. Accordingly, French trials may

have facilitated or primed naming in English trials, leading to reduced switch costs in English, though this was only a unidirectional influence. Another possible explanation is that participants may have relied on naming strategies in this task which were most observed between the weakest and strongest languages (i.e., English and French, respectively). However, this is speculative and difficult to establish, as the ICM accounts for the overall findings including the asymmetry in the English-French pair—it is only the directionality of the switch costs which is at odds with the ICM.

A final possible explanation is the effect of 'backward inhibition' (Philipp & Koch, 2009; Philipp et al. 2007), which might have influenced the results. Backward inhibition suggests that performance in n-2 repetition trials (e.g., A-B-*A*) is less accurate and slower than in no repetition trials (e.g., C-B-*A*) due to persisting inhibition of a previously relevant task which then may hamper performance when the same task is reactivated in an n-2 trial. That is, for A-B-*A* and C-B-*A*, the prior activation of 'A' (bold-faced) in the **A**-B-*A* task might lead to a less accurate and slower response of '*A*'. Thus, *A* is named more accurately in C-B-*A* than A-B-*A*. Backward inhibition has not been consistently reported in the literature (Baus, Branzi, & Costa, 2015) and is generally based on RT rather than accuracy data, and it was not controlled for in the present study. However, it may have influenced the results of this task, although again, it is peculiar that such an influence would only be observed in this language pair.

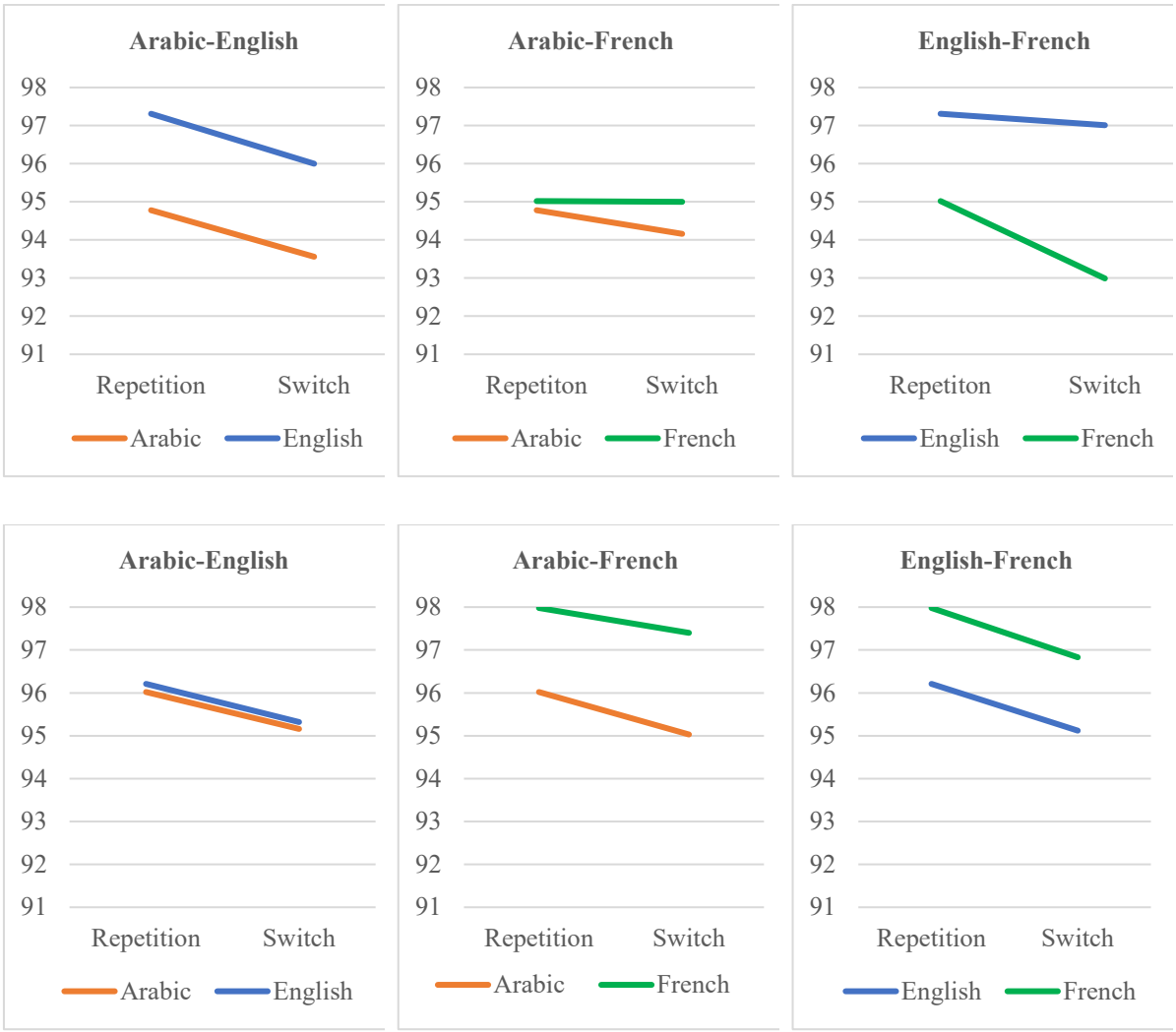


Figure 5.2 TPNT Mean Accuracy Rates (in %) by Language Pairs (Top Row=EMI, Bottom Row=FMI)

For the FMI group, accuracy rates on repetition trials the Arabic-English language pair, repetition trials were significantly more accurate than switch trails ($z=2.11$, $p<0.05$), and English switch costs were symmetrical due to the non-significant interaction between Preceding Language and Trial Type ($p>0.05$). For the Arabic-French language pair, responses were marginally, but not significantly, more accurate in French than in Arabic on switch trials ($p=0.07$), but as this did not reach significance, the switch costs were symmetrical. For the English-French language pair, responses were significantly more accurate in French than in English ($z=3.89$, $p<0.01$), and switch costs were marginally yet not significantly smaller for English than for French.

Accordingly, switch costs were symmetrical for the three language pairs. Examining prior language effects further through language pairing, switching into Arabic from either English or French did not influence accuracy rates ($p>0.05$). Similarly, switching into French from Arabic or English resulted did not influence accuracy rates, nor did switching from either Arabic or French into English ($p>0.05$ for both contrasts). The results are depicted in Figure 5.2 (bottom row).

5.2.1.4 Summary of TPNT Accuracy Analysis

To summarise the results of TPNT accuracy rates, for both groups, repetition trials were significantly more accurate than switch trials across the three languages. Furthermore, FMI participants were significantly more accurate in French and Arabic than EMI participants, with no significant difference in English between the two groups. Planned-contrast analyses revealed that accuracy rates across the different language pairs significantly influenced the magnitude of switch costs for EMI participants only. EMI participants showed symmetrical switch costs for the Arabic-English language pair, yet asymmetrical switch costs for the Arabic-French and English-French language pairings. For the EMI group, English switch costs were larger when switching from Arabic than French. Arabic switch costs were significantly larger when switching from English than French. Finally, French switch costs were significantly larger when switching from English than Arabic. FMI participants, on the other hand, exhibited symmetrical switch costs in all the language pairings.

5.2.1.5 Reaction Times (RTs)

Reaction time data were not normally distributed, as indicated by the Shapiro-Wilk Test ($p<0.01$) and visual inspection of the normal Q-Q plots and histograms. Table 5.5 offers descriptive statistics of the raw RT scores, and Figure 5.3 illustrates the distribution of raw RTs (correct responses only) for each language and group. To

approximate normality, an inverse transformation for the RT data was conducted through the Box-Cox function of the MASS package in R (Venables & Ripley, 2002). Upon applying this transformation, visual inspection of the distribution of the entire dataset, as well the distribution of reaction times for each participant, indicated that the reciprocal transformation was indeed the most appropriate transformation of the data. A multilevel (hierarchical) linear model from the nlme package was then selected to analyse the data, as indicated in the general analytical strategy. The model was fitted with transformed RTs as the response variable and with Group, Trial Type, and Language, along with their interactions, as predictor variables.

Table 5.5 TPNT Mean RTs (in ms) and Standard Deviations

	Arabic		**English**		**French**	
	M	**SD**	**M**	**SD**	**M**	**SD**
EMI						
Repetition	921	(57)	828	(99)	1003	(114)
Switch	1159	(154)	1050	(164)	1136	(123)
Switch Cost	238		222		133	
FMI						
Repetition	909	(82)	828	(105)	824	(141)
Switch	1170	(141)	1110	(147)	1115	(110)
Switch Cost	261		282		291	

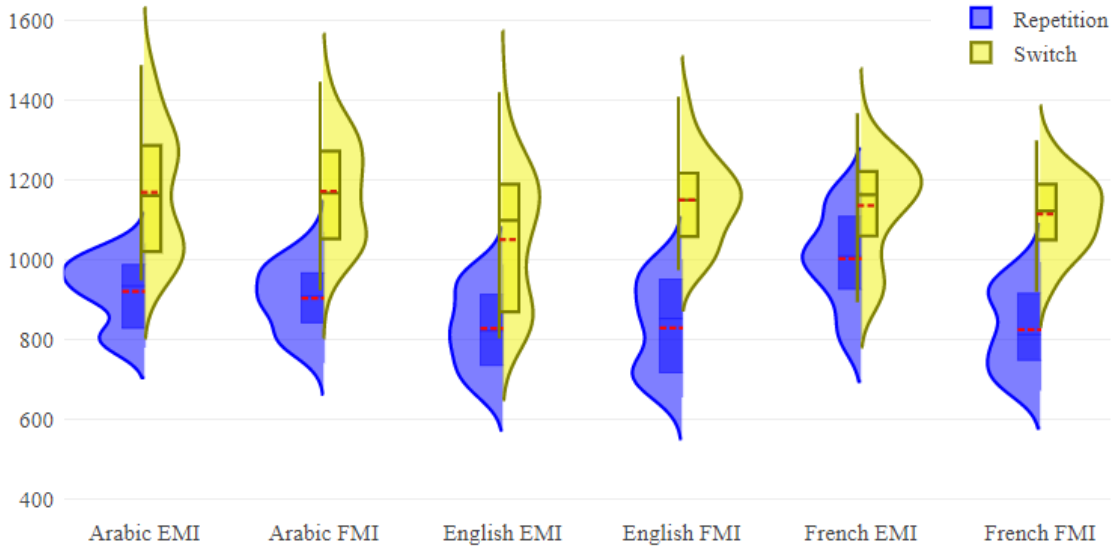


Figure 5.3 Violin plots showing the distribution of the RTs of the correct TPNT responses for each language and group. The plot outline shows the density of data points at the different RTs. The boxplot shows the interquartile range, and the thin vertical line indicates minimum and maximum scores for each language. The median is indicated by the solid horizontal line, while the dotted horizontal red line represents the mean.

The overall model results, summarised in Table 5.6, revealed significant main effects for Trial Type and Language, as well as significant results for all two-way interactions: (1) Group x Trial Type (2) Group x Language; and (3) Trial Type x Language. Most importantly, there was a significant three-way interaction between Group, Trial Type, and Language ($F(2,530)=5.19$, $p<0.0001$). A schematic representation of the three-way interaction is presented in Figure 5.4.

Table 5.6 MEM ANOVA (Type III) Summary of TPNT RT Data

	F-Statistic	DF1	DF2	p-value
Intercept	8871.68	1	530	<0.0001
Group (EMI vs. FMI)	0.04	1	58	0.84
Trial Type (Repetition vs. Switch)	884.05	1	530	<0.0001
Language (Arabic vs. English vs. French)	29.06	2	530	<0.0001
Group*Trial Type	39.70	1	530	<0.0001
Group*Language	33.62	2	530	<0.0001
Trial Type *Language	13.62	2	530	<0.0001
Group*Trial Type*Language	5.19	2	530	<0.001

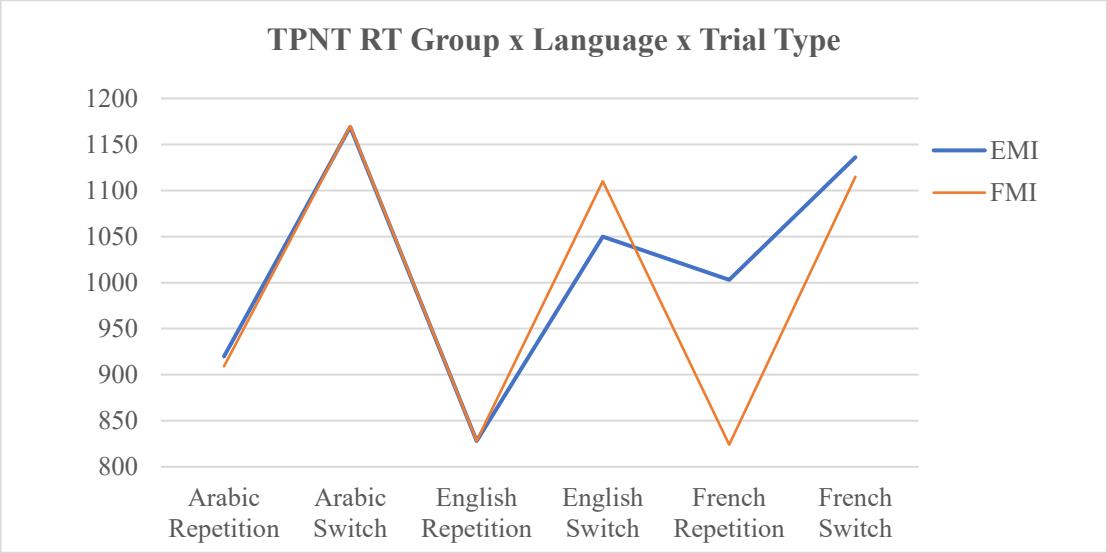


Figure 5.4 Interaction Plot of TPNT RT data (Group x Language x Trial Type)

In order to investigate the significant three-way interaction between Group, Language, and Trial Type, simple main effects were analysed by changing the factor contrast coding for the mixed model, and *post hoc* contrasts were analysed by changing the reference (intercept) for each level of Group, Language, and Trial Type. A sample of the output from the GLM is presented in Table 5.7, where the baseline Group is FMI, baseline language is Arabic, and baseline Trial Type is Repetition.

Table 5.7 MEM Output Summary for TPNT RT Data (Estimated coefficients, standard errors (SE), t-values, and p-values) Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p<0.001(∗∗∗) and p< 0.0001 (∗∗∗∗)

Predictor	Coefficient	SE	t-value	p-value
Fixed Effects				
Intercept	1.13	0.02	51.78	0.00****
Group	0.03	0.03	0.93	0.36
Switch	-0.26	0.02	-12.54	0.00****
English	0.10	0.02	4.39	0.00****
French	0.08	0.02	3.41	0.00***
Group*Switch	0.03	0.03	1.18	0.24
Group*English	0.03	0.03	0.77	0.44
Group*French	-0.17	0.03	-4.90	0.00****
Switch*English	-0.11	0.03	-3.72	0.00***
Switch*French	-0.03	0.03	-0.88	0.38
Group*Switch*English	0.08	0.04	2.04	0.04*
Group*Switch*French	0.13	0.04	3.18	0.00**
Random Effects				
Participants	0.07			
Items	0.00			
Residuals	0.09			

*Formula: RT ~ Group*Language*Trial Type + (1 + Language*Trial Type| Participant) + (1+Language*Trial Type| item)*

The *post-hoc* analyses revealed that the three-way interaction was driven by a significant Trial Type x Language where response latencies on French repetition trials were significantly larger for EMI than FMI participants (β=0.19, SE=0.03, t=6.41, p<0.0001), but this effect was not observed for Arabic or English repetition trials (p>0.05 for both contrasts) or for switch trials in any language.

The three-way interaction was also driven by a significant Group x Language interaction for French and English, where the effect was larger for French than English in both RTs and switch costs. The analyses revealed that FMI participants were significantly faster than EMI participants in French responses (β=0.08, SE=0.03, t=2.92, p<0.01). On the other hand, EMI participants were significantly faster on English trials than the FMI group (β=0.08, SE=0.03, t=2.68, p<0.01). The two groups had comparable RTs on Arabic trials (p>0.05). Furthermore, French switch costs were significantly larger for the FMI than EMI group (β=0.43, SE=0.06, t=7.14, p<0.0001).

and English switch costs were significantly larger for EMI participants (β=0.17, SE=0.06, t=2.76, p<0.0001). The two groups displayed similar switch costs in Arabic. In other words, French switch costs were significantly smaller for EMI participants compared to Arabic and English switch costs, and significantly smaller for EMI compared to FMI participants.

Furthermore, the two groups differed in RTs across the three languages. EMI participants responded significantly faster in English than Arabic (β=0.11, SE=0.02, t=4.77, p<0.0001) and French (β=0.13, SE=0.02, t=5.34, p<0.0001). There was no difference in response latencies between Arabic and French trials. On the other hand, FMI participants were significantly faster in French than Arabic (β=0.06, SE=0.02, t=2.69, p<0.01), but there were no significant differences between Arabic and English or between English and French (p>0.05 for both contrasts).

5.2.1.6 Prior Language Effects

As with the planned contrasts conducted for accuracy data, detailed analyses were performed on RT data to disentangle the effects of previous trial language on switch costs. The data were divided into three language pairs (Arabic-English, Arabic-French, and English-French) for each group. Descriptive data for each language pair is presented in Table 5.8.

Table 5.8 TPNT mean RTs in milliseconds and Standard Deviations (in parentheses) for Language Pairs

	EMI				FMI			
Arabic & English	**Arabic**		**English**		**Arabic**		**English**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Repetition	921	(82)	828	(100)	909	(83)	829	(112)
Switch	1155	(163)	1026	(82)	1181	(99)	1126	(99)
Switch Cost	234		198		272		297	
Arabic & French	**Arabic**		**French**		**Arabic**		**French**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Repetition	921	(82)	1003	(115)	909	(83)	825	(120)
Switch	1163	(132)	1107	(139)	1162	(154)	1094	(91)
Switch Cost	242		104		253		269	
English & French	**English**		**French**		**English**		**French**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Repetition	828	(100)	1003	(115)	829	(112)	825	(120)
Switch	1075	(173)	1165	(99)	1173	(96)	1135	(76)
Switch Cost	247		162		344		310	

As depicted in Figure 5.5, overall, participants in both groups were significantly faster on repetition than switch trials. Analyses of the planned contrasts revealed that for the Arabic-English language pair, EMI participants were significantly faster on repetition trials in English than in Arabic ($\beta=0.13$, $SE=0.02$, $t=5.59$, $p<0.0001$). Further, repetition trials were significantly faster than switch trials ($t=4.20$, $p<0.0001$). There was no significant interaction between Language and Trial Type ($p>0.05$); accordingly, Arabic-English switch costs were symmetrical. For the Arabic-French language pair, EMI participants were significantly faster on repetition than switch trials (Arabic: $t=4.03$, $p<0.001$; French: $t=3.97$, $p<0.001$). Switch costs were asymmetrical for Arabic and French, with Arabic switch costs being significantly larger than French switch costs as indicated by the significant interaction between Trial Type and Preceding Language ($t=5.32$, $p<0.0001$). For the English-French language pair, responses were faster in English than in French ($t=3.11$, $p<0.01$) and in repetition trials compared to switch trials ($t=2.86$, $p<0.05$). There was a significant interaction of Language x Trial Type ($t=4.74$,

$p<0.0001$), indicating that switch costs for English were larger than those for French. That is, for the English-French language pair, switch costs were asymmetrical.

Overall, EMI participants responded faster in Arabic and English than in French, which is in line with the assumptions of the ICM. Furthermore, switch costs were symmetrical for the Arabic-English language pair, but asymmetrical for both the Arabic-French and English-French language pairs, with larger switch costs for Arabic and English compared to the weaker French. These results are also commensurate with the results of the OPT which indicated that EMI participants are similarly proficient in Arabic and English but have a lower proficiency in French.

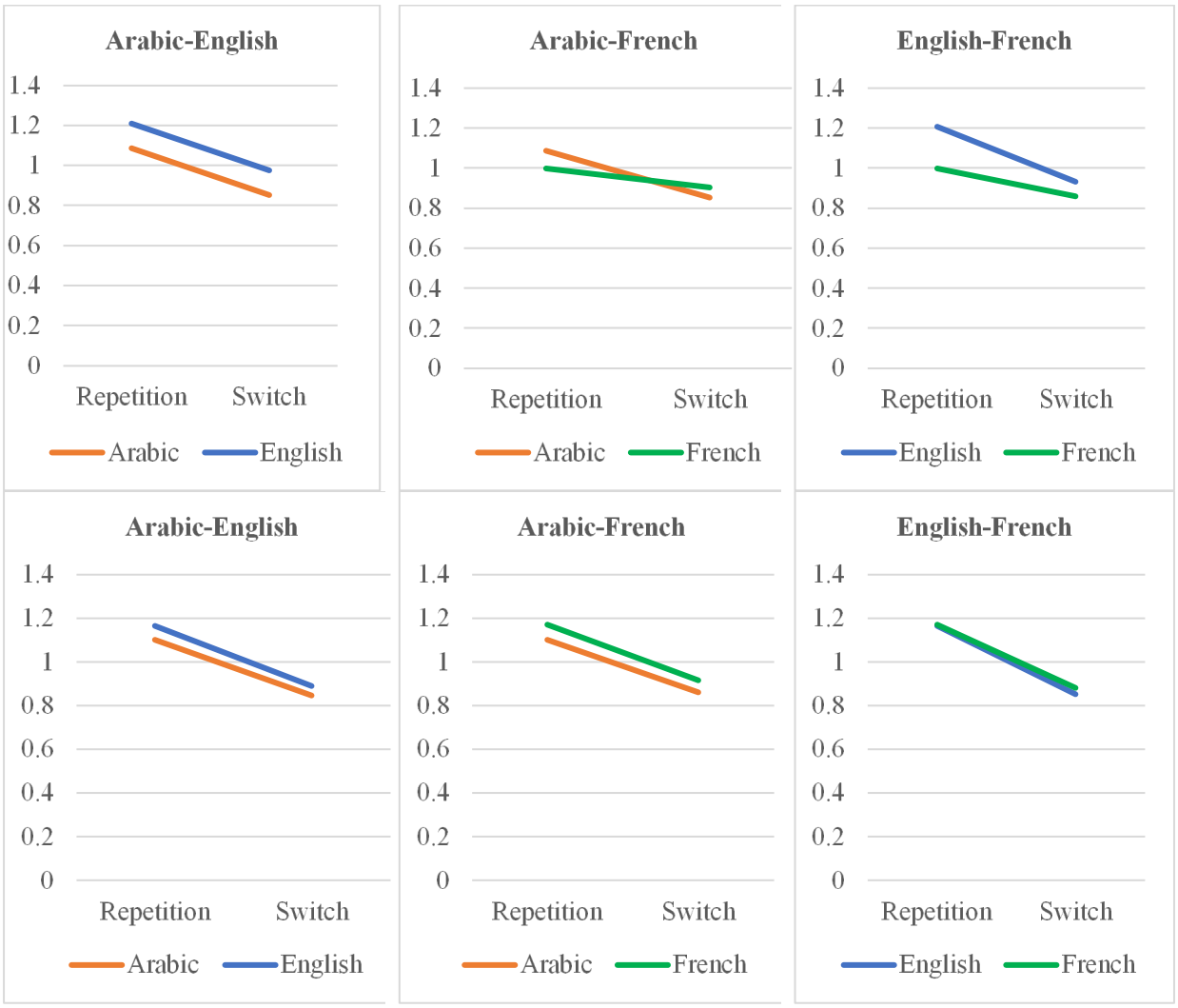


Figure 5.5 Schematic representation of RT switch costs across the language pairs. The y-axis values represent reciprocal (log-transformed) RTs, thus leading to the negative slopes in the interaction plots. Top row=EMI, bottom row=FMI.

FMI participants, on the other hand, revealed different patterns of switching behaviour compared to their EMI peers. The post-hoc analyses revealed that RTs in repetition trials were significantly faster than responses in switch trials in all three languages, as indicated by the non-significant interactions between Language and Trial Type ($p>0.05$ for all language pair contrasts). Furthermore, responses in French were faster than in Arabic in the Arabic-French pair and faster than English in the English-French pair, but neither contrast reached significance ($p>0.05$). There was also no significant difference between Arabic and English response latencies in the Arabic-

English language pair. Taken together, these results revealed symmetrical switch costs between the three languages across the three language sets. For the FMI group, therefore, switch costs were commensurate with the proficiency results obtained from the OPT. These results also offer support to the ICM, which predicts symmetrical switch costs between the three languages of balanced bi/multilinguals.

5.2.1.7 Summary of TPNT Analyses

To summarise, participants in both EMI and FMI groups responded significantly more accurately and significantly faster on repetition trials than on switch trials in all languages. However, the TPNT revealed group differences, where EMI participants were significantly less accurate in French and Arabic compared to FMI participants, with no difference between the groups in English. However, accuracy data analyses revealed unexpected findings for the EMI group who exhibited (a) smaller switch costs in English compared to French in the English-French language pair; and (b) symmetrical switch costs between Arabic and French. Several explanations were offered, but these findings were mainly attributed to participants' very high performance on this task. RT data and RT switch costs, however, revealed patterns which were consistent with the hypotheses of this study and with the findings of previous studies such as those of Linck et al. (2012) and Meuter and Allport (1999). Analyses of RTs revealed symmetrical switch costs for FMI participants across the three languages (and language pairs), which was consistent with this group's classification as balanced trilinguals. EMI participants, however, sustained smaller switch costs for the weaker French whereas they exhibited larger—and symmetrical—switch costs in Arabic and English. These results also parallel the OPT scores which identified EMI participants as highly proficient English-Arabic speakers with a lower proficiency in French. The findings from the TPNT, therefore, support the predictions of the ICM for both groups of participants in terms of RT switch costs.

5.2.2 The Narration Task (NT)

RQ3B. Do EMI and FMI participants exhibit the same frequency and length of pauses in the single-language and trilingual conditions of the narration task?

The second part of RQ3 explored whether EMI and FMI participants differed in their fluency on the narration task and whether this fluency paralleled RTs in the TPNT. In the single-language condition, participants were instructed to produce narrations in the cued language only, i.e., without switching to the non-target languages. In the trilingual-condition, participants were instructed to use the three languages freely and to switch between the three languages as (in)frequently as they wished, as long as they used each language at least once. The following sections present the general analytical strategy of the task, followed by a detailed description of the results from each condition.

5.2.2.1 General Analytical Strategy

Data Preparation

For each participant, output from the NT was obtained from four linguistic conditions: Arabic-only, English-only, French-only, and Arabic-English-French. Data from each condition were transcribed verbatim, including all hesitation phenomena. The audio-recordings and transcriptions were then analysed using the PRAAT software (version 6.0.40, Boersma & Weenink, 2018) using the TextGrid silences programme extension. Using this software permitted the segmentation of the narration output into *speech* (or 'sounding' segments) and *pauses* (or 'silent'[15] segments). The PRAAT software produced the following output (section A of Table 5.9), which permitted performing further calculations (section B of Table 5.9).

[15] As indicated earlier, 'pauses' in this study comprised silent pauses as well as filled pauses which contained 'meaningless' lexicalisations such as 'uh' or 'um'. Accordingly, some of the 'sounding' segments contained 'pauses'. This is further discussed in this section.

Table 5.9 PRAAT Output and Computations

	Description
Section A	
Pause frequency	Total number of pauses (at least 250ms each)
Pause duration	Length of each pause in milliseconds
Total pause duration	Total length of all pauses in a speech sample in milliseconds
Total phonation duration	Total speaking time calculated as total speech sample duration – total pause duration
Total speech sample duration	Overall duration of the entire production which included speech and silences
Section B	
Pause Frequency Ratio	Total number of (silent and filled) pauses divided by the total phonation duration of the speech sample
Mean pause duration	Total duration of all pauses (of 250 milliseconds and above) divided by the number of pauses in a given speech sample

Data Cleaning

Although PRAAT was expected to facilitate the analysis of NT data, a manual check of the output revealed that the programme over-reported the number of pauses. For each participant, *pause frequency* was calculated by measuring the number of non-lexicalised pauses in production in single-language conditions (for Arabic, English, and French) and in the trilingual condition. As mentioned earlier, non-lexicalised pauses include silent pauses, where no speech is uttered, and filled pauses which include hesitations such as 'um' and 'er'. However, the PRAAT software identified filled pauses as 'sounding' segments. Accordingly, the programme over-reported the number of speech segments as well as pause segments, as filled pauses were typically preceded and followed by silences. As such, a filled pause was segmented into three parts: (1) a silent pause, (2) a sounding segment, and (3) a silent pause. A manual adjustment of these pitfalls in the software was then conducted to ensure the correct number (and length) of pauses was reported.

Another adjustment which was made, but which was not software-driven, was the removal of inter-slide pauses. Pauses which were incurred when participants transferred

their gaze from one picture slide of the story board to the next were excluded from analyses. These pauses varied in duration, and they stemmed from participants' processing of the visual information from the pictures before they proceeded with narrating the stories. Accordingly, participants' pauses indexed looking at specific details in the slides, as opposed to being disfluent due to a lexical gap or planning their speech, for instance. Although using eye-tracking technology would have assisted with determining the gaze duration and location more accurately, it was clear to the researcher who was both present at the audio-recording of the narrations and was highly familiar with the content of each slide when a pause occurred due to an inter-slide transition.

Pause Frequency

As the total number of pauses varied as a function of the total length of the speech for each participant, it was expected that participants producing longer narrations would have a higher number of pauses. In order to analyse pause frequency, therefore, the total number of pauses over 250 seconds for each participant was divided by the total duration of his/her speech sample, measured in seconds. For instance, a participant whose total number of pauses in Arabic was 24 in a speech sample of 480 seconds (8 minutes) received a ratio of 0.05 (24 pauses/total speech sample duration) as 'pause frequency in Arabic'. A lower ratio indicated that there were few pauses throughout a participant's speech, whereas a higher ratio indicated that the number of pauses was high given the overall duration of the speech sample. Pause frequency ratios were calculated for each participant for each language in the single-language condition, and in the trilingual condition, and these ratios were then used in all subsequent analyses.

Pause Duration

Upon obtaining the duration of each pause (of 250ms and above), the total duration of pauses in every language condition was calculated for every participant. For each

condition, *mean pause duration* was then calculated by dividing the total length of pauses in the respective language by the total number of pauses in that language condition. For example, if a participant paused four times for 250ms, 300ms, 350ms, and 400ms, the mean pause duration was computed as (250+300+350+400)/4 = 325ms.

Model Structuring

In order to investigate pause frequency ratios and mean pause duration as the outcome variables, a series of linear mixed effects models were fitted with Group (2 levels: EMI and FMI) and Language Condition (4 levels: Arabic-only, English-only, French-only, and Trilingual), as well as interactions between the two variables as predictors. Group was coded as -0.5 and 0.5, whereas Language Condition was dummy coded. All models had fully random slopes and intercepts by participants and items, i.e., they had maximal random-effects structures (Barr et al., 2013; Bates et al., 2015). The same analytical procedure of TPNT data was adopted, where model comparisons were performed and the best-fit model was determined by Akaike's information criterion (AIC; Akaike, 1998) and its significance was determined by the likelihood ratio test (Matuschek et al., 2015). The model was then checked for quality-of-fit through visual inspections of the distribution of the residuals and transformations were made to approximate residuals' normality as necessary to satisfy the homoscedasticity condition (Baayen, 2008). Details of data transformations and results of the analyses are offered next.

5.2.2.2 Pause Frequency Analysis

In order to investigate pause frequency ratios as the outcome variable, a linear mixed effects model was fitted with Group (2 levels: EMI and FMI) and Condition (4 levels: Arabic-only, English-only, French-only, and Trilingual), along with interactions between them, as predictors. Tests for normality (Shapiro-Wilk $p<0.00$) and inspections

of the residual plots indicated that the pause frequency ratio data were non-normally distributed for either group. Therefore, a transformation to approximate normality was conducted through the Box-Cox function of the MASS package in R. The best transformation of raising the ratios to the power of 0.25 was adopted, and visual inspections of the residuals' QQ-plots indicated that normality was achieved. While Table 5.10 summarises the raw mean scores and standard deviations[16] of the frequency ratios of the two groups across the languages, the analyses report on the transformed data.

Table 5.10 Mean and SD of Pause Frequency Ratios

	EMI		**FMI**	
	M	**SD**	**M**	**SD**
Arabic	0.0404	(0.0071)	0.0452	(0.0096)
English	0.0232	(0.0024)	0.0329	(0.0060)
French	0.0747	(0.0136)	0.0289	(0.0035)
Trilingual	0.0230	(0.0024)	0.0287	(0.0035)

The results of the MEM indicated significant main effects for Group and Language Condition (Table 5.11), as well as a significant interaction between Group x Language Condition ($F(3,174)=333.71$, $p<0.0001$). A visual illustration of this interaction is offered in Figure 5.6.

Table 5.11 MEM ANOVA (Type III) Summary of Narration Task -- Pause Frequency

	F-Statistic	DF1	DF2	p-value
Intercept	76853.84	1	174	<0.0001
Group	10.68	1	58	<0.01
Language Condition	324.70	3	174	<0.0001
Group*Language Condition	333.71	3	174	<0.0001

[16] Although a standard two-decimal format is adopted throughout this thesis, it was important to report four-decimal numbers in this table due to the low ratio values.

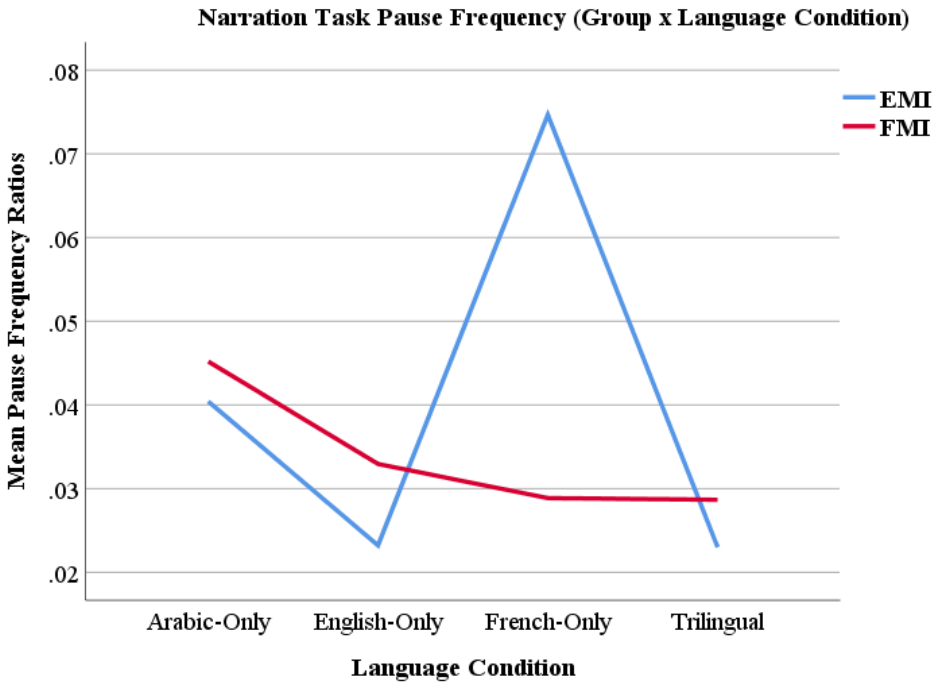


Figure 5.6 Pause Frequency x Group x Language Condition Interaction Plot

In order to explore the significant interaction, simple main effects were analysed by changing the factor contrast coding for the mixed model, and *post hoc* contrasts were analysed by changing the reference (intercept) for each level of the Group and Language Condition variables. A sample of the MEM output (baselines: EMI and Arabic-only condition) is presented in Table 5.12.

Table 5.12 MEM Output for Pause Frequency Ratio (Estimated coefficients, standard errors (SE), t-values, and p-values) Asterisks (*) indicate: p< .05 (*), p< .01 (**), p<. 001(***) and p< .0001 (****).

Predictor	Coefficient	SE	t-value	p-value
Fixed Effects				
Intercept	4.48	0.03	142.26	0.00****
FMI	0.12	0.04	2.73	0.00**
English	0.57	0.04	15.62	0.00****
French	0.74	0.04	20.21	0.00****
Trilingual	0.58	0.04	15.87	0.00****
FMI: English	0.23	0.05	4.37	0.00****
FMI: French	0.12	0.05	23.51	000****
FMI: Trilingual	0.09	0.05	1.88	0.06
Random Effects				
Participants	0.10			
Residuals	0.14			

*Formula: Pause Frequency Ratio ~ Group*Language + (1+Language|subject)*

The results of the analyses indicated that for the two groups differed in every language condition (see Table 5.13 for group contrasts). Specifically, EMI participants paused significantly fewer times than FMI participants in the Arabic-only, English-only, and the trilingual-only conditions, whereas FMI participants paused significantly less frequently than EMI participants in the French-only condition. The effect of language conditions was largest in this condition. Although the latter finding was expected, the two groups were not expected to differ in the Arabic-only and trilingual conditions, especially as their proficiency in Arabic was comparable. Thus, the finding that EMI having significantly fewer pauses in Arabic is unexpected. This difference may offer yet further evidence to the notion that Arabic may be the weakest language of the FMI group, despite the length of exposure to this language.

Table 5.13 Narration Task Pause Frequency Group Contrasts

EMI vs. FMI	β	SE	t-value	p-value
Arabic	0.12	0.04	2.73	<0.01
English	0.34	0.04	7.83	<0.0001
French	1.10	0.04	24.65	<0.0001
Trilingual	0.22	0.04	4.92	<0.0001

Additionally, each group differed in their pause frequency across the language conditions, as summarised in Table 5.14 and illustrated in Figure 5.7. Overall, both groups had significantly fewer pauses in the trilingual condition and their respective immersion languages compared to the other two languages. That is, EMI participants had significantly fewer pauses in the English-only and the trilingual conditions compared to Arabic-only ($t=15.62$, $p<0.0001$ for English; $t=15.87$, $p<0.0001$ for trilingual) and French-only ($t=35.83$, $p<0.0001$ for English; $t=36.08$, $p<0.0001$ for trilingual) conditions. There were no significant differences in pause frequency between the trilingual and English-only conditions ($p>0.05$). Furthermore, the Arabic-only condition had significantly fewer pauses than French ($t=20.21$, $p<0.0001$), which is expected given

participants' lower proficiency in French.

On the other hand, FMI participants had significantly fewer pauses in the French-only and trilingual conditions compared to the Arabic-only ($t=13.02$, $p<0.0001$ for French; $t=13.21$, $p<0.0001$ for trilingual) and English-only ($t=3.59$, $p<0.01$ for French; $t=3.78$, $p<0.001$ for trilingual) conditions. There were no significant differences between French and the trilingual conditions ($p>0.05$). Furthermore, the English-only condition had significantly fewer pauses than the Arabic-only condition ($t=9.43$, $p<0.0001$). An unexpected finding is that FMI participants reported using Arabic more frequently in their social interactions compared to the other languages; however, it seems that language use—at least in social interactions—does not predict pausing frequency. This is further corroborated by findings of RQ4.

Table 5.14 NT Pause Frequency Ratios Post-hoc Analyses by Group

	B	SE	t-value	p-value
EMI				
Arabic vs. English	0.57	0.04	15.62	<0.0001
Arabic vs. French	0.74	0.04	20.21	<0.0001
English vs. French	1.31	0.04	35.83	<0.0001
Arabic vs. Trilingual	0.58	0.04	15.87	<0.0001
English vs. Trilingual	0.00	0.04	0.25	0.80
French vs. Trilingual	1.32	0.04	36.08	<0.0001
FMI				
Arabic vs. English	0.35	0.04	9.43	<0.0001
Arabic vs. French	0.48	0.04	13.02	<0.0001
English vs. French	0.13	0.04	3.59	<0.001
Arabic vs. Trilingual	0.48	0.04	13.21	<0.0001
English vs. Trilingual	0.14	0.04	3.78	<0.001
French vs. Trilingual	0.00	0.04	0.19	0.85

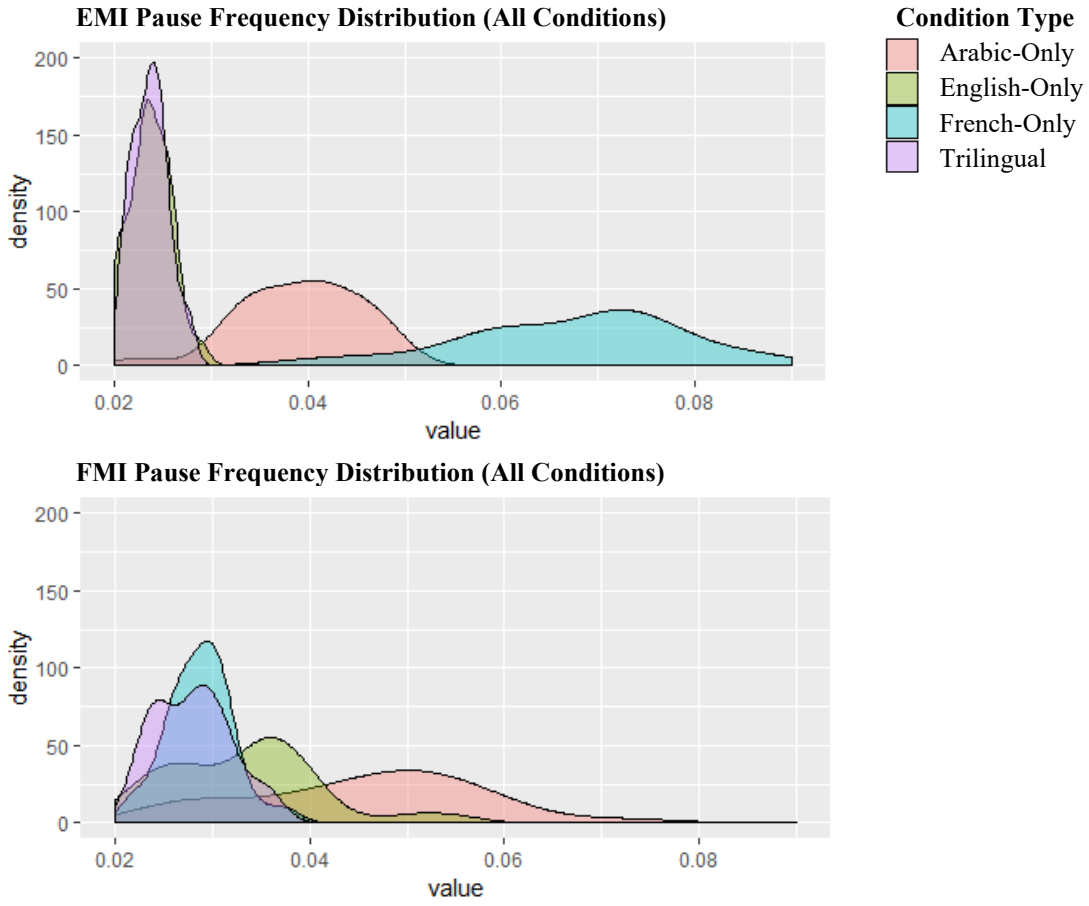


Figure 5.7 Density distribution of Pause Frequency by group

To summarise, these results suggest that speech in the trilingual condition and immersion-language contexts was the most fluent. While EMI participants were expected to pause most frequently in the French-only condition given their low proficiency in French, they were expected to have similar pause frequency in the Arabic-only and English-only conditions as they were equally proficient in the two languages. One possible explanation for their significantly fewer pauses in English is that participants are accustomed to switching to English and/or French when conversing in Lebanese Arabic, as this frequent codeswitching is a staple of Lebanese speech. Accordingly, the participants in this group are not used to speaking in Arabic only. In the English-only condition, however, EMI participants are used to speaking in English-only especially given their immersion experience. As such, and based on these results taken alone, it

could be argued that participants are most fluent in English. This may also explain why FMI participants paused most frequently in Arabic. It may therefore be that the participants were not accustomed to producing narrations in Arabic-only conditions.

5.2.2.3 Pause Duration

In order to analyse whether participants differed in pause durations, mean pause duration' was the outcome variable in a mixed-effects model, whereas Language (Arabic, English, and French) and Group (EMI and FMI), along with interactions between them, were fitted as predictors. The data were not normally distributed for either group; therefore, a transformation to approximate normality was conducted through the Box-Cox function of the MASS package in R. The best transformation of dividing the durations by a half was adopted, and visual inspections of the QQ-plots indicated that normality was achieved. While Table 5.15 summarises the raw mean scores and standard deviations of the two groups across the languages, the analyses report on the transformed data.

Table 5.15 Mean and SD of NT Pause Durations (in milliseconds)

	EMI		**FMI**	
	M	**SD**	**M**	**SD**
Arabic	849.60	(195.06)	950.20	(236.98)
English	579.43	(76.90)	760.93	(153.80)
French	3038.93	(621.51)	668.83	(95.49)
Trilingual	554.81	(81.49)	637.20	(93.25)

The model revealed significant main effects for Group and Language (Table 5.16), as well as a significant interaction between the two variables (β=0.84, SE=0.03, t=28.21, p<0.0001), illustrated in Figure 5.8, and summarised in Table 5.16.

Table 5.16 MEM ANOVA (Type III) Summary of Narration Task Pause Duration

	F-Statistic	DF1	DF2	p-value
Intercept	61227.33	1	171.58	<0.0001
Group (EMI vs. FMI)	88.85	1	56.44	
Language Condition (Arabic, English, French, and Trilingual)	339.44	3	1	<0.0001
Group*Language Condition	400.22	3	174	<0.0001

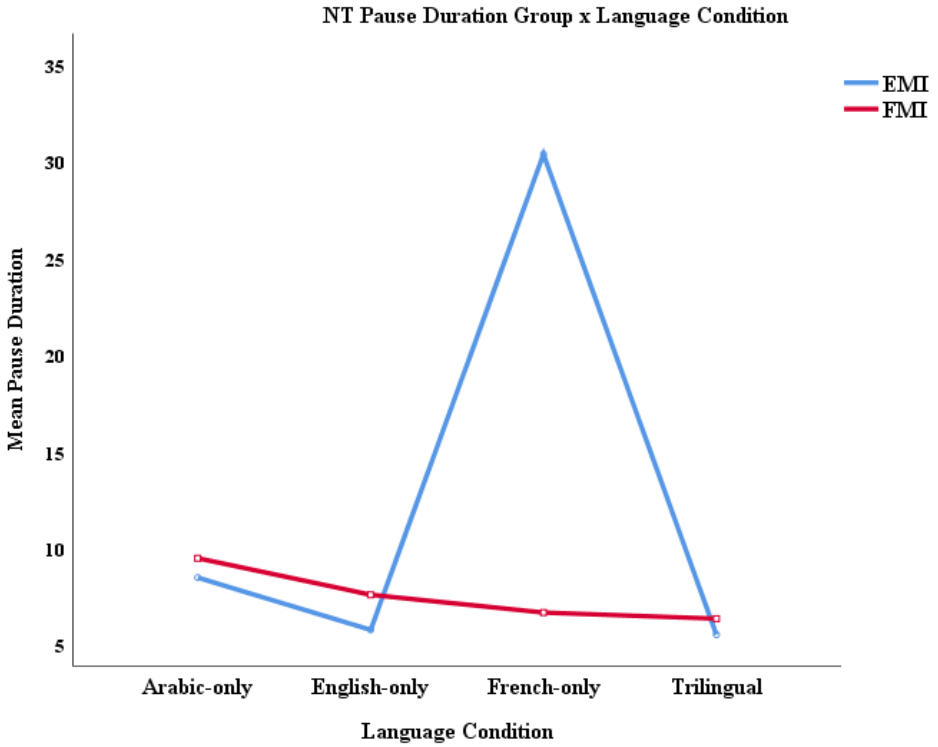


Figure 5.8 Pause Duration x Group x Language Condition Interaction plot

In order to further examine the interaction, the model was releveled to examine simple main effects, as demonstrated in Table 5.17 where EMI and the Arabic-only condition are the baselines.

Table 5.17 MEM Output for Pause Duration (Estimated coefficients, standard errors (SE), t-values, and p-values) Asterisks (*) indicate: p< .05 (*), p< .01 (**), p<. 001(***) and p< .0001 (****)

Predictor	Coefficient	SE	t-value	p-value
Fixed Effects				
Intercept	2.09	0.03	75.28	0.00****
FMI	0.09	0.04	2.24	0.03*
English	0.25	0.04	7.00	0.00****
French	1.19	0.04	33.11	0.00****
Trilingual	0.28	0.04	7.79	0.00****
FMI: English	0.09	0.05	1.81	0.07
FMI: French	1.44	0.05	28.24	0.00****
FMI: Trilingual	0.00	0.05	0.04	0.97
Random Effects				
Participants	0.06			
Residuals	0.14			

*Formula: Mean Pause Duration ~ Group*Language + (Group*Language|Participant)*

The results indicated that for each language condition, the two groups differed in pause duration across the four language conditions. However, the two groups differed most significantly in pause duration in the French-only condition ($t=32.94$, $p<0.0001$). Specifically, EMI participants paused for significantly shorter durations than their FMI peers in Arabic-only, English-only, and Trilingual conditions, whereas FMI participants paused for significantly shorter durations in the French-only condition. Group contrasts are summarised in Table 5.18.

Table 5.18 NT Mean Pause Duration -- Group Contrasts

EMI vs. FMI	B	SE	t-value	p-value
Arabic	0.05	0.02	2.31	<0.05
English	0.11	0.02	4.76	<0.0001
French	0.73	0.02	32.94	<0.0001
Trilingual	0.05	0.02	2.42	<0.05

While the two groups were predicted to differ in the French-only condition given EMI's lower French proficiency, differences in the other conditions, especially Arabic-only, were unexpected as both groups had a high proficiency in Arabic. As FMI participants reported using Arabic significantly more frequently than the other two languages, however, it was expected that any difference between the two groups would stem from FMI participants being more fluent in the Arabic-only condition. In terms of English, it was uncertain how the FMI group would perform in the English-monolingual condition. On the one hand, FMI participants achieved high scores in English proficiency and RTs on the TPNT which were not significantly different from those attained by the EMI group. This would have led to the prediction that FMI participants would pause similarly in duration (and frequency) to the EMI group. On the other hand, EMI participants are expected to outperform their FMI counterparts in the English-only condition. The two groups' difference in performance on the trilingual condition was also an unexpected finding, as both groups were expected to be equally fluent in a

condition which they typically engage and which they can linguistically exploit. It remains unclear why EMI participants were more fluent than their FMI peers in this condition.

Planned contrasts were performed to examine each group's fluency across the different conditions. The results, summarised in Table 5.19 and illustrated in Figure 5.9, indicated that EMI participants exhibited significantly shorter pauses in the English-only and trilingual conditions, with no significant difference between these two conditions ($p>0.05$). Furthermore, the French-only condition revealed significantly longer pauses than any other condition, indicating that the EMI group incurred disfluencies when asked to speak in French only. The FMI group, on the other hand, exhibited significantly shorter pauses in the French-only and trilingual conditions, with no significant difference between these conditions ($p>0.05$). Furthermore, the Arabic-only condition was the least fluent for FMI participants, as speakers incurred the longest pauses in this condition compared to English-only, French-only, and trilingual conditions ($p<0.0001$ for all contrasts, as indicated in Table 5.19).

Table 5.19 NT Pause Duration Planned-Contrast Analyses by Group

	β	**SE**	**t-value**	**p-value**
EMI				
Arabic vs. English	0.15	0.02	7.28	<0.0001
Arabic vs. French	0.64	0.02	31.94	<0.0001
English vs. French	0.79	0.02	39.22	<0.0001
Arabic vs. Trilingual	0.16	0.02	8.14	<0.0001
English vs. Trilingual	0.02	0.02	0.86	0.39
French vs. Trilingual	0.81	0.02	40.08	<0.0001
FMI				
Arabic vs. English	0.09	0.02	4.57	<0.0001
Arabic vs. French	0.14	0.02	7.05	<0.0001
English vs. French	0.05	0.02	2.48	<0.05
Arabic vs. Trilingual	0.16	0.02	8.02	<0.0001
English vs. Trilingual	0.07	0.02	3.45	<0.001
French vs. Trilingual	0.02	0.02	0.97	0.33

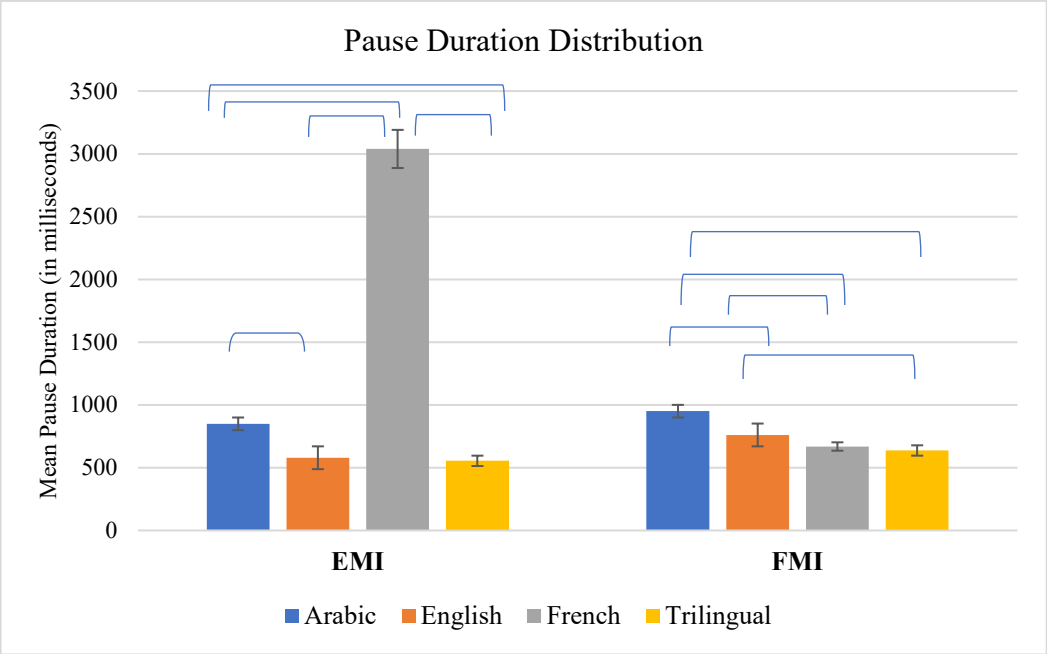


Figure 5.9 Pause Duration by Group and Language Condition, bars= standard error of the mean, brackets=significant difference

5.2.2.4 Summary of NT Analyses

The pattern of results observed in pause frequency was replicated in pause duration analyses[17]. Between-group analyses revealed that EMI participants paused less frequently and for shorter durations than their FMI counterparts in all conditions except in French-only contexts. Furthermore, post-hoc analyses revealed that both EMI and FMI participants were most fluent in their immersion languages, English and French, respectively, as well as in the trilingual context. Despite the two groups' having comparable proficiency scores in Arabic, and despite FMI participants' self-report of frequent Arabic use, not only were EMI participants more fluent than FMI in Arabic, but Arabic-only was the least fluent condition for FMI participants among the other language conditions.

This finding may be attributed to the methodological limitation of using self-

[17] As the response patterns were highly similar, it was of interest to investigate whether participants who paused more frequently also exhibited longer responses. Correlation analyses using Pearson's R revealed that this was indeed the case ($r(1,58)=0.928$, $p<0.0001$) (see Appendix I).

reports, where participants in the FMI group over-reported the frequency of their Arabic use. An alternative explanation is that language use in social settings is not as strong a predictor than initially believed. That is, assuming that FMI participants had accurately reported their language use, it is perhaps their language use in academic settings which highly influenced their performance on the narration task. This is commensurate with the findings where the immersion language was the least taxing for the participants—even more so than the trilingual condition which speakers could opportunistically exploit. This explanation also fits in neatly with the EMI group's performance and self-reports. For the EMI group, language use in social interactions mimics language use in academic settings, due to the ubiquity of English in social contexts. Accordingly, EMI participants' high fluency in their immersion language may stem from their language use in academic settings which overlapped with their language use in social settings. It is this latter explanation which seems to be a more reasoned account of the unexpected findings.

Another significant point to note was that the Arabic-only condition was significantly taxing for both groups. This may be attributed to the fact that codeswitching is highly frequent in social settings in which Lebanese speakers converse. As such, speaking in an Arabic-only condition was 'atypical' for both groups, as evidenced by their high frequency and length of pauses. As such, speaking in a monolingual mode in this language was arduous.

5.2.3 Correlation between NT and TPNT

RQ3C. Do EMI and FMI participants' pause durations in the narration task correlate with their RTs in the picture-naming task?

The last subsection of RQ3 investigated whether producing one-word utterances in the TPNT overlapped with producing prolonged utterances in the NT. To correlate performance on both tasks, response latencies in the picture-naming task were mapped

onto pause durations in the narration task. Specifically, RTs on the repetition trials for each language of the TPNT were correlated with pause durations in the single-language conditions of the narration task. For example, RTs on English repetition trials were correlated with mean pause durations in the English-only condition. For the trilingual condition, mean RTs on switch trials from the three languages were correlated with mean pause durations in the trilingual-condition of the narration task. As pause durations in the trilingual condition were relatively short and similar in length, and given the variable number of switches between languages for each participant, RTs from switch trials and trilingual pauses were correlated without distinguishing language pairs.

The distributions of RTs from both tasks were not normal; therefore, transformed RTs were used in the data analysis. However, as the data were continuous and not ordinal, non-parametric tests such as Spearman's Rho and Kendall's Tau could not be performed (Fields et al., 2012). Accordingly, Pearson's product-moment correlation coefficients were computed to assess the relationship between response latencies for each language. The R statistical 'stats' package was then used to perform correlations. As the two groups differed in their performance on both the TPNT and NT, the correlations were performed for each group separately.

For EMI participants, there was a positive correlation between response latencies on English-repetition trials and English-only narrations, $r(1,29)=0.52$, $p<0.01$. Furthermore, a positive correlation was found between RTs on switch trials and the trilingual-condition of the narration task, $r(1,29)=0.43$, $p<0.05$. No other significant correlations were observed for the EMI group. A summary of all the correlation analyses for the EMI group is offered in Table 5.20, whereas Figure 5.10 offers a schematic representation of the significant correlations.

Table 5.20 EMI Correlation Analyses – TPNT RTs and NT Pause Durations

	Pearson Correlation Coefficient	P-value
Arabic	0.24	0.21
English	0.52	<0.01
French	0.15	0.44
Trilingual	0.43	<0.05

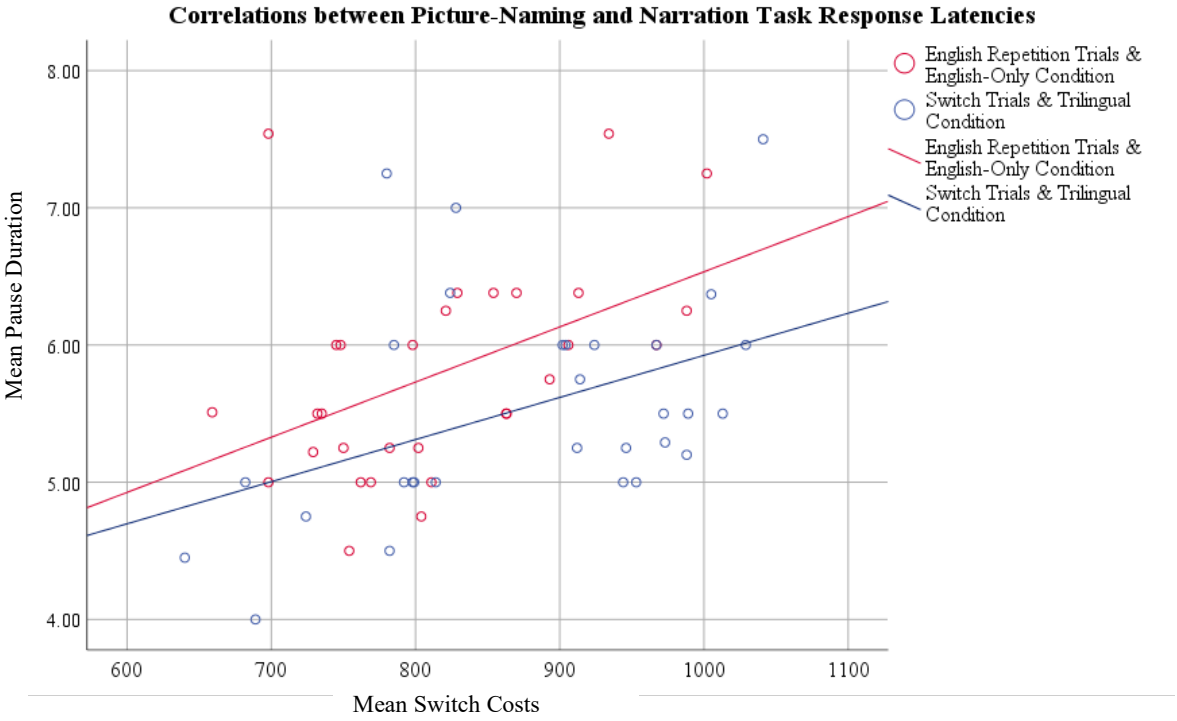


Figure 5.10 Correlations between TPNT and NT (EMI)

For FMI participants, response latencies on the French repetition trials positively correlated with response latencies (pauses) in the French-only condition of the narration task, $r(1,29)=0.51$, $p<0.01$. Furthermore, RTs on switch trials positively correlated with pauses in the trilingual condition of the narration task $r(1,29)=0.42$, $p<0.01$. A summary of all the correlation analyses for the FMI group is offered in Table 5.21, whereas Figure 5.11 offers a schematic representation of the significant correlations.

Table 5.21 FMI Correlation Analyses – TPNT RTs and NT Pause Durations

	Pearson Correlation Coefficient	p-value
Arabic	0.06	0.75
English	0.31	0.09
French	0.51	<0.01
Trilingual	0.42	<0.01

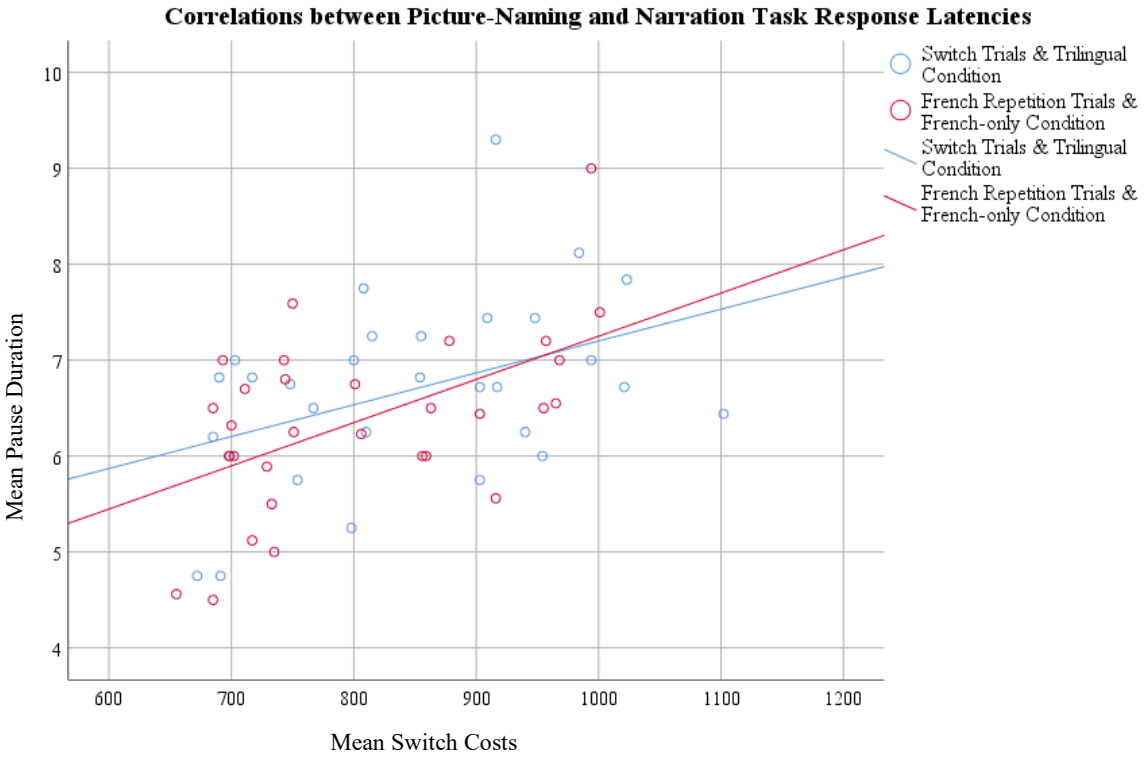


Figure 5.11 Correlations between TPNT and NT (FMI)

These findings therefore indicate that participants in both groups who had the shortest pauses in the trilingual condition of the narration task also had the fastest RTs on switch trials in the picture-naming task. Furthermore, RTs on repetition trials for each group's immersion language correlated with the participants' fluency on the narration task in that language. That is, for the EMI group, participants who exhibited smaller RTs on English repetition trials also had shorter pauses when narrating in English. FMI participants who were more fluent when speaking in French in the French-only condition were faster at naming French repetition trials.

The assumption that performance across both tasks is correlated hinges on an exploratory attempt to extend participants' one-word utterances to long narrations. While

significant correlations are not observed for every language or language condition, the results offer a positive outlook into investigating language production outside the laboratory-based designs and beyond one-word utterances. As such, the results of this research question—and the narration task generally—offer support to the predictions of the ACH. Furthermore, these results offer a more nuanced picture of individual differences, where highly fluent speakers respond fastest on trials (both repetition and switch) on the picture-naming task.

5.3 RQ4- PREDICTORS OF LINGUISTIC TASK PERFORMANCE

> *RQ4. Do language proficiency and language experience predict performance within and across EMI and FMI participants in the picture-naming task and narration task?*

The fourth research question explored whether language proficiency and language experience were significant predictors of linguistic task performance. A series of linear mixed-effect models was used to analyse the data in R using the lme4 package (Bates et al., 2014). RT switch costs[18] from the TPNT and pause frequencies and pause durations from the NT were entered as outcome variables in separate models. The log-transformed data from these tasks were used to satisfy the normality assumption. To reduce collinearity, all continuous fixed effects were z-scored. The predictors included different groupings of explanatory variables, as well as interactions between these variables, described in detail in the relevant sections. The full list of predictors included Group (2 levels: EMI and FMI), Language (3 levels: Arabic, English, and French), Language

[18] As accuracy was close to ceiling, with some participants having 100% accuracy rates, a generalised linear model was fitted with the binary data of responses as outcome variables and with the language proficiency and language use as predictors. The residuals did not meet the prerequisite assumption of homoscedasticity and revealed a negative dispersion with fitted values approaching 1. This indicated that the ceiling effect created a bias in the model, so no further analyses were performed on accuracy switch costs.

Condition (4 levels: Arabic-only, English-only, French-only, and trilingual), OPT scores (3 levels: Arabic, English, and French), VFT scores (3 levels: Arabic, English, and French), HPBT (2 levels: Bilingual vs. Trilingual), Codeswitching Frequency (CSF), Salient-Cue detection (SALC), Language Use (3 levels: Arabic, English, and French), and Language Exposure (3 levels: Arabic, English, and French). Participants and items (when applicable) were included in the models as random effects.

For each research question, models were developed by adopting a full structure which included all fixed effects of interest, random intercepts for participants, and random slopes for all predictors (i.e., a maximal structure, Barr et al., 2013). Keeping the random effects structure constant, as analyses at the participant-level were of interest to this project, the best-fit model was then selected through a stepwise procedure. First, non-significant predictors and interactions were removed, and a comparison between the simpler model and the full model was determined by means of the Akaike's Information Criterion (AIC) (Baayen, 2008) and likelihood ratio tests for significance. If the simpler model did not significantly differ from the full model, then the simpler model which had fewer predictors was selected. After that, each significant predictor was removed from the selected simple model to examine whether the model improved. Comparisons of the models with different fixed effects being removed were conducted with parameters estimated with Maximum Likelihood for fixed effects and Restricted Maximum Likelihood for random effects comparisons (Pinheiro & Bates, 2006). For each linguistic task, a discussion of the overall best-fit model is presented, followed by further (post-hoc) analyses conducted when necessary.

5.3.1 Language Proficiency and Linguistic Performance

RQ4A: Does language proficiency influence switch costs in the TPNT and pause frequency and duration in the NT for the two groups?

5.3.1.1 Language Proficiency and Switch Costs

In order to explore whether having higher language proficiency predicts smaller switch costs on the TPNT, a linear mixed-effects model was fitted with a maximal random structure. The predictors included Group, Language, OPT Scores, VFT scores, and HPBT. All interactions between these variables were also included as predictors. The dependent variable was the log-transformed switch costs. The full model did not converge; accordingly, simpler models were selected, as outlined in the previous section. The results of the models are presented next.

Model 1: Group x Language x OPT Scores

The first model investigated whether having a higher proficiency in any specific language or set of languages would influence switch costs. The best-fit model included Group, (switch-cost) Language, and OPT scores, with interactions between Language and OPT scores. The results of the model's ANOVA, summarised in Table 5.22, revealed significant main effects for Language and OPT scores, but not for Group ($p>0.05$). The model also revealed a significant interaction between OPT Scores x Language ($F(4,174)=11.39$, $p<0.0001$).

Table 5.22 MEM ANOVA (Type III) Summary for Switch Costs (Model 1)

	F-Statistic	DF1	DF2	p-value
Intercept	563.19	1	174	<0.0001
Group (EMI vs. FMI)	0.58	1	58	0.44
Language (Arabic vs. English vs. French)	13.72	2	174	<0.0001
OPT Scores (Arabic vs. English vs. French)	6.59	2	174	0.01
Language*OPT Scores	11.39	4	174	<0.0001

In order to explore the interaction, simple main effects were analysed by changing the factor contrast coding for the mixed model, and *post hoc* contrasts were analysed by changing the reference (intercept) for each level of Language and OPT Scores. As there was no main effect for Group or significant interaction with this

variable, data from all participants were aggregated. The post-hoc results, summarised in Table 5.23, indicated that OPT scores in each language significantly predicted participants' switch costs in that respective language, however this effect was significantly larger for French than English ($t=4.18$, $p<0.001$) and Arabic ($t=3.62$, $p<0.001$). This was due to the larger range of proficiency scores in French as data from both EMI and FMI groups were analysed jointly, whereas the significance was smaller in English and Arabic as proficiency scores were generally high for all participants. As evidenced by the positive slope for each language (Figure 5.12), participants who were highly proficient in a language incurred the largest switch costs in that language. This is in line with previous findings (e.g., Schweiter & Sunderman, 2008, 2009, 2016), as well as with the predictions of the ICM.

Table 5.23 MEM Output Summary for Switch Costs and OPT (Estimated coefficients, standard errors (SE), t-values, and p-values) Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	Switch Costs			
Predictors	**Estimate**	**SE**	**t-value**	**p-value**
Fixed Effects				
Intercept	9.09	18.47	0.49	0.62
Arabic Language	4.23	0.78	5.42	0.00****
English Language	3.57	0.78	4.58	0.00****
French Language	4.31	0.78	5.52	0.00****
OPT Arabic	0.26	0.12	2.14	0.04*
OPT English	0.32	0.12	2.62	0.02*
OPT French	0.58	0.12	4.76	0.00***
Arabic Language*OPT Arabic	0.28	0.09	2.96	0.00**
English Language*OPT English	0.31	0.09	3.15	0.00**
French Language*OPT French	0.43	0.09	4.41	0.00***

*Formula: Switch Costs ~ Switch Language*OPT Scores (1+Switch Language|Participant) + (1+OPT Scores|Participant)*

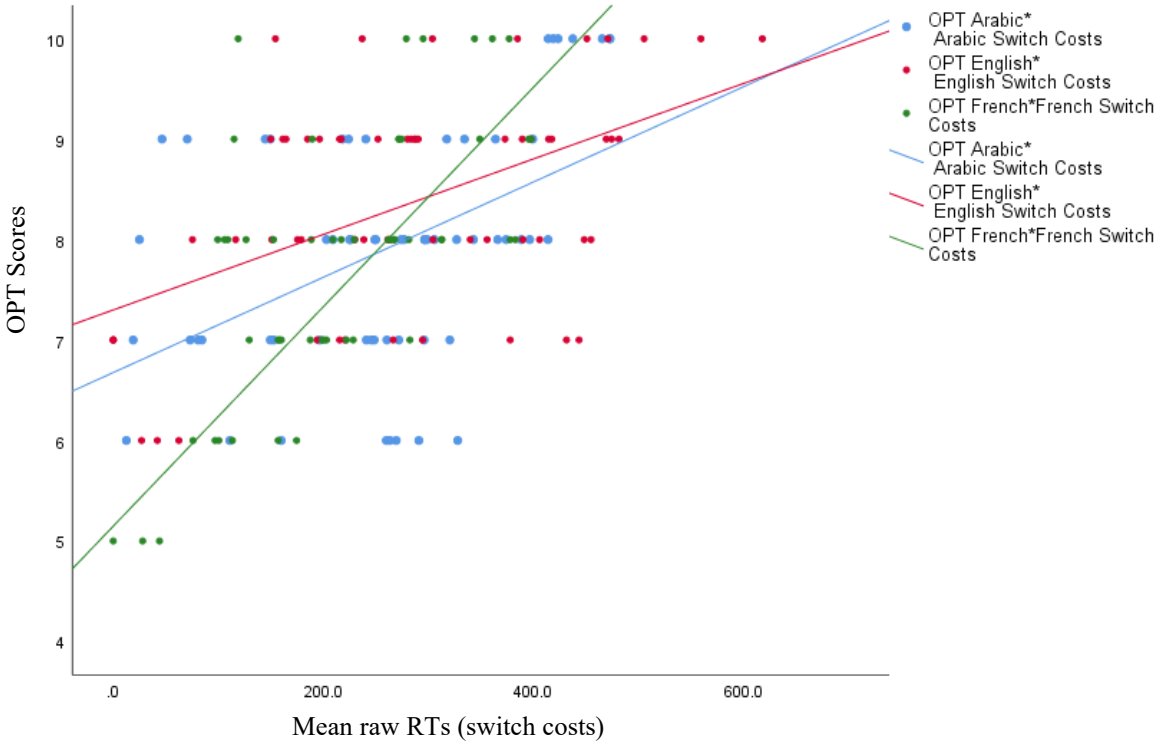


Figure 5.12 OPT scores x Switch Costs (per language)

Model 2: Group x Language x VFT Scores

The second model was fitted with the log-transformed RT switch costs as the dependent variable and with Group, Language, and VFT scores as the predictors, along with all interactions between the predictors. The model indicated significant main effects for Group and Language Condition, as well as significant interactions between Group and VFT scores, Group and Language Condition, and a significant three-way interaction between Group x VFT scores x Language ($F(6,174)=4.96$, $p<0.01$). A summary of the MEM's ANOVA results is presented in Table 5.24.

Table 5.24 MEM ANOVA (Type III) Summary for Switch Costs and VFT (Model 2)

	Switch Costs			
	F-statistic	DF1	DF2	p-value
Intercept	5878.66	1	174	<0.0001
Group	6.86	1	58	0.01
Language Condition	14.70	3	174	<0.001
VFT Scores	1.03	2	174	0.31
Group*Language Condition	12.20	3	174	<0.001
Group*VFT Scores	4.26	2	174	<0.01
Language Condition*VFT Scores	2.47	6	174	0.06
Group*Language Condition*VFT Scores	4.96	6	174	<0.01

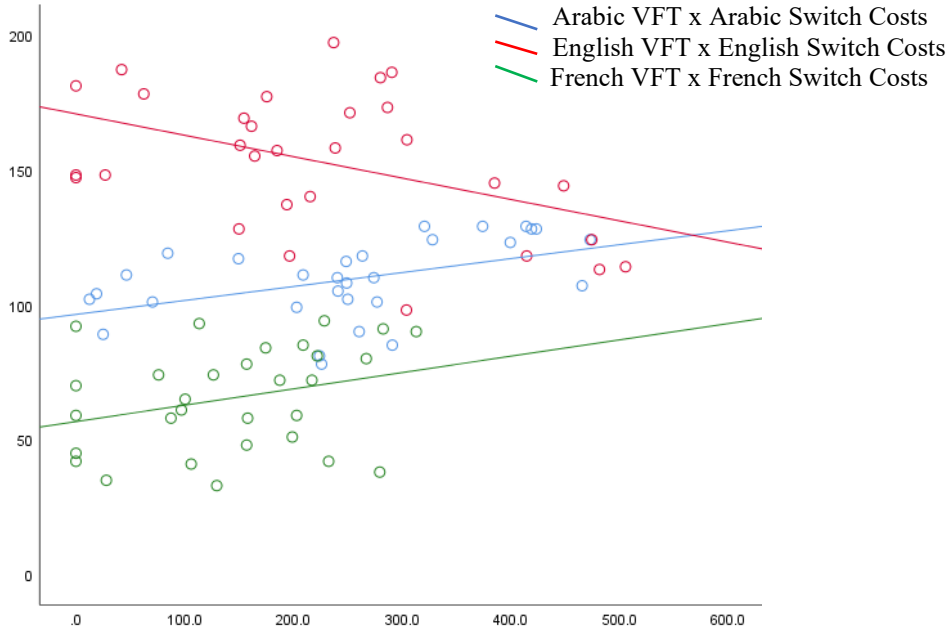
In order to investigate the significant Group x Language interaction, simple main effects were analysed by changing the factor contrast coding for the mixed model, and *post hoc* contrasts were analysed by changing the reference (intercept) for each explanatory factor. Results of the analyses, illustrated in Figure 5.13., revealed that this interaction was driven by English VFT scores significantly predicting *smaller* switch costs for EMI participants ($t=3.62$, $p<0.001$) but not for the FMI participants. Furthermore, for the EMI group, there was no significant relationship between lexical robustness scores and Arabic and French switch costs ($p>0.05$ for both contrasts). For FMI participants, VFT scores significantly predicted *larger* switch costs in French ($t=2.98$, $p<0.01$), but not for English or Arabic ($p>0.05$ for both contrasts).

The results of the analysis revealed several unexpected results. First, the VFT, which is a purported measure of language proficiency, was not analogous to the OPT in terms of predicting switch costs. As discussed in Chapters 2 and 3, the VFT may not be a suitable measure of general language proficiency, as it is not a 'pure' measure of this construct. Rather, it taps into specific areas of proficiency (lexical access and retrieval), as well as into working memory. The difference between OPT scores and VFT scores in predicting switch costs, therefore, fortifies the present study's criticism of research

designs using the VFT as a sole measure of overall language proficiency.

Another unexpected result pertains to the divergent results of the VFT in predicting switch costs for both groups. Analyses of the VFT in Chapter 4 classified EMI participants as English-dominant unbalanced trilinguals whose Arabic and French differed significantly. On the other hand, FMI participants were identified by VFT scores as French-dominant (unbalanced) trilinguals whose Arabic and English proficiencies were similar. Accordingly, analyses of the VFT as a predictor of switch costs should have yielded larger switch costs for EMI participants, as they switch into their dominant language (English). Surprisingly, the negative relationship between the two variables offers counter-evidence either to the notion that EMI participants are indeed English-dominant or to the notion that the VFT is a reliable measure of language proficiency (and/or dominance). As results from all other tasks have indicated the high proficiency of EMI participants in English, it is reasonable to assume that the VFT is not a valid or reliable tool in measuring the construct it is said to measure. The notion of the VFT's 'reliability' is further called into question given the lack of replicability in findings for the EMI and FMI groups. The VFT predicted larger switch costs in French for FMI participants—an expected result for this group which the VFT identified as French-dominant. Accordingly, it remains unclear why the VFT revealed a contradictory pattern for the two groups.

EMI VFT Scores x Switch Costs



FMI VFT Scores x Switch Costs



Figure 5.13 Switch Costs x VFT scores (y-axis: VFT scores, x-axis, switch costs)

Model 3: HPBT x Language

The third model included HPBT and Language as the predictor variables, along with all interactions between the levels of the two variables. When this model was first fitted, OPT scores and VFT scores were included (separately) with the HPBT variable. Upon comparing the models, however, the removal of OPT and VFT resulted in a better

fit, as determined by the AICs and likelihood ratio tests. Accordingly, the model which has a less complex structure was selected. Despite the use of a simpler model, HPBT did not predict switch costs, and there was no significant interaction between Language and HPBT, as summarised in Table 5.25.

Table 5.25 MEM ANOVA (Type III) Summary for Switch Costs and HPBT (Model 3)

	F-Statistic	DF1	DF2	p-value
Intercept	6616.15	1	178	<0.0001****
HPBT	2.68	1	58	0.11
Language	18.12	2	178	<0.0001
HPBT*Language	0.504	2	178	0.48

Exploratory analyses were then performed by creating an *ad hoc* variable (hereafter, 'SM') which classified participants as having either symmetrical or asymmetrical switch costs across the three languages. For each participant, having a significant difference in switch costs in any of the three language pairs classified them as 'asymmetrical switchers'. For example, if a participant exhibited symmetrical English-French switch costs and symmetrical English-Arabic switch costs, but *asymmetrical* Arabic-French switch costs, they were classified as 'asymmetrical switchers'. 'Symmetrical switchers', on the other hand, had non-significant differences in switch costs across the three language pairs.

Correlations were then performed to explore whether participants who were highly proficient in two languages would exhibit a similar pattern of switch cost (a)symmetry to participants who were highly proficient in three languages, irrespective of group membership. As HPBT and SM were binary variables based on 'continuous' measures (ordinal OPT scores and continuous switch costs, respectively), a tetrachoric correlation was initially selected for analysis. However, as the test's assumption of normal distribution was not met ($p<0.05$), a Phi (ϕ) (fourfold point) correlation was

selected. The results of the analysis indicated a high correlation between HPBT and SM ($r(1,58)=0.815$, $p<0.01$). Being highly proficient in three languages was correlated with symmetrical switches, whereas being highly proficient in two languages correlated with asymmetrical switch costs. In other words, participants who were classified as balanced trilinguals exhibited symmetrical switch costs irrespective of whether they were EMI or FMI participants. This finding is consistent with the previous finding where FMI participants exhibited symmetrical switch costs, since the majority of this group were highly proficient trilinguals.

5.3.1.2 Language Proficiency and Pause Frequency

To investigate whether language proficiency modulates pause frequency across the different language conditions, four models were constructed to avoid multicollinearity and model overfitting. The outcome variable was the log-transformed pause frequency ratio, whereas the predictors were Group (2 levels: EMI and FMI), OPT scores (3 levels: Arabic, English, and French), VFT scores (3 levels: Arabic, English, and French), HPBT (Bilingualism vs. Trilingualism), and narration task Language Condition (4 levels: Arabic, English, French, and Trilingual). Interactions between the predictor variables were also included when relevant and possible. All models had maximal random effects structures. The results of each model are described next.

Model 1: Group x OPT Scores x Language Condition

The first model included Group, OPT Scores, and Language Condition as well as interactions between the three variables. The ANOVA results of the MEM, summarised in Table 5.26, revealed significant results, including a significant three-way interaction between Group, OPT Scores, and Language Condition ($F(6,150)=3.62$, $p=0.04$). A schematic representation of the interaction is offered in Figure 5.14. Accordingly, *post-*

hoc analyses were performed to explore simple main effects as well as contrasts of interest by changing the reference level for Group, OPT scores, and Language Condition. A summary of the model's *post-hoc* analyses of interest is offered in Table 5.27.

Table 5.26 MEM ANOVA (Type III) Summary for Pause Frequency and OPT (Model 1)

	F-statistic	DF1	DF2	p-value
Intercept	4596.32	1	150	<0.0001
Group	34.27	1	56	<0.0001
Language Condition	219.580	3	150	<0.0001
OPT Scores	27.42	2	150	<0.001
Group*Language Condition	250.83	3	150	<0.0001
Group*OPT Scores	2.56	2	150	0.05
Language Condition*OPT Scores	5.13	6	150	0.03
Group*Language Condition*OPT Scores	3.62	6	150	0.04

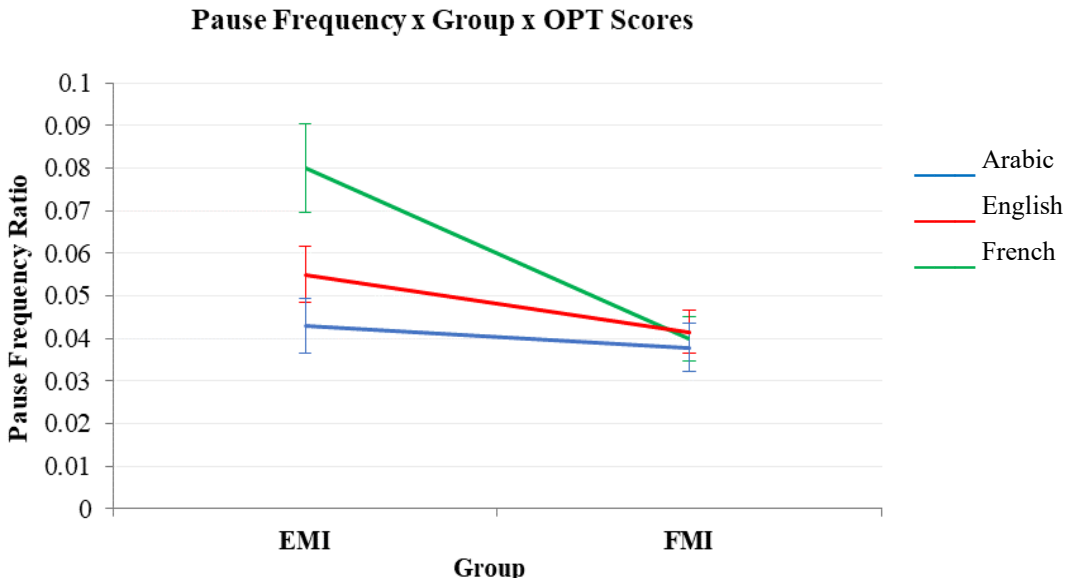


Figure 5.14 Pause Frequency as predicted by OPT Scores (EMI vs. FMI)

Table 5.27 MEM Output Summary for Pause Frequency *post-hoc* analyses (Estimated coefficients, standard errors (SE), t-values, and p-values) Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	Pause Frequency			
Predictor	**Coefficient**	**SE**	**t-value**	**p-value**
Intercept	0.45	0.03	13.35	0.00****
Group	0.09	0.04	2.54	0.03*
OPT Arabic	0.07	0.02	3.65	0.00***
OPT French	0.08	0.02	4.05	0.00****
OPT English	0.06	0.02	3.07	0.00**
Arabic-only	0.07	0.02	3.44	0.00***
English-only	0.08	0.02	4.21	0.00****
French-only	0.09	0.02	3.69	0.00***
Trilingual	0.07	0.02	3.05	0.00**
Group*OPT Arabic	0.11	0.04	2.85	0.00**
Group*OPT French	0.16	0.04	4.03	0.00***
Group*OPT English	0.10	0.04	2.33	0.02*
Group*Arabic-only	0.08	0.04	2.39	0.02*
Group*English-only	0.10	0.04	2.55	0.01*
Group*French-only	0.11	0.04	2.72	0.00**
Group*Trilingual	0.04	0.04	1.04	0.27
OPT Arabic*Arabic-only	0.01	0.03	2.32	0.02*
OPT English*English only	0.08	0.03	2.67	0.00**
OPT French*French-only	0.02	0.03	5.69	0.00****
Group*OPT Arabic*Arabic-only	0.09	0.03	2.69	0.00**
Group*OPT English*English only	0.07	0.03	2.33	0.02*
Group*OPT French*French only	0.13	0.03	5.04	0.00****
Group*OPT Arabic*Trilingual	0.04	0.04	1.07	0.29
Group*OPT English*Trilingual	0.86	0.04	2.14	0.04*
Group*OPT French*Trilingual	0.84	0.04	2.09	0.05*

*Formula: Pause Frequency ~ Group + Language Condition + OPT Scores + Group*OPT scores + Group*Language Condition + Group*OPT(by level)*Language Condition(by level) + (1|Participant)*

The model revealed that the three-way interaction was driven by EMI's French proficiency scores predicting pause frequency in the French-only condition significantly more than the English (t=5.59, p<0.0001) and Arabic (t=3.67, p<0.001) proficiency scores predicting pause frequency in English-only and Arabic-only conditions, respectively. FMI participants' French OPT scores also predicted pause frequency in the French-only condition significantly more than English OPT (t=3.29, p<0.001) and Arabic OPT (t=5.33, p<0.0001) scores predicted pause frequency in their respective

language contexts. Arabic OPT scores did not significantly predict pause frequency in the Arabic-only condition for FMI participants ($p>0.05$).

Furthermore, EMI participants paused significantly more frequently in French compared to their FMI peers, and the effect of proficiency in French was significantly larger for EMI than FMI participants ($t=5.11$, $p<0.0001$). The three-way interaction was also driven by the Arabic condition, where Arabic language proficiency predicted pause frequency more significantly for the EMI than the FMI group ($t=2.23$, $p<0.05$). For the English-only condition, the results of the analyses revealed that EMI proficiency scores predicted pause frequency for EMI participants significantly more than they did for FMI participants ($t=2.67$, $p<0.01$). The interaction was also driven by EMI participants' English OPT scores significantly predicting trilingual pause frequency significantly more than the OPT scores of their other two languages (Arabic: $t=3.20$, $p<0.001$; French: $t=6.17$, $p<0.0001$). On the other hand, in the trilingual condition, FMI pause frequency was predicted by French OPT scores significantly more than by English OPT ($t=2.14$, $p<0.05$) and Arabic OPT ($t=2.26$, $p<0.05$) scores.

These results highlight the role of proficiency, especially in French, in modulating fluency for the EMI group. While higher proficiency scores predicted fewer pauses in almost every language condition, the magnitude of the influence of proficiency was prominent as EMI participants produced significantly more pauses in French than in Arabic and English. FMI participants, on the other hand, were more fluent in French than English and Arabic, the latter being the only condition where proficiency did not predict pause frequency. Another noteworthy observation is that in the trilingual condition, EMI participants' English proficiency predicted their pause frequency, whereas FMI participants' French proficiency predicted their fluency. This is in line with the analysis of pause frequency ratios in the NT, where performance on

the trilingual condition was analogous to that of the immersion language in each group.

Model 2: Group x VFT Scores x Language Condition

The second model included Group, VFT scores in each language, and Language Condition in the four linguistic contexts as predictor variables. Two-way interactions between the three variables were also included as explanatory variables of pause frequency ratios, as selected after model comparisons. The results of the model's full estimates, offered in Table 5.28, revealed no significant main effects for VFT scores or for interactions between VFT scores and the other variables.

Table 5.28 MEM Output Summary for Pause Frequency and VFT (Estimated coefficients, standard errors (SE), t-values, and p-values) Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	Pause Frequency			
Predictor	**Coefficient**	**SE**	**t-value**	**p-value**
Fixed Effects				
Intercept	5.22	0.41	12.73	0.00****
Group	1.14	1.15	0.99	0.33
VFT Arabic	0.03	0.04	0.78	0.94
VFT French	0.07	0.04	1.71	0.09
VFT English	0.05	0.04	1.21	0.23
Arabic-only	0.04	0.04	1.21	0.22
English-only	0.12	0.04	3.32	0.00**
French-only	0.25	0.04	6.97	0.00****
Trilingual	0.16	0.04	4.53	0.00****
Group*VFT Arabic	0.07	0.04	1.66	0.10
Group*VFT French	0.03	0.03	1.09	0.28
Group*VFT English	0.07	0.04	1.81	0.08
Group*Arabic-only	0.04	0.02	1.61	0.11
Group*English-only	0.01	0.03	0.38	0.70
Group*French-only	0.02	0.00	0.72	0.45
Group*Trilingual	0.03	0.00	0.39	0.69
VFT Arabic*Arabic-only	0.05	0.00	1.11	0.27
VFT English*English only	0.01	0.00	1.04	0.30
VFT French*French-only	0.00	0.00	0.06	0.99

*Formula: Pause Frequency ~ Group + VFT scores + Language Condition + VFT*Group + VFT*Language Condition + Group*Language Condition +(1|Participant)*

As this result was unexpected, and having tested some of the different simulations obtained from the 'glmulti' package which included the removal of Group and

interactions with Group, VFT scores as predictors of pause frequency were analysed further through splitting the data by levels. That is, VFT scores were split into three variables (by language) and Pause Frequency was split into four variables (by language condition). The outcome variable was no longer a multi-level variable; rather, pause frequency ratios in each language condition were analysed as separate dependent variables in separate linear (non-nested) models. The independent variables included VFT scores in only one language (and the trilingual condition was examined in four separate models (see Appendix I)).

Only one model which included French VFT and Group as predictors revealed significant results, as indicated in Table 5.29. According to this model, whose fit was acceptable as examined by the residual distribution plot, there was a significant interaction between Group and VFT scores in French. Upon releveling the model, it was observed that VFT scores were a significant predictor of French pause frequency for EMI participants (β=0.14, SE=0.02, t=6.12, p<0.0001), but not for FMI participants (t=1.79, p>0.05). That is, higher VFT French scores predicted fewer pauses in the French-only condition, but this effect was only observed in the EMI group.

Table 5.29 MEM ANOVA (Type III) Summary of French-only Condition and French VFT Scores

	F-Statistic	DF1	DF2	p-value
Intercept	62649	1	56	<0.0001
Group	859.39	1	56	<0.0001
VFT French	28.93	1	56	<0.0001
Group*VFT French	17.88	1	56	<0.0001

*Formula: French Pause Frequency~Group*French VFT + (1|Participant)*

Although the results of this model (and the exploratory models) were generally non-significant, they offer insights into the two tasks which are examined in this model: the VFT and the narration task. The validity of the VFT in measuring proficiency is once again demurred. Model 1 in this section indicated that pause frequency was

modulated by proficiency scores obtained from the OPT. The overall lack of significant results when the VFT was examined as a predictor highlights the difference between the OPT and VFT—both of which are said to be measures of language proficiency. The exploratory model deconstructed both Pause Frequency and VFT scores, and the observation that only French VFT scores for the EMI group predicted pause frequency for this group indicates that the VFT is only able to discern differences when the results are striking (i.e., highly significant). However, it is not a sensitive measure in conditions where there are subtle differences in performance—unlike the OPT.

Furthermore, considering the constructs which the VFT and NT are designed to measure, one could argue that the VFT signals lexical access and retrieval and that (at least some) pauses stem from lexical gaps, or slower lexical access and retrieval. Accordingly, participants with high performance on the VFT task were expected to pause less frequently in the respective languages. The non-significant relationship between these two tasks, however, may be ascribed to the notion that the pauses by participants did not mainly stem from slower lexical access, but from other sources such as difficulty in constructing grammatical structures. Like the picture-naming task, the VFT involves one-word utterances, whereas the narration task taps into a much larger knowledge of language, and a dissociation between performance on these tasks is conceivable.

Another component of the VFT which was believed to overlap with the narration task is that of working memory. In both tasks, participants needed to remember the task instructions (i.e., target language) and constantly monitor their speech (for the NT, in the single-language conditions only) to ensure proper language choice. However, the divergent results of the two tasks revealed that monitoring may not have played a critical role in participants' performance when the two tasks were

measured against each other. The non-significant relation between VFT scores and working memory, as examined in Chapter 4, further supports this claim.

Model 3: HPBT x Language Condition

This model investigated whether being highly proficient in two or three languages influenced the frequency of pauses in each of the language conditions. The predictors were HPBT (2 levels), Language Condition (4 levels), and interactions between these two variables. The outcome variable was the transformed pause frequency ratios. Two initial models which included additional fixed effects such as OPT scores and VFT scores (separately) yielded non-significant results. By removing these two variables, the fit of the model improved as tested by the AIC and likelihood ration tests. The results of the model's output are presented in Table 5.30.

Table 5.30 MEM Output Summary for Pause Frequency and HPBT (Estimated coefficients, standard errors (SE), t-values, and p-values) Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	Pause Frequency			
Predictor	**Coefficient**	**SE**	**t-value**	**p-value**
Fixed Effects				
Intercept	0.45	0.05	9.07	0.00**
HPBT	0.32	0.08	4.06	0.00****
Arabic-only	0.25	0.07	3.64	0.00***
English-only	0.27	0.07	3.88	0.00****
French-only	0.31	0.07	4.29	0.00****
Trilingual	0.24	0.07	3.45	0.00***
HPBT*Arabic-only	0.00	0.01	0.43	0.66
HPBT*English-only	0.00	0.01	0.15	0.88
HPBT*French-only	0.07	0.01	6.57	0.00****
HPBT*Trilingual	0.00	0.01	0.42	0.67

*Formula: Pause Frequency ~ HPBT*Language Condition + (1|Participant)*

The model revealed significant main effects for HPBT and Language Condition, as well as a significant interaction between HPBT and the Language Condition (β=0.02, SE=0.01, t=2.21, p<0.05), illustrated in Figure 5.15. *Post-hoc* analyses revealed that the simple main effect of the HPBT variable was significant in the French-only condition,

where participants who were highly proficient in three languages paused significantly less frequently than participants who were highly proficient in two languages only (t=6.57, p<0.0001). These findings correspond to the earlier analyses which noted a difference in FMI vs. EMI group performance in the French-only condition, with the FMI group significantly pausing fewer times than the EMI group. As the majority of the highly-proficient trilingual group are FMI participants, this finding was expected.

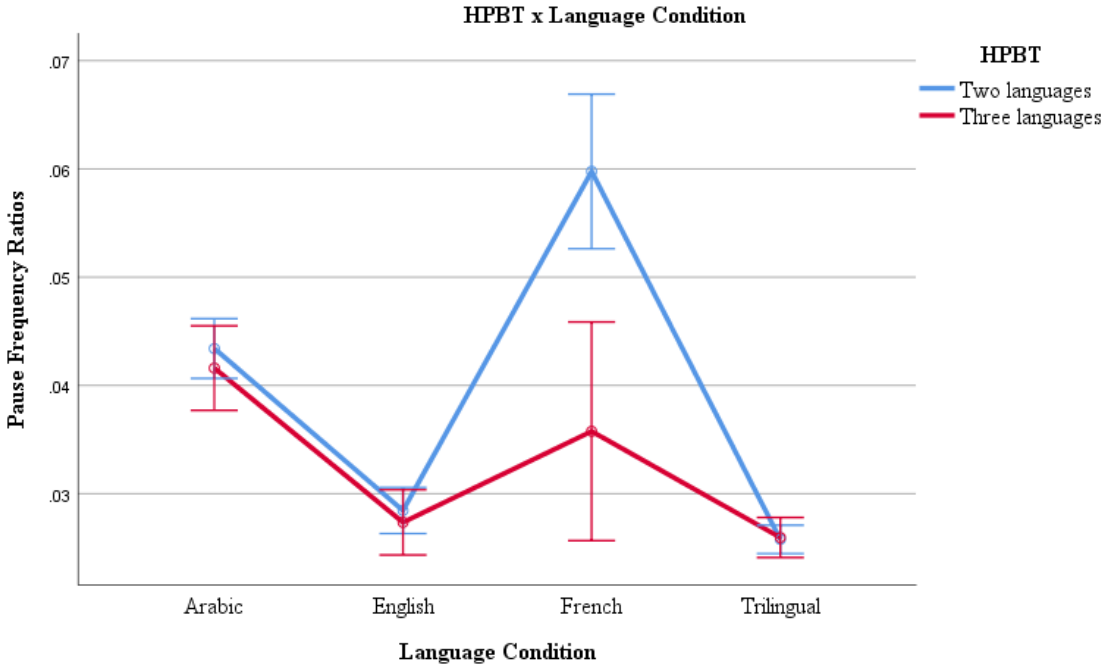


Figure 5.15 HPBT x Language Condition Interaction Plot, error bars=CI (95%)

5.3.1.3 Language Proficiency and Pause Duration

Analyses of pause frequency and pause duration revealed a similar pattern across the different language contexts. It was therefore hypothesised that a similar pattern of results would be observed when investigating the influence of language proficiency on pause duration. Several models, similar to those constructed for pause frequency analyses, were created; the outcome variable was mean pause duration and the predictors were Group (2 levels: EMI and FMI), OPT scores (3 levels: Arabic, English, and French), VFT scores (3 levels: Arabic, English, and French), HPBT

(Bilingualism vs. Trilingualism), and narration task Language Condition (4 levels: Arabic-only, English-only, French-only, and Trilingual). Interactions between the predictor variables were also included when relevant and possible. The results of each model are described next.

Model 1: Group x OPT Scores x Language Condition

The best-fit model included Group, OPT Scores, and Language Condition as well as interactions between the three variables. The model revealed significant main effects for Group, OPT scores, and Language Condition, as well significant interactions between the three variables ($F(6,174)=7.41$, $p<0.01$), as summarised by the MEM's ANOVA results in Table 5.31. Figure 5.16 illustrates the interaction between OPT scores and pause durations for the two groups.

Table 5.31 MEM ANOVA (Type III) Summary for Pause Duration and OPT

	F-statistic	DF1	DF2	p-value
Intercept	73071.01	1	174	<0.0001
Group	106.03	1	58	<0.0001
Language Condition	340.28	3	174	<0.0001
OPT Scores	3.40	2	174	0.02
Group*Language Condition	58.52	3	174	<0.0001
Group*OPT Scores	401.21	2	174	<0.0001
Language Condition*OPT Scores	4.56	6	174	0.04
Group*Language Condition*OPT Scores	7.41	6	174	<0.01

The significant Group x OPT scores x Language condition interaction was further examined through analyses of simple main effects. *Post-hoc* contrasts were obtained by changing the reference (intercept) for each level of Group, OPT scores, and/or Language condition as needed. Of interest was to investigate whether language proficiency in each language predicted switch costs in that language. For the trilingual condition, proficiency scores from all three languages were examined, as the potential source(s) of influence for each group was uncertain. The results of the model's output, summarised in Table 5.32, highlight the analyses of interactions and simple main effects as relevant.

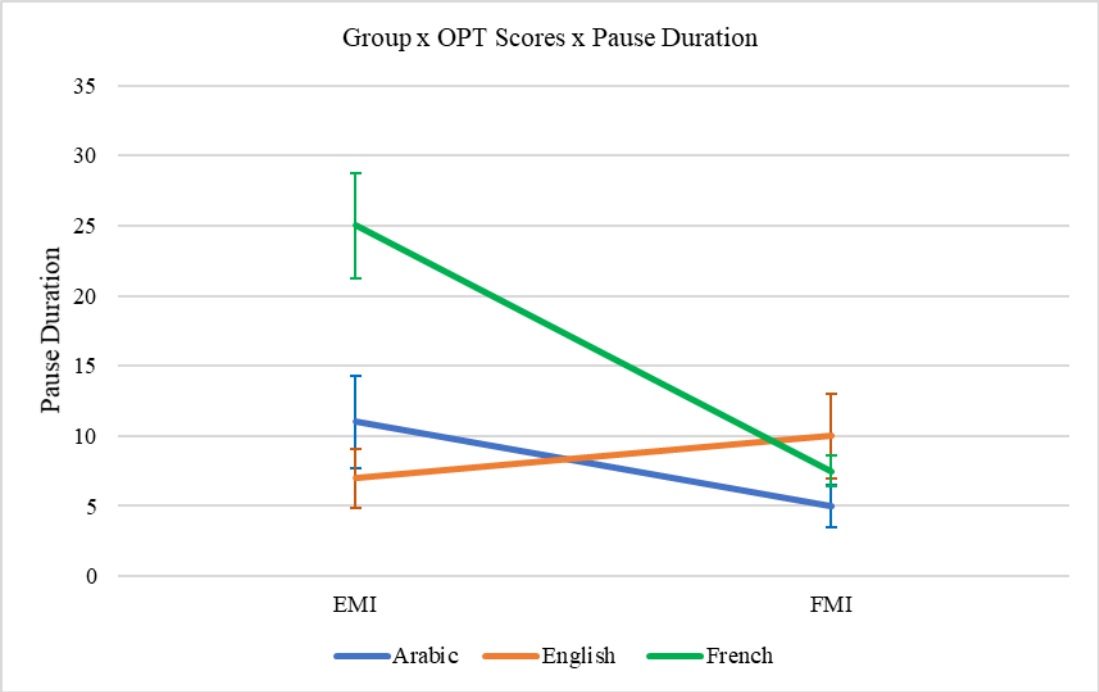


Figure 5.16 Interaction plot for Group x OPT Scores x Pause Duration; error bars=CI(95%)

Table 5.32 Overall Summary of post-hoc analyses for Pause Duration

	Pause Duration			
Predictor	**Coefficient**	**SE**	**t-value**	**p-value**
Fixed Effects				
Intercept	2.75	0.06	48.79	0.00****
Group	0.31	0.02	2.53	0.01**
OPT Scores	0.53	0.22	2.43	0.02*
Language Condition	0.33	0.15	2.28	0.02*
Group*OPT Scores	0.47	0.10	4.69	0.00****
Arabic OPT *Arabic Condition	0.02	0.01	1.48	0.14
English OPT*English Condition	0.06	0.02	3.67	0.00***
French OPT*French Condition	0.71	0.12	5.90	0.00****
Arabic OPT*Trilingual Condition	0.01	0.02	0.45	0.65
English OPT*Trilingual Condition	0.06	0.02	3.59	0.00***
French OPT*Trilingual Condition	0.05	0.02	2.46	0.02*
Group*Arabic OPT *Arabic Condition	0.87	0.33	2.63	0.01*
Group*English OPT*English Condition	0.03	0.02	1.98	0.05
Group*French OPT*French Condition	1.03	0.15	7.08	0.00****
Group*Arabic OPT*Trilingual Condition	0.02	0.02	1.02	0.31
Group*English OPT*Trilingual Condition	0.34	0.14	2.42	0.02*
Group*French OPT*Trilingual Condition	0.33	0.14	2.28	0.02*

*Formula: Pause Duration~Group*OPT Scores*Condition+(1+OPT* Condition|Participant)*

The analyses indicated that the three-way interaction was driven by OPT French proficiency scores predicting significantly longer pause durations in French for EMI than FMI participants ($t=3.07$, $p<0.01$). Arabic OPT scores also significantly predicted pause durations in the EMI group, but not in the FMI group ($p>0.05$). Furthermore, OPT scores in English predicted EMI pause durations in the English-only condition more strongly than in the FMI group ($t=2.98$, $p<0.05$).

The significant interaction was also driven by differences within each group. For EMI participants, French OPT scores predicted pause durations in French significantly more than English OPT scores predicted pause durations in the English-only condition ($t=3.78$, $p<0.001$) and significantly more than Arabic OPT scores predicted pause durations in the Arabic-only condition ($t=4.17$, $p<0.0001$). Furthermore, for this group, higher English OPT scores predicted shorter pause durations in the trilingual condition marginally more than Arabic OPT ($t=2.08$, $p<0.05$) and significantly more than French OPT ($t=5.21$, $p<0.0001$) scores. For the FMI group, OPT scores in French predicted pause durations significantly more than the English OPT scores ($t=3.19$, $p<0.001$) and Arabic OPT scores ($t=3.08$, $p<0.001$) in their respective monolingual conditions. Additionally, French OPT scores predicted pause durations in the trilingual condition significantly more than English OPT ($t=2.87$, $p<0.05$) and Arabic OPT ($t=2.12$, $p<0.05$) scores.

These findings generally parallel those of pause frequency. The significance of OPT scores in predicting pause duration was expected, as pauses index disfluencies which are attributed to several sources such as lexical gaps and difficulty in formulating grammatical structures. Accordingly, the magnitude of pause durations was hypothesised to be influenced by the oral proficiency scores of the participants.

The observation that Arabic OPT scores for the FMI group did not predict their

pause durations on the narration task, however, is contrary to this prediction, and is attributed to the unexpectedly long(er) pauses exhibited by this group. A possible explanation is that some of the words which are typically considered loans may have been perceived by the FMI participants as French words (i.e., switches). Participants would have then tried to find the lexical equivalent for such items, which then resulted in increased pausing (both in frequency and duration). As the origin of these words is French, FMI participants would have noticed that, whereas EMI participants may have been oblivious to the fact that these words are of French origin. This conjecture is based on the narration task transcriptions, where some pauses preceded Arabic lexical items. Although the present study does not delve into data from the transcriptions, this was a salient observation, as participants who were unable to retrieve a Lebanese Arabic word resorted to using Modern Standard Arabic, which is a written—not a spoken—variety or to translating words or phrases into Lebanese Arabic. When other participants performed the same narration storyboard task in the trilingual condition, they used either English or French words to express these items. An example of this is the use of the French 'chaise longue' (translation: 'beach chair' or 'deck chair') which a few participants struggled to name in Lebanese Arabic. Generally, participants translated the lexical phrase literally to Arabic after a pause (by saying 'long chair', whereas the Lebanese equivalent is 'beach chair'). This observation suggests, therefore, that word frequency and the prevalence of loans (and switches) in participants' Lebanese Arabic repertoire play a critical role in speech production. It also suggests that the participants were aware of their language use and monitored their speech to ensure only the target language was produced.

Model 2: Group x VFT x Language Condition

The second model included Group, VFT scores, and Language Condition in the four linguistic contexts as predictor variables of mean pause durations. Interactions

were removed as explanatory variables; thus, the best-fit model included the main predictors only. No significant results were obtained (all p>0.05), as summarised by the model's ANOVA in Table 5.33. Accordingly, exploratory analyses were performed through simpler and more complex models (Appendix I), including deconstructed variables, as was done in Model 2 of Pause Frequency x VFT. No model offered significant results, indicating that performance on the VFT (i.e., lexical robustness) was dissociated from pause durations in the narration task across all conditions.

Table 5.33 MEM ANOVA (Type III) Summary for Pause Duration and VFT (Model 2)

	F-statistic	DF1	DF2	p-value
Intercept	18236.39	1	177	<0.0001
Group	53.407	1	58	<0.0001
Language Condition	26.46	3	177	<0.0001
VFT Arabic	0.006	1	55	0.93
VFT English	0.284	1	55	0.59
VFT French	1.27	1	55	0.26

Formula: Pause Durations ~ Language Condition + Group + VFT Scores (+1|Participant)

Model 3: HPBT x Language Condition

The third model included HPBT, Language Condition, and interactions between these two variables as explanatory variables, whereas the response variable was the pause duration. This was the best-fit model as determined by the AIC and likelihood ratio tests. The results of this model's ANOVA, summarized in Table 5.34, revealed significant main effects for HPBT and Language Condition, and a significant interaction between the two variables ($F(3,132)=12.10$, $p<0.0001$).

Table 5.34 MEM ANOVA (Type III) Summary for Pause Duration and HPBT (Model 3)

	F-statistic	DF1	DF2	p-value
Intercept	25508.10	1	132	<0.0001
HPBT	14.05	1	56	<0.001
Language Condition	74.70	3	132	<0.0001
HPBT*Language Condition	12.10	3	132	<0.0001

*Formula: Pause Durations ~ HPBT*Language Condition (+1|Participant)*

To investigate the significant interaction, *post-hoc* analyses were performed, and are presented in Table 5.35. The results of these analyses revealed that the simple main effect of the HPBT variable was significant in the French-only condition, where participants who were highly proficient in three languages paused for significantly shorter durations than participants who were highly proficient in two languages only (t=8.18, p<0.0001) (see Figure 5.17). As with the observation noted regarding the significant HPBT x Pause Frequency interaction, the present findings are in line with the finding that the majority of the highly proficient trilinguals were FMI participants.

Table 5.35 MEM Output Summary for Pause Duration and HPBT post-hoc analysis (Estimated coefficients, standard errors (SE), t-values, and p-values)
Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗)

	Pause Duration			
Predictor	**Coefficient**	**SE**	**t-value**	**p-value**
Fixed Effects				
Intercept	2.08	0.03	68.08	0.00****
HPBT	0.32	0.05	6.10	0.00****
Arabic-only	0.12	0.04	2.79	0.00**
English-only	0.15	0.04	3.89	0.00***
French-only	0.29	0.04	7.23	0.00****
Trilingual	0.10	0.04	2.55	0.00**
HPBT*Arabic-only	0.01	0.07	0.17	0.87
HPBT*English-only	0.02	0.07	0.23	0.82
HPBT*French-only	0.57	0.07	8.18	0.00****
HPBT*Trilingual	0.00	0.07	0.06	0.95

*Formula: Pause Duration ~ HPBT*Language Condition + (1|Participant)*

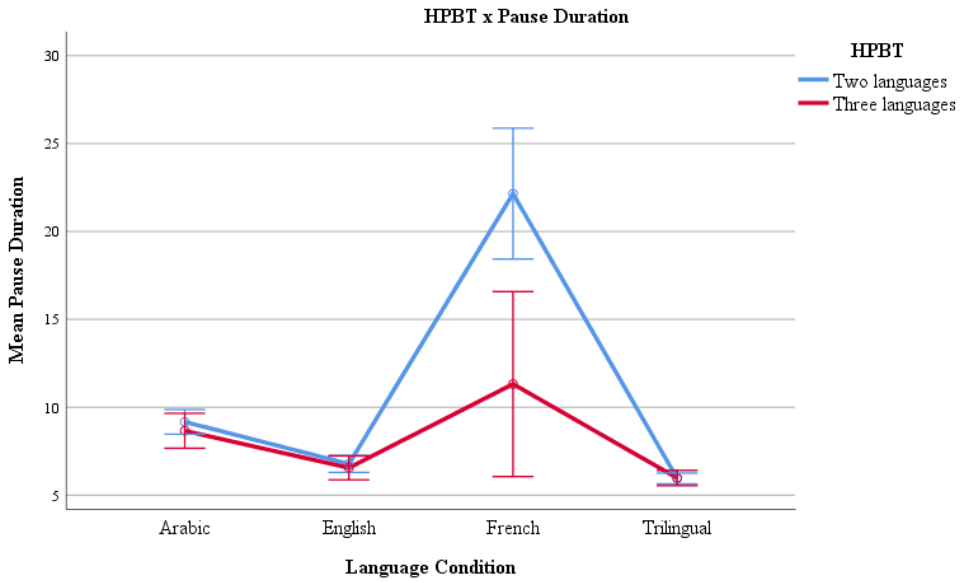


Figure 5.17 HPBT x Language Condition x Pause Duration Interaction Plot, error bars=CI(95%)

5.3.1.4 Summary of Language Proficiency as a Predictor of Linguistic Performance

To summarise, OPT scores, VFT scores, and the HPBT variable were explored as predictors of linguistic task performance. Switch costs from the picture-naming task were strongly predicted by OPT scores for both groups, where participants' higher proficiency in a language predicted larger switch costs. This finding is in line with the ICM. VFT scores, however, predicted switch costs for the FMI group only in French, with an opposite (or reverse) finding for EMI participants. That is, according to this task, a higher proficiency in English for EMI participants predicted smaller English switch costs. Furthermore, the VFT did not significantly predict switch-costs for 'weaker' languages. Finally, for switch costs, the HPBT and *ad hoc* variable SM revealed that participants who were highly proficient in three languages exhibited symmetrical switch costs across the three languages, whereas participants who were highly proficient in two languages generally exhibited asymmetrical switch costs.

For the narration task, pause frequency was predicted by OPT scores for both groups and across all conditions except in the Arabic-only condition of the FMI group, as participants in this group had the highest number of pauses in this condition. VFT scores failed to predict pause frequency in any condition except in the French-only condition where French VFT scores predicted EMI participants' high number of pauses. Finally, for the HPBT variable, participants who were highly proficient in three languages paused less frequently in the trilingual condition of this task than their peers whose proficiency was high in only two languages. Pause durations on the narration task were also predicted by OPT scores for almost all conditions and for both groups. The only exception was that OPT scores did not predict FMI pause durations in the Arabic-only condition. The VFT scores did not predict pause durations in any condition for either group, whereas the HPBT variable predicted shorter pauses in participants who were highly proficient in three languages in the French-only condition.

5.3.2 Language Experience as a Predictor of Linguistic Task Performance

RQ4B: Does language experience influence switch costs in the TPNT and pause frequency and duration in the NT for the two groups?

In order to investigate the influence of language experience on switch costs, pause frequency, and pause duration, a series of mixed-effects models was selected for analysis. As indicated at the beginning of this section, the models had a maximal random-effects structure and were checked for multicollinearity and singularity. For the language experience variable, the four sub-constructs were analysed as predictors of linguistic task performance, including interactions when necessary and possible. The models' selection and quality-of-fit were determined by following the procedures

outlined throughout the chapter (AIC, likelihood ratio tests, and residual plots). The results of the models are presented for each of the outcome variables next.

5.3.2.1. Language Experience and Switch Costs

To examine the influence of language experience on switch costs, three models were created, in order to avoid model overfitting and multicollinearity, especially between Language Use and Language Exposure. As with all previous analyses, model selection and inspection were performed to ensure that the best-fit models were selected for analysis.

Model 1: Group x Language Use x Language

The first model included Group (2 levels: EMI vs. FMI), Language Use (3 levels: Arabic, English, and Arabic), switch Language (3 levels: Arabic, English, and French), along with interactions between them, as predictors. The outcome variable was RT switch costs. The ANOVA summary of the model, presented in Table 5.36, indicated significant main effects and interactions for all predictors, including a significant interaction between the three predictor variables ($F(4,112)=4.46$, $p<0.05$), illustrated in Figure 5.18.

Table 5.36 MEM ANOVA (Type III) Summary of Switch Costs and Language Use

	F-statistic	DF1	DF2	p-value
Intercept	579.80	1	112	<0.0001
Group	15.70	1	58	<0.0001
Language	12.10	2	112	<0.0001
Language Use	6.53	2	112	0.01
Group*Language	3.72	2	112	0.03
Group*Language Use	6.59	2	112	0.01
Language*Language Use	3.84	4	112	0.03
Group*Language*Language Use	4.46	4	112	0.02

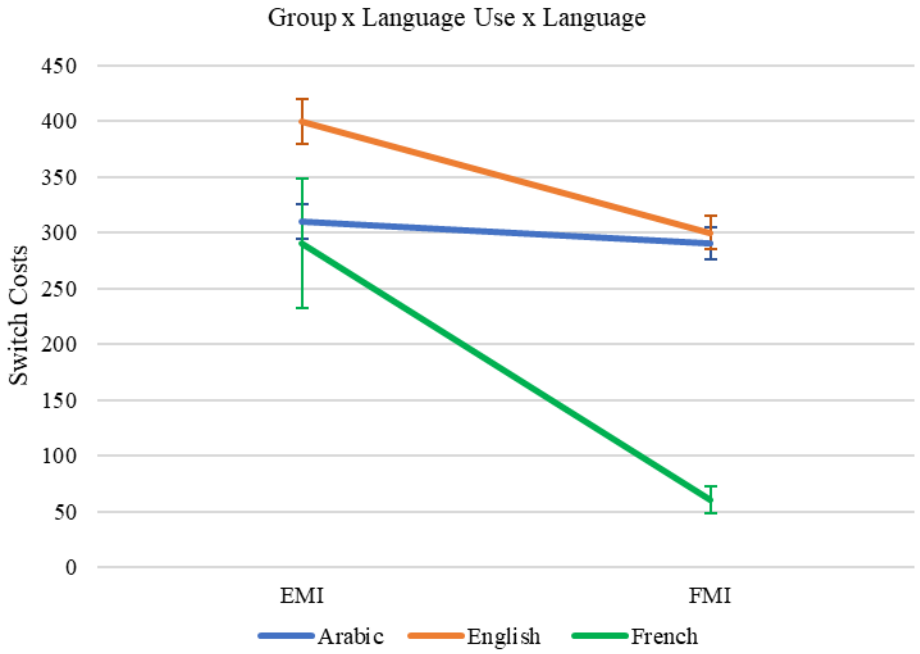


Figure 5.18 Switch Costs x Language Use x Group Interaction Plot

In order to explore the interaction, simple main effects were examined through changing the factor coding and contrasts of relevance were obtained through releveling the intercept for Group, Language, and Language Use. The results of the *post hoc* analyses revealed that language use in English predicted EMI and FMI participants' switch costs (EMI: β=3.04, SE=1.48, t=2.06, $p<0.05$; FMI: β=3.23, SE=1.56 t=2.08, $p<0.05$), where an increased self-reported use of English predicted larger English switch costs. Furthermore, French language use significantly predicted switch costs for EMI participants (β=6.52, SE=1.28, t=5.09, $p<0.0001$), but not for FMI participants ($p>0.05$). Specifically, EMI participants reported minimal French use in their daily interactions, and this predicted the small switch costs in French. Arabic language use did not predict switch costs for either group ($p>0.05$ for both group analyses).

Language Use was believed to be a significant predictor of switch costs, yet it did not seem to account for performance on this task, especially for the FMI group. FMI participants reported using Arabic more frequently than English and French, yet their

switch costs were the smallest, though not significantly lower, in Arabic (262 compared to 282 and 291, respectively). FMI participants who reported using English more frequently, however, had larger switch costs in English, yet this effect was not as strong a predictor as language proficiency. As for French, language use was reported to be minimal, yet this did not predict switch costs in French, which were the highest in French. Perhaps it is the relatively more equal distribution of language use (across the three languages) by this group that led to the marginal influence of 'Language Use' as a predictor. This is contrasted with the EMI group who reported using English almost two-thirds of the time and using French minimally (5% of the time). This unbalanced use of languages was therefore reflected in the more significant findings for the EMI group compared to their FMI peers.

Model 2: Group x Language Exposure x Language

The second model included Group, Language Exposure, and Language, along with interactions between the three variables, as predictors. The outcome variable was RT switch costs. The results of the model indicated no significant main effects or interactions, so a stepwise procedure was followed. The final model excluded Group and all interactions and retained a by-participant random effects structure to control for individual variability. The results of the model are summarised in Table 5.37.

Table 5.37 MEM Output Summary for Switch Costs and Language Experience (Model 2) (Estimated coefficients, standard errors (SE), t-values, and p-values)
Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗)

	Switch Costs			
Predictor	**Coefficient**	**SE**	**t-value**	**p-value**
Fixed Effects				
Intercept	261.67	66.45	3.94	0.00***
Arabic Exposure	0.00	1.58	0.00	0.96
English Exposure	2.75	1.38	1.99	0.05
French Exposure	7.84	1.45	5.40	0.00****

The results of the model indicated main effects for French (t=5.40, p<0.0001), where switch costs for French were significantly influenced by the amount of French input participants reported receiving in social settings. Although English exposure was statistically non-significant, as its p-value was slightly above the 0.05 cut-off, it may also be seen as an important variable which predicts English switch cost performance. Accordingly, for both EMI and FMI participants, language exposure in French predicted switch cost performance, where an increased amount of input predicted larger switch costs.

Model 3: Group x CSF x SalC x Language Condition

The third model included Group (2 levels), self-reported Codeswitching Frequency (CSF) scores, self-reported Salient-Cue detection (SalC) scores, and Language (3 levels) as predictors. Interactions between the four variables were also included in the initial structuring of the model. The 'Group' variable did not significantly predict the magnitude of switch costs in this model, and removing it improved the overall fit as indicated by the AICs and likelihood ratio tests. Accordingly, a simpler model which excluded Group was adopted.

Table 5.38 MEM ANOVA (Type III) Summary for Switch Costs and Language Experience (Model 3)

	F-statistic	DF1	DF2	p-value
Intercept	415.76	1	112	<0.0001
Language	11.43	2	112	<0.0001
CSF	5.49	1	56	0.02
SalC	4.62	1	56	0.04
Language*CSF	0.10	2	112	0.75
Language*SalC	0.79	2	112	0.46
CSF*SalC	3.49	1	56	0.07
Language*CSF*SalC	0.84	2	112	0.44

The results of the model, presented in Table 5.38, indicated significant main effects for Language ($F(2,112)=11.44$, $p<0.0001$), CSF ($F(1,56)=5.49$, $p<0.05$) and SalC ($F(1,56)=4.62$, $p<0.05$) but no significant interactions between any variables.

These results indicate that higher codeswitching frequency predicted smaller switch costs, and that this effect was language independent. That is, frequent codeswitchers had smaller switch costs. Similarly, participants who reported being attentive to cues exhibited smaller switch costs on the TPNT across the different languages.

5.3.2.2 Language Experience and Pause Frequency

To investigate the influence of language experience on pause frequency in the narration task, mixed-effects models were fitted with the log-transformed pause frequency ratios as the outcome variable. The predictors included Group (2 levels: EMI vs. FMI), Language Condition (4 levels: Arabic-only, English-only, French-only, and Trilingual), Language Exposure (3 levels: Arabic, English, and French), Language Use (3 levels: Arabic, English, and French), self-reported code-switching frequency (CSF), and self-reported salient-cue detection (SalC). Due to the presence of many predictors and to multicollinearity between language exposure and language use, the model was overfitted and did not converge. Accordingly, correlations and non-significant interactions were removed, and separate models were used to investigate the influence of the factors on pause frequency. The final models which were selected converged with all subject and item slopes. The fit of each model was assessed according to the procedure outlined at the beginning of this section.

Model 1: Group x Language Condition x Language Use

The first model was fitted with Group (2 levels), Language Condition (4 levels), and Language Use (3 levels) as predictor variables, and with the log-transformed pause frequency ratios as the outcome variable. The results of the model, summarised in Table 5.39, indicated a significant interaction between Group, Language Use, and Language Condition ($F(6,168)=3.49$, $p<0.01$).

Table 5.39 MEM ANOVA (Type III) Summary for Pause Frequency and Language Use (Model 1)

	F-statistic	DF1	DF2	p-value
Intercept	76774.33	1	168	<0.0001
Group	10.67	1	58	0.00
Language Condition	333.86	3	168	<0.0001
Language Use	0.38	2	168	0.78
Group*Language Condition	343.09	3	168	<0.001
Group*Language Use	1.98	2	168	0.16
Language Condition*Language Use	1.46	6	168	0.23
Group*Language Condition*Language Use	3.49	6	168	<0.01

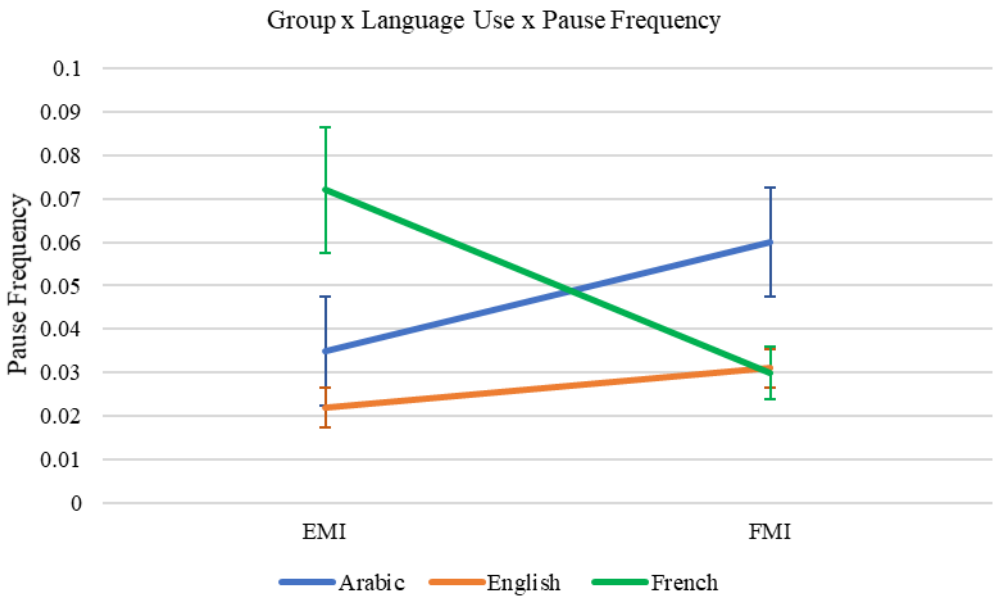


Figure 5.19 Language Use x Pause Frequency x Group Interaction plot (error bars=95%CI)

In order to investigate the significant three-way interaction, simple main effects were analysed by changing the factor contrast coding for the mixed model, and *post hoc* contrasts were analysed by changing the reference (intercept) for each level of Group, Language Condition, and Language Experience. A summary of the model's output is presented in Table 5.40.

Table 5.40 MEM Output Summary for Pause Frequency and Language Use (Model 2) (Estimated coefficients, standard errors (SE), t-values, and p-values) Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗)

	Pause Frequency			
Predictor	**Coefficient**	**SE**	**t-value**	**p-value**
Intercept	30.13	1.42	21.22	<0.0001
Group	0.03	0.01	3.06	0.00**
Arabic Condition	0.06	0.01	5.11	0.00****
English Condition	0.08	0.01	7.52	0.00****
French Condition	0.13	0.01	11.55	0.00****
Trilingual Condition	0.13	0.01	12.01	0.00****
Arabic Use	0.07	0.04	1.75	0.08
English Use	0.04	0.04	1.24	0.21
French Use	0.08	0.04	2.09	0.04*
Arabic Use*Arabic Condition	0.07	0.03	2.33	0.02*
English Use*English Condition	0.05	0.03	1.58	0.11
French Use*French Condition	0.11	0.03	3.82	0.00****
Group*English Use	0.04	0.01	3.91	0.00****
Group*French Use	0.14	0.01	9.55	0.00****
Group*Arabic Use	0.01	0.01	0.69	0.49
Group*Arabic Use*Arabic Condition	0.03	0.02	1.41	0.16
Group*English Use*English Condition	0.05	0.02	2.37	0.02*
Group*French Use*Arabic Condition	0.09	0.02	4.43	0.00****
Group*Arabic Use*Trilingual Condition	0.04	0.02	1.76	0.08
Group*English Use*Trilingual Condition	0.00	0.02	0.09	0.93
Group*French Use*Trilingual Condition	0.00	0.02	0.07	0.94

*Formula: Pause Frequency ~Group*Language Condition*Language Use+(1+Condition*Use|Participant)*

The results of the *post-hoc* analyses revealed that the influence of language use differed by language and by group. For Arabic, EMI participants' self-reported language use significantly predicted their pausing frequency on the narration task (t=2.68, p<0.01). That is, the higher the language use in Arabic, the fewer the pauses in the Arabic-only condition. However, for FMI participants, Arabic language use did not reach significance (p>0.05). This was an unexpected finding as FMI participants reported using Arabic significantly more than their other languages (at home, outside class at school, and in social contexts). However, as noted in the analysis of pause frequency, FMI participants

paused significantly more frequently in Arabic than any other language condition.

For English, there was a significant interaction between Group and English use. Language use significantly predicted the pausing frequency in the English-only condition for both groups. However, it predicted pause frequency significantly more for the EMI group than the FMI group ($t=2.19$, $p<0.05$). That is, for both groups, the higher the self-reported use of English, the more fluent participants were in their narrations in the English-only condition. However, the influence of English use was a stronger predictor for EMI participants than FMI participants.

For French, a group difference was observed, where EMI participants' self-reported language use significantly predicted their frequency of pausing in the French-only condition ($t=6.28$, $p<0.0001$). However, for FMI participants, self-reported French use did not predict their pausing frequency on the narration task ($p>0.05$). This finding, though unexpected, is attributed to the low frequency of language use of French as reported by the FMI participants (approximately 21% of total language use). However, the few pauses which FMI exhibited in the French condition may be attributed to the high proficiency which participants had attained in their French language and to the immersion context in which they are accustomed to producing speech in a French-only context. Accordingly, their pausing frequency in this condition may be ascribed to other explanatory variables (such as OPT French scores).

For the trilingual condition, the interaction between Language Use and Group did not reach significance for any of the three languages ($p>0.05$ for all contrasts). This may be due to the notion that for the two groups, pausing in this condition did not stem from a specific use of any one language. Rather, it may be due to the more balanced use of the three languages, or to other variables such as self-reported codeswitching, as is examined in Model 3 of this section. In order to examine whether a relatively balanced use of the

three languages predicted pause frequency in the trilingual condition, exploratory analyses were performed using the post-hoc analyses of Language Use (Appendix I). There were only four participants (out of 60) who had reported equal use of the three languages, all of whom were FMI participants. Descriptive analyses of their pause frequency revealed that the participants varied in the number of pauses they each produced on the trilingual condition, indicating that balanced language use did not influence their fluency on this condition differently to the other participants.

Model 2: Group x Language Condition x Language Exposure

In order to investigate the influence of language exposure in each language for each group, a model similar in structure to Model 1 was adopted with the replacement of Language Use with Language Exposure. Accordingly, the model included Group (2 levels), Language Condition (4 levels), and Language Exposure (3 levels). The best-fit model excluded all interactions between the variables. The results of the model, summarised in Table 5.41, revealed significant main effects for the three variables.

Table 5.41 MEM Output Summary for Pause Frequency and Language Exposure (Model 2) (Estimated coefficients, standard errors (SE), t-values, and p-values)
Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗)

	Pause Frequency			
Predictor	**Coefficient**	**SE**	**t-value**	**p-value**
Intercept	50.40	13.25	3.80	0.00***
Group	0.71	0.22	3.25	0.00**
Arabic Condition	1.46	0.24	6.11	0.00****
English Condition	1.47	0.24	5.94	0.00****
French Condition	0.89	0.24	3.62	0.00***
Trilingual Condition	1.69	0.24	6.85	0.00****
Arabic Exposure	0.03	0.02	1.56	0.10
English Exposure	0.07	0.02	3.31	0.00**
French Exposure	0.07	0.02	3.25	0.00**

Formula: Pause Frequency ~Group + Language Condition + Language Exposure + (1|Participant)

The results of the analyses revealed that for both EMI and FMI participants,

self-reported exposure to English predicted their pause frequency in the English condition of the narration task (EMI: t=3.73, p<0.001; FMI: t=3.12, p<0.01). Similarly, language exposure to French significantly influenced both groups on the French-only condition (EMI: t=3.29, p<0.01; FMI: t=3.64, p<0.001). There was no influence of Arabic exposure on pause frequency on either group. Similarly, trilingual pause frequency was not predicted by exposure to any of the three languages.

Model 3: Language Condition x Group x CSF x SalC

The third model included Language Condition (4 levels), Group, CSF scores, and SalC scores as explanatory variables. The best-fit model included only main effects, with no interactions between the variables. The results of the model, summarised in Table 5.42, indicated no significant results for any of the predictor variables, except for Language Condition and Group (which is, of course, expected). The removal of Language Condition and Group did not change the outcome of the analysis (see Appendix I).

Table 5.42 MEM ANOVA (Type III) Summary of Pause Frequency, Codeswitching Frequency, and Salient Cue Detection (Model 3)

	Pause Frequency			
	F-statistic	DF1	DF2	p-value
Intercept	38907.62	1	177	<.0001
Group	5.41	1	56	0.0237
Language Condition	56.80	3	177	<.0001
CSF	1.95	1	56	0.17
SalC	0.00	1	56	0.99

Formula: Pause Frequency~Language+Group+CSF+SalC + (1|Participant)

As pause frequency in the monolingual conditions of the narration task may have masked the influence of codeswitching frequency and salient cue detection, exploratory analyses were conducted given the assumption that participants who switch frequently would have fewer pauses in the trilingual condition. The data was split, and pause frequency in the trilingual condition was fitted as the outcome variable in a linear

mixed-effects model. 'Group' was excluded from the model as that improved the model's fit. The only two predictors were codeswitching frequency and salient cue detection, as well as the interaction between them. The results of the model, summarised in Table 5.43, revealed that in the trilingual condition, the two variables were significant predictors, but there was no significant interaction between them. These results suggest that participants who reported codeswitching frequently and who reported being attentive to cues exhibited fewer pauses than participants who reported less codeswitching and detection of cues in this context.

Table 5.43 MEM ANOVA (Type III) Summary of Codeswitching Frequency, Salient Cue Detection, and Trilingual Pause Frequency

	Trilingual Pause Frequency			
	F-statistic	DF1	DF2	p-value
Intercept	2833.79	1	56	<.0001****
CSF	5.79	1	56	0.02*
SalC	8.56	1	56	0.00**
CSF*SalC	0.01	1	56	0.91

Formula: Pause Frequency ~Language+Group+CSF+SalC + (1|Participant)

5.3.2.3 Language Experience and Pause Duration

As with the analysis of the influence of language experience on pause frequency, pause duration was analysed in three LMEs, where language use, language exposure, codeswitching frequency, and salient cue detection were fitted as predictors. The results of each model are presented in the next subsections.

Model 1: Language Use x Group x Pause Duration

The first model was fitted with Group (2 levels), Language Condition (4 levels), and Language Use (3 levels), along with interactions between these variables, as predictor variables. The outcome variable was the log-transformed mean pause durations. The results of the model, summarised in Table 5.44, indicated a significant

interaction between Group, Language Use, and Language Condition ($F(1,168)=4.32$, $p<0.0001$), presented in Figure 5.20.

Table 5.44 MEM ANOVA (Type III) Summary of Pause Duration and Language Use (Model 1)

	F-statistic	DF1	DF2	p-value
Intercept	1781.17	1	168	<0.0001
Group	32.66	1	58	<0.0001
Language Condition	301.47	3	168	<0.0001
Language Use	1.68	2	168	0.17
Group*Language Condition	347.52	3	168	<0.0001
Group*Language Use	1.63	2	168	0.19
Language Condition*Language Use	1.06	6	168	0.37
Group*Language Condition*Language Use	2.24	6	168	<0.05

*Formula: Pause Duration ~ Group*Language Condition*Language Use + (1|Participant)*

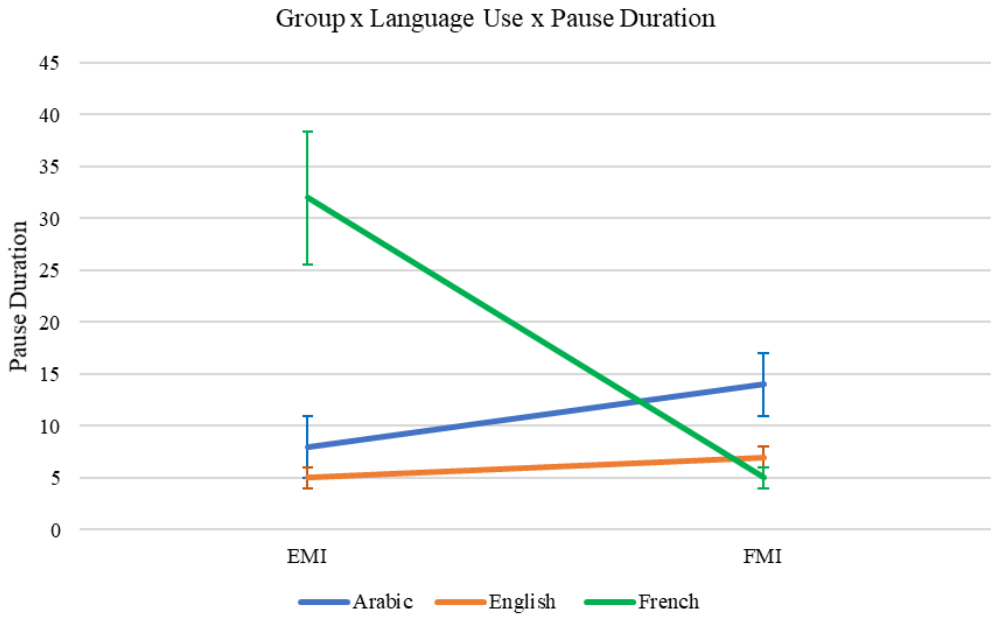


Figure 5.20 Language Use x Pause Duration x Group Interaction Plot

In order to investigate the significant three-way interaction, simple main effects were analysed by changing the factor contrast coding for the mixed model, and *post hoc* contrasts were analysed by changing the reference (intercept) for each level of Group, Language Condition, and Language Use. The results of the analysis indicated that the interaction was driven by French language use where FMI participants' increased

language use significantly predicted their short pause durations compared to EMI participants whose minimal use of French predicted much longer pauses (β=0.79, SE=0.26, t=3.03, p<0.01). Language Use in Arabic and English did not influence pause durations in the Arabic-only, English-only, or trilingual conditions of the task (all p>0.05). These results are in line with the participants' OPT scores, where EMI and FMI participants had comparable proficiency levels in English and Arabic, but differed in their French proficiency. The participants' self-reported language use is commensurate with their proficiency levels, and it is not surprising that pause durations on the narration task reflect the difference in language use on fluency.

Model 2: Language Condition x Language Exposure

This model was fitted with Language Condition and Language Exposure as predictor variables, and with the log-transformed pause durations as the outcome variable. The best-fit model excluded all interactions between the variables as well as the Group variable which was initially included; therefore, only main effects were examined in this model. The results of the linear mixed-effects model revealed no significant main effects for Language exposure except in French. The ANOVA summary of the model is presented in Table 5.45.

Table 5.45 MEM ANOVA (Type III) Summary of Duration and Language Exposure (Model 2)

	Pause Duration			
	F-statistic	DF1	DF2	p-value
Intercept	595.96	1	177	<.0001
Language Condition	51.76	3	177	<.0001
Arabic Exposure	0.03	1	56	0.87
English Exposure	0.79	1	56	0.38
French Exposure	13.28	1	56	<0.001

Formula: Pause Duration ~ Language Condition + Language Exposure + (1|Participant)

Accordingly, language exposure in French was the only significant predictor of pause durations (in the French-only condition) for both groups. That is, increased self-

reported exposure to French in social contexts predicted shorter pause durations in participants from both groups. This effect was not observed for Arabic or English for the Arabic-only and English-only conditions, or for exposure to any language in the trilingual condition.

As pause duration differed significantly between EMI and FMI participants on the French-only condition of this task, a post-hoc model explored whether there was a significant Group x French Exposure interaction in predicting pause durations. The results of the model revealed a significant interaction between Group and French Exposure ($F(1,56)=5.14$, $p<0.05$), where exposure to French influenced EMI participants' pause durations significantly more than FMI participants' ($\beta=19.17$, $SE=3.21$, $t=5.98$, $p<0.0001$). This result, along with the non-significant influence of language exposure in Arabic and English, underscores the critical role of exposure to a language in which proficiency is not high. Both EMI and FMI participants reported having the least amount of input in social settings from French, yet a difference in exposure highly influenced pause durations by EMI participants on this task.

Model 3: CSF x SalC

When fitting the third model, Language Condition was included as a fixed effect, along with Group, CSF, and SalC. However, the model did not converge; therefore, Group and Language Condition, which were non-significant predictors, were removed following the stepwise procedure, resulting in a better fit for the model. The model which was selected included CSF, SalC, and the interaction between these two variables, as fixed effects. It explored whether having codeswitching frequently and being attentive to cues predicted shorter pause durations. The results, presented in the ANOVA summary in Table 5.46, revealed no significant main effects or interaction.

Table 5.46 MEM ANOVA (Type III) Summary of Pause Duration, Codeswitching Frequency and Salient Cue Detection (Model 3)

	Pause Frequency			
	F-statistic	DF1	DF2	p-value
Intercept	595.96	1	180	<.0001
CSF	0.06	1	56	0.81
SalC	2.56	1	56	0.12
CSF*SalC	0.20	1	56	0.66

*Formula: Pause Duration ~ Codeswitching Frequency * Salient Cue Detection + (1|Participant)*

Although it was predicted that higher codeswitching frequency would predict shorter pause durations, this was not observed. A potential explanation is that pause durations do not relate to the frequency of switching between languages but to the use of each language, as revealed in the analysis of Model 1. However, as with pause frequency, it was of interest to see whether frequent codeswitchers would exhibit shorter pause durations on the trilingual condition of the narration task. To explore this hypothesis, pause duration data were split, and codeswitching frequency and salient cue detection were fitted as predictors in a linear mixed effects model, with trilingual pause durations as the outcome variable. The results of the model revealed no significant main effects for either variable; CSF: $F(1,57)=1.59$, p=0.21 and SalC: $F(1,57)=3.75$, p=0.06).

5.3.2.4 Summary of Language Experience as a Predictor of Linguistic Task Performance

To summarise the results of this section, high self-reported English language use predicted larger switch costs in English for both EMI and FMI participants. Furthermore, French language use predicted switch costs for the two groups, where the more frequent use of French predicted larger switch costs for participants in both groups. Arabic language use did not predict switch cost performance in Arabic for either group. Furthermore, language exposure in French predicted larger switch costs for participants in both groups, but Arabic and English self-reported language exposure

did not predict switch costs for either group. As for codeswitching frequency and salient cue detection, for both groups, the higher the self-reported frequency and attentiveness to cues, the smaller the switch costs irrespective of the language.

As for the narration task, EMI participants who reported using Arabic more frequently exhibited fewer pauses in the Arabic condition; however, this finding was not observed for FMI participants. Similarly, EMI participants who reported using French more frequently in their daily interactions were more fluent (i.e., paused less frequently) than their peers who reported less use of French. This was also not observed in the FMI group. As for English, a higher language use predicted fewer pauses for both groups. Language exposure also predicted fewer pauses in English and French for both groups, but not in Arabic. As for codeswitching frequency and salient cue detection, post-hoc analyses revealed that in the trilingual-condition only, higher codeswitching frequency and self-reported salient cue detection predicted fewer pauses in this condition.

Pause durations on the narration task were also influenced by language experience. Increased self-reported language use in French significantly predicted shorter pause length for both EMI and FMI participants. However, language use in Arabic and English did not predict pause duration for either group. Similarly, increased input in French predicted shorter pauses by participants in both groups in the French-only condition, but language exposure in Arabic and English did not predict pause duration in these languages. Finally, codeswitching frequency and salient cue detection did not predict pause duration in any language condition and for either group.

5.4 RQ5 – RELATIONSHIP BETWEEN COGNITIVE AND LINGUISTIC TASK PERFORMANCE

RQ5. What is the relationship between cognitive mechanisms and temporal switching costs in the linguistic tasks?

The final research question in this project aimed to examine the potential overlap between cognitive and linguistic task performance. Specifically, it examined whether enhanced inhibitory skills, conflict monitoring skills, and working memory capacity predict smaller *switching costs*. Switching costs were operationalised as RT switch costs in the picture naming task and as switch costs (i.e., long pause durations) of codeswitching in the narration task. The same analytical strategy which was adopted throughout the study was applied in the analyses of this question. That is, models were constructed with maximal random effects structures, and model comparison and selection were performed by applying a stepwise procedure. Models which did not converge were excluded from comparisons, and non-significant predictors were removed from the models. Model selection was then determined by Akaike's Information Criterion (AIC; Akaike, 1998) and the likelihood ratio chi-square tests. Visual inspections of the residuals' distribution plots confirmed that the model which was selected had met the homogeneity of variance condition. The results of the two subsections of this research question are presented next.

5.4.1. RQ5A – Cognitive Performance and the Picture-Naming Task

RQ5A. Do inhibitory control capacities, conflict monitoring skills, and working memory capacity predict RT switch costs in the picture-naming task?

In order to investigate whether enhanced cognitive abilities influenced the magnitude of switch costs, and by extension, whether there is an overlap between cognitive and linguistic control processes, a generalised mixed-effects model was fitted with the transformed RT switch costs as the outcome variable. Predictor variables included the z-scored Simon Effect scores, Conflict Effect scores, Backward Digit-Span

Task scores, Group (2 levels: EMI vs. FMI), and Language (3 levels: Arabic, English, and French). Interactions between these variables were also included as fixed effects.

Model comparisons revealed that Conflict Effect scores did not predict switch costs in any language ($p>0.05$ for all models tested); therefore, this variable was excluded from further analyses. Although separate models for each Simon Effect and Working Memory were tested, a model which combined both predictors resulted in similar outcomes, but more interestingly, revealed a significant interaction between the two predictors. Accordingly, the selected model included both Simon Effect and BDST scores. The ANOVA summary of the model is offered in Table 5.47.

Table 5.47 GLM ANOVA (Type III) Summary of Switch Costs and Cognitive Tasks

	Switch Costs			
	F-statistic	DF1	DF2	p-value
Intercept	1067.22	1	100	<0.0001
Language	23.98	2	100	<0.0001
Group	28.92	1	56	<0.0001
Simon Effect	2.36	1	100	0.13
BDST	3.33	1	56	0.07
Language*Group	6.86	2	100	0.00
Language*Simon Effect	28.62	2	100	<0.0001
Group*Simon Effect	4.69	1	100	0.03
Group*BDST	0.94	1	56	0.34
Language*BDST	0.52	2	100	0.60
Simon Effect*BDST	64.19	1	100	<0.0001
Language*Group*Simon Effect	22.51	2	100	<0.0001
Language*Group*BDST	1.17	2	100	0.31
Language*Simon Effect*BDST	6.88	2	100	0.00
Group*Simon Effect*BDST	2.49	1	100	0.12
Language*Group*Simon Effect*BDST	0.99	2	100	0.37

*Formula: Switch Costs ~ Language*Group*Simon Effect*BDST Scores+ (1+Simon Effect* BDST|Participant)*

The model revealed significant main effects for Group, Language, the Simon Effect, and BDST, as well as several significant interactions. Of interest were two significant three-way interactions between: (1) Language, Group, and Simon Effect scores ($F(2,100)=22.51$, $p<0.0001$) (Figure 5.21); and (2) Language, Simon Effect, and

BDST scores ($F(2,100)=6.88$, $p<0.01$).

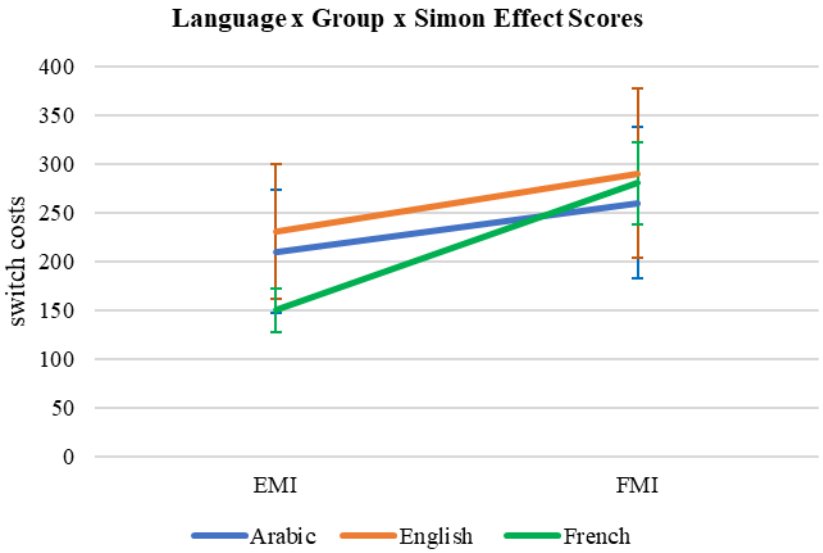


Figure 5.21 Switch Costs by Language x Group x Simon Effect score Interaction plot (error bars: 95%CI)

In order to investigate the significant Group x Language x Simon Effect interaction, simple main effects were analysed by changing the factor contrast coding for the mixed model, and *post hoc* contrasts were analysed by changing the reference (intercept) for each level of Group and Language. The results of the analysis indicated that Inhibitory Control (IC) was a significant predictor of switch costs, where a lower Simon Effect score predicted smaller switch costs, yet this was modulated by Language for each group. Specifically, Simon Effect scores predicted lower switch costs for EMI participants in English ($\beta=0.17$, $SE=0.03$, $t=5.83$, $p<0.0001$) and Arabic ($\beta=0.11$, $SE=0.02$, $t=5.27$, $p<0.0001$), but not French ($p>0.05$). There was no significant difference in the influence of IC on switch costs in Arabic and English ($p>0.05$). On the other hand, for FMI participants, Simon Effect scores predicted lower switch costs in all three languages, in French ($\beta=0.29$, $SE=0.05$, $t=5.80$, $p<0.0001$), English ($\beta=0.30$, $SE=0.05$, $t=6.02$, $p<0.0001$), and Arabic ($\beta=0.28$, $SE=0.03$, $t=5.82$, $p<0.0001$). The influence of IC on switch costs across the three languages for the FMI group was similar ($p>0.05$ for all contrasts). Furthermore, IC was a stronger predictor of switch costs in the

FMI group than the EMI group in French ($t=7.92$, $p<0.0001$), but not in English or Arabic ($p>0.05$ for both contrasts). Figure 5.22 illustrates the relationship between Simon Effect scores and switch costs of each language by group.



Figure 5.22 Inhibitory Control x Switch Costs x Language Interaction Plot (y axis: transformed Simon Effect scores, x-axis: switch costs)

Another significant interaction which was observed was the Language x Simon Effect scores x BDST scores interaction. Post-hoc analyses were performed to examine simple main effects and contrasts between the different languages, by changing the reference level for each language. The results of the analyses indicated that the interaction was driven by high BDST scores which significantly predicted smaller switch costs in Arabic (β=65.54, SE=5.34, t=12.28, p<0.0001), English (β=95.06, SE=6.72, t=14.14, p<0.0001), and French (β=44.69, SE=14.71, t=3.04, p<0.01). Furthermore, smaller Simon Effect scores predicted smaller switch costs in Arabic (β=0.08, SE=0.01, t=4.86, p<0.0001) and English (β=0.07, SE=0.01, t=5.39, p<0.0001), but not in French (p>0.05).Also, the influence of BDST scores in predicting switch costs across all languages was significantly larger than that of Simon Effect scores (t=5.98, p<0.0001).

These results indicate that smaller Simon Effects scores and higher BDST scores predicted smaller switch costs for participants in Arabic and English. In other words, participants who exhibited better inhibitory control and higher working memory capacity incurred significantly smaller costs when switching between these two languages on the TPNT. Furthermore, both cognitive measures significantly influenced switch costs, but working memory was a stronger predictor of switch cost performance. The dissociation between French switch costs and cognitive task performance highlights the role of proficiency. For both groups, proficiency in Arabic and English was high, and this was reflected in switch cost performance. IC and working memory were able to predict performance in languages in which proficiency was high. The significant relationship between IC and switch costs is in line with the predictions of ICM and parallel findings of previous research which has demonstrated a link between IC and switch costs in the L1 (but not in weaker languages) (e.g., Linck et al., 2012). Furthermore, the findings highlight the importance of working memory and attentional control in language control.

5.4.2 Cognitive Tasks and the Narration Task

RQ5B. Do inhibitory control capacities, conflict monitoring skills, and working memory capacity predict pause duration in the trilingual condition of the narration task?

In order to answer this research question, pauses in the trilingual condition were broken down by type: intra-language pauses and inter-language pauses. Intra-language pauses comprised pauses which occurred within each language and were therefore excluded from analysis, as they do not index the temporal cost of a switch. Inter-language pauses were then categorised by directionality (6 switch types), and the mean pause duration of each switch type was calculated for every participant. It should be noted, however, that inter-language and intra-language pauses were similar in terms of duration; that is, participants' pausing did not index a switch 'cost' in the trilingual condition. It is for this reason, as mentioned in the previous analyses of the narration task, that pauses from the trilingual condition were not distinguished further. However, for this research question, the distinction between switches based on directionality was deemed necessary to examine the influence of directionality on pause duration.

The directionality of switching between languages, along with the number of switches, differed across and within the two groups (see Table 5.48). However, the mean pause duration for each condition did not significantly differ for either group. This may be explained by the proposition that voluntary switching is cost-free. As a reminder, in this task, participants were asked to use the three languages at least once, yet the amount of each language used and the frequency of codeswitching were not cued. As mean pause duration was not influenced by switch directionality, mean pause durations of all switch types were not distinguished further in the analyses.

Table 5.48 Mean Number and Mean Duration (in seconds) of Trilingual Narration Task Switch Directionality

	EMI		FMI	
Switch Directionality	**Number**	**Duration**	**Number**	**Duration**
Arabic-to-English	6	5.21	3	5.29
Arabic-to-French	3	5.23	5	5.42
English-Arabic	5	5.17	4	5.37
English-French	2	5.25	4	5.31
French-Arabic	1	5.22	3	5.44
French-English	2	5.19	3	5.31

In order to assess the relationship between pause duration and cognitive task performance, transformed Simon Effect scores, Conflict Effect scores, Backward Digit-Span Task scores, Group, and interactions between these variables were entered as predictor variables in a linear mixed-effects model. Log-transformed mean pause durations in the trilingual condition were entered as the outcome variable. The model aimed to assess whether pausing for shorter periods of time before switching languages parallels RTs in language switching.

The results of the model, summarised in Table 5.49, revealed no significant main effects for any predictor except BDST scores. No other significant main effects or interactions were observed. Although all predictor variables were of interest and their inclusion was theoretically driven, the exclusion of the non-significant predictors did not significantly improve the fit of the model. Separate models for each of the cognitive predictors were also tested but revealed no significant results. Accordingly, the initial model was retained along with the only significant finding: larger working memory capacity predicted shorter pause durations by participants from both EMI and FMI groups.

Table 5.49 GLM Summary of Narration Task and Cognitive Predictors (Estimated coefficients, standard errors (SE), t-values, and p-values) Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗)

	Coefficient	SE	t-value	p-value
Intercept	16.19	5.29	3.06	0.00**
Group (EMI vs. FMI)	7.87	4.76	1.65	0.10
Simon Effect	0.97	0.80	1.21	0.23
BDST	0.98	0.46	2.09	0.04*
Group*BDST	0.74	0.52	1.42	0.15
Group*Simon Effect	0.84	0.49	1.71	0.09
Simon Effect*BDST	0.81	0.67	1.21	0.23
Group*Simon Effect*BDST	0.71	0.43	1.66	0.10

Both the significant and non-significant results of this research question offer valuable insights. The role of working memory in speech production, and particularly fluency, is once again underscored. The trilingual condition of the narration task, as stipulated by the ACH and CPM, entailed opportunistic planning and cooperative task schemas to exploit all resources available to the participants when producing utterances. Accordingly, there was no need to inhibit a language (or lexical item) or to monitor potential interference. As such, the non-significant relationship between conflict monitoring (Conflict Effect) and Inhibitory Control (Simon Effect) on the narration task may be ascribed to the trilingual context.

5.5 CHAPTER SUMMARY

This chapter examined performance on the two linguistic tasks, namely the picture-naming task and narration task. Results of these tasks, which were addressed in RQ3 were analysed and correlated. The independent variables, Language Proficiency and Language Experience, were then analysed as predictors of linguistic task performance. The chapter then concluded with the analysis of RQ5 which aimed at exploring the overlap between cognitive and linguistic task performance.

CHAPTER 6: GENERAL DISCUSSION

6.1 INTRODUCTION

While Chapters 4 and 5 offered the results and discussions of the specific research questions, the present chapter offers a general discussion of the overall findings, linking these to previous studies and to the present study's aims and theoretical framework. The chapter offers the discussions in the same order as the literature review and research questions. Cognitive performance (RQ1 and RQ2) is first discussed in terms of cognitive advantages of multilingualism and the trilingual advantage hypothesis. The following section of the chapter discusses linguistic performance (RQ3 and RQ4), where the ICM, ACH, and CPM are discussed. The overlap between cognitive and linguistic performance is then addressed (RQ5). Summaries of unexpected findings and a synthesis of the role of language proficiency and language experience in modulating cognitive and linguistic performance are offered within the overall discussions. The chapter then lists the methodological and analytical limitations of the present study, followed by a brief conclusion of the chapter.

6.2 SUMMARY AND DISCUSSION OF OVERALL FINDINGS

This project aimed at investigating cognitive and linguistic task performance of Arabic-English-French adolescent trilingual speakers who were divided into two groups based on the immersion programme in which they were enrolled. As the study acknowledged the potential influence of individual differences in language proficiency and language experience, trends in cognitive and linguistic behaviour were analysed based on both the group- and the individual-levels. That is, general group differences were investigated, but participants' cognitive and linguistic performance was also analysed irrespective of their group membership.

6.2.1 Executive Functions and the Trilingual Advantage

6.2.1.1 The Absence of the Trilingual Advantage

Participants' performance on the flanker task, the Simon task, and Backward Digit-Span task was examined with the hypothesis that FMI participants would outperform their EMI peers given the former group's high proficiency in the three languages. Contrary to predictions, participants in both groups performed similarly on all three tasks. The results of the first research question, therefore, did not lend support to the notion of a tri/multilingual advantage for the participants in this study. That is, cognitive performance did not distinguish the two groups on domain-general tasks, and post-hoc analyses of all participants based on the HPBT grouping (and even no grouping) further verified that having a higher proficiency in a third language did not confer cognitive advantages. These results, therefore, parallel those of Poarch and van Hell (2012), Vega-Mendoza et al. (2015), and Schroeder and Marian (2017) who did not observe qualitative differences between the bilingual and/or trilingual participants in their studies.

The similarity in performance by both groups, whether the distinction was based on the EMI or FMI grouping, or on having high proficiency in two or three languages, may be attributed to participants' high performance on these tasks. The study adopted more complex versions of the tasks through the manipulation of monitoring and congruence conditions in the flanker task, the use of the arrow version of the Simon task, and the use of a backward-digit span task for working memory. Despite attempting to distinguish performance through the use of these complex tasks, participants performed close to ceiling in several conditions of these tasks. This has been ascribed to the notion that the participants were at peak performance given their age (17-18 years), making differences between them undiscernible, as suggested by Bialystok (2017). It is

plausible, therefore, that a cognitive difference would have been observed if a different age group were recruited, such as in Kavé et al. (2008) who noted a trilingual advantage in older adults. Alternatively, it may be speculated that the present study's non-significant findings in terms of cognitive performance may have stemmed from the low variability between participants. Also, it is also possible that participants, even those who had low proficiency in a third language, had reached a certain threshold of proficiency beyond which no cognitive advantages could be observed. That is, the initial hypothesis which assumed that having a higher proficiency in the L3 would confer additional cognitive benefits is not supported.

While the findings of this study did not provide evidence for a trilingual advantage, it is important to emphasise that the study aimed at distinguishing an advantage between two groups of multilinguals. Typically, studies which examine a bilingual or multilingual advantage do so by comparing bi/multilingual participants to monolinguals (e.g., Morales et al., 2013, 2015; Prior & MacWhinney, 2010). However, including monolingual speakers in this study would have entailed recruiting participants from a different socio-economic status or from a different country. Doing so would have compromised the study's design by including variables which have been observed to influence (or confound) cognitive task performance (e.g., Morton & Harper, 2007). Thus, any differences in cognitive task performance would have been difficult to categorically attribute to L3 proficiency as opposed to other variables.

6.2.1.2 Cognitive Task Correlations

Another important finding which was noted in Chapter 4 was the difference in correlations between the different cognitive tasks. While Conflict Effect scores and Simon Effect scores were each correlated with BDST scores, they were not correlated with each other. As indicated in the literature review, the flanker task and the Simon task

are two measures of inhibitory control, yet a lack of convergent validity between these two measures has been observed in the literature (Paap & Sawi, 2014). The manipulation of the tasks, however, seems to have successfully created a difference in task design, thus distinguishing the two tasks in terms of the cognitive processes which they engender. That is, the use of trial manipulations in the flanker task induced a greater demand for conflict monitoring abilities, whereas the use of the arrow Simon task required increased interference suppression and response inhibition.

In light of the further analyses which examined the relationship between cognitive and linguistic task performance, a discrepancy in findings (i.e., null results) would have been observed if the flanker task was the only task of measuring cognitive performance. This gives credence to the notion that differences in findings of executive functions may stem from methodological disparities given the types—and perhaps versions—of the tasks used across studies, as suggested by Bak (2016). More importantly, however, the partial overlap(s) between the flanker task, the Simon task, and the BDST reveal that these instruments may tap into different cognitive mechanisms and processes. It is consistent with Miyake and Friedman's (2012) 'unity and diversity' proposal which suggests that cognitive processes, while overlapping, are distinguishable abilities which share an underlying mechanism that extends to working memory.

6.2.1.3 Inhibitory Control, Conflict Monitoring, and Working Memory

The manipulation of congruent-to-incongruent trials in the flanker task, as put forth by Costa et al. (2009), was intended to create contexts of different conflict monitoring demands. The participants in this study were observed to perform differently based on these manipulations, with heightened accuracy and significantly slower reaction times in the high-monitoring conditions. Accordingly, the present study suggests that participants heeded the increased monitoring demands of this condition and were capable

of detecting the high frequency of trial switching, thus producing accuracy rates similar to those in the low-monitoring condition. This explanation of a shift in attention when congruent and incongruent trials were equally distributed is reinforced by the relationship which was observed between salient cue detection scores and Conflict Effect scores. As the only predictor of Flanker task performance, self-reported attention to changes in cues by participants predicted smaller Conflict Effect scores. It may be inferred, therefore, that similar processes of directing attention to situations where a conflict needs resolving and where monitoring demands are high (in cognitive tasks) are shared in conversational situations where speakers are attuned to cues which require a switch in language. This further underscores the role which attentional control plays in tasks measuring high-level cognitive tasks. As suggested by Grundy and Timmer (2017) and Piolat et al. (2006), high working memory capacity is observed to be a reliable predictor of performance on complex cognitive tasks, including ones measuring interference suppression. Accordingly, it is possible to ascribe participants' performance on the high-monitoring condition to attentional control which was triggered by a shift in monitoring conditions. The correlation between the flanker task and BDST scores is thus further attested. This overlap, however, did not extend to further analyses which included both the Flanker and BDST. When analysing linguistic tasks, a disparity in the influence of the Conflict Effect and BDST on performance was observed. Notably, Conflict Effect scores did not predict performance on any linguistic task, whereas BDST scores predicted smaller switch costs and higher fluency. It can therefore be assumed that while an overlap between the flanker task and BDST is observed, it is only partial.

While attentiveness to cues is notably associated with increased conflict monitoring, it is important to consider the language experience of the participants and the context(s) in which they converse, as suggested by the ACH and CPM. As mentioned

earlier, Lebanon is a multilingual country which not only allows codeswitching across almost all social situations, but it encourages it. This would therefore entail a lesser need to constantly monitor one's speech or to inhibit production in a language to avoid a breakdown in communication. However, some participants in this study reported being sensitive to changes in language due to topic, an interlocutor's language proficiency, or other cues, and this has been reflected in their performance on the flanker task. The discussion in Chapter 4 suggested that the the flanker task's trial manipulations may have mirrored the cognitive processes involved in the micro-management of languages, as suggested by Hofweber et al., 2016. Accordingly, participants who are from a linguistic context where codeswitching is ubiquitous and where they report being sensitive to cues may possess enhanced conflict monitoring skills. As indicated earlier, this is contrary to Green and Li's (2014) CPM which postulates that participants from dense-codeswitching contexts are more susceptible to interference as the cognitive processes they typically engage do not merit inhibition or conflict monitoring.

The Simon Task, as evidenced by the Simon Effect scores, was a more complex cognitive task than the flanker task. More importantly, however, it was observed to predict linguistic performance whereas the flanker task failed to do so. Although the Simon task did not distinguish the two groups, it was influenced by self-reported codeswitching frequency and salient cue detection. This finding signals an important aspect of the bilingual experience which has been said to modulate executive functions, and it parallels the findings of Verreyet et al. (2016) who noted that bilingual participants who frequently codeswitched outperformed their monolingual and 'infrequent codeswitching bilingual' counterparts on cognitive task performance. This is therefore contrary to the findings of Soveri et al. (2011) and the claims of Kang and Lust (2018) which suggest that codeswitching frequency may not play a significant or direct role in

cognitive task performance. Furthermore, the interaction between salient cue detection and Simon Effect scores further elucidates the interplay and complexity of the bilingual experience. It also introduces a new facet of the bilingual experience which interacts with codeswitching frequency in influencing inhibitory control. As no previous study has explored these two variables within language experience, it is not possible to discuss these findings in terms of previous research. However, as both constructs are central to the ACH and CPM, this study offers a unique contribution in providing support for the postulations of these proposals.

Furthermore, Simon Effect scores were predicted by English proficiency scores, which was the only predictor of BDST scores as well. This finding was unexpected, as participants in both groups who had a higher proficiency score in English exhibited significantly enhanced inhibitory control performance and large working memory capacity. It is highly unlikely that the English language is vastly typologically distinct from Arabic and French or that it possesses intrinsic characteristics which enhance executive functions once high proficiency is achieved. A more sensible, though speculative, explanation for this finding is that it may be high (and specific) English language use and exposure, which correlate with English language proficiency. That is, it may have been the specific prevalence of English in the context of technology (online, computer, video games, etc.), as indicated in the LBQ, which has enhanced a specific aspect of executive functions that underlies these two cognitive measures. This finding, however, remains open for further investigation.

Performance on the cognitive tasks is discussed further in the subsequent sections of this chapter in terms of the overlap between cognitive and linguistic performance. However, to conclude this section, three important notions must be highlighted. First, despite recent calls advocating the compartmentalisation of executive functions in order

to identify the specific aspects of cognition which underlie the bi/multilingual advantage, the overlapping processes measured by the tasks render this difficult, as suggested by Bialystok et al. (2008) and Calabria et al. (2011). Furthermore, while both the Flanker and Simon tasks are said to be measures of inhibitory control, and more specifically interference suppression, the lack of correlation between the two tasks and the difference in the two tasks predicting linguistic performance further supports the notion that task type and task design may underlie the discrepancy in findings in the field. Finally, the role of attentional control, which is inherently tied to working memory, seems to play a critical role in higher-order cognitive tasks.

6.2.2 Linguistic Task Performance and Control Model Applications

6.2.2.1 Switch Costs

In addition to comparing EMI and FMI participants' performance on cognitive tasks, the present study was designed to examine the two groups' switching performance on the linguistic tasks. The picture-naming task revealed that participants responded more accurately and significantly faster on repetition than switch trials, which was expected. In terms of switch costs, EMI participants exhibited symmetrical switch costs when switching between English and Arabic, yet asymmetrical switch costs when switching into/from French. On the other hand, FMI participants exhibited symmetrical switch costs across the three languages. The hypothesis that the two groups would differ in performance on this task was confirmed and is in line with the predictions of the ICM and CPM. For EMI participants, asymmetrical switch costs indexed a different amount of inhibition for Arabic and English compared to the weaker French, leading to different reactivation costs, which are observed in larger reaction times. For FMI participants, symmetrical switching costs are assumed to reflect similar amounts of suppression and thus similar reactivation costs.

The (a)symmetry of switch costs is rooted in the assumption that language proficiency modulates switching performance, and this lends support to the basic premise of the ICM. As observed in the analyses of RQ4 (variables influencing switch cost performance), language proficiency was a significant predictor of language switching for both groups, and the findings of the present study parallel those of Meuter and Allport (1999), Costa and Santesteban (2004), Schwieter and Sunderman (2008, 2011), among others. For EMI, the symmetrical switch costs between English and Arabic and the asymmetrical switch costs between each of these languages and the much weaker French support the dominance-based model. Similarly, the FMI group's symmetrical switch costs reflect their high proficiency levels across the three languages, as measured by the OPT. This finding resonates with the post-hoc analysis of the HPBT variable, where high-proficiency in three languages, as opposed to two, predicted symmetrical switch costs.

It is imperative to reiterate the importance of objectively measuring language proficiency when drawing conclusions about whether a model or a theoretical framework is supported. As indicated in the discussion of RQ4, VFT scores did not provide an adequate measure of proficiency or dominance. Further, VFT scores in English offered an opposite result, as higher VFT scores predicted smaller switch costs for EMI participants. From the onset of this project, the VFT was approached with caution given its conflation of working memory and verbal production. Although highly advocated as a useful and simple instrument to measure proficiency (Linck et al., 2012), it has been observed as a non-valid measure of oral or general proficiency in the present study. If VFT scores were adopted as the sole measure of overall proficiency, then results of the present study would have been interpreted as offering counter-evidence to the ICM.

Another significant predictor of switch costs is language experience. For both groups, a high self-reported use of a language generally predicted larger switch costs in that language. While language use was not as significant a predictor as language proficiency in modulating switch costs, both language proficiency and language use predicted performance on the TPNT in a comparable way. This finding supports examining factors which influence performance on picture-naming tasks, among others. In terms of language exposure, for both groups only French input significantly predicted switch cost performance. This finding may be attributed to the wider range of switch costs exhibited by participants in both groups. That is, the extent to which Arabic and English switch costs differed was much smaller than that of French, and the influence of language exposure was therefore more pronounced in the French condition.

The observation that high self-reported codeswitching frequency also significantly predicted smaller switch costs contributes to the literature on the role of individual differences in language experience in modulating linguistic task performance. Though the influence of codeswitching frequency on switch costs has been beset with controversy, this study offers support that participants who codeswitch more frequently incur smaller switch costs. This further lends support to the notion that the constant management of the two languages when both are coactivated enhanced linguistic performance. These findings resonate with those of Prior & Gollan (2011) and Hartanto and Yang (2016) but are contrary to those reported by Yim and Bialystok (2012), among others. The findings of the present study also lend support to the ACH and CPM which postulate that participants who codeswitch frequently are accustomed to benefiting from open-control and cooperative task schemas and would therefore outperform bilinguals who do not engage in frequent switching due to being in single- or dual-language contexts. Although the present study does not (and cannot) distinguish participants in

terms of their linguistic contexts, it offers support to the notion that frequency of codeswitching is a significant variable to consider in examining language control.

Furthermore, self-reported salient cue detection was also observed to be a significant predictor of switch costs, where higher attention to cues predicted significantly smaller switch costs. The implications of this finding are twofold: First, this result lends support to the notion that constant monitoring of one's and others' speech (among other cues) enhances language control. Further, this finding lends *partial* support to the ACH and CPM which suggest that participants in the dual-language context would possess heightened conflict monitoring skills given the demands of their linguistic environment. While the present study offers support to the significance of salient cue detection as a cognitive process underlying language control, it challenges the ACH and CPM's predictions regarding the restrictive nature of the linguistic context in which salient cue detection is observed. Salient cue detection may be regarded as being akin to conflict monitoring, or to entail similar processes of constantly monitoring and noticing cues. As such, it may not necessarily be limited to dual-language contexts.

In addition to language proficiency and language experience modulating switch cost performance, Simon Task and BDST scores were also observed to be significant predictors of linguistic task performance. Lower Simon Effect scores predicted smaller switch costs, yet this was constrained by language proficiency. That is, higher inhibitory control skills predicted more efficient language switching when switching was performed between languages of high (or perhaps, similar) proficiency. In line with Linck et al.'s (2012) findings, therefore, the relationship between IC and switch costs may be limited in scope. On the other hand, higher working memory capacity predicted smaller switch costs across the three languages for both groups. It could be argued then that while IC is more susceptible to language proficiency because it is narrower in scope,

working memory is agnostic to proficiency as it reflects a more general cognitive function. It is also possible that there is greater reliance on inhibitory control when the language is more dominant, whereas working memory is constantly activated throughout a switching task. In other words, and as evidenced by the interaction between the two cognitive tasks in predicting switch costs, both inhibitory control and working memory underlie language control. However, working memory may evoke global control through monitoring and managing task-schema engagement and disengagement, whereas inhibitory control may play a more localised and transitory role as it varies across trials given the difference in proficiency in each trial's language. This is partially congruent to Gollan and Ferreira's (2009) proposal that balanced bilinguals may exhibit IC, whereas unbalanced bilinguals exhibit localised, momentary IC. In the present study, however, the suggestion is that for both groups, IC may act as a more localised process to control language production, whereas it is complemented by working memory which plays the more global role in regulating the task schemas.

6.2.2.2 Fluency

Fluency, as measured by pause frequency and pause duration, has revealed several findings regarding EMI and FMI participants' linguistic performance on a task exceeding one-word utterances. The two groups varied significantly across the single-language and trilingual conditions of this task, where EMI participants were significantly more fluent in the English-only, Arabic-only, and trilingual conditions of this task, whereas FMI participants were significantly more fluent in the French-only condition of this task. This was observed for both pause frequency and pause durations. Aside from the latter finding, the predictions of this task were in opposition to the results obtained from the analyses. That is, FMI participants were expected to outperform their EMI peers in Arabic, French, and the trilingual conditions. This prediction was based on the higher

proficiency of FMI participants in French and the more ubiquitous use of and exposure to Arabic by this group, as reported in the LBQ. Regarding the trilingual condition, it was assumed that FMI participants who were accustomed to switching between the three languages would incur fewer and shorter pauses than EMI participants in this condition. However, the findings of the trilingual condition revealed a multifarious pattern of codeswitching, where EMI participants used French only minimally, thus attenuating the cost of switching in their speech. It is possible to interpret the findings from the perspective of switch costs analysis, where the high(er) proficiency of FMI participants across the three languages resulted in increased response latencies in the narration task. However, this interpretation seems counter-intuitive given the cost-free nature of codeswitching in naturalistic conversation. It also offers counter evidence to the analyses of predictor variables, where increased proficiency predicted fewer and shorter pauses.

In terms of the predictors of fluency, both pause frequency and pause durations were significantly predicted by language proficiency scores for EMI and FMI participants. However, there was only one exception, which is that of Arabic proficiency scores which did not predict the longer, more frequent pauses of FMI participants in the Arabic-only condition. The general hypothesis that increased language proficiency predicts enhanced performance on the narration task, and specifically fluency, was still upheld. This finding supports the suggestion of Green and Li (2014) who propose examining fluency and measures such as the narration task which simulate naturalistic speech.

However, as indicated in the discussion of switch costs, the analysis and interpretations of the findings are contingent on the way in which proficiency is operationalised and measured. The VFT task, for instance, did not predict fluency on the narration task. It was only after splitting the data that VFT scores discerned pause

frequency in French only, and this is generally attributed to the task detecting differences when they are striking. In other words, the VFT is unable to detect subtle variation in participants' pause frequency. Despite the different simulations and model structuring, VFT scores did not predict pause durations. This further consolidates the argument about the limited utility of the VFT as a measure of proficiency.

Similar to VFT scores' relation to fluency, the HPBT distinction did not prove to be a robust variable in distinguishing fluency performance by the two groups. Specifically, HPBT analyses only revealed differences between highly proficienct bilinguals and trilinguals in the French-only condition, where EMI participants generally incurred larger disfluencies as they constituted the majority of the 'highly proficient in two languages' group. However, it was not sensitive to measure the difference between the two groups in the Arabic condition.

As for language experience, language use significantly predicted fluency in the single-language conditions, where increased use of a language was related to fewer and shorter pauses when speaking in that language. However, as with VFT and HPBT, the difference in performance in the French-only condition was more evident given the variation among participants in fluency on this task. Similarly, language exposure was a significant predictor of pause frequency in the English and French conditions and for pause duration in the French-only condition. These findings indicate that language experience is a significant predictor of fluency to a lesser extent than language proficiency.

An important observation is participants' performance on the trilingual condition of the narration task, which was highly fluent for both groups in terms of pause frequency and pause duration. Taken together, pause frequency and pause duration indexed that while the two groups differed in fluency in the single-language mode, they were both

highly fluent in the trilingual condition. Furthermore, an unexpected finding was the similarly long duration of Arabic pausing in both groups, which is a reflection of the notion that Lebanese speech is mixed with the other languages. This supports the notion that in such a linguistic context, speakers are more fluent when they can codeswitch freely, whereas speaking in their 'mother tongue' may be more taxing. The influence of the immersion context in which the participants have engaged on a regular basis can be observed in the low frequency and duration of pauses in English for EMI and in French for FMI participants. Although the topics of the slides were of a non-academic nature, as universal themes were selected, the fluency with which participants spoke in the immersion language indicate the relative ease with which they can produce speech on a variety of topics. This is further corroborated by the scores of the OPT which involved open-ended questions relating to numerous subjects.

More importantly, however, is that high self-reported codeswitching frequency and high salient cue detection significantly predicted fewer pauses in the trilingual condition. This finding provides evidence for the important role of these two predictors in modulating language control in naturalistic speech data. Furthermore, it overlaps with the findings of switch costs, which were similarly predicted by codeswitching frequency and salient cue detection. However, pause duration was not predicted by either variable. Rather, it was predicted by BDST scores, where higher working memory capacity predicted shorter pause durations. Inhibitory control, whether measured by the flanker task or the Simon task, did not predict pause durations. However, this may be explained by the notion that in the trilingual condition of this task, there is no need to inhibitor or monitor speech, and thus IC may not necessarily predict fluency in a condition where participants may use all the linguistic resources available to them to produce utterances. This supports the ACH and CPM in terms of the open-control mechanisms where task

schemas cooperate by allowing elements from different languages to be selected. Furthermore, the high fluency in the trilingual condition compared to the monolingual conditions supports Hlavac's (2011) and Kroll et al.'s (2015) observation that voluntary switching may be cost-free. Contrasted with the monolingual conditions, the trilingual condition was observed to be the least taxing, along with the immersion language of each group. Accordingly, it provides support for the ACH postulation that voluntary switching is cost-free for this study's participants.

6.2.2.3 Switch Cost and Fluency Overlap

Although the overlap in performance on the picture-naming task and the narration task was discussed at length in Chapter 5, it is important to note this significant correlation in terms of the methodological and theoretical underpinnings of the present study. The overall findings of the task revealed that for both EMI and FMI groups, participants who exhibited the shortest pauses in the trilingual condition of the narration task had the fastest RTs on switch trials in the picture-naming task. Additionally, RTs on repetition trials in the TPNT for each group's immersion language correlated with the participants' pause duration in the single-language condition of that language in the narration task. Though the overlap between the two tasks is partly supported, these findings are highly promising in terms of extending the claims of the ACH to be tested in a different language control paradigm. However, it seems that the overlap is observed only for the conditions of the narration task where pause duration is short. The approximation of fluency to response latencies in the other conditions may be obfuscated by the magnitude of response latencies. The lack of correlation between the two tasks ultimately stems from the different nature of the tasks, since the TPNT does not require high proficiency in naming the cued pictures, whereas the narration task involves

numerous facets of linguistic knowledge and assumes a minimum threshold of proficiency.

6.2.3 Applicability of the Theoretical Models

The main aim of the study was to explore the applicability of language control models in a linguistic context where codeswitching is ubiquitous and where participants speak three typologically distant languages. While the ICM has generally been supported in the literature, the ACH and CPM have not been thoroughly investigated. The findings of the present study are commensurate with the predictions of the ICM, and they offer partial support for the two recent models. To recapitulate, the present study provided evidence for the ICM in terms of switch cost asymmetry, and the findings suggest that the ICM is applicable in simultaneous multilinguals whose proficiency is generally high. Furthermore, the ACH and CPM have been generally supported, as individual differences in language proficiency and language experience have been observed to modulate both cognitive and linguistic performance. However, the present study has revealed the limitations of the ACH and CPM in accounting for participants who typically engage in the three different linguistic contexts (single-language, dual-language, and dense-codeswitching). As the participants in this study produced different types of code-switches and were not limited to insertions, alternations, or congruent lexicalisations, the CPM and ACH fall short in accounting for participants from a linguistically diversified background.

This study has also explored the applicability of the ACH and CPM in examining the processes underlying language control. The design of the study simulated the single-language and dense-codeswitching conditions which the ACH and CPM propose and in which the participants of this study regularly engage. In their immersion programmes, participants are usually in a single-language mode (of their immersion language),

whereas in social settings, they typically codeswitch. Accordingly, it is important to examine the linguistic performance of participants who are experienced in conversing in both types of linguistic contexts. However, the performance of participants on linguistic and domain-general tasks cannot be categorically ascribed to the enhanced cognitive processes which are involved in one specific contexts. Furthermore, it is difficult to ascertain whether the participants in this study have qualitatively different types of cognitive processes than the ones outlined by the ACH and CPM, but it is possible that they combine these processes as they constantly engage in the different contexts. Accordingly, for participants whose linguistic contexts are variable, the models fail to demarcate the specific aspects of the linguistic contexts and/or cognitive processes which underlie performance on language and cognitive tasks. The applicability of these models may therefore be more justified in contexts where participants engage in distinct linguistic contexts.

The ACH and CPM predict that participants who come from dense-codeswitching contexts may struggle to produce monolingual utterances, and this has been observed for the EMI and FMI participants in different languages (French for EMI and Arabic for FMI). However, the difficulty in producing monolingual utterances appears to be modulated by language proficiency to some extent. Furthermore, the models propose that participants who come from contexts where codeswitching is frequent would easily and fluidly switch between languages. This has indeed been observed in the trilingual condition of the narration task. The cognitive processes underpinning the linguistic production of participants could not be disentangled for the present study's sample. The models suggest that in the single-language context, inhibitory control is global and constant, whereas in the dual-language context, conflict monitoring and a narrowing of attention is in highest demand. Finally, in the dense codeswitching context, the models

suggest there is no demand for inhibitory control, leaving speakers to being susceptible to interference from activated lexical items from the different languages. This could not be tested or disentangled by the present study.

Furthermore, the models propose that control processes vary according to context, such as participants' speaker ecology or the general codeswitching behaviour of the speech community. This finding was supported in the present study as participants were generally attentive to the constant switching of languages. The CPM also predicted that voluntary codeswitching, which could be measured by fluency, does not incur a conversational cost. Although the present study did not employ a conversational task, the narration task bridged the psycholinguistic one-word utterances and natural speech production. In the trilingual condition, participants incurred the least amount disfluency, and this therefore offers support for the CPM. However, Green and Li (2014) and Green and Abutalebi (2013) assume a high level of proficiency by the speakers, but as evidenced in this study, disfluencies occurred in the single-language mode. Despite having global, constant control to ensure that only one language is used, the influence of language proficiency plays a critical role in speech production, even if speakers are habitually accustomed to separating their languages.

As mentioned in the critique of the models in Chapter 2, it is not only the switching between languages which enhances cognitive and linguistic performance—it is also the constant switching between linguistic contexts which may confer benefits. The present study's finding that salient cue detection was a significant predictor of cognitive and linguistic task performance suggests that this may indeed be measurable. The ACH and CPM suggest that salient cue detection is only (or, more conservatively, most highly) present in dual-language contexts when a speaker must heed potential cues which alert them to the need to switch. However, the present study suggests that salient cue detection

may be closely tied to the linguistic contexts at a macro-level. That is, salient cue detection may be perceived as an 'external' element of language control which alerts participants to potential language switches, even in situations where trilingual switching is permissible. The ICM and CPM account for this external element—the supervisory attentional system in the ICM, and the conceptual-intentional system in the CPM—which manages the language task schemas to ensure that the communicative goal is achieved. Green and Abutalebi (2013) suggest that repeated use leads to a given configuration becoming a habit of control which is then triggered by contextual cues. That is, the management of task schemas enhances cognitive function, and this is reiterated by Poarch and van Hell (2012) who suggest that the constant management of two languages leads to enhanced general attentional control mechanisms.

In the present study, participants who reported being highly attentive to external and internal cues in language choice and codeswitching exhibited significantly smaller switch costs than participants who reported not being attuned to such cues. This finding suggests that participants who constantly monitor their and their interlocutors' speech incur smaller costs when switching languages. This may be due to anticipatory action and an enhanced ability to proactively and reactively manage their language task schemas in both a top-down and a bottom-up manner. The management of their task schemas is achieved through conflict monitoring. This also highlights the role of attentional control in the constant monitoring of incoming information and managing the task schemas. It is therefore possible that language control models are equally dependent on inhibition and conflict monitoring, which is regulated by attentional control. This is also tied to task shifting in Miyake et al.'s (2000) framework. Future research may wish to measure salient cue detection as a variable just as codeswitching frequency is measured, and examine whether this construct predicts performance on task-shifting

tasks such as those in Prior and MacWhinney's (2010) study. This also aligns with Hernandez et al. (2012) and Cockcroft et al.'s (2017) claim that working memory plays a central role in bilingualism, leading to enhanced top-down shifting and facilitating higher-order processes in complex cognition.

.

6.3 METHODOLOGICAL AND ANALYTICAL LIMITATIONS

While the findings have addressed the aims of the study, there are some methodological and analytical limitations to note. First, while the participants in this study were selected specifically to examine the potential of a trilingual advantage on both cognitive and linguistic tasks, the lack of a significant finding may stem from not having a monolingual or bilingual group of participants. However, as noted earlier, recruiting monolingual or even bilingual speakers would entail introducing several covariates or confounding variables (such as socioeconomic status) which may have influenced the findings. Furthermore, the participants in this study belong to a very specific stratum of the Lebanese community; they may be representative of a population consisting of high socioeconomic children but not that of their age-matched peers throughout the Lebanese population. This, therefore, limits the generalisability of the findings.

Additionally, while the study criticises the use of subjective measures in assessing language proficiency, it relied on self-reported data for language experience. In order to overcome this caveat, the study thoroughly investigated the constructs in 'language experience' by asking participants to report on their language experience in a range of contexts with different speakers and across various communication mediums. However, given the general trends observed in the self-reports and the complexity of the analyses, the study did not probe further into potential idiosyncrasies in language

experience. In order to account for variability among participants and to avoid making generalisations about participants, the use of (generalised) linear mixed-effects models and inclusion of full random maximal structures allowed for a more accurate representation of the findings. Despite the attempt to mitigate the influence of self-reported data, the use of objective measures of codeswitching frequency and salient cue detection would have provided a more accurate description of participants' linguistic and attentional/monitoring behaviour. Furthermore, the study did not distinguish between participants' codeswitching in terms of whether it involved two or three languages—a distinction which may have influenced in the findings. Accordingly, a more fine-grained analysis of the frequency of codeswitching which includes the number of language switches would allow for a more nuanced analysis of the results.

In terms of the trilingual picture-naming task, the inclusion of mixing cost analysis by blocking languages would have allowed for additional data to be obtained and for further parallels to be drawn between the narration task in the single-language conditions and the picture-naming task in the blocked-language conditions. However, as the focus was that of switching, and given the complexity and diversity of tasks in the present study, the picture-naming task was not blocked by language. Furthermore, the analysis of n-2 effects in the picture-naming task may have offered evidence for (or against) persisting inhibition, yet this was not controlled for in the initial design of the experiment given the controversy regarding the findings of this effect. However, it remains a limitation to be acknowledged.

In terms of the narration task, operationalising fluency using only pauses facilitated the contrasts obtained between this task and the picture-naming task. However, a more fine-grained analysis of the narration task, which includes repetitions, switch directionality, and the amount of language(s) used measured through

syllables/utterance, may offer a more precise reflection of fluency. Furthermore, it is possible that by focusing only on pauses, what was construed as a disfluency was actually a covert error that does not reach the level of articulation but is reflected in a pause. De Jong (2018:246) gives an example of a speaker saying 'Go to the…left' (ellipses reflecting a pause) due to the planned—but unarticulated—speech 'Go to the right'. Accordingly, a pause in this case reflects participants' constant monitoring of their speech to ensure correct utterances are made, as opposed to a disfluency.

Furthermore, as noted in the analysis of the narration task, a potential confounding variable is the difference in language use among participants in the trilingual condition. As participants were able to switch freely using the three languages in different combinations in the trilingual task, pause frequency and duration in the trilingual condition might have been influenced by the amount of language and combination types. That is, participants may have used only the two languages they were more proficient in, leading to the high fluency in the trilingual condition. While this condition was designed to elicit free codeswitching, not accounting for this variability is an analytical limitation.

6.4 CHAPTER SUMMARY

This chapter presented a discussion of the overall findings of the study, offering discussions of the theoretical models and highlighting similarities between the present study's findings and those of previous studies. Cognitive and linguistic task performance, along with the variables which predicted both, were discussed. A critical assessment of the theoretical models was offered, followed by a brief discussion of the methodological and analytical limitations of the study.

CHAPTER 7: CONCLUSION

7.1 INTRODUCTION

The final chapter of this thesis summarises the general conclusions of the study and offers a brief discussion of the pedagogical implications based on the main findings. The chapter then highlights the study's contributions to the field and concludes by offering suggestions for future research.

7.2 CONCLUSIONS AND PEDAGOGICAL IMPLICATIONS

The present study aimed to demarcate the specific aspects of bilingualism which pertain to advanced cognitive processing. The main findings of the study identified codeswitching frequency and salient cue detection as significant individual differences which modulate performance on cognitive tasks. Working memory was indexed as an underlying mechanism which permeates both inhibitory and conflict monitoring processes and is argued to be equally important in language control models and potentially the locus of linguistic and cognitive advantages. Moreover, the findings highlight the importance of taking into account participants' language experience given their speaker ecologies and individual differences in language proficiency and cognitive functions.

While the present study aimed to investigate whether a trilingual advantage could be observed when comparing two groups of multilingual speakers who vary in their L3 proficiency, the results indicated that having a high(er) proficiency in a third language did not confer cognitive advantages for this sample. This finding was attributed to the possibility that the participants had a reached a certain proficiency threshold beyond which an advantage was no longer discernible. Other potential explanations were also offered such as task type and task complexity. The study further examined participants' switch costs across their three languages, and the findings were in support of the

inhibitory control models (ICM and CPM) which were adopted in this study. Symmetrical switch costs were observed among participants who were balanced, and asymmetrical switch costs were exhibited by participants with a low L3. Furthermore, switching into the more dominant language(s) resulted in larger response latencies, which is in line with the predictions of the models.

The findings in this study revealed a positive relationship between working memory capacity and language proficiency. A recent interest in attentional control and working memory has surged in psycholinguistic research, and the implications of the amenability of WM to training are important to note. The role of working memory has been investigated in second-language learning, with results lending support to the notion that participants with larger WM capacities perform better on a variety of linguistic tasks, such as vocabulary learning, writing, and resolving ambiguities in reading comprehension (Atkins & Baddeley, 1998; Linck et al., 2015; Linck & Weiss, 2011; Michael & Gollan, 2005)(. While a positive relationship between WM and language attainment has generally been observed, it is difficult to ascertain the extent to which WM training would benefit different populations. As most studies examining WM training involve younger participants and/or late L2 learners, it is possible that these populations may benefit from training WM alongside L2 learning more than populations similar to the one selected for the present study (i.e., young-adult simultaneous multilinguals). The prominence of the role of WM in language attainment and the effects of training WM on the different facets of language learning across populations should be further explored. If WM training is revealed to be consistently conducive to L2 (or L3, etc.) learning across a variety of contexts, then perhaps educators may wish to consider allocating a specific amount of time to training this cognitive function.

Another important observation made in the study was the high overlap between language exposure and language use. The amount of input in each language highly matched the amount of language use, as reported by the participants. However, the role of language exposure and language use extends beyond the input-output element of language acquisition. The sample in this study were simultaneous multilinguals whose exposure to the three languages varied across contexts. Although differences in language exposure did not predict cognitive performance, it is perhaps the exposure to the three languages which has conferred cognitive advantages to this sample as a whole.

Furthermore, language exposure played a similar, though limited, role in predicting switch costs to that of language proficiency. While this was true for French only, where participants in both groups who had reported receiving more input in French incurred larger switch costs in French, it is an important observation. It was also observed in the narration task, as increased input in French predicted shorter pauses by participants in both groups in the French-only condition. Accordingly, it is important to examine or control for the different variables which influence linguistic (and cognitive) task performance.

Although tangential to the central aims of the study, an important observation is the role to which the supposed native language of this sample, Arabic, has been relegated. Participants in both immersion programmes struggled to produce narrations in Arabic, and this has been explained by the abundance of loan words and switches in Lebanese Arabic which makes speaking in an Arabic-monolingual mode arduous. However, participants were most fluent in their immersion languages, suggesting that the immersion language is the most dominant language for these participants. In an earlier study by Zantout (2014), Lebanese children and their parents had highly favourable opinions about English being the more dominant language (instead of Arabic). While the

linguistic context of this particular stratum of the Lebanese population does not denounce the lower proficiency in Arabic, this may have ramifications in other contexts where the native language is perceived as having a higher status than an immersion language. It may also have negative implications in other segments of the Lebanese community which differ from the one selected for the present study. Further, ideological beliefs and notions of self and identity, along with attitudes towards language(s) and codeswitching, play a pivotal role in whether this is accepted or not in other speech communities.

Another important implication involves the selection of the immersion programme given the ambient languages in a multilingual society. Although the study did not observe a trilingual advantage in FMI participants over their EMI peers, a linguistic difference was observed in speaking a second 'foreign' language in the narration task, as well as in switch costs in the picture-naming task. FMI participants received a substantially large amount of English-language input from their environment(s). Accordingly, they did not struggle when speaking in English, whereas EMI participants found speaking in French taxing. Accordingly, the selection of the immersion programme may be influenced by the amount of input or exposure participants receive outside the classroom setting. The extent of the influence of the immersion language in determining overall proficiency should take into account external exposure.

Furthermore, although the present study did not find evidence for a trilingual advantage, learning an additional language and multilingualism, needs to be highlighted as a non-detrimental phenomenon. Similarly, switching between languages involves different cognitive processes to the ones in monolingual conditions. Accordingly, a

positive attitude towards codeswitching is important as it is a typical feature of multilingualism and does not denote a lack of language proficiency.

7.3 CONTRIBUTIONS TO THE FIELD

The study aimed at expanding the boundaries of current codeswitching research by developing a single-language and trilingual narration task which was analysed through a psycholinguistic approach and mapped onto psycholinguistic-task production (the TPNT). Although the narration task did not elicit a fully accurate simulation of multilinguals' naturalistic codeswitching in conversational settings, it allowed for parallels to be drawn between laboratory-based one-word utterances and longer speech production. The study has also aimed to synthesise findings from both production tasks within the theoretical language and cognitive control frameworks. This was conducted by taking into account other variables which influence both cognitive and linguistic task performance, to offer a more holistic picture of the processes and other aspects of the bilingual experience which influence production. Indeed, individual variation can inform current debates on the link between bilingualism and executive functions, code-switching practices, and language control. This has been done in line with recent suggestions calling for a more thorough examination of speaker profiles (Bialystok, 2009; de Bruin, 2019; Green & Abutalebi, 2013). Specifically, the study has highlighted the role of working memory and conflict monitoring alongside inhibitory control, by offering support for both the ICM and CPM. It has also identified codeswitching frequency and salient cue detection as significant predictors of both linguistic and cognitive task performance.

The study has also examined the applicability of the ACH and CPM, and the much researched and supported ICM, in a multilingual context where three distant and understudied languages (in combination) are examined. In so doing, it offered partial

support to these proposals while underscoring the need to develop models of speech production which apply to multilingual contexts. This is especially necessary as multilingualism is on the rise and has already been reported to supersede monolingualism.

Correlating cognitive task performance, which is not typically done, further elucidated the discrepancy in findings in the literature, especially in terms of providing support for a bi/multilingual advantage. The study has also tested the applicability of the 'Monitoring Cost' as proposed by Costa et al. (2009) and provided counter-evidence to its utility among proficient multilinguals. It also highlighted the lack of significant results when performance on the flanker task is examined as a predictor of cognitive and linguistic task performance and perhaps as the sole measure of a bi/multilingual advantage.

Another task which was examined and discussed in depth was the Verbal Fluency Task. In the present study, VFT scores did not predict or correlate with switch costs, BDST scores, Simon Effect scores, or Conflict Effect scores. Accordingly, the VFT may be useful linguistic task and a measure of lexical robustness, but extending its narrow scope to measuring language proficiency and/or balance (or dominance) is critiqued by the study.

A final contribution of the present project is the use of mixed-effects models in analysing the data. As discussed in Chapters 4 and 5, MEMs overcome the limitations of other analytical approaches, such as ANOVAs, by controlling for participant- and item-based variance within the same explanatory model(s). As such, MEMs can cater for the complex hierarchical nature of the data and reveal more accurate results. The use of MEMs has been recently recommended for the analysis of psycholinguistic data (e.g., Linck et al., 2012), especially when the data comprises and/or combines non-normally

distributed continuous, ordinal, and categorical variables. While the analytical strategy is strongly advocated, it is important to note that its application is at times exploratory, as in the structuring and fitting of random effects. Furthermore, the inclusion of numerous variables may lead to model overfitting. Additionally, as indicated by Linck (2016), effect sizes are not typically reported in MEM analyses, as effect sizes (and statistical power calculations) can only be applied to models which have one fixed effect and two random effects or to models where the number of observations per participant is very high (>1000s) (see Brysbaert & Stevens, 2018, for a detailed discussion). Nevertheless, MEMs have demonstrated to be far more robust in analysing quantitative data and are strongly recommended due to their purported superiority to traditional, limited methods such as ANOVAs.

7.4 SUGGESTIONS FOR FUTURE RESEARCH

The present study suggests that the lack of a trilingual advantage may be attributed to the age group selected for the study. Future research may recruit younger and older participants, where a bilingual advantage has generally been observed, to examine the potential influence of high proficiency in the third language. If replicability of the study were to be conducted with the same or a similar age group as the one in the present study, then developing more complex tasks is strongly recommended as a means of discerning a significant difference between the participants under investigation.

It is also recommended that future studies examine codeswitching types in order to further examine the predictions of the ACH and CPM. It would be of interest then to examine whether producing a specific type of switch reflects or is reflected by the specific cognitive processes outlined by the theoretical models. In examining the types of switches produced by participants, it is suggested that naturalistic conversations are recorded and analysed to obtain naturalistic, uncued codeswitching data.

Further studies may also wish to examine whether inhibitory control and monitoring load differ across languages by examining cross-linguistic influence and the number of errors made in productive tasks, whether they are one-word utterances or longer narration or conversational utterances. Furthermore, while the present study incorporated the examination of a novel variable, salient cue detection, which was observed to be a significant predictor of both cognitive and linguistic tasks, it relied on self-reported data to measure this construct. Future research may wish to objectively measure this construct through behavioural (or neuroimaging) tasks. A final recommendation for future research pertains to the analytical strategy adopted in studies similar to the present project. In analysing quantitative data, especially those obtained from tasks such as the ones employed in this study, researchers may wish to adopt mixed-effects models to ensure that generalisations and overall findings are more robust by including and controlling for variation in participants and task items. Doing so would not only offer more accurate analyses and interpretations of the data but also potentially reduce the inconsistency in findings which may stem from the different analytical methods used.

References

Abutalebi, J., & Green, D. W. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, *20*(3)*,* 242–275.

Abutalebi, J., & Green, D. W. (2008) Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and Cognitive Processes*, *23*(4), 557-582.

Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition*, *19*(4), 689-698.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In Parzen, E., Tanabe, K., and Kitagawa, G. (Eds.), *Selected Papers of Hirotugu Akaike* (pp. 199–213). Springer, New York.

Allport, D.A., Styles, E.A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.). Attention and Performance XV: Conscious and nonconscious information processing (pp.422-452). Cambridge, MA: MIT Press.

Anderson, J. A., Chung-Fat-Yim, A., Bellana, B., Luk, G., & Bialystok, E. (2018). Language and cognitive control networks in bilinguals and monolinguals. *Neuropsychologia*, *117*, 352-363.

Astheimer, L. B., Berkes, M., & Bialystok, E. (2015). Differential Allocation of Attention During Speech Perception in Monolingual and Bilingual Listeners. *Language, Cognition and Neuroscience*, *31*(2), 196-205.

Atkins, P. W., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics*, *19*(4), 537-552.

Auer, P. (1984). *Bilingual conversation*. Amsterdam: John Benjamins.

Ayneto, A., & Sebastian-Galles, N. (2017). The influence of bilingualism on the preference for the mouth region of dynamic faces. *Developmental Science*, *20*(1), published online.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3,* 12–28.

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

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Cambridge, MA: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York, NY: Cambridge University Press.

Bak, T. (2016). Cooking pasta in La Paz: Bilingualism, bias and the replication crisis. *Linguistic Approaches to Bilingualism*, *6*(5), 699-717.

Bak, T., Nissan, J., Allerhand, M., & Dreary, I. (2014). Does Bilingualism Influence Cognitive Aging? *Annals of Neurology*, *75*(6), 959-963.

Baker, C., & Athanasopoulos, P. (2011). Bilingualism and Cognition. In C. Baker

(Ed.)*Foundations of Bilingual Education and Bilingualism*, 5th Edition (pp. 138-162). Bristol, UK: Multilingual Matters.

Barac, R., Moreno, S., & Bialystok, E. (2016). Behavioral and electro-physiological differences in executive control between monolingual and bilingual children. *Child Development, 87*(4), 1277–1290.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Bates, D. (2010). lme4: Mixed-effects modeling with R. Springer. Retrieved from: http://lme4.r-forge.r-project.org/book .

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv:1506.04967*. Retrieved from: https://arxiv.org/abs/1506.04967.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using S4 classes. R package version 1.1–7. Retrieved from: http://CRAN. R-project.org/package=lme4.

Baus, C., Branzi, F. M., & Costa, A. (2015). On the mechanism and scope of language control in bilingual speech production. In J. W. Schwieter (Ed.), *The Cambridge Handbook of bilingual processing*. (pp. 508-526). (Cambridge handbooks in language and linguistics). Cambridge: Cambridge University Press.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development, 70*(3), 636–644.

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 60*(1), 68-79.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*(1), 3–11.

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, *143*(3), 233-262.

Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, *122*(1), 67–73.

Bialystok, E., Craik, F. I. M., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage, 24*(1)*,* 40-49.

Bialystok, E., Craik, F., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*(2)*,* 290-303.

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 859-875.

Bialystok, E., Craik, F. I. & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–250.

Bialystok, E., Martin, M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, *9*(1), 103-119.

Bizri, F. (2013). Linguistic Green Lines in Lebanon. *Mediterranean Politics, 18*(3), 444-459.

Blain-Brière, B., Bouchard, C., & Bigras, N. (2014). The role of executive functions in the pragmatic skills of children age 4–5. *Frontiers in Psychology*, *5*(240), 1-14.

Blanco-Elorrieta, E., & Pylkkänen, L. (2017). Bilingual language switching in the laboratory versus in the wild: The spatiotemporal dynamics of adaptive language control. *Journal of Neuroscience*, *37*(37), 9022-9036.

Blumenfeld, H.K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in Stimulus-Stimulus inhibition. *Bilingualism: Language and Cognition*, *17*, 610-629.

Boersma, P., & Weeninik, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.40. Retrieved 30 May 2018, from http://www.praat.org/.

Bonfieni, M., Branigan, H. P., Pickering, M. J., & Sorace, A. (2019). Cognitive control in bilinguals: Effects of language experience and individual variability. *Bilingualism: Language and Cognition*, First View, 1-12.

Botvinick, M., Carter, C. S., Braver, T. S., Barch, D. M., & Cohen, J. D. (2001). Conflict- monitoring and Cognitive Control. *Psychological Review*, *108*(3), 624-652.

Botvinick, M., Cohen, J., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*(12), 539–546.

Branzi, F. M., Calabria, M., Boscarino, M. L., & Costa, A. (2016). On the overlap between bilingual language control and domain-general executive control. *Acta Psychologica*, *166*, 21-30.

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113.

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, *1*(1), 9-29.

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. (2002). Immature frontal lobe contributions to cognitive control in children: evidence from fMRI. *Neuron*, *33*(2), 301-311.

Calabria, M., Branzi, F. M., Marne, P., Hernandez, M., Costa, A. (2015). Age-related effects over bilingual language control and executive control. *Bilingualism: Language and Cognition, 18*(01), 65-78.

Calabria, M., Hernandez, M., Branzi, F. M., Marne, P., Costa, A. (2012). Qualitative differences between bilingual language control and executive control: Calabria, M., Marne, P., Romero-Pinel, L., Juncadella, M., Costa, A. (2014). Losing control of your languages: A case study. *Cognitive Neuropsychology*, *31*(3), 266–286.

Calcagno, V., & de Mazancourt, C. (2010). glmulti: An R package for easy automated model selection with (generalized) linear models. *Journal of Statistical Software*, *34*(12), 1-29.

Calvo, N., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*(3), 278–288.

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11*(2)*,* 282–298.

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism, 7*(1), 71-89.

Clyne, M. (1997). Some of the things trilinguals do. *The International Journal of Bilingualism, 1*(2)*,* 95-116.

Clyne, M. (2003). *Dynamics of Language Contact*. Cambridge: Cambridge University Press.

Cockcroft, K., Wigdorowitz, M., & Liversage, L. (2017). A multilingual advantage in the components of working memory. *Bilingualism: Language and Cognition*, *22*(1): 1-15.

Coderre, E.L., & van Heuven, W.J.B. (2014). The effect of script similarity on executive control in bilinguals. *Frontiers in Psychology*, *5*(1070), 1-16.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). New York, NY: Routledge/Taylor & Francis Group.

Collet, D. (2002). *Modelling Binary Data*. London: Chapman & Hall/CRC Press.

Colomé, A. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language, 45*(4)*,* 721–736.

Comrie, B. (1991). On the importance of Arabic for general linguistic theory. In B. Comrie & M. Eid (Eds.) *Perspectives on Arabic Linguistics III: Papers from the Third Annual Symposium on Arabic Linguistics* (pp. 3-30). Amsterdam/Philadelphia: John Benjamins.

Costa, A. (2005). Lexical Access in Bilingual Production. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 289-307). New York, NY, US: Oxford University Press.

Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language- specific? Further evidence from Spanish–English and English–Spanish bilinguals. *Bilingualism: Language and Cognition, 2*(3)*,* 231–244.

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*(1), 59-86.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*(2), 135-149.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*(4), 491-511.

Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1057-1074.

Cowan, N. (1999). An embedded-processes model of working memory. In Miyake, A., & Shah, P. (Eds.). *Models of working memory: Mechanisms of active maintenance and executive control.* New York, NY: Cambridge University Press.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*, 323-338.

Cowan, N. (2016). *Working Memory Capacity: Classic Edition*. London: Routledge.

Crowe, S. F. (1998). Decrease in performance on the verbal fluency test as a function of time: Evaluation in a young healthy sample. *Journal of Clinical and Experimental Neuropsychology*, *20*(3), 391-401.

Daneman, M., & Hannon, B. (2007). What do working memory span tasks like reading span really measure? In N. Osaka, R. H. Logie, & M. D'Esposito (Eds), *The cognitive neuroscience of working memory* (pp. 21-42). Oxford & New York, NY: Oxford University Press.

De Baene, W., Duyck, W., Brass, M., & Carreiras, M. (2015). Brain circuit for cognitive control is shared by task and language switching. *Journal of Cognitive Neuroscience, 27*(9), 1752-1765.

de Bot, K. (2004). The multilingual lexicon: Modelling Selection and Control. *The International Journal of Multilingualism, 1*(1)*,* 17–32.

de Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioural Sciences*, *9*(3), 33-46.

de Bruin, A., & Della Sala, S. (2016). The importance of language use when studying the neuroanatomical basis of bilingualism. *Language, Cognition and Neuroscience*, *31*(3), 335–339.

de Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2018). Voluntary language switching: When and why do bilinguals switch between their languages? *Journal of Memory and Language*, *103*, 28-43.

Declerck, M., Grainger, J., Koch, I., & Philipp, A. M. (2017). Is language control just a form of executive control? Evidence for overlapping processes in language switching and task switching. *Journal of Memory and Language, 95,* 138-145.

De Groot, A., & Christoffels, I. (2006). Language Control in bilinguals: Monolingual tasks and simultaneous interpreting. *Bilingualism: Language and Cognition, 9*(2), 189 – 201.

De Groot, A., & Starreveld, P. (2015). Parallel language activation in bilinguals' word production and its modulating factors: A review and computer simulations. In J.W. Schwieter (Ed.), *The Cambridge handbook of bilingual processing* (pp. 389-415). Cambridge, UK: Cambridge University Press.

De Jong, N.H. (2018). Fluency in second language testing: insights from different disciplines, *Language Assessment Quarterly*, *15*(3), 237-254.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*(4), 533-557.

Deuchar, M., Muysken, P., & Wang, S. L. (2007). Structured variation in codeswitching: towards an empirically based typology of bilingual speech patterns. *International Journal of Bilingual Education and Bilingualism*, *10*(3), 298-340.

Dong, Y., & Xie, Z. (2014). Contributions of second language proficiency and interpreting experience to cognitive control differences among young adult bilinguals. *Journal of Cognitive Psychology*, *26*(5), 506-519.

Duñabeitia, J.A., Hernàndez-Cabrera, J.A., Estévez, A., Macizo, P., Anton, E., Fuentes, L.J. (2014) The Inhibitory Advantage in Bilingual Children Revisited: Myth or Reality? *Experimental Psychology, 61*(3), 1-18.

Eager, C., & Roy, J. (2017). Mixed effects models are sometimes terrible. *arXiv preprint arXiv:1701.04858*. Retrieved from https://arxiv.org/abs/1701.04858.

Edwards, J. (1994). *Multilingualism.* London: Routledge.

Edwards, J. (2019). Multilingual Individuals. In D. M. Singleton & L. Aronin (Eds.), *Twelve Lectures on Multilingualism* (pp.135-162). Bristol, UK: Multilingual Matters.

Edwards, M., & Dewaele, J. M. (2007). Trilingual conversations: A window into multicompetence. *International Journal of Bilingualism, 11*(2), 221-242.

Engel de Abreu, P. M., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism enriches the poor: Enhanced cognitive control in low-income minority children. *Psychological Science*, *23*(11), 1364-1371.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1)*,* 19 –23.

Engle, R. W., & Kane, M. J. (2004). Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. In B. H. Ross (Ed.), *The*

psychology of learning and motivation: Advances in research and theory, Vol. 44, pp. 145-199). New York, NY, US: Elsevier Science.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Attention, Perception, & Psychophysics*, *16*(1), 143-149.

Falotico, R., & Quatto, P. (2015). Fleiss' kappa statistic without paradoxes. *Quality & Quantity, 49*(2), 463-470.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340-347.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191.

Fay, M.P., & Shaw, P.A. (2010). Exact and Asymptotic Weighted Logrank Tests for Interval Censored Data: The interval R package. *Journal of Statistical Software, 36*(2):1-34. Retrieved from http://www.jstatsoft.org/v36/i02/.

Festman, J. & Mosca, M. (2016). Influence of preparation time on language control. In J. W. Schwieter (Ed.) *Cognitive Control and Consequences of Multilingualism*, (pp. 145-171). Amsterdam/Philadelphia: John Benjamin.

Festman, J., & Münte, T. F. (2012). Cognitive control in Russian-German bilinguals. *Frontiers in Psychology*, *3*(115), 1-7.

Festman, J., Poarch, G. J., & Dewaele, J. M. (2017). *Raising multilingual children*. Bristol UK: Multilingual Matters.

Festman, J., Rodriguez-Fornells, A., Muente, T.F. (2010). Individual differences in control of language interference in late bilinguals are mainly related to general

executive abilities. *Behavioral and Brain Functions*, *6*(5), 1-12.

Festman, J., & Schwieter, J.W. (2015). Behavioural measures of language control: Production and comprehension. In J. W. Schwieter (Ed.) *The Cambridge Handbook of bilingual language processing* (pp. 527–547). Cambridge: Cambridge University Press.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. 4th Ed. London: Sage Publications, Ltd.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R.* London: Sage Publications, Ltd.

Filippi, R., D'Souza, D., & Bright, P. (2018). A developmental approach to bilingual research: The effects of multi-language experience from early infancy to old age. *International Journal of Bilingualism*, 1-13.

Fink, A., & Goldrick, M. (2015). Pervasive benefits of preparation in language switching. *Psychonomic Bulletin and Review*, *22*(3)*,* 808–814.

Finkbeiner, M., Gollan, T. H., & Caramazza, A. (2006). Lexical access in bilingual speakers: What's the (hard) problem? *Bilingualism: Language and Cognition*, *9*(2), 153-166.

Fisk, J. E., & Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *Journal of Clinical and Experimental Neuropsychology*, *26*(7), 874-890.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*(3), 613-619.

Frank, A. F. (2011). VIF.mer Package in R. Retrieved from https://github.com/aufrank/R-hacks/blob/master/mer-utils.R

Gafaranga, J. (2005). Demythologising language alternation studies: Conversational structure vs. social structure in bilingual interaction. *Journal of Pragmatics*, *37*(3), 281-300.

Gardner-Chloros, P. (2009). *Code-switching*. New York: Cambridge University Press.

Gardner-Chloros, P., McEntee-Atalianis, L., & Paraskeva, M. (2013). Code-switching and pausing: an interdisciplinary study. *International Journal of Multilingualism*, *10*(1), 1-26.

Gathercole, V. C., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Viñas Guasch, N., . . . Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology, 5*(11), 1-14.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Gollan, T. H., & Ferreira, V. S. (2009). Should I stay or should I switch? A cost–benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 640-665.

Gollan, T.H., Montoya, R.I., & Werner, G.A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, *16*(4), 562-576.

Grainger, J. (1993). Visual word recognition in bilinguals. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 11–26). Amsterdam/Philadelphia: John Benjamins.

Green, D.W. (1986). Control, Activation, and Resource: A Framework and a Model for the Control of Speech in Bilinguals. *Brain and Language, 27*(2), 210–223.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2)*,* 67–81.

Green, D. W. (2017). Trajectories to third-language proficiency. *International Journal of Bilingualism*, *21*(6), 718–733.

Green, D. W. (2018). Language control and code-switching. *Languages*, *3*(8), 1-16.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515-530.

Green, D.W. & Li, W. (2014). A control process model of code-switching. *Language, Cognition and Neuroscience*, *29*(4), 499-511.

Grosjean, F. (1982). *An introduction to bilingualism.* Cambridge, MA: Harvard University Press.

Gross, M., & Kaushanskaya, M. (2015). Voluntary language switching in English–Spanish bilingual children. *Journal of Cognitive Psychology*, *27*(8), 992-1013.

Grundy, J.G., Timmer, K. (2017). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research, 33*(3), 325–340.

Gullifer, J. W., Chai, X. J., Whitford, V., Pivneva, I., Baum, S., Klein, D., & Titone, D. (2018). Bilingual experience and resting-state brain connectivity: Impacts of L2 age of acquisition and social diversity of language use on control networks. *Neuropsychologia*, *117*, 123-134.

Guo, T., Liu, H., Misra, M., & Kroll, J.F. (2011). Local and global inhibition in bilingual word production: fMRI evidence from Chinese-English bilinguals. *Neuroimage*, *56*(4), 2300–2309.

Guo, G., & Zhao, H. (2000). Multilevel modeling for binary data. *Annual Review of Sociology, 26,* 441–462.

Hakuta, K. (1986). *Mirror of language: The debate on bilingualism*. New York, NY: Basic Books.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23-34.

Hammarberg, B. (2001). Roles of L1 and L2 in L3 production and acquisition. In J. Cenoz, B. Hufeisen & U. Jessner (Eds.), *Cross-linguistic Influence in Third Language Acquisition* (pp. 21–41). Bristol: Multilingual Matters.

Hartanto, A., & Yang, H. (2016). Disparate bilingual experiences modulate task-switching advantages: A diffusion-model analysis of the effects of interactional context on switch costs. *Cognition, 150,* 10–19

Hernandez, M., Costa, A., Fuentes, L.J., Vivas, A.B., & Sebastian-Galles, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Bilingualism: Language and Cognition*, *13*(3), 315 – 325.

Hernández, M., Costa, A., & Humphreys, G. W. (2012). Escaping capture: bilingualism modulates distraction from working memory. *Cognition*, *122*(1), 37-50.

Hernandez, M., Martin, C.D., Barcelo, F., Costa, A. (2013). Where is the bilingual advantage in task-switching? *Journal of Memory and Language*, *69*(3), 257-276.

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin and Review, 18*(4)*,* 625-658.

Hlavac, J. (2011). Hesitation and monitoring phenomena in bilingual speech: A consequence of code-switching or a strategy to facilitate its incorporation? *Journal of Pragmatics*, *43*(15), 3793-3806.

Hoffmann, C., & Stavans, A. (2007). The evolution of trilingual codeswitching from infancy to school age: The shaping of trilingual competence through dynamic language dominance. *International Journal of Bilingualism, 11*(1), 55-72.

Hofweber, J., Marinis, T., & Treffers-Daller, J. (2016). Effects of dense code-switching on executive control. *Linguistic Approaches to Bilingualism*, *6*(5), 648-668.

Hoshino, N., & Kroll, J. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, *106*(1), 501-511.

Huensch, A., & Tracy-Ventura, N. (2017). Understanding second language fluency behaviour: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics, 38*(4), 755-785.

Humphrey, A.D., & Valian, V.V. (2012). Multilingualism and cognitive control: Simon and flanker task performance in monolingual and multilingual young adults. *Paper presented at the 53rd Annual Meeting of the Psychonomic Society*. Minneapolis, MN.

Jacobson, R. (Ed.). (2011). *Codeswitching Worldwide. II* (Vol. 126). Berlin: Walter de Gruyter.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4)*,* 434–446.

Joseph, J. (2004). *Language and identity: National, ethnic, and religious.* New York: Palgrave.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*, 601-625.

Jylkkä, J., Lehtonen, M., Lindholm, F., Kuusakoski, A., & Laine, M. (2018). The relationship between general executive functions and bilingual switching and monitoring in language production. *Bilingualism: Language and Cognition*, *21*(3), 505-522.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), 809-854.

Kang, C., & Lust, B. (2018). Code-switching does not predict Executive Function performance in proficient bilingual children: Bilingualism does. *Bilingualism: Language and Cognition*, *22*(2), 1-17.

Kave, G., Eyal, N., Shorek, A., & Cohen-Mansfield, J. (2008). Multilingualism and cognitive state in the oldest. *Psychology of Aging, 23*(1), 70–78.

Kerrigan, L., Thomas, M., Bright, P., & Filippi, R. (2017). Evidence of an advantage in visuo-spatial memory for bilingual compared to monolingual speakers. *Bilingualism: Language and Cognition*, *20*(3), 602-612.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin, 136*(5), 849-874.

Kleinman, D., & Gollan, T. H. (2016). Speaking Two Languages for the Price of One: Bypassing Language Control Mechanisms via Accessibility-Driven Switches. *Psychological Science*, *27*(5), 700–714.

Koch, I., Gade, M., Schuch, S., & Philipp, A. M. (2010). The role of inhibition in task switching - A review. *Psychonomic Bulletin and Review, 17*(1)*,* 1–14.

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, *25*(5), 497–514.

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current Directions in Psychological Science*, *23*(3), 159–163.

Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: evidence for inhibitory processes. *Acta Psychologica*, *128*(3), 416–430.

Kroll, J. F., & Gollan, T. H. (2014). Speech planning in two languages: What bilinguals tell us about language production. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), *Oxford library of psychology: The Oxford handbook of language production* (pp. 165-181). New York, NY, US: Oxford University Press.

Kroll J. F., Gullifer, J., McClain, R., Rossi, E., & Cruz Martin, M. (2015). Selection and control in bilingual comprehension and production. In Schwieter, J. W. (Ed.), *Cambridge Handbook of Bilingual Processing* (pp. 485-507). New York: Cambridge University Press.

Kupisch, T., & Weijer, J. V. D. (2016). The role of the childhood environment for language dominance: A study of adult simultaneous bilingual speakers of German and French. In C. Silva-Corvalán & J. Treffers-Daller (Eds.) *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 174-194). Cambridge: Cambridge University Press.

Kutas, M., Moreno, E.M., Wicha, N.Y.Y. (2009). Code Switching and the Brain. In B. Bullock & A.J. Toribio (eds.), *The Cambridge Handbook of Linguistic Code Switching* (pp. 1–22). New York, NY: Cambridge University Press.

Leeser, M., & Sunderman, G. (2016). Methodological Issues of Working Memory Tasks for L2 Processing Research. In Y. Yilmaz, D. O. Jackson, & G. Granena (Eds.), *Cognitive Individual Differences in Second Language Processing and Acquisition* (pp. 89-104), [Bilingual Processing and Acquisition 3]. Amsterdam/Philadelphia: John Benjamins.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *3*, 387-417.

Linck, J. A. (2016). Analyzing individual differences in second language research. In G. Granena, D. O. Jackson, Y. Yilmaz (Eds.), *Cognitive Individual Differences in Second Language Processing and Acquisition*, (pp. 105-128), [Bilingual Processing and Acquisition 3]. Amsterdam/Philadelphia: John Benjamins.

Linck, J., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*(1), 185-207.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*(4), 861-883.

Linck, J.A., Schwieter, J.W., & Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition*, *15*(3), 651-662.

Linck, J. A., & Weiss, D. J. (2011). Working memory predicts the acquisition of explicit L2 knowledge. In L. Sanz. *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism*, (pp. 101-114). Washington, DC: Georgetown University Press.

Liu, C., Jiao, L., Wang, Z., Wang, M., Wang, R., & Wu, Y. (2018). Symmetries of bilingual language switch costs in conflicting versus non-conflicting contexts. *Bilingualism: Language and Cognition*, 22(3), 624-636.

Liu, H., Rossi, S., Zhou, H., & Chen, B. (2014). Electrophysiological evidence for domain-general inhibitory control during bilingual language switching. *PLoS One*, *9*(10), 1-10.

Locke, A., Ginsborg, J., & Peers, I. (2002). Development and disadvantage: implications for the early years and beyond. *International Journal of Language & Communication Disorders*, *37*(1), 3-15.

Long, J. S., & Freese, J. (2006). *Regression Models for Categorical Dependent Variables Using Stata*. 2nd ed. College Station, TX: Stata Press.

Logan, G. D., & Bundesen, C. (2003). Clever homunculus: Is there an endogenous act of control in the explicit task-cuing procedure? *Journal of Experimental Psychology: Human Perception and Performance*, *29*(3), 575-599.

Lu, C. H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review*, *2*(2), 174-207.

Luk, G. & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605–621.

Luk, G., De Sa, E., Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, *14*(4), 588-595.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19-40.

Macnamara, J. (1967). The bilingual's linguistic performance: a psychological overview. *Journal of Social Issues, 23*(2)*,* 58–77.

Madrazo, A. R., & Bernardo, A. B. (2012). Are three languages better than two? Inhibitory control in trilinguals and bilinguals in the Philippines. *Philippine Journal of Psychology*, *45*(2), 225-246.

Madrazo, A. R., & Bernardo, A. B. (2018). Measuring Two Types of Inhibitory Control in Bilinguals and Trilinguals: Is There a Trilingual Advantage? *Psychological Studies*, *63*(1), 52-60.

Marian, V., Blumenfeld, H. K., Mizrahi, E., Kania, U., & Cordes, A. K. (2013). Multilingual Stroop performance: Effects of trilingualism and proficiency on inhibitory control. *International Journal of Multilingualism*, *10*(1), 82-104.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *11*(1), 81–93.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H. & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, *22*(3), 276-282.

McKendry, M. & Murphy, V.A. (2011). A comparative study of listening comprehension measures in English as an additional language and native English-speaking primary school children. *Evaluation and Research in Education, 24*, 17–40.

Meisel, J.M. (1989). Early differentiation of languages in bilingual children. In K. Hyltenstam & L. Obler, (Eds.), *Bilingualism across the lifespan: Aspects of*

acquisition, maturity, and loss, (pp. 13-40). Cambridge: Cambridge University Press.

Mertler, C., & Reinhart, R. V. (2016). *Advanced and multivariate statistical methods: Practical application and interpretation: Sixth edition*. Taylor and Francis.

Meuter, R., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*(1)*,* 25–40.

Michael, E. B., & Gollan, T. H. (2005). Being and becoming bilingual. *Handbook of bilingualism: Psycholinguistic approaches*, (pp. 389-407). New York, NY: Oxford University Press.

Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language*, *67*(1), 224-237.

Miyake, A. & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49-100.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York, NY: Cambridge University Press.

Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of experimental child psychology*, *114*(2), 187–202.

Morales, J., Gómez-Ariza, C. J., & Teresa Bajo, M. (2013). Dual mechanisms of cognitive control in bilinguals and monolinguals. *Journal of Cognitive Psychology*, *25*(5), 531-546.

Morales, J., Yudes, C., Gomez-Ariza, C.J., & Teresa Bajo, M. (2015). Bilingualism modulates dual mechanisms of cognitive control: Evidence from ERPs. *Neuropsychologia*, *66,* 157-169.

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, *10*(6), 719-726.

Mosca, M. (2019). Trilinguals' language switching: A strategic and flexible account. *Quarterly Journal of Experimental Psychology*, *72*(4), 693-716.

Mosca, M. & Clahsen, H. (2016). Examining language switching in bilinguals: The role of preparation time. *Bilingualism: Language and Cognition, 19*(2), 415-424

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing.* Cambridge: Cambridge University Press.

Myers-Scotton, C. (2006). *Multiple voices: An introduction to bilingualism.* Cornwall: Blackwell Publishing.

Nashef, H. A. (2013) 'أهلاً', hello and bonjour: a postcolonial analysis of Arab media's use of code switching and mixing and its ramification on the identity of the self in the Arab world. *International Journal of Multilingualism, 10*(3), 313-330.

Nieuwenhuis, R., Grotenhuis, M., & Pelzer, B. (2017). Weighted Effect Coding for Observational Data with wec. *The R Journal, 9*(477), 477-485.

Noble, K.G., Norman, M.F., & Farah, M.J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental Science, 8*(1), 74–87.

Olsthoorn, N. M., Andringa, S., & Hulstijn, J. H. (2014). Visual and auditory digit-span performance in native and non-native speakers. *International Journal of Bilingualism*, *18*(6), 663-673.

Paap, K. R. (2014). The role of componential analysis, categorical hypothesising, replicability and confirmation bias in testing for bilingual advantages in executive functioning. *Journal of Cognitive Psychology*, *26*(3), 242–255.

Paap, K. R., Darrow, J., Dalibar, C., & Johnson, H. A. (2015). Effects of script similarity on bilingual advantages in executive control are likely to be negligible or null. *Frontiers in Psychology*, *5*(1539), 1-3.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232–258.

Paap, K.R., Johnson, H.A., & Sawi, O. (2015). Should the search for bilingual advantages at executive functioning continue? *Cortex, 74,* 305-314.

Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R. & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology*, *29*(2), 89–112.

Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology*, *5*(962), 1-15.

Peal, E., & Lambert, W. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: General and Applied, 76*(27), 1-23.

Philipp, A. M., Gade, M., & Koch, I. (2007). Inhibitory processes in language switching? Evidence from switching language-defined response sets. *European Journal of Cognitive Psychology, 19*(3)*,* 395–416.

Philipp, A. M., Kalinich, C., Koch, I., & Schubotz, R. I. (2008). Mixing costs and switch costs when switching stimulus dimensions in serial predictions. *Psychological Research, 72*(4)*,* 405–414.

Philipp, A. M., & Koch, I. (2006). Task inhibition and task repetition in task switching. *European Journal of Cognitive Psychology, 18*(4)*,* 624–639.

Philipp, A., & Koch, I. (2009). Inhibition in language switching: What is inhibited when switching between languages in naming tasks? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1187–1195.

Pinheiro, J., & Bates, D. (2006). Mixed-effects models in S and S-PLUS. Switzerland: Springer Science & Business Media.

Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology*, *19*(3), 291-312.

Poarch, G. J., & Bialystok, E. (2015). Bilingualism as a model for multitasking. *Developmental Review, 35,* 113–124.

Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *113*(4), 535–551.

Posner, M.I. (2012). Imaging attention networks. *Neuroimage, 61*(2), 450-456.

Prior, A., Goldwasser, N., Ravet-Hirsh, R., & Schwartz, M. (2016). Executive functions in bilingual children: Is there a role for language balance? In J. W.

Schwieter (Ed.) *Cognitive Control and Consequences of Multilingualism*, (pp. 323-350). Amsterdam/Philadelphia: John Benjamin.

Prior, A., & Gollan, T.H. (2011). Good language-switchers are good task-switchers: evidence from Spanish-English and Mandarin-English bilinguals. *Journal of the International Neuropsychological Society, 17*(4), 682 – 691.

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition, 13*(2), 253-262.

Revesz, A., Ekiert, M., & Torgersen, E. (2016). The Effects of Complexity, Accuracy, and Fluency on Communicative Adequacy in Oral Task Performance. *Applied Linguistics*, *37*(6), 828-848.

Ritchie, W.C., & Bhatia, T.K., (2013). Social and psychological factors in language mixing. In T.K. Bhatia & W.C. Ritchie, (Eds.), *The Handbook of Bilingualism and Multilingualism,* (2nd ed, pp. 375-390). Oxford: Blackwell Publishing.

Rodriguez-Fornells, A., Kraemer, U.M., Urbano, L., Festman, J., & Muente, T.F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology*, 2(388): 1-15.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology*, *124*(2), 207-231

Rosselli, M., Ardila, A., Lalwani, L., N., & Velez-Uribet, I. (2016). The effect of language proficiency on executive functions in balanced and unbalanced Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *19*(3), 489–503.

Rossiter, M., Derwing, T., & Jones, V. (2008). Is a Picture Worth a Thousand Words? *TESOL Quarterly, 42*(2), 325-329.

Rothman, J., & Niño-Murcia, M. (2008). Three brothers, Tres lenguas, Tre identitá linguistiche? A case of English, Spanish, Italian childhood trilingualism. In M. Niño-Murcia & J. Rothman, (Eds.), *Linguistic identity and bilingualism in Different Hispanic Contexts.* Amsterdam/Philadelphia: John Benjamins.

Runnqvist, E., Strijkers, K., Alario, X-F., & Costa, A. (2012). Cumulative semantic interference is blind to language: Implications for models of bilingual speech production. *Journal of Memory and Language, 66*(4), 850-869.

Runnqvist, E., Strijkers, K., & Costa, A. (2014). Bilingual word access. In M. Goldrick, V. Ferreira, & M. Miozzo, (Eds.), *The Oxford handbook of language production*, (pp. 182-198). Oxford: Oxford University Press.

Schroeder, S. R., & Marian, V. (2017). Cognitive consequences of trilingualism. *International Journal of Bilingualism*, *21*(6), 754-773.

Schwieter, J. W., & Ferreira, A. (2016). Effects of cognitive control, lexical robustness, and frequency of codeswitching on language switching. In J. W. Schwieter (Ed.) *Cognitive Control and Consequences of Multilingualism*, (pp. 193-216). Amsterdam/Philadelphia: John Benjamin.

Schwieter, J.W., & Sunderman, G. (2008). Language switching in bilingual speech production: In search of the language-specific selection mechanism. *The Mental Lexicon*, *3*(2), 214-238.

Schwieter, J.W., & Sunderman, G. (2009). Concept selection and developmental effects in bilingual speech production. *Language Learning*, *59*(4), 897-927.

Schwieter, J. W., & Sunderman, G. (2011). Inhibitory control processes and lexical access in trilingual speech production. *Linguistic Approaches to Bilingualism*, *1*(4), 391-412.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. London: Routledge.

Shaaban, K., & Ghaith, G. (1999). Lebanon's language-in-education policies: From bilingualism to trilingualism. *Language Problems and Language Planning, 23*(1), 1-16.

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*(772), 1-10.

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2015). Working memory capacity and the scope and control of attention. *Attention, Perception, & Psychophysics, 77*(6), 1863–1880.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420.

Simon, J. R., & Rudell, A. P. (1967). Auditory SR compatibility: the effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*(3), 300.

Silva-Corvalan, C., & Treffers-Daller, J. (2016). *Language Dominance in Bilinguals: Issues of Measurement and Operationalization.* Cambridge: Cambridge University Press.

Skehan, P. 2003. Task-based instruction. *Language Teaching, 36*, 1–14

Sorace, A., & Serratrice, L., 2009. Internal and external interfaces in bilingual language development: beyond structural overlap. *International Journal of Bilingualism, 13*(2), 195–210.

Sorge, G. B., Toplak, M. E., & Bialystok, E. (2017). Interactions between levels of attention ability and levels of bilingualism in children's executive functioning. *Developmental Science*, *20*(1), 1-28.

Soveri, A., Rodriguez-Fornells, A., Laine, M. (2011). Is there a relationship between

language switching and executive functions in bilingualism? Introducing a within- group analysis approach. *Frontiers in Psychology*, *2*(183), 1-8.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174-215.

Stavans, A. (1992). Sociolinguistic factors affecting codeswitches produced by trilingual children. *Language, Culture and Curriculum*, *5*(1), 41-53.

Suleiman, Y. (2003). *The Arabic language and national identity*. Edinburgh: The Cromwell Press.

Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition*, *28*(3), 387-422.

Tavakoli, P., & Foster, P. (2008). Task Design and Second Language Performance: The Effect of Narrative Type on Learner Output. *Language Learning*, *58*, 439-473.

Tavakoli, P., & P. Skehan. 2005. Strategic planning, task structure, and performance testing. In R. Ellis (Ed.) *Planning and Task-performance in a Second Language* (pp. 239–273). Amsterdam/Philadelphia: John Benjamins.

Timmer, K., Calabria, M., & Costa, A. (2019). Non-linguistic effects of language switching training. *Cognition*, *182*, 14-24.

Titone, D., Gullifer, J., Subramaniapillai, S., Rajah, N., & Baum, S. (2017). History-inspired reflections on the bilingual advantages hypothesis. In E. Bialystok and M. D. Sullivan (Eds.), *Growing Old with Two Languages: Effects of Bilingualism on Cognitive Aging* (pp. 265-295), [Studies in Bilingualism *53*]. Amsterdam/Philadelphia: John Benjamins.

Treffers-Daller, J. (2009). CS and transfer: an exploration of similarities and differences.

In: Bullock, B.E. & Almeida, J.T., eds. (2009) *The Cambridge Handbook of Linguistic Codeswitching* (pp.58-74). Cambridge: Cambridge University Press.

Treffers-Daller, J. (2016). Language Dominance: The construct, its measurement and operationalisation. In C. Silva-Corvalan, J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp.174-195). Cambridge: Cambridge University Press.

Tsai, N., Au, J., & Jaeggi, S. M. (2016). Malleability of working memory and implications for second language acquisition. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in L2 processing and acquisition*, (pp. 69-88). Amsterdam/Philadelphia: John Benjamins.

Tse, C. S., & Altarriba, J. (2012). The effects of first-and second-language proficiency on conflict resolution and goal maintenance in bilinguals: Evidence from reaction time distributional analyses in a Stroop task. *Bilingualism: Language and Cognition*, *15*(3), 663-676.

Tse, C. S., & Altarriba, J. (2014). The relationship between language proficiency and attentional control in Cantonese-English bilingual children: evidence from Simon, Simon switching, and working memory tasks. *Frontiers in Psychology*, *5*(954), 1-18.

Valian, V. (2015). Bilingualism and cognition: A focus on mechanisms. *Bilingualism: Language and Cognition, 18*(11), 47–50.

van Assche, E., Duyck, W., & Brysbaert, M. (2013). Verb processing by bilinguals in sentence contexts. *Studies in Second Language Acquisition*, *35*(2), 237–259.

van Heuven, W., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language conflict in the bilingual brain. *Cerebral Cortex*, *18*, 2706-2716.

Venables, W. N., & Ripley B. D. (2002). Modern Applied Statistics with S. Fourth Edition. New York, NY: Springer, New York. Available from http://www.stats.ox.ac.uk/pub/MASS4.

Vega-Mendoza, M., West, H., Sorace, A., & Bak, T. H. (2015). The impact of late, non-balanced bilingualism on cognitive performance. *Cognition*, *137*, 40-46.

Verhoef, K., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition*, *110*(1), 84-99.

Verreyt, N., Woumans, E., Vandelanotte, D., Szmalec, A., & Duyck, W. (2016). The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition*, *19*(1), 181–190.

Wang, Y., Xue, G., Chen, C., Xue, F., & Dong, Q. (2007). Neural bases of asymmetric language switching in second-language learners: An ER-fMRI study. *NeuroImage*, *35*(2), 862-870.

Yang, S., & Yang, H. (2016). Bilingual effects on deployment of the attention system in linguistically and culturally homogeneous children and adults. *Journal of Experimental Child Psychology, 146,* 121–136.

Yang, S., Yang, H., & Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism: Language and Cognition, 14*(3)*,* 412–422.

Yim, O., Bialystok, E. (2012). Degree of conversational code-switching enhances verbal task-switching in Cantonese-English bilinguals. *Bilingualism: Language and Cognition, 15*(4)*,* 873–883.

Yoshida, H., Tran, D. N., Benitez, V., & Kuwabara, M. (2011). Inhibition and

adjective learning in bilingual and monolingual children. *Frontiers in Psychology, 2*(210), 1-14.

Zantout, M. (2014). An investigation of the codeswitching behaviour of trilingual children in Lebanon. Unpublished Master's Thesis, University of Oxford.

Zentella, A. C. (1999). *Growing up bilingual: Puerto Rican children in New York.* Oxford: Blackwell Publishers, Ltd.

Appendix A

Language Background Questionnaire

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries'education.ox.ac.uk www.education.ox.ac.uk

Questionnaire

Bridging Sociopragmatic and Psycholinguistic Theories on Trilingual Codeswitching

Dear Participant,

Thank you for choosing to participate in this study and for taking the time to answer the following questions about the use of different languages. Your answers will not be shared or discussed with anyone other than me and my supervisor, and you do not have to provide your name.

I will be present while you answer the questions, so if there is anything you are not sure about, you may ask me. Once you have finished answering the questionnaire, please hand it back to me.

Thank you very much for your time and for volunteering in this project. Your participation is greatly appreciated.

Sincerely,
May Zantout

May Zantout
Department of Education
15 Norham Gardens, Oxford, UK, OX2 6PY
E-mail: may.zantout@lmh.ox.ac.uk

Supervisor: Prof. Victoria Murphy
Department of Education
15 Norham Gardens, Oxford, UK, OX2 6PY
E-mail: victoria.murphy@education.ox.ac.uk

Section I: General Information

Name: _____

A. Please check the box(es) that apply.

1. Gender

☐ Male ☐ Female

2. Age

☐ 11 ☐ 12 ☐ 17 ☐ 18

3. Grade

☐ 6th ☐ 12th

4. Section

☐ English ☐ French

5. First Language

☐ Arabic ☐ English ☐ French ☐ Other

*If Other, please specify _____

6. Second Language

☐ Arabic ☐ English ☐ French ☐ Other

*If Other, please specify _____

7. Third Language

☐ Arabic ☐ English ☐ French ☐ Other

*If Other, please specify _____

8. Other Language(s), if any: _____

9. Nationality(ies)

☐ Lebanese ☐Other _____

B. Please answer the following questions:

1. Have you read the invitation letter for this project? Yes / No

2. Are both your parents Lebanese? Yes / No

3. Have you always been enrolled at [Name of School Deleted] (since Nursery)? Yes / No

4. Have you taken any language courses at a formal institution, with a private tutor, at a language center, or online (e.g. an app)? Yes / No

   If yes, please indicate the language(s) taken, the duration of the course(s), the name of the place/medium of language learning (e.g. name of language center), etc.

Language(s)	Course Duration	Name of Place/Medium	Beginning Date	End Date	Level Started	Level Attained

5. Are you right-handed? Yes / No

6. Are you colorblind? Yes / No

7. Do you have any visual or mental impairments? Yes / No

8. Have you lived in or travelled to other countries for longer than three weeks? Yes / No

If yes, please indicate the country/ies, your length of stay, and the language(s) you used on a daily basis.

Country	Language(s)	Length of Stay	Year(s) of travel

9. School Grades: Indicate your final average for this year in:

Arabic

English

French

Section II: Interpersonal Language Use

A. Languages at Home

1. **Which language(s) does your mum speak at home?**	**All the time**	**Most of the time**	**Half the Time**	**A little**	**Never**
Arabic	1	2	3	4	5
English	1	2	3	4	5
French	1	2	3	4	5
Other (please specify)	1	2	3	4	5

2. **Which language(s) does your dad speak at home?**	**All the time**	**Most of the time**	**Half the Time**	**A little**	**Never**
Arabic	1	2	3	4	5
English	1	2	3	4	5
French	1	2	3	4	5
Other (please specify)	1	2	3	4	5

3. **Which language(s) do you use when you talk with your mom?**	**All the time**	**Most of the time**	**Half the Time**	**A little**	**Never**
Arabic	1	2	3	4	5
English	1	2	3	4	5
French	1	2	3	4	5
Other (please specify)	1	2	3	4	5
When talking to your mom, how frequently do you mix or switch languages?	1	2	3	4	5

4. **Which language(s) do you use when you talk with your dad?**	**All the time**	**Most of the time**	**Half the Time**	**A little**	**Never**
Arabic	1	2	3	4	5
English	1	2	3	4	5
French	1	2	3	4	5
Other (please specify)	1	2	3	4	5
When talking to your dad, how frequently do you mix or switch languages?	1	2	3	4	5

5. **Which language(s) do you use when you talk with your brothers and sisters?**	**All the time**	**Most of the time**	**Half the Time**	**A little**	**Never**
Arabic	1	2	3	4	5
English	1	2	3	4	5
French	1	2	3	4	5
Other (please specify)	1	2	3	4	5
When talking to your siblings, how frequently do you mix or switch languages?	1	2	3	4	5

6. **Which language(s) do you use when you talk with your grandparents?**	**All the time**	**Most of the time**	**Half the Time**	**A little**	**Never**
Arabic	1	2	3	4	5
English	1	2	3	4	5
French	1	2	3	4	5
Other (please specify)	1	2	3	4	5
When talking to your grandparents, how frequently do you mix or switch languages?	1	2	3	4	5

7. **Overall, what language do you hear the most in your home?**	**All the time**	**Most of the time**	**Half the Time**	**A little**	**Never**
Arabic	1	2	3	4	5
English	1	2	3	4	5
French	1	2	3	4	5
Other (please specify)	1	2	3	4	5

B. Languages with Peers

8. **Which language(s) do you speak with your friends at school?**	**All the time**	**Most of the time**	**Half the Time**	**A little**	**Never**
Arabic	1	2	3	4	5
English	1	2	3	4	5
French	1	2	3	4	5
Other (please specify)	1	2	3	4	5
When talking to your friends at school, how frequently do you mix or switch languages?	1	2	3	4	5

9. **Which language(s) do you speak with your friends outside school (e.g. at the mall, restaurant, etc.)?**	**All the time**	**Most of the time**	**Half the Time**	**A little**	**Never**
Arabic	1	2	3	4	5
English	1	2	3	4	5
French	1	2	3	4	5
Other (please specify)	1	2	3	4	5
When talking to your friends outside school, how frequently do you mix or switch languages?	1	2	3	4	5

10. **Which language(s) do you use when you talk with friends who are in the French program?**	**All the time**	**Most of the time**	**Half the Time**	**A little**	**Never**
Arabic	1	2	3	4	5
English	1	2	3	4	5
French	1	2	3	4	5
Other (please specify)	1	2	3	4	5

11. **Overall, what language do you hear the most at your school?**	**All the time**	**Most of the time**	**Half the Time**	**A little**	**Never**
Arabic	1	2	3	4	5
English	1	2	3	4	5
French	1	2	3	4	5
Other (please specify)	1	2	3	4	5

C. Language Use in Everyday Interactions

	Very True	True	Not Sure	Not Very True	Not at all True
12. When my friends speak to me in English, I answer in English.	1	2	3	4	5
13. When my friends start mixing languages, I start mixing languages too.	1	2	3	4	5
14. If I ask my friends a question in English, and they answer in Arabic, I will switch to Arabic.	1	2	3	4	5
15. If I am speaking in Arabic, and my friends use other languages, I switch to those languages.	1	2	3	4	5
16. I use English with my friends or parents when I don't want other people to understand what I am saying.	1	2	3	4	5
17. I use French with my friends or parents when I don't want other people to understand what I am saying.	1	2	3	4	5
18. When someone doesn't understand what I mean in English, I repeat it in Arabic.	1	2	3	4	5
19. When someone doesn't understand what I mean in French, I repeat it in Arabic.	1	2	3	4	5
20. I find it strange if I speak to someone in English and they respond in Arabic.	1	2	3	4	5
21. I find it strange if I speak to someone in French and they respond in Arabic.	1	2	3	4	5
22. I think it is not okay to use different languages when a person cannot speak these languages.	1	2	3	4	5
23. My parents don't like it when I use different languages in the same sentence.	1	2	3	4	5
24. If I know that someone only speaks Arabic, I try to be careful not to use any English.	1	2	3	4	5

25. If I know that someone only speaks Arabic, I try to be careful not to use any French.	1	2	3	4	5

Section III: Media Language Exposure

	All the time	**Most of the time**	**Half the Time**	**A little**	**Never**
26. How often do you watch television or films in Arabic?	1	2	3	4	5
27. How often do you watch television or films in English?	1	2	3	4	5
28. How often do you watch television or films in French?	1	2	3	4	5
29. How often do you listen to music in Arabic?	1	2	3	4	5
30. How often do you listen to music in English?	1	2	3	4	5
31. How often do you listen to music in French?	1	2	3	4	5
32. How often do you read books, magazines, newspapers, etc. in Arabic?	1	2	3	4	5
33. How often do you read books, magazines, newspapers, etc. in English?	1	2	3	4	5
34. How often do you read books, magazines, newspapers, etc. in French?	1	2	3	4	5

Section IV: Computer Use

	All the time	**Most of the time**	**Half the Time**	**A little**	**Never**
35. How often do you use the computer for your studies?	1	2	3	4	5
36. How often do you use Arabic when working on the computer for homework, projects, etc.?	1	2	3	4	5

37. How often do you use English when working on the computer for homework, projects, etc.?	1	2	3	4	5
38. How often do you use French when working on the computer for homework, projects, etc.?	1	2	3	4	5
39. How often do you use the computer to stay in touch with friends (e.g. e-mail, facebook, twitter, skype, etc.)?	1	2	3	4	5
40. How often do you use Arabic when using these social networks to talk to friends?	1	2	3	4	5
41. How often do you use English when using these social networks to talk to friends?	1	2	3	4	5
42. How often do you use French when using these social networks to talk to friends?	1	2	3	4	5
43. How often do you use the computer to stay in touch with family (e.g. e-mail, facebook, twitter, skype, etc.)?	1	2	3	4	5
44. How often do you use Arabic when using these social networks to talk to family?	1	2	3	4	5
45. How often do you use English when using these social networks to talk to family?	1	2	3	4	5
46. How often do you use French when using these social networks to talk to family?	1	2	3	4	5
47. How often do you use Arabic when looking at websites (e.g. youtube, google)?	1	2	3	4	5
48. How often do you use English when looking at websites (e.g. youtube, google)?	1	2	3	4	5
49. How often do you use French when looking at websites (e.g. youtube, google)?	1	2	3	4	5
50. How often do you type on the computer in Arabic?	1	2	3	4	5

51. How often do you type on the computer in English?	1	2	3	4	5
52. How often do you type on the computer in French?	1	2	3	4	5

Section V. Language Choice

	Very True	True	Not Sure	Not Very True	Not at all True
53. I use English and/or French when I cannot find the right word in Arabic to use.	1	2	3	4	5
54. I find it difficult to speak in Arabic only for a whole conversation.	1	2	3	4	5
55. I find it difficult to speak in English only for a whole conversation	1	2	3	4	5
56. I find it difficult to speak in French only for a whole conversation,	1	2	3	4	5
57. I feel comfortable using English all the time without the need to speak in Arabic.	1	2	3	4	5
58. I prefer using English instead of Arabic to talk about certain topics.	1	2	3	4	5
59. When I am doing a task and thinking out loud, I talk to myself in English.	1	2	3	4	5
60. When I am studying subjects taught in English, I always use English.	1	2	3	4	5
61. I use English to get what I want when using Arabic doesn't work.	1	2	3	4	5
62. I use French to get what I want when using Arabic doesn't work.	1	2	3	4	5
63. I don't notice when I mix or switch languages	1	2	3	4	5
64. I use English instead of Arabic because I feel my Arabic is not as good as my English.	1	2	3	4	5

65. When my friends or family change languages, I don't notice it.	1	2	3	4	5
66. I mix and switch languages within the same conversation.	1	2	3	4	5
67. I mix and switch languages within the same sentence or utterance.	1	2	3	4	5
68. I mix and switch languages within the same word.	1	2	3	4	5
69. I want to speak English fluently.	1	2	3	4	5
70. I want to speak Arabic fluently.	1	2	3	4	5
71. I want to speak French fluently.	1	2	3	4	5

Section VI: Opinions about Language

	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
72. A person should be able to speak the language of his/her home country.	1	2	3	4	5
73. Knowing more than one language is important for education.	1	2	3	4	5
74. Some languages are more important to learn than others.	1	2	3	4	5
75. Some languages are more difficult to learn than others.	1	2	3	4	5
76. A person who speaks English or French in addition to Arabic has better thinking skills than a person who only speaks Arabic.	1	2	3	4	5
77. Once a person has learned two languages, it is easier to learn a third one.	1	2	3	4	5
78. Knowing more than one language confuses the speaker about which language to choose.	1	2	3	4	5

79. If a person changes languages in the middle of sentences or from one sentence to another all the time, he/she is not fluent in either/any language.	1	2	3	4	5
80. If a person changes languages in the middle of sentences or from one sentence to another, it is because he/she is confused about which language to speak.	1	2	3	4	5
81. People who speak only Arabic cannot achieve the same educational or professional levels as people who speak more languages.	1	2	3	4	5
82. It is perfectly normal for people to use words from two or more languages in one sentence.	1	2	3	4	5
83. It is best to begin learning more than one language as young as possible.	1	2	3	4	5
84. I want to be able to speak two or three languages fluently.	1	2	3	4	5
85. If I always codeswitch (use different languages in a conversation) or speak in English, I may lose my ability to speak Arabic.	1	2	3	4	5
86. If I always codeswitch or speak in French, I may lose my ability to speak Arabic.	1	2	3	4	5
87. It is more important to speak English than Arabic.	1	2	3	4	5
88. Speaking French is not as important as speaking English.	1	2	3	4	5

89. It is important to show that you can speak more than one language if you want to impress someone.	1	2	3	4	5
90. It is important to show that you speak English when talking about technology.	1	2	3	4	5
91. It is important to show that you speak French when talking about food.	1	2	3	4	5
92. It is important to speak in Arabic when discussing religion.	1	2	3	4	5
93. English is more expressive than Arabic.	1	2	3	4	5
94. French is more expressive than Arabic.	1	2	3	4	5
95. It is okay to use English with friends and family but not with strangers (e.g. when shopping).	1	2	3	4	5
96. It is okay to use French with friends and family but not with strangers (e.g. when shopping).	1	2	3	4	5
97. Almost everyone in Lebanon uses English in their everyday speech.	1	2	3	4	5
98. Almost everyone in Lebanon uses French in their everyday speech.	1	2	3	4	5
99. If someone speaks in English all the time although they know Arabic, they are showing off.	1	2	3	4	5
100. If someone speaks in French all the time although they know Arabic, they are showing off.	1	2	3	4	5
101. People might use English because they don't know certain words in Arabic.	1	2	3	4	5

102. People might use French because they don't know certain words in Arabic.	1	2	3	4	5
103. Lebanese people change languages intentionally to show that they are multilingual (can speak many languages).	1	2	3	4	5
104. Using English along with Arabic is part of the Lebanese culture.	1	2	3	4	5
105. Using French along with Arabic is part of the Lebanese culture.	1	2	3	4	5
106. Using Arabic, English, and French in one conversation or sentence is part of the Lebanese culture.	1	2	3	4	5

Section VII. Proficiency Self-Ratings

Given the evaluation criteria (see scale) below, what rating would you give yourself in

English _____

Arabic _____

French _____

1 = Novice Low = I need some time and cues to exchange greetings, introduce myself, and name a number of familiar objects. I cannot participate in a true conversation.
2 = Novice Middle = I can communicate only very minimally and with great difficulty using a number of isolated words and memorized phrases.

3 = Novice High = I can communicate with some success about simple topics only. I rely heavily on memorized phrases, or on words provided by the person I am speaking with. My sentences are short and incomplete, and miscommunication occurs frequently.

4 = Intermediate Low = I can successfully talk about a limited number of simple topics by using and combining into short statements what I know and what the person I am speaking with says.

5 = Intermediate Middle = I can successfully talk about a variety of simple topics (food, travel, family, daily activities and personal preferences). I can speak in full sentences and even with some strings of sentences.

6 = Intermediate High = I can successfully handle many uncomplicated tasks and social situations where I need to exchange basic information related to school, recreation, particular interests and areas of competence. I sometimes hesitate, make errors, and there are certain words I cannot remember or do not know.

7 = Advanced Low = I can actively participate in most informal and a limited number of formal conversations on activities related to school, home, and leisure activities and, to a lesser degree, those related to events of current, public, and personal interest or individual relevance. I can rarely speak formal or professional language, and I cannot speak at a professional level for an extended period of time.

8 = Advanced Middle = I can easily and confidently handle a large number of communicative tasks such as informal and some formal exchanges on a variety of concrete topics relating to school, home, and leisure activities, as well as to events of current, public, and personal interest or individual relevance. I can sometimes function at a formal or professional level of language but not consistently and not with a broad range of topics.

9 = Advanced High = I can participate fully and effectively in conversations on a variety of topics in formal and informal settings from both concrete and abstract perspectives. I can speak at a formal or professional level of language usually without difficulty. When speaking at a formal or professional level, I make some errors but these do not interfere with communication.

10 = Superior = I speak like a highly educated native speaker. I can participate fully and effectively in conversations on a variety of topics in formal and informal settings from both concrete and abstract perspectives with accuracy and fluency using formal and professional quality language. I might make some occasional errors but these do not interfere with communication.

Appendix B

Oral Proficiency Test Sample

This first section of the OPT serves as the introduction to the test. It consists of greetings and conversation openers (which will be excluded from analysis) and then moves on to questions pitched at the Intermediate Level, as the participants' formal instruction and informal exposure predicts at least this level of proficiency in the L3 (and higher for the L1 and L2). When a participant seems reasonably comfortable with the target language, the examiner will move to the next section. The examiner will have a range of questions to select from for each of the two sections. The examiner will mark the questions used, in order to make sure the same questions are not used in the other languages the participant is examined in.

Section I Questions:

A. Where did you grow up/travel to last summer? How is it similar/different from City X?

B. Tell me about your first day of your senior year. What was it like?

C. I'd like to move on and ask you some questions about your friends.

   1. Who is/are your closest friend(s)?
   2. How long have you been friends?
   3. What's the characteristic you admire most about your best friend?
   4. What do you both/all do for fun?

D. I'd like to move on and ask you some questions about your free time
   1. What type of activities do you like to do in your free time?
   2. How long have you been interested in these activities?
   3. Do you like to do these activities alone or with other people? (Why)?
   4. Do you think people have enough free time? (Why/why not)?

E. Tell me about next week. Where will you be, and what will you be doing each day?

F. Let's change the topic and talk about your neighbours

   1. Do you know the people who live next door to you?
   2. How often do you see each other?
   3. What kind of relationship do you have?
   4. How can neighbours be helpful?
   5. What kind of problems can people have with their neighbours in a big city?

G. I'd now like to discuss dreams with you.

1. Do you dream much at night?
2. Do you often remember your dreams?
3. Do you think we can learn anything from dreams? (Why)?
4. Do people in your country talk about their dreams? (Why)?
5. Do you think that dreams can come true?

H. Describe someone who has had an important influence on your life. You should say:
   Who the person is
   How long you have known him/her
   What qualities this person has
   Explain why they have had such an influence on you.

Section II Questions:

A. Some people believe that exposing young children to technology, and particularly tablets, may have negative consequences on their development. Do you think young children should be allowed to use tablets? Explain your answer.

B. Let's role-play: Imagine that you arrived at the airport in City X (depending on the language), your luggage is not in the baggage claim area. You speak with a service representative, explain why you and your luggage did not arrive in the same flight, and make arrangements to have the bags delivered to your hotel.

C. Let's role-play: You have just received an award for 'Innovator of the Year'. Decide on the field in which you have made this innovation, and make a brief speech accepting your award.

D. You are leading a discussion at your book club on a book you have just finished reading. Describe the main plot and themes of the book, and discuss the significance of one of the themes to society.

Appendix C

Oral Proficiency Test Scoring Rubric

1 = Novice Low = No real functional ability. Given lots of time and cues may be able to exchange greetings, give identity and name a number of familiar objects. Cannot participate in a true conversational exchange.

2 = Novice Middle = Can communicate only very minimally and with great difficulty using a number of isolated words and memorized phrases.

3 = Novice High = Can communicate with some success about simple topics only. Heavy reliance on memorized phrases, or on words provided by person speaking with. Speaks in short or incomplete sentences, and frequent miscommunications occur.

4 = Intermediate Low = Can successfully handle a limited number of uncomplicated communicative tasks by combining and recombining into short statements what they know and what the person speaking with says.

5 = Intermediate Middle = Can successfully handle a variety of uncomplicated communicative tasks about simple topics (food, travel, family, daily activities and personal preferences). Speaks in full sentences and even with some strings of sentences.

6 = Intermediate High = Can successfully handle many uncomplicated tasks and social situations requiring an exchange of basic information related to work, school, recreation, particular interests and areas of competence. Some hesitation, errors, and gaps in communication may still occur.

7 = Advanced Low = Can participate actively in most informal and a limited number of formal conversations on activities related to school, home, and leisure activities and, to a lesser degree, those related to events of work, current, public, and personal interest or individual relevance. Can rarely function at the level of formal or professional language, and cannot speak at a professional level for an extended period of time.

8 = Advanced Middle = Can handle with ease and confidence a large number of communicative tasks such as informal and some formal exchanges on a variety of concrete topics relating to work, school, home, and leisure activities, as well as to events of current, public, and personal interest or individual relevance. Can sometimes function at a formal or professional level of language but not consistently and not with a broad range of topics.

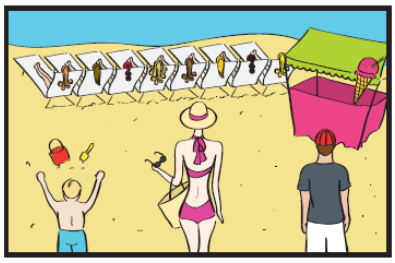
9 = Advanced High = Can participate fully and effectively in conversations on a variety of topics in formal and informal settings from both concrete and abstract perspectives. Can speak at a formal or professional level of language usually without difficulty. When speaking at a formal or professional level some patterns of errors may still appear but these do not interfere with communication.

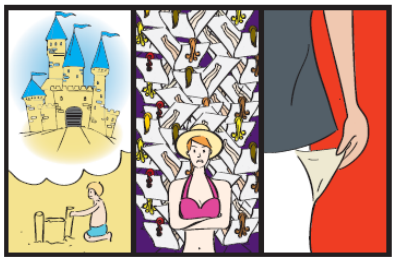
10 = Superior = Speaks like a highly educated native speaker. Can participate fully and effectively in conversations on a variety of topics in formal and informal settings from both concrete and abstract perspectives with accuracy and fluency using formal and professional quality language. Occasional errors may still occur but these do not interfere with communication.

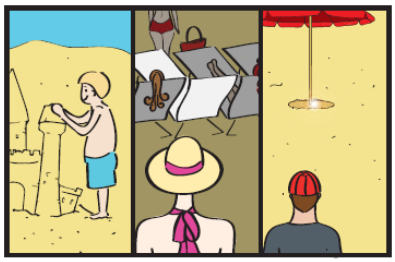
Appendix D

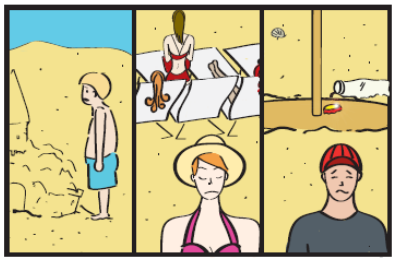
Narration Task – Storyboard and Transcription Samples

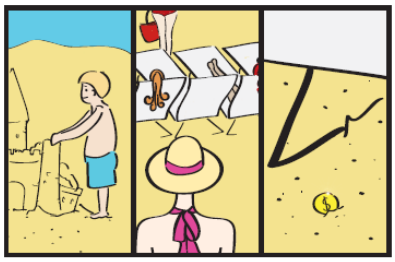
The next section offers three storyboards, each distributed across three pages and followed by sample transcriptions of narrations by EMI or FMI participants. The first storyboard is followed by both EMI and FMI participant production. The two transcriptions are representative of the differences between the two groups' overall patterns of codeswitching behaviour. Particularly, the minimal use of French by the EMI participant in the trilingual condition is in stark contrast with the FMI participant's use of English (and generally similar amount of language use of the three languages) in this condition. The black font denotes English, red denotes Arabic, and green denotes French. For each picture in the storyboard, an English translation (italicised and in brackets) immediately follows the original transcription of the participants' production. Elipses signify long pauses (exceeding 500ms), and they are included in the transcriptions to better represent the fluency of the participants. Hesitation phenomena such as 'um' and 'eh' are not assigned to any language, as they constitute 'filled pauses'. In the transcriptions, they are coloured based on the language that follows them, but this does not signify that they are assigned to any specific language.

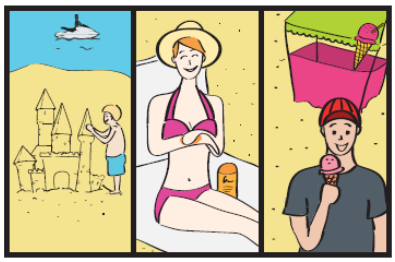


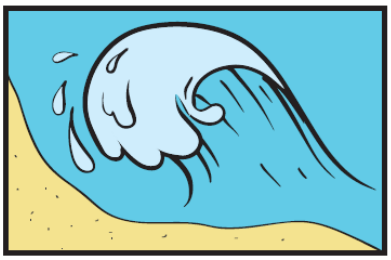


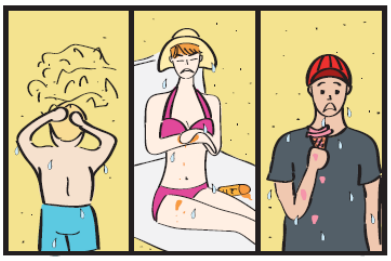


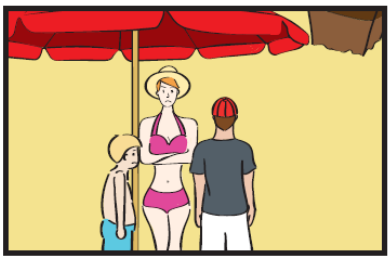


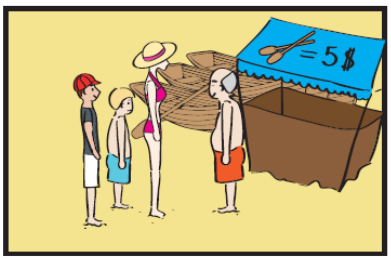


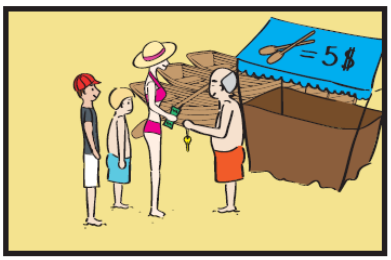


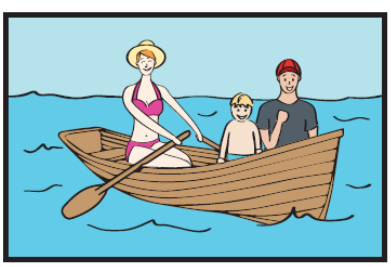


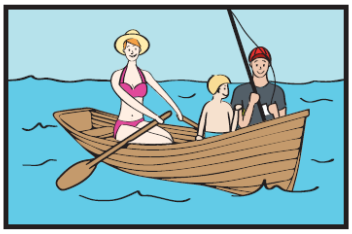


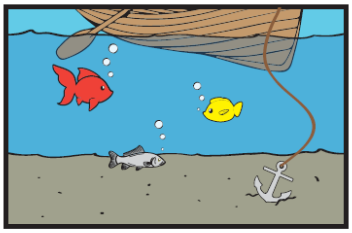


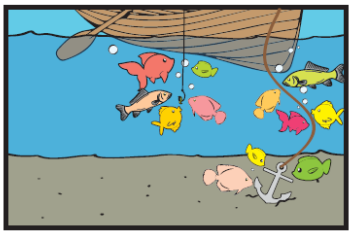

=5$

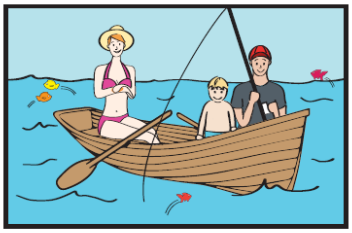

=5$

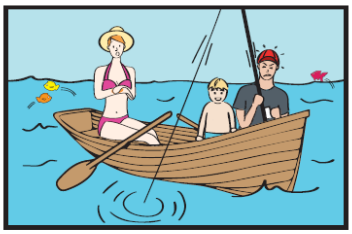


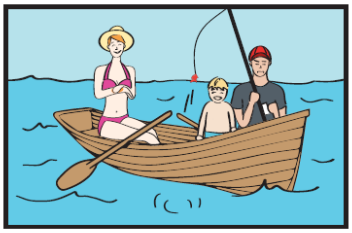












Story 1: EMI participant – Trilingual condition

There's a woman with a boy يمينا ع and another boy شمالا ع. One of the boys أزغر من اتاني . Eh… They are at the beach و they- they are all لبسين maillots. و the-the woman لابسه chapeau. كمان the older boy porte une casquette. The younger boy معو bucket and spade شكلو كبن عالرمل. His arms are up so that's probably what happened. في ice cream stand on the right بس شكلو مسكر لانه ما في حداً ورا ال- stand. ال ال مرأة شكلا بدا تعمل bronzage بس eh كل ال-chaises longues مأخودين. What else…um… شكلا لابسه بكيني وشلحت ال lunettes لتشوف أحسن…. Th- The older boy is wearing a t-shirt الصبي الزغير لابس مايوه أزرق. رماديه وشورت أبيض.

[*There's a woman with a boy to her right and another boy to her left. One of the boys is younger than the other. Eh…They are at the beach and they-they are all wearing swimsuits. And the- the.-woman is wearing (a) hat. Also the older boy is wearing a cap. The younger boy has (a) bucket and spade. It looks like he threw them on the sand. His arms are up so that's probably what happened. There's (an) ice cream stand on the right but it looks like it's closed because there's no one behind the stand. The-The woman looks like she wants to do (a) suntan but..eh.. all the beach chairs are taken. What else.. um… she looks like she's wearing a bikini and she took off the glasses to see better. Th-The older boy is wearing a grey t-shirt and white shorts. The younger boy is wearing a blue swimsuit.*]

Okay…So the boy was dreaming of making a sandcastle .. اه هلأ عم نشوف كل واحد شو كان ب بالو وبلش يساوي. The woman, ….ما بعرف maybe she's his mom, seems to be very upset لانو she still can't find a place to sit. She..ummm.. she… I don't know.. She's crossing her arms …ووعابسه ب her lips.... والصبي الكبير, is he the husband?... Anyway, we see انو he's emptying his pocket وشكلو طفران… ما معو… ما معو مصاري وجيبتو فاضيه

[*Ahhh now we are seeing what each one had in mind… Okay…So the boy was dreaming of making a sandcastle and he started making it. The woman… I don't know maybe she's his mom, seems to be very upset because she still can't find a place to sit. She..ummm.. she… I don't know.. She's crossing her arms and…and she's frowning with her lips… And the older boy, is he the husband?... Anyway, we see that he's emptying his pocket and he looks broke…He doesn't have…he doesn't have money and his pocket is empty.*]

Okay.. Okay.. هلأ… هلأ عم نشوف انو انو the boy is still building his sandcastle and it's starting to become almost like what he imagined in the previous picture. The woman spotted a seat where another woman has gotten off of فشكلا حتروح وتقعد عالكرسي finally. والصبي الكبير… الصبي الكبير…is standing إدام an umbrella وتحتا there's something shining.

[*Okay.. Okay.. Now…now we can seet that.. that…the boy is still building his sandcastle and it's starting to become almost like what he imagined in the previous picture. The woman spotted a seat where another woman has gotten off of so it looks like she's going to sit in the chair finally… And the big boy…the big boy is standing in front of an umbrella and under it there's something shining.*]

Oh.. okay… so the small boy is sad لانو وقع his sandcastle. فرط إل sand castle. So he's upset…Eh… المرة كامينا كان شكلا ز علانه لانو another woman beat her to the… chaise longue. So شكلا… شكلا she turned around وهلأ في .و…وز علانه close-up على what was shining و it seems to be a bottle cap… فا ال- older boy…the older boy كمان ز علان بس ما فهمت ليه انو what..what was he expecting it to be? ما بعرف

[*Oh.. okay… so the small boy is sad because his sandcastle fell over. The sand castle crumbled. So he's upset…Eh… The woman also looks upset because another woman beat her to the… chaise longue. So it seems… it seems she turned around and is upset. And… and now there's a close-up on what was shining and it seems to be a bottle cap… so the older boy.. the older boy is also upset but I don't understand why, like what…what was he expecting it to be? I don't know.*]

ويزبطو ما… ما بعرف إذا ز علان أو his sandcastle عم يرجع يعمر small boy… هلأ عم نشوف انو ال ….مبسوط وشكلا the lady that sat and took the empty spot هلأ is getting up ورايحة… So I guess that's good news لل woman... وبي the third section of the picture… في coin with a dollar sign… ما بعرف as if under another sign أو شي.

[*Now we see that the small boy… is rebuilding his sandcastle and re-fixing it…I don't… I don't know if he's upset or happy…. and the lady that sat and took the empty spot looks as if now is getting up and leaving… So I guess that's good news for the woman…. And in the third section of the picture… there's a coin with a dollar sign…I don't know.. as if under another sign or something*]

Ah okay هلأ ال- young boy has finally built the sandcastle he planned. وفي jet ski بل background بالبحو… the woman is finally happy وعم بتحط sunscreen ع جسما من شين تعمل bronzage. والصبي الكبير شكلو he's happy لا انو عم ياكل ice cream… آه okay هلأ فهمت ما كان معو مصاري so he found the gold coin و he bought ice cream with it. Okay okay.

[*Ah okay now the young boy has finally built the sandcastle he planned. And there's a jet ski [loan word] in the background in the sea…the woman is finally happy and is putting sunscreen on her body to suntan. And the older boy looks like he's happy because he's eating ice cream… Ah okay [loan word] now I understand. He didn't have any money so he found the gold coin and he bought ice cream with it. Okay okay.*]

Okay so there's big wave.

Oh no! The wave destroyed everything. The small boy's sandcastle is completely ruined. The woman is all wet and she looks furious and the older boy's ice cream is gone.

Okay هلأ كلن واقفين under the umbrella. They seem to be very upset, and it's as if, as if they're talking or they're trying to find a solution or they are discussing something I don't know.. they're not happy.

[*Okay [loan word] now they are all standing under the umbrella. They seem to be very upset, and it's as if, as if they're talking or they're trying to find a solution or they are discussing something I don't know.. they're not happy*.]

راحو عند an old man. He's wearing red shorts. He looks like he's the owner of some boat rental place… oars are for five dollars. The woman is talking to him, and the two boys are standing behind her.

[*They went to an old man. He's wearing red shorts. He looks like he's the owner of some boat rental place… oars are for five dollars. The woman is talking to him, and the two boys are standing behind her*.]

Now the woman is giving money to the old man and he's giving her a golden key. They seem happy. They are all smiling but eh- the- younger boy doesn't seem to be pleased.

Okay now they're in the boat and they look very happy. The woman is rowing the boat and the two boys are very happy.

اه هلأ شكلو ال- older boy طلع صنارة صيد. So I think they will do some fishing. The younger boy is helping his older brother. The woman is still rowing the boat and looks happy. They all look happy.

[Ah now it looks like the older boy took out a fishing rod. So I think they will do some fishing. The younger boy is helping his older brother. The woman is still rowing the boat and looks happy. They all look happy.]

Okay ال نزلو-…anchor و they are surrounded by fish. There is a red fish, a yellow fish, and a silver fish. It looks like they will stop and do some fishing.

[Okay [loan word], they threw the… anchor and they are surrounded by fish. There is a red fish, a yellow fish, and a silver fish. It looks like they will stop and do some fishing.]

Oh okay now they threw the bait وشكلو there are lots and lots more fish in the water under the boat. They are probably going to catch a lot of fish now. The fish are near both the anchor and the fishing line.

[Oh okay now they threw the bait and it seems there are lots and lots more fish in the water under the boat. They are probably going to catch a lot of fish now. The fish are near both the anchor and the fishing line.]

Now we go back up to the surface and we see the family in the boat. They all look happy and the fish are jumping all around them.

The fish are still jumping all around them .. the boy- the older boy seems straining. He's pulling on the fishing rod like uh- like he caught a really big fish. The woman or the mother seems shocked and the younger boy seems happy.

Haha. He pulled the rod وطلعت سمكه ز غيره كتير. The woman is laughing and putting sunscreen on again. The small boy كمان is laughing بس the older boy is upset.

[Haha. He pulled the rod and a very small fish came out. The woman is laughing and putting sunscreen on again. The small boy also is laughing but the older boy is upset.]

Story 1: FMI participant – Trilingual Condition

بشوف بهالصورة a woman qui porte un chapeau.. eh.. avec deux enfants. They are all looking على على girls while tanning. على اليمين je vois a stand qui vende du ice cream.

[I see in this picture a woman who is wearing a hat.. eh.. with two children. They are all looking at- at girls while tanning. To the right, I see a stand which sells ice cream.]

هلأ تاني صورة... الصبي is building a castle. Il pense qu'il va construire un chateau aussi grand.. que l'image. Alors que la femme is jealous من كل هودي البنات يلي عم يتشمسو ول ول بي is.. doesn't have enough money, he's broke. L'enfant est en train de construire un chateau..eh.. Il est heureux. Eh.. La fille..eh.. in the-in this image is also looking at girls tanning.. oh! She found an empty seat and et l'homme il a trouvé محل فيو يقعد تحت الشمسيه.

[*Now the second picture… the boy is building a castle. He thinks he will build a big castle as (in) the picture. Whereas the woman is jealous of all these girls who are sitting and sunbathing. And-and the dad is… doesn't have enough money, he's broke. The child is building a castle...eh.. he is happy. Eh.. the girl.. eh.. in the- in this image is also looking at girls tanning…Oh! She found an empty seat and and the man, he found a place to sit under the umbrella.*]

L'enfant a- a cassé son chateau. Eh.. Malheuresement اجت مرة تقعد بالمحل الفاضي . And the man broke his beer bottle. I'm not sure.

[*The child has- has broken his castle.. Eh.. unfortunately, a(nother) woman came to sit in the empty place. And the man broke his beer bottle. I'm not sure.*]

The-The boy is still building his sandcastle. He looks sad now... I think. La femme revoit la chaise بس في فهمت شو كتير ما وبهالصورة ... في و .يلي كان بدا تقعد عليها coin آه لأى .عالارض coin عالارض . The guy found a coin on the floor- on the ground.

[*The-The boy is still building his sandcastle. He looks sad now... I think. The woman sees the chair again which she wanted to sit in. And…and in this picture I don't really understand what's happening but there's a coin on the ground… Ah he found a coin on the ground. The guy found a coin on the floor- on the ground.*]

The boy constructed successfully the castle. La femme est en train de mettre de l'huile وشكلا مبسوط كتير و et elle porte encore son chapeau donc j'ai parle au debut. Et- Et l'homme qui a trouvé le coin عم ياكل ice cream وشكلو كتير مبسوط and is now happily eating his ice cream cone.

[*The boy constructed successfully the castle. The woman is putting on oil and she looks very happy and she's still wearing her hat just like I said at the beginning. And- and the man who found the coin is eating ice cream and he looks very happy and is now happily eating his ice cream cone.*]

Suddenly a huge wa- a huge wave came and destroyed the boy's castle et a noyée la femme et la femme est maintenant malheureuse والزلمي هلأ كمين صار زعلين لا أن ال- ice cream تبعو إنتزع .فصارو هلأ كلن upset because they were finally having a good time واجيت ال- wave وغيرت كل شي

[*Suddenly a huge wa- a huge wave came and destroyed the boy's castle and drowned (soaked) the woman and the woman is unhappy now. And the man now is also sad because his ice cream is ruined. So now they are all upset because they were finally having a good time and the wave came and changed everything.]*

Then…umm… there was a family meeting and they were all sad… Maybe they're telling each other what happened.

So they have reunited and they… et ils ont décidé de parler à un vieil homme.. قررو انو يعطوا مصاري وهوي يعطيين une clé كرميل يشترو a small bateau. Ah non…. the boat- the boat is probably chained so the key is to unlock the boat. Is it a boat? It's not a boat, it's a… a… a kayak.

[*So they have reunited and they… and they decided to talk to an old man.. They decided to give him money and he gives them a key to buy a small boat. Ah no….the boat- the boat is probably chained so the key is to unlock the boat. Is it a boat? It's not a boat, it's a… a… a kayak*].

And they went.. ils sont partis... Ils sont montés sur le bateau…le kayak. Ils sont heureux et ils sourient... They look much happier than before.

[*And they went.. they left… they climbed onto the boat…the kayak. They are happy and they are smiling. They look much happier than before.*]

L'homme is now deciding to fish…. Oh not fish... Oh yes fish. He decided to catch some fish. He threw his fishing stick. Il a jeté son صنارة ولقط هلأ سمكه .

[*The man is now deciding to fish…. Oh not fish... Oh yes fish. He decided to catch some fish. He threw his fishing stick. He threw his fishing rod and now caught a fish.*]

Here, we can see qu'il y a beaucoup de poissons.

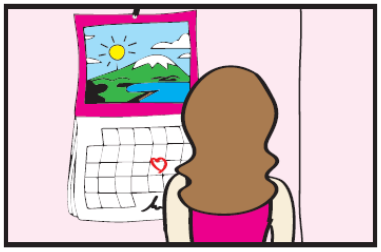
[*Here, we can see that there are lots of fish*].

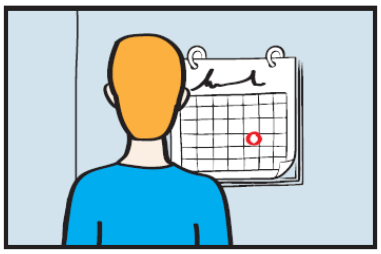
ب cette image, l'homme pêche et il transpire…He- He caught a fish I think.

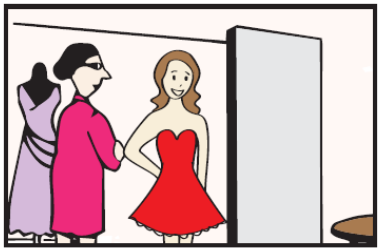
[*In this image, the man is fishing and is sweating. He- He caught a fish I think.]*

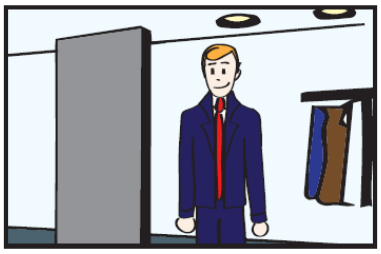
إيه ...اه...he did. He caught a very small fish… It's so small, and everyone is laughing at him.

[*Ah…yes... he did. He caught a very small fish… It's so small, and everyone is laughing at him*.]

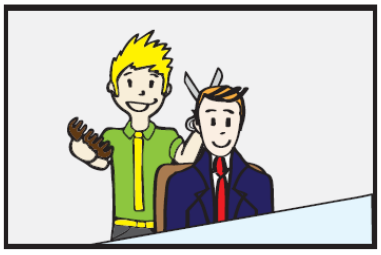


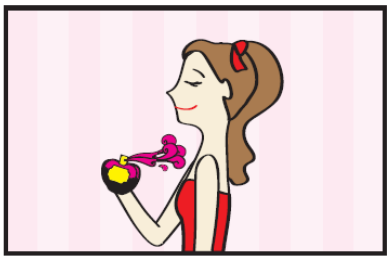




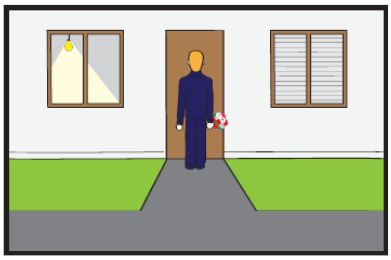




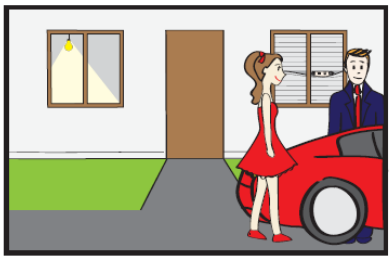


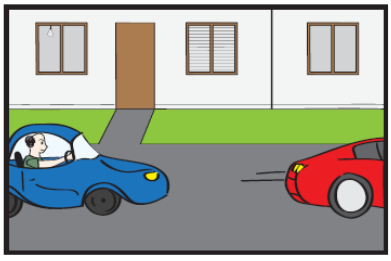


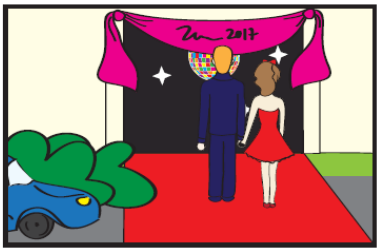


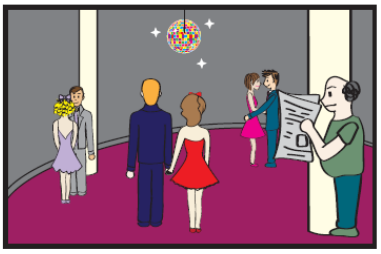






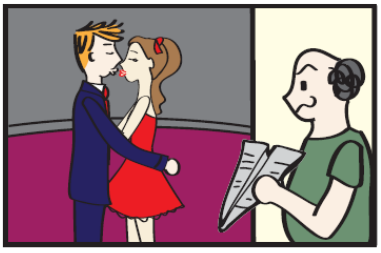


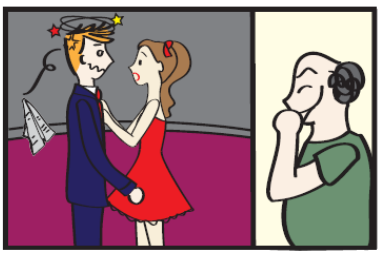

2017











Story 2: FMI student – English-only condition

There's a girl standing in front of a.. a calendar… There's a heart circled around a specific date, but I can't see what date it is. The calendar's background has a sun and mountains and a lake… or ocean and some clouds. The…The girl has brown hair and she's wearing a pink top.

Now there's a boy in the second picture..eh.. who is also looking at a calendar. It's a different calendar… There's as circle around a certain date, but we don't know which date or if it's the same date as the girl's but I think that it is the same date… He's in a different room because the background is blue, and in the picture above it was pink...Ummm… He has orange hair and he's wearing a blue t-shirt.

Now we see the girl at a dress shop and there's an older woman standing next to her... It…uh.. It's either a sales woman or her mom. The girl's wearing a red dress… and looking at herself in the mirror… The woman next to her is wearing a pink dress and has eyeglasses. She has black hair… There's also a purple dress on display in the background….Eh… There's also a brown table on the side.

Now…uh… now it's the boy… or a- a man who's wearing a suit and tie and he's standing in front of a mirror… I don't know if he's in a shop or in his house. He is smiling and I think he's excited about something… I think he's going out on a date with the girl in the red dress… I don't know what's behind him… on the wall, but it might be where he has his clothes hanging up.

Eh…We go back to the girl, and she's at the hair salon now. The hairdresser has fixed her hair. It's an old guy with a hair brush in his hand. He's wearing a pink shirt and a pink watch… and…he looks very happy with how the girl's hair looks. The girl is sitting down… I think she's also sitting on a pink chair. She has a…a bow in her hair and now she's wearing make-up. She has red lipstick and mascara. She looks happy with her hairstyle.

The boy is also styling his hair. He is at a..a barbershop. The barber is wearing a green shirt and yellow tie and he's holding scissors and a comb. The- The boy is also happy with his hairstyle which looks spiky in the front. The barber has yellow hair, and his hair looks stylish too.

I think… Eh- I think now the girl is back home and she's putting on perfume. The bottle is pink and black, and we can see the spray coming out of the bottle... She is still wearing her red dress and her hair looks the same as before. She's also still wearing her makeup.

And..uh..and the boy is now putting on some perfume. I think they are both done getting ready for their date. The boy is uh at his house I think because the background is blue… and when he was looking at the calendar before, it was also blue. His perfume bottle is also blue.

The boy is.. eh.. standing in front of a brown door and..uh..and he's holding a bouquet of flowers. We see there's a light on, so someone's inside the house. I think he's picking up the girl. Yeah..I think he's standing outside the girl's door to pick her up. There are two windows, one with closed shutters and one which is open and shows the lightbulb… There's also green grass on both sides of the entrance.

Yeah.. yeah.. it's the girl's house. She opens the door for him and she's so happy to see him…She takes the flowers from him and smiles at him. Ah… I didn't notice before, but the flowers are red and white and match her red dress… There's another man in the background sitting on an armchair. He looks old because he's half bald… He's wearing a green shirt and he's smiling.

The girl and boy are now going out on their date. They uh are getting into the boy's red car. It looks like a cool sports car. Ahhh in the background between the shutters there's the other man,… I think it's her father,… and he's spying on them.

The uh… the girl and boy are driving away but now the dad is following them… in his blue car. He- he looks angry or not happy at all. Maybe the boy's car is speeding because we can see the two lines in the picture.

Now…uh… the boy and girl are going inside a…a..uh.. aballroom. There's a disco ball hanging from the ceiling….Maybe it's their prom because there's 2017 written on the sign… They're going to their prom and and- we can see the dad's car parked in front of a bush outside.

Now we are inside the prom. We can..uh… see the coloured disco ball and other couples who are dancing. One couple is wearing grey or light purple. The girl has curly blond hair and has a grey dress on, and… the boy has brown hair and is wearing a grey tux- tuxedo. The other couple is also dressed up for prom. The girl is wearing a dress, and the boy is wearing a blue suit. The… the.. main girl and boy are going on to the dancefloor and..uh… the dad is standing behind a wall or… a pole staring from behind a newspaper. So…so… he's hiding behind the newspaper so they don't see that he's spying on them.

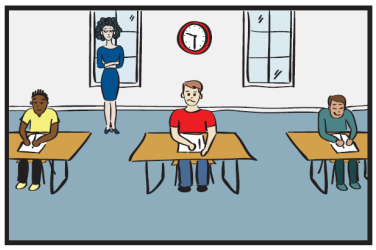
Now…uh…they are dancing and the father looks very angry. He- He's not happy that they are dancing so close together, and…that the boy's arms are around his girl's waist…. The boy and girl don't notice that the dad is there and they are smiling at each other.

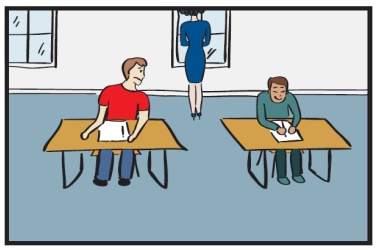
Now they are about to kiss, and the father is shocked and his mouth is open.

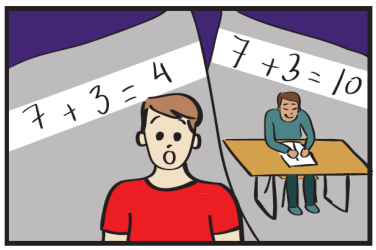
They are almost kissing now,… so the father… turned the newspaper to a paper airplane. He looks very upset and angry… but also like he's planning something,… like he's determined to stop them.

Ah… he threw the paper airplane at the boy to stop him from kissing his daughter,… so the boy has stars around his head… so it hit his head and he's dizzy. The girl is shocked because

she doesn’t know what happened, and the father is laughing and is very happy he stopped them from kissing…Poor boy.

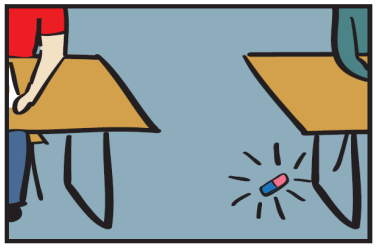


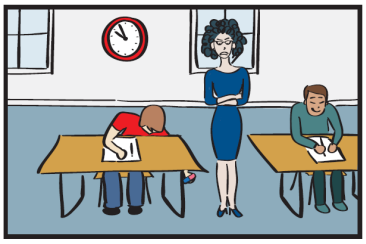



7 + 3 = 4
7 + 3 = 10







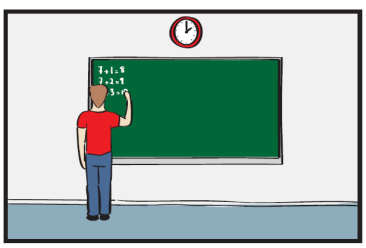


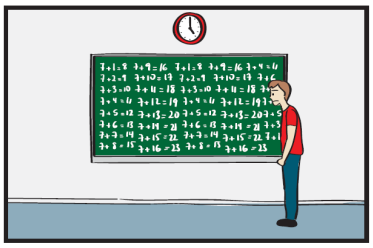


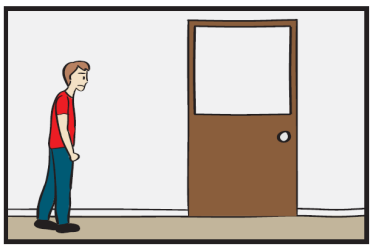


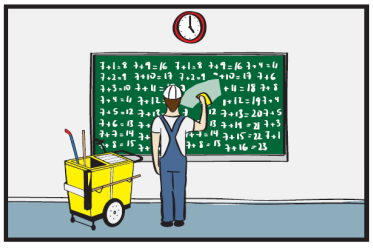




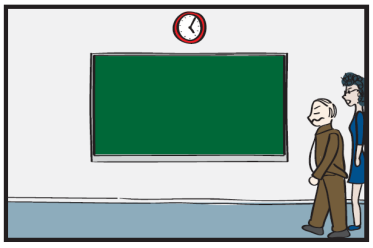














Story 3: EMI student – French-only condition

*Errors are indicated in the translation in blue.

Il y a trois étudiants,….. trois garçons,…. dans une classe. Ils font un examen….. Il.. Il y a une prof qui porte…qui porte… une robe bleue…. avec des mains croisées…. Elle…elle.. porte des lunettes. Il y a aussi une…une uhhhh… horlage. Il fait neuf heures et demie… Il est, Il est neuf heures et demie, neuf heures trente.

[*There are three students…three boys… in a class. They are doing an exam. There…There is a teacher who is wearing… who is wearing… a blue dress….with crossed hands. She…She.. is wearing glasses. There is also a,,, a….. clock. It is doing nine [o'clock] and a half….it is ,it is nine and a half, nine thirty*.]

Ella a… elle a… elle est…elle a tournée…. Le garçon a la droite écrit mais le garçon a la gauche…. veut….copier les réponses.

[*She has…she has…she is.. she has turned…. The boy on the right is writing but the boy on the left…. wants…to copy the answers*.]

C'est un test de mathématiques…. le garçon qui… uhhh… qui… porte une chemise… rouge a découvrir que sa réponse est fauteIl a écrit que sept plus trois égal quatre mais son ami a écrit que le réponse est dix.

[*It is a mathematics exam and the boy who uhhhh….who… is wearing a red shirt discovered [gr.] that his response is wrong [gr.]. He wrote that seven plus three equals four but his friend wrote that the answer is ten.*]

Il cherche pour une gomme dans…uh.. dans… sa trousse….Il est maintenant dix heures moins quart.

[*He's looking for [gr.] an eraser in…uh…in… his kit (pencil case). It is now quarter to ten.*]

Il a… il a… il a mis toutes les choses dehors mais il n'a pas trouvé de gomme. Il fait de panique..

[*He has… he has.., he has put all the things outside but he did not find an eraser. He made a panic [he panicked]*].

Il y a une gomme sur le….. sur le.. le…sur.. sous la table.

[*There is an eraser on the… on the.. the.. on…under the table*].

Il prit la gomme sous la table mais…uh…mais la prof pense qu'il fait quelque chose mal.

[*He takes [gr.] the eraser under the table but…uh…but the teacher thinks he's doing something wrong [gr.]*].

Oh non! Elle est très…. très … ell est en colère et elle… elle… elle tore up his test paper.

[*Oh no! She is very.. very.. she is angry and she.. she… she tore up his test paper [English switch]*].

Il pleure.

[*He's crying*].

Ils…Ils sont maintentant avec le directeur d'école…. Le…le garçon….attends dehors et…uh…et la prof parle avec le directeur.

[*They…they are now with the school director (or principal). The….the boy is waiting outside and…..uh…and the teacher is talking with the director*.]

Le directeur crie au garçon… et le garçon… est triste.

[*The director screams at the boy… and the boy… is sad*].

Il est deux heures…. Le garçon écrit sur le tableau…eh…il écrit…eh… les maths.

[*It's two o'clock…. The boy is writing on the board…eh… he is writing…eh the maths [word choice]*].

Il est cinq heures… et il a fini ses… devoirs.

[*It's five o'clock… and he finished his…. homework*].

Il… il… retourne chez le directeur.

[*He…he… goes back to the director*].

Oh non! Il ya le… le…. L'homme de menage? Eh…. Il ya un homme qui….qui…uh..quitravail a l'ecole. Le janitor. Il…il efface le tableau.

[*Oh non! There is the.. the.. the cleaning man? Eh…There's a man who…who…uh..who works at the school. The janitor [English switch]. He….he is erasing the board*.]

Oh! Le garçon marche…. avec le directeur et la prof.

[*Oh! The boy is walking…. with the director and the teacher*.]

Ils…eh….ils… ont trouvé que..uh…que le garçon n'a pas fait les devoirs.

[*They…eh..they… found that…uh..that the boy did not do his homework.*]

Ils.. ils.. sont.. f-fâchés et le directeur crie… et.. et le garçon… pleure.

[*They.. they.. are..a-angry and the director is screaming… and… and the boy.. is crying.]*

Appendix E

CUREC Form and CUREC Approval Letter (Notification Email)

University of Oxford

CENTRAL UNIVERSITY RESEARCH ETHICS COMMITTEE (CUREC)

CUREC/1A Checklist for the Social Sciences and Humanities

The University of Oxford places a high value on the knowledge, expertise, and integrity of its members and their ability to conduct research to high standards of scholarship and ethics. The research ethics clearance procedures have been established to ensure that the University is meeting its obligations as a responsible institution. They start from the presumption that all members of the University will take their responsibilities and obligations seriously and will ensure that their research on human subjects is conducted according to the established principles and good practice in their fields and in accordance, where appropriate, with legal requirements. Since the requirements of research ethics review will vary from field to field and from project to project, the University accepts that different guidelines and procedures will be appropriate. Please check the CUREC website to ensure that you have the correct form for your project.

This form does not cover research governance, satisfactory methodology, compliance with the requirements of publishers when administering their tests or questionnaires, or the health and safety of employees and students. As principal investigator, it is your responsibility to ensure that requirements in these areas are met. Please carry out a risk assessment of the project, in consultation with all researchers involved, using the checklist and CUREC's other documentation.

The use of an asterisk in this form indicates a phrase defined in the glossary. The glossary and further information on the University's research ethics procedures are available from the CUREC website:

www.admin.ox.ac.uk/curec

This form is designed largely for research that falls within the Divisions of Social Sciences and Humanities and which does not involve a high level of risk to the subjects. Elite interviews, field work and oral history are included in the CUREC process. Please take a moment to read through it and if you have any questions or doubts as to whether it is the appropriate form, please review Section A or consult the CUREC website.

Note on anonymised data and audit: you do not need to obtain ethical approval for your study if:

- you are using previously collected **anonymised data** about people which neither you nor anyone else involved in your study can trace back to the individuals who provided them (e.g. census data, administrative data, secondary analysis). Please refer to the definition of *personal data in the glossary and FAQ A4 for further guidance; or
- you are conducting research on behalf of or at the request of a service provider that matches the definition of *audit in the glossary.

If your research is audit or uses prior-anonymised data, please check this box: ☐

You do not need to seek ethical approval from CUREC, and you do not need to complete any more of this form. However, please check with your department's requirements, as some departments require you to lodge this form with them.

Office use only: IDREC Ref. No. _____

Date of confirmation that checklist accepted on behalf of IDREC: // //

SECTION A	**Yes**	**No**
1) Are you using research methodologies commonly used in biomedical or behavioural laboratory sciences?		✓
2) Is there a significant risk that the research will induce anxiety, stress or other harmful psychological states in participants that might persist beyond the duration of any test or interview in which they are participating?		✓
3) Will the research involve human participants recruited by means of their status as present or past NHS patients or their relatives or carers?		✓
4) Does the research involve *human participants aged 16 and over who do not have *capacity to consent for themselves? See the Mental Capacity Act 2005		✓
5) Is the study to be funded by the US National Institutes of Health or another US federal funding agency?		✓

If you answered 'yes', please **stop** work on this checklist and

- **for questions 1 and 2**, complete CUREC/1 instead (available from www.admin.ox.ac.uk/curec);
- **for questions 3 and 4**, submit your proposal to the appropriate NHS ethics committee (see www.nres.nhs.uk and http://www.admin.ox.ac.uk/researchsupport/ctrg/ for further information);

- **for question 5**, or **if you answered 'yes' to questions 1, 2 or 4 and your research will take place outside the EU and is a biomedical study**, submit your proposal to OXTREC, which uses separate documentation. **Applications to OXTREC using this form will not be accepted.** If your research is not a biomedical study and does not have US funding, but will take place outside the EU, you may use this form to submit your application for approval to the Social Sciences and Humanities IDREC.

If you have answered 'no' to all questions in Section A, please complete Sections B-E. This form and any supporting materials should be typewritten.

SECTION B

Principal investigator/ supervisor/student researcher (title and name):	Mrs. May Zantout
Name of supervisor (STUDENT RESEARCH PROJECTS ONLY):	Prof. Victoria A. Murphy, Dr. Catherine Walter
Degree programme, e.g. DPhil, MPhil, MSc (STUDENT RESEARCH PROJECTS ONLY):	DPhil (Education)
Department or institute:	Department of Education
Address for correspondence (if different):	Summertown House, Flat 68, 369 Banbury Road, OX2 7QZ, Oxford, UK
Email and phone contact:	May.zantout@lmh.ox.ac.uk +44(0)7454269922
Title of research project:	Bridging sociopragmatic and psycholinguistic theory on trilingual codeswitching

Brief lay description of *research (about 150 words), plus description (about 200 words) of the nature of participants (including the criteria for inclusion/exclusion & method of recruitment, how professional guidelines are being applied (if applicable) and use to which the results/data will be put.

Please describe how you will obtain informed consent, citing and attaching, where applicable, documentation produced in support of your application such as generic recruitment and advertisement material, participant information sheet(s), consent form(s) and defriefing document(s).

This study aims to explore the social and cognitive motivations for and mechanisms involved in codeswitching by trilingual youth enroled in English- and French-immersion programs in Lebanon. As several tasks (described below) are administered in this study, a breakdown of these tasks per session is provided in Appendix A. Each session will take approximately 45-90 minutes, which is not uncommon for participants in the target age groups who typically sit for exams that last between 1-3 hours at their school. However, unlike the exams which may be stressful in nature, the tasks provided in this study are not stressful in nature and do not subject the participants to the pressures of testing.

Taking place entirely in each participant's place of residence, the study will first ask participants to fill out a language background questionnaire (Appendix B) to determine language (oral) proficiency, dominance, and use. The questionnaire will take approximately 10-15 minutes to complete. The participants will also be asked to perform language proficiency tests in Arabic, English, and French to have a more objective, external measure of their proficiencies in their three languages. The exact tests are yet to be determined, but the types of questions on these tests are similar to those of the IELTS speaking section, indicated in Appendix C. Each of these tests will take about 30 minutes. In order to distinguish general cognitive processes from language-specific cognitive processes, the Simon task will be administered on the researcher's computer screen. Participants will be asked to press the right or left buttons, depending on the colour of the stimulus they are shown on the screen, and reaction time and accuracy will be measured. The Simon Task takes about 3 to 5 minutes to complete. A series of language tasks will then be provided, such as a verbal fluency task (where participants have 60 seconds to name as many items as possible belonging in a category in each language. For e.g., they have 60 seconds to name as many items as they know that belong to the category 'fruit'). Performing the verbal fluency task in each languages will take approximately 10 minutes. Participants will then perform a trilingual picture-naming task, where participants are asked to name objects (e.g. book, ball, pen) they see on the researcher's computer screen in the language(s) specified by the researcher. Accuracy and reaction times will be measured to examine how language dominance and proficiency affect performance in this task. The picture-naming task requires approximately 15 minutes for completion. A narrative task will then be provided, and participants will be asked to narrate a story they see in a storyboard in the language(s) selected by the researcher. The accuracy and ability to produce long speech utterances in languages in which participants are proficient will be examined. Narrative tasks in different languages and language combinations will take approximately 2-5 minutes each, or a total of 14-35 minutes, depending on the participants' performance. Participants will then partake in interactions with peers from different language backgrounds (e.g. Arabic monolinguals, English-Arabic bilnguals, etc.) in a one-to-one context to examine the frequency of codeswitching and the influence of the interlocutor on the participants' language choice and behaviour. Each interaction will last approximately 15-20 minutes.

Participants will be asked to undertake the recalls immediately after the interactions. Stimulated recalls (Appendix D) take 10-15 minutes to be conducted. By exploring the cognitive processes involved in codeswitching, and taking into account proficiency in the three languages and social motivations based on interlocutors, etc., the study will examine how and why participants' codeswitching behaviour is exhibited. It will also explore participants' ability to produce mono/bi/trilingual utterances when prompted in the various tasks and from data obtained in the stimulated recalls, language background questionnaires, and proficiency tests.

Participant Description

Upon receiving CUREC approval, the researcher will contact the president of a school in Beirut (Appendix E), Lebanon, to request permission for distributing information sheets (Appendix F) and consent forms (Appendix G) to the students meeting the inclusion criteria. These students and their parents will have no difficulty understanding the documents that are written in English and French. The inclusion criteria require that all children in this study be enrolled at this school from nursery till their current grade in the same curriculum type (Lebanese curriculum). The first age group, consisting of 80 children (male and female), aged 11 and 12, are in the 1st intermediate class (6th Grade), 40 of whom are in an English-immersion program and 40 are in a French-immersion program. The second group of children comprises 80 (male and female) participants in third secondary (12th Grade) who are 17 and 18 years of age. Half of these participants will be in an English-immersion program, and the other half will be in a French-immersion program.

Data collection, storage and disposal

The project will be explained orally to the participants, who will be invited to provide their assent to participate in this research, and children's parents will be fully informed in writing (Appendix F) and invited to provide their consent to allow their children to participate (as per Protocol 25). All these data collection methods are deemed non-invasive, and they therefore comply with the requirements set by Protocol 25. If parents indicate their interest in receiving a summary of the main findings of this study, they may select the relevant box in the consent form (Appendix G). A pilot study (Appendix H) will be carried out with children of the same ages, grades, and school. The pilot study participants will undertake the same tasks, questionnaires, and stimulated recalls upon providing their assent and parents' consent.
All participants, parents, and the school would be made anonymous in all research reports. The data collected would be stored on the researcher's computer and an external hard-drive, which will be locked away. All data will be kept strictly confidential, available only to the researcher and her supervisors, and it will not be used other than specified. All audio-recordings will be destroyed at the end of the research period.

As the data will be collected in participants' homes, safety measures aligned with the risk assessment form protocols will be adopted. Along with the regular timings of contact with the supervisor, the dates and timings of the sessions, as well as participants' addresses, will be recorded prior to the sessions and sent as a detailed itinerary to the supervisor.

List actual or probable location(s) where project will be conducted, if known:	Beirut, Lebanon
If your research involves overseas travel or fieldwork, have you completed and returned a travel risk assessment form? (Bear in mind that this may be necessary to ensure that the travel or fieldwork is covered by the University's travel insurance – see http://www.admin.ox.ac.uk/finance/insurance/travel/.)	YES / ~~NO / N/A~~
Anticipated duration of project:	18 months
Anticipated start date:	20 / 01 / 2016
Anticipated end date:	31 / 07 / 2017
Name and status (e.g. 3rd year undergraduate; post-doctoral research assistant) of others taking part in the project:	Not Applicable
Please indicate what training on research ethics the researchers involved with this study have received, e.g. the title of the online or in-person course, and date completed (online training available at www.admin.ox.ac.uk/researchsupport/integrity/human/):	May completed the Epigeum course on Research Integrity: Social and Behavioural Sciences at https://weblearn.ox.ac.uk/access/content/group/78538594-13fc-4982-884b-bbf340ddbbd9/index.html on 17 February 2014 and scored 17/20.

SECTION C

Methods to be used in the study (**tick** as many as apply: this information will help the committee understand the nature of your research and may be used for audit).

	Please tick
Interview	
Questionnaire	✓
Analysis of existing records	
Participant performs verbal/paper and pencil/computer based task	✓
Measurement/recording of motor behaviour	
Audio recording of participant	✓
Video recording or photography of participant	
Physiological recording from participant	
Participant observation	
Covert observation	
Systematic observation (Narration and Interaction/Conversation Tasks)	✓
Observation of specific organisational practices	
***Other (please specify)*: Stimulated Recalls**	✓

SECTION D

Have you read one or more of the following professional guidelines and do you undertake to use the principles listed there as a guide for your own work? Please note that this is not intended to be an exhaustive list. Links to the guidelines listed below are included on the CUREC website.

	Please tick
***British Society of Criminology: Code of Ethics for Researchers in the Field of Criminology* [http://britsoccrim.org/new/?q=node/22**	
British Educational Research Association Ethical Guidelines for Educational Research [www.bera.ac.uk/guidelines]	✓
Academy of Management's Code of Ethics [www.aomonline.org/aom.asp?ID=&page_ID=242]	
Association of American Geographers Statement on Professional Ethics [www.aag.org/cs/resolutions/ethics]	
***Oral History Society of the UK Ethical Guidelines* http://www.ohs.org.uk/ethics.php**	

American Political Science Association (APSA) Guide to Professional Ethics in Political Science (Section H) [www.apsanet.org/content_9350.cfm]	
Political Studies Association Guide to Good Professional Conduct [www.psa.ac.uk/psa]	
British Psychological Society Code of Ethics and Conduct [www.bps.org.uk/what-we-do/ethics-standards/ethics-standards]	✓
Ethics Guidelines of the Association of Social Anthropologists of the UK and Commonwealth [www.theasa.org/ethics/guidelines.shtml]	
Social Research Association: Ethical Guidelines [http://the-sra.org.uk/research-ethics/ethics-guidelines/]	✓
Statement of Principles of Ethical Research Practice from the Socio-Legal Studies Association [www.slsa.ac.uk/content/view/247/270/]	
Statement of Ethical Practice for the British Sociological Association [www.britsoc.co.uk/about/equality/statement-of-ethical-practice.aspx]	✓
Other professional guidelines (please specify): British Association for Applied Linguistics Recommendations for Good Practice in Applied Linguistics Student Projects [http://baal.org.uk/dox/goodpractice_stud.pdf]	✓

SECTION E

Please put a tick in the yes/no column as appropriate to indicate your response.

1) Will you obtain informed consent according to good practice in your discipline before participation?	**Yes**	**No**
	✓	
2) Will you ensure that *personal data collected directly from participants or via a *third party is held and processed in accordance with the provisions of the Data Protection Act?	**Yes**	**No**
	✓	
3) Does the research involve as participants *people whose ability to give free and informed consent is in question? ***Note: participants aged under 16 are generally considered to require consent of a parent or guardian (only answer 'no' to this question if you can cite one of the protocols listed under 'children'). For participants aged 16-17, consult FAQ C13***	**Yes**	**No**
	✓	
	Yes	**No**

4) As a consequence of taking part in the research, will participants be at serious risk of rendering themselves liable to criminal prosecution (e.g. by providing information on drug abuse or child abuse)?		✓
5) Does the research involve the *deception of participants, as part of the investigation/experiment?	**Yes**	**No**
		✓
If any of your answers above are in a shaded box, please indicate whether those aspects of your project are fully covered by the following.		
6) Research protocol(s) which has/ve received IDREC/CUREC approval?	**Yes**	**No**
If yes, please give protocol number(s): Protocol 25	✓	
7) Professional guidelines that you will be following, as noted under Section D?	**Yes**	**No**
	✓	

If any of your answers in Section E are in a shaded box and are not covered by a protocol or by professional guidelines, please complete CUREC/2, available to download from the CUREC website. Then submit both this form (you need not complete section F) and the CUREC/2 to the Social Sciences and Humanities IDREC.

If all your answers in Section E are in the unshaded boxes or your answers in shaded boxes are covered by a protocol or professional guidelines, complete Section F and submit this form and any accompanying documents to the Social Sciences and Humanities IDREC or to the relevant officer/committee at departmental level (see notes and address below).

FINAL CHECK

Please check each of the following before submitting the checklist. **If the appropriate supporting documentation is not included with your application, you will then be asked to provide this separately. This may well delay the ethical review process, and thus the start of your research.**

Have you completed Sections A-E? ☒

Have you defined all technical terms and abbreviations used? ☒

If you have produced any documentation in support of your application (which might include questionnaires, participant information, consent forms/form or note of procedure for recording oral consent, advertisements and surveys), have you attached a copy of these? ☒

Are all pages (including appendices and attachments) numbered? ☒

SECTION F

You can submit this checklist by email and/or as a signed hard copy; if it is being sent by email only, the checklist, and any email from the head of department or nominee separately endorsing its submission, must be sent from a University of Oxford email address (i.e. as a minimum, the checklist and supporting documents must be submitted by the head of department or nominee indicating his/her approval from a University of Oxford email account).

Complete this section only if you do not need to submit form CUREC/2.

I understand my responsibilities as principal researcher/supervisor/student researcher as outlined in the CUREC glossary and guidance on the CUREC website.

I declare that the answers above accurately describe my research as presently designed and that I will submit a new checklist should the design of my research change in a way which would alter any of the above responses so as to require completion of CUREC/2 (involving full scrutiny by an IDREC). I will inform the relevant IDREC if I cease to be the principal researcher on this project and supply the name and contact details of my successor if appropriate.

Signed by principal researcher/supervisor/student researcher:…May Zantout……

Date: 06/02/2015……

Print name (block capitals)…MAY ZANTOUT……

Signed by supervisor:……(for student projects)

Date:……

Print name (block capitals)……

I understand the questions and answers that have been entered above describing the research, and I will ensure that my practice in this research complies with these answers, subject to any modifications made by the principal researcher properly authorised by the CUREC system.

Signed by associate/other researcher:
…….

Print name (block capitals)……

Date ……

I have read the research project application named above. On the basis of the information available to me, I:

(i) consider the principal researcher/supervisor/student researcher to be aware of her/his ethical responsibilities in regard to this research;

(ii) consider that any ethical issues raised have been satisfactorily resolved or are covered by relevant professional guidelines and/or CUREC approved protocols, and that it is appropriate for the research to proceed without further formal ethical scrutiny at this stage (noting the principal researcher's obligation to report should the design of the research change in a way which would alter any of the above responses so as to require completion of a CUREC/2 full application);

(iii) am satisfied that the proposed project has been/will be subject to appropriate *peer review and is likely to contribute something useful to existing knowledge and/or to the education and training of the researcher(s) and that it is in the *public interest.

(iv) [FOR DEPARTMENTS/FACULTIES WITH A DEPARTMENTAL RESEARCH ETHICS COMMITTEE (DREC) OR EQUIVALENT BODY - PLEASE DELETE IF NOT APPLICABLE] confirm that this checklist (and associated research outline) has been reviewed by the Department's Research Ethics Committee (DREC)/equivalent body, and attach the associated report from that body.

Signed:……..
(Head of department or nominee e.g Chair of DREC, Director of Graduate Studies for postgraduate student projects)

Print name (block capitals)……

Date:……

If your research involves participants recruited by means of their status as current or former NHS staff, or the research will, in whole or in part, be carried out on NHS premises, use NHS facilities or assess NHS facilities or services, please see FAQ B3 (www.admin.ox.ac.uk/curec/faqs/).

Please check with your department about its procedures for the approval of CUREC forms. If your department indicates that the checklist should be submitted directly to the IDREC, please send it, together with any supporting documentation, to the following address(es), keeping a copy for yourself:

Secretary of the Medical Sciences IDREC Email: ethics@medsci.ox.ac.uk
Research Services
University of Oxford
Wellington Square
Oxford, OX1 2JD

Secretary of the Social Sciences and Humanities IDREC Email: ethics@socsci.ox.ac.uk
Research Services
University of Oxford
Wellington Square
Oxford, OX1 2JD

IDRECs and/or CUREC will review a sample of completed checklists and may ask for further details of any project.

Dear May,

Application Approval

Title: *Bridging sociopragmatic and psycholinguistic theory on trilingual codeswitching*

The above application has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants. Thank you for submitting the requested amendments. I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

If your research involves participants whose ability to give free and informed consent is in question (this includes those under 18 and vulnerable adults), then it is advisable to read the following NSPCC professional reporting requirements for cases of suspected abuse: http://www.nspcc.org.uk/Inform/research/briefings/reporting_child_abuse_wda74908.html

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to research.office@education.ox.ac.uk for consideration.

Good luck with the project!

Yours sincerely,
Nikki Luke
Member of DREC

Nikki Luke
Research Fellow
Rees Centre for Research in Fostering and Education
University of Oxford Department of Education
15 Norham Gardens
Oxford OX2 6PY

nikki.luke@education.ox.ac.uk
+44 (0)1865 274050
http://reescentre.education.ox.ac.uk

Information Sheet and Consent Form

Information Sheet

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries'education.ox.ac.uk www.education.ox.ac.uk

Director Professor Ernesto Macaro

Invitation

Your child's school has agreed to take part in assisting me, May Zantout, a DPhil student at the University of Oxford, in gathering information for a project entitled *Bridging Sociopragmatic and Psycholinguistic Theory on Trilingual Codeswitching* that studies how and why children residing in Lebanon use and mix different languages. I would like to invite your child to be part of this project. I very much hope that you would like to take part, but before you decide, it is important that you understand why the project is being done and what it will involve.

What am I trying to find out?

This study is an exploration of how and why children to use English, Arabic, and French, and whether motivations for switching between languages are social or cognitive (mental). Whether your child uses one language more than the other because of the person he/she is talking to or whether it is easier/faster for your child to switch from English to Arabic to French are some questions I will address in my study.

What will happen if your child takes part?

If you agree that your child participate in this study, he/she will be first asked to complete a paper-based questionnaire and language tests (lasting about 1.5 hours in total). Then he/she will perform a series of computer-based tasks and speaking tasks which will be video-recorded. All the tasks will take place at your residence. The computer sessions will last approximately 1 hour in total, and the speaking will last approximately 3 hours, which will be spread out over the course of two weeks, depending on your child's availability. Typically, I would visit your residence five times for about 1.5 hours each time over those two weeks.

Please note that all data collected will be kept strictly confidential.

Is your child obliged to take part? What are the benefits and risks of taking part?

Your child does not have to participate in this project. If he/she does not participate, there will be no consequences for his/her education. If you change your mind about your child's participation, or if your child changes his/her mind about participating at any time, you may withdraw him/her from the study without having to give a reason or justification, simply by telling me.

There are no known risks to participating in this study, nor are there direct benefits. However, your child's participation in this study may be of value to research investigating trilingual children's language choice and use.

What happens to the results of the study?

The results of this research will be presented in my doctoral thesis. The results may also be published in academic publications or be presented at academic conferences in the future. Your child will not be identifiable in any of the publications and/or presentations. No one but my supervisor and me will have access to your name or your child's name. All electronic data, audio recordings, digital transcripts and paper materials through which your identity could be revealed, will be stored either on a password-locked computer (recordings and their digital transcripts) or in a keylocked safe (paper materials) in a private, secure residence to which only I have access. The recordings and papers will be destroyed (shredded or permanently deleted) once the data have been analysed. Anonymised electronic data (for example, transcripts) will be kept indefinitely in order to be used for later research by other researchers or by me. However, please note that the confidentiality of your data is subject to normal legal requirements.

If you wish to receive a summary of the main findings of this study, kindly select the relevant box in the consent form. The school will receive an anonymised summary of the main findings of the study.

Who is organising this research?

This research is organised by me as an independent doctoral research project in conjunction with the Department of Education, University of Oxford, under the supervision of Prof. Victoria Murphy.

Kindly note that this project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee.

What should you do next?

Should you have any questions about this research, please do not hesitate to contact me, May Zantout, at the contact details provided below. If you agree to participate and allow your child to participate in this study, kindly complete the consent form.

Contact for further information or follow up

At any time during or after the study, please feel free to contact me, May Zantout at the contact details provided below. Should you have any comments or concerns about this study at any time, and you are not satisfied with the answers I have given you, you can contact my supervisor, Prof. Victoria Murphy, by e-mailing her at the following address: victoria.murphy@education.ox.ac.uk

May Zantout
Department of Education
15 Norham Gardens, Oxford, UK, OX2 6PY
Tel: +961(3)800903, e-mail: may.zantout@lmh.ox.ac.uk

Research Consent Form

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION
15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Director Professor Ernesto Macaro

Bridging Sociopragmatic and Psycholinguistic Theory on Trilingual Codeswitching

Researcher: May Zantout

Declaration of Consent:

I have read the participant information sheet and have had the opportunity to ask questions about the study and receive satisfactory answers to questions. I understand that I may withdraw my child from the study without penalty at any time by advising the researchers, and any data already recorded will be discarded immediately. I understand that this project has been reviewed by, and received ethics clearance through, the University of Oxford's Central University Research Ethics Committee. I understand who will have access to the data I provide, that my personal data will be treated in total confidence and kept securely in a password-controlled server, and what will happen to the data at the end of the project. I understand how to raise a concern and make a complaint, and agree to participate and/or for my child to participate in this study as indicated below.

Please fill in the following information.

Child Participation

Child's Name

Child's Grade

Child's Immersion Program

☐ I give permission for my child to take part in the study.

Parent's Signature

Child's Signature

Date

Please tick the following boxes as you see fit and include relevant information.

☐ I confirm that I have read the associated information sheet and understand the intent and purpose of this research.

☐ I agree that data captured by this research can be shared among the research team on this project within the limits specified in the information sheet.

☐ I would like to receive a summary of the findings at the following e-mail address: _____

*Please note that you may find out more about this study by contacting the researcher at her e-mail address: may.zantout@lmh.ox.ac.uk or her contact number: +961(3)800903. Please feel free to ask any questions or raise any concerns you may have. If you feel that there are still some concerns you wish to address further, you may contact my supervisor, Prof. Victoria Murphy, at victoria.murphy@education.ox.ac.uk.

Name of researcher: May Zantout

Signature: _____

Date: _____

Letter to the School President

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk
www.education.ox.ac.uk

Director Prof. Ernesto Macaro

President's Name
School Location, Ras Beirut
January __, 2016

Dear President X,

I am writing to enquire about conducting some research in your school in the current academic year. I am a DPhil student at the University of Oxford, supervised by Prof. Victoria Murphy (Professor of Applied Linguistics). In my research project, entitled *Bridging Sociopragmatic and Psycholinguistic Theory on Trilingual Codeswitching*, I will explore how and why children residing in Lebanon use and mix different languages when engaging in tasks and social interactions.

Description of the Study

This study aims to explore the social and cognitive motivations for codeswitching by trilingual Lebanese youth enroled in English- and French-immersion programs. The study will first ask participants to fill out a language background questionnaire and take part in a series of tasks, which will take place at the participants' homes and be audiorecorded. These tasks include a trilingual picture-naming task and a narrative task, where participants are asked to name objects (e.g. book, ball, pen) and narrate stories in the language(s) specified by the researcher. Participants will then partake in interactions with peers from differnet language backgrounds in a one-to-one context to examine the frequency of codeswitching and the influence of the interlocutor on the participants' language choice and behaviour. Optional stimulated recalls are then offered to ask participants what motivated their language switching.

Role of the School

The research aims to recruit students in grade 6 (1st Intermediate) and grade 12 (Baccalaureate II) enrolled in the Lebanese program and who have been students at this school since nursery. Your help would be highly appreciated if you could

identify students meeting these criteria. Volunteering participants will provide their assent, as well as their parental consent, to take part in the study which will be conducted at their places of residence, not on campus.

By participating in the research, your school, my alma mater, would be contributing to a project that will deepen our understanding of how and why multilingual children make linguistic choices and employ codeswitching strategies.

The commitment from the school would be to allow the dissemination of information sheets and consent forms by the class advisors to students in these grades meeting the criteria mentioned above, as identified by the teachers. The distribution of these documents could take place during the normal class time when the advisors are teaching these grades.

Ethical Considerations

The University of Oxford has strict ethical procedures on conducting ethical research with young people, consistent with current British Educational Research Association guidelines. Before beginning the research, I will inform parents about the research and answer any questions they may have. The participants may change their minds about participating at any point in time during the study with no consequence.

All participants, including students, parents, and the school, would be made anonymous in all research reports. The data collected would be kept strictly confidential, available only to my supervisor and me, and it will not be used other than specified. All audio-recordings will be destroyed at the end of the research period and kept in locked conditions until then. Kindly note that this project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee. I have enclosed copies of the information sheet and consent form to be given to the parents with this letter.

If you are interested in receiving a summary of the main findings of the study, an anonymised summary can be provided after the project is submitted.

If you agree to take part in the study, or need more information about what is involved, please contact me. Whether or not you feel it would be appropriate for your school to participate, I would be grateful if you would complete the pro-forma below, and return it to me via e-mail. If at any time you have concerns about this study, you can also contact my supervisor at victoria.murphy@education.ox.ac.uk

Thank you for your time and attention. I look forward to hearing from you.

Yours sincerely,
May Zantout
may.zantout@education.ox.ac.uk

Bridging Sociopragmatic and Psycholinguistic Theory on Trilingual Codeswitching

May Zantout

Oxford University, Department of Education

School Name

School Location, Beirut

President's Name

☐ We do not wish to participate in this project.

☐ We would like to find out more about this project.

☐ We would like to take part in this project.

If you would like further information, or are interested in taking part, please give the name of a contact person for your school, and details of the best way to contact him or her.

Contact name: _____

Contact email: _____

Contact telephone number: _____

Please return this form via e-mail.

Thank you for your help.

Appendix F

Breakdown of the Tasks

***Session Time does not include breaks between tasks.**

Please note that this breakdown of tasks was based on the initial data collection, yet some of the tasks have been excluded from the analysis of the present project.

Session Number	Task	Total Duration/Task (in minutes)
1	Language Background Questionnaire	15-20
Total Duration of Session 1: 15-20 minutes		
2	English OPT	20-30
2	Simon Task	10-15
2	Conversation Task (uncued)	10-15
Total Duration of Session 2: 33-50 minutes		
3	French OPT	20-30
3	Arabic Verbal Fluency Task	10
3	Conversation Task (English only)	10-15
Total Duration of Session 3: 45-65 minutes		
4	Arabic OPT	20-30
4	Trilingual picture-naming task	15
4	Narration Task (uncued)	5-15
4	French Verbal Fluency Task	10
Total Duration of Session 4: 50-70 minutes		
5	Narration Task (English-French-Arabic)	5-15
5	Conversation Task (Arabic only)	10-15
5	Stimulated Recall (post-narration)	10-15
Total Duration of Session 5: 25-45 minutes		
6	Narration Task (French only)	5-10
6	Backward Digit-Span Task	5
6	Conversation Task (Arabic-English)	10-15
6	Stimulated Recall (post-conversation)	10-15
Total Duration of Session 6: 30-40 minutes		
7	Narration Task (Arabic only)	5-15
7	Narration Task (French-English)	5-15
7	Conversation Task (French only)	10-15

7	English Verbal Fluency Task	10
Total Duration of Session 7: 30-55 minutes		
8	Conversation Task (French-English)	10-15
8	Narration Task (English only)	5-15
8	Conversation Task (English-Arabic-French)	10-15
Total Duration of Session 8: 25-45 minutes		
9	Conversation Task (Arabic-French)	10-15
9	Narration Task (English-Arabic)	5-15
9	Narration Task (French-Arabic)	5-15
Total Duration of Session 8: 25-45 minutes		

Session/Task Counterbalancing Schemes

*These schemes were based on the data collection as set out initially, yet some of the tasks included in this schedule are not analysed in the current thesis.

Session Number	**Scheme 1**	**Scheme 2**	**Scheme 3**	**Scheme 4**	**Scheme 5**	**Scheme 6**
1	Language Background Questionnaire					
2	English OPT	Conversation Task (uncued)	French OPT	Arabic OPT	Trilingual picture-naming task	English Verbal Fluency Task
2	Simon Task	Arabic Verbal Fluency Task	Narration Task (uncued)	Conversation Task (uncued)	English Verbal Fluency Task	Conversation Task (uncued)
2	Conversation Task (uncued)	Narration Task (uncued)	Narration Task (English-Arabic)	Narration Task (uncued)	Narration Task (uncued)	Stimulated Recall (post-conversation)
2		Backward Digit-Span Task	English Verbal Fluency Task	Conversation Task (English-Arabic-French)	Narration Task (English only)	Arabic OPT
3	French OPT	Narration Task (English-Arabic)	Conversation Task (uncued)	Conversation Task (French only)	Conversation Task (uncued)	Conversation Task (French only)
3	Arabic Verbal Fluency Task	English OPT	Arabic OPT	Conversation Task (Arabic only)	Simon Task	French Verbal Fluency Task
3	Conversation Task (English only)	Trilingual picture-naming task	Narration Task (English only)	Narration Task (English-Arabic)	Backward Digit-Span Task	Narration Task (uncued)
3		Narration Task (English-French-Arabic)	Narration Task (French only)	Conversation Task (French-English)	Narration Task (French-Arabic)	Narration Task (Arabic only)
4	Arabic OPT	Narration Task (Arabic only)	Arabic Verbal Fluency Task	Stimulated Recall (post-conversation)	Conversation Task (English only)	French OPT

4	Trilingual picture-naming task	Conversation Task (English only)	Narration Task (English-French-Arabic)	English OPT	Narration Task (Arabic only)	Narration Task (English-French-Arabic)
4	Narration Task (uncued)	Stimulated Recall (post-conversation)	Conversation Task (Arabic-French)	French Verbal Fluency Task	Conversation Task (French-English)	Conversation Task (Arabic only)
4	French Verbal Fluency Task	Conversation Task (English-Arabic-French)		Narration Task (French only)		Conversation Task (English-Arabic-French)
5	Narration Task (English-French-Arabic)	Narration Task (French-English)	Trilingual picture-naming task	French OPT	Narration Task (French only)	Narration Task (English-Arabic)
5	Conversation Task (Arabic only)	Narration Task (English only)	Conversation Task (Arabic only)	Narration Task (Arabic only)	Conversation Task (French only)	Conversation Task (English only)
5	Stimulated Recall (post-narration)	Conversation Task (Arabic-French)	English OPT	Narration Task (French-Arabic)	Arabic OPT	Narration Task (French only)
5			Conversation Task (English-Arabic-French)		Conversation Task (English-Arabic-French)	
6	Narration Task (French only)	Simon Task	Backward Digit-Span Task	Arabic Verbal Fluency Task	English OPT	Narration Task (French-English)
6	Backward Digit-Span Task	Arabic OPT	Conversation Task (English only)	Conversation Task (Arabic-French)	French Verbal Fluency Task	Conversation Task (French-English)
6	Conversation Task (Arabic-English)	Conversation Task (Arabic-English)	Narration Task (French-Arabic)	Narration Task (English-French-Arabic)	Narration Task (French-English)	Trilingual picture-naming task
6	Stimulated Recall (post-conversation)		Stimulated Recall (post-narration)	Stimulated Recall (post-narration)	Stimulated Recall (post-narration)	

7	Narration Task (Arabic only)	French OPT	French Verbal Fluency Task		French OPT	Simon Task
7	Narration Task (French-English)	Conversation Task (Arabic only)	Narration Task (Arabic only)	Conversation Task (English only)	Conversation Task (Arabic-French)	Conversation Task (Arabic-English)
7	Conversation Task (French only)	Narration Task (French only)	Conversation Task (French-English)	Trilingual picture-naming task	Conversation Task (Arabic only)	Narration Task (English only)
7	English Verbal Fluency Task	Stimulated Recall (post-narration)		English Verbal Fluency Task	Stimulated Recall (post-conversation)	
8	Conversation Task (French-English)	Conversation Task (French only)	Narration Task (French-English)	Backward Digit-Span Task	Arabic Verbal Fluency Task	English OPT
8	Narration Task (English only)	English Verbal Fluency Task	Conversation Task (Arabic-English)	Narration Task (French-English)	Conversation Task (Arabic-English)	Narration Task (French-Arabic)
8	Conversation Task (English-Arabic-French)			Simon Task	Narration Task (English-Arabic)	Stimulated Recall (post-narration)
8						Arabic Verbal Fluency Task
9	Conversation Task (Arabic-French)	French Verbal Fluency Task	Conversation Task (French only)	Conversation Task (Arabic-English)	Narration Task (English-French-Arabic)	Backward Digit-Span Task
9	Narration Task (English-Arabic)	Conversation Task (French-English)	Stimulated Recall (post-conversation)	Narration Task (English only)		Conversation Task (Arabic-French)
9	Narration Task (French-Arabic)	Narration Task (French-Arabic)	Simon Task			

APPENDIX G
PILOT STUDY

To pilot the study's instruments, the researcher recruited 17 (11 EMI and 10 FMI) participants belonging to the same age group and grade as the targeted sample. Upon receiving ethical clearance and the assent and consent of the adolescents and their parents, the researcher collected data over the course of three weeks (July-August 2015), adopting the main study's methodology. Noting the importance of amending the Verbal Fluency Task (to modify the instructions) and Narration Task (to include additional pictures), the researcher re-tested the modified tasks with four additional participants (2 EMI, 2 FMI) in September 2015. A summary of the modifications and partial descriptive analyses are provided below.

The following table (Table 1) summarises the issues that arose in the first pilot data collection phases and how these were addressed through modifications. These tasks were then retested with four new participants (two EMI and two FMI).

Table 1 Summary of Pilot Study Modifications

Task Name	**Problematic Items**	**Modification**
Verbal Fluency Task	Standardised categories, such as 'words starting with 'a' and 'e'', were problematic, as these two vowels represent the same letter and sound in Arabic, but they are different in English/French. Phrasing of 'words starting with' (a letter) elicited MSA words, as LA does not have a standardised script.	A non-standardised letter using the consonants 'm' and 'k' replaced the vowels. The category phrasing was modified to 'words *sounding* with' along with the phonetic sound of a letter.
Narration Task	Storyboards originally consisted of eight detailed pictures/story, yet the task duration ranged between 1.5-6 minutes with the 17 participants.	A storyboard with 18 pictures was piloted with the four additional participants in September. Their production ranged between 7 and 12 minutes, thus overcoming the challenge in comparing performance across tasks.

By testing the tasks' ability to elicit speech from the participants, as well as show differences within and between the groups, the pilot study indicated that the instruments are suitable. Partial descriptive analyses revealed preliminary, yet promising, indications regarding the possible findings of the main study. Descriptive analyses of three EMI and three FMI participants' results on the Verbal Fluency Task and Narration Task reveal that FMI participants were able to produce more items and sustain longer utterances in all three languages than their EMI counterparts. For the Verbal Fluency Task (Fig. 1, showing mean number of items/language), EMI participants listed more items in English than Arabic or French (in absolute terms), whereas their FMI counterparts performed better on all categories yet no statistical analyses were performed to confirm that.

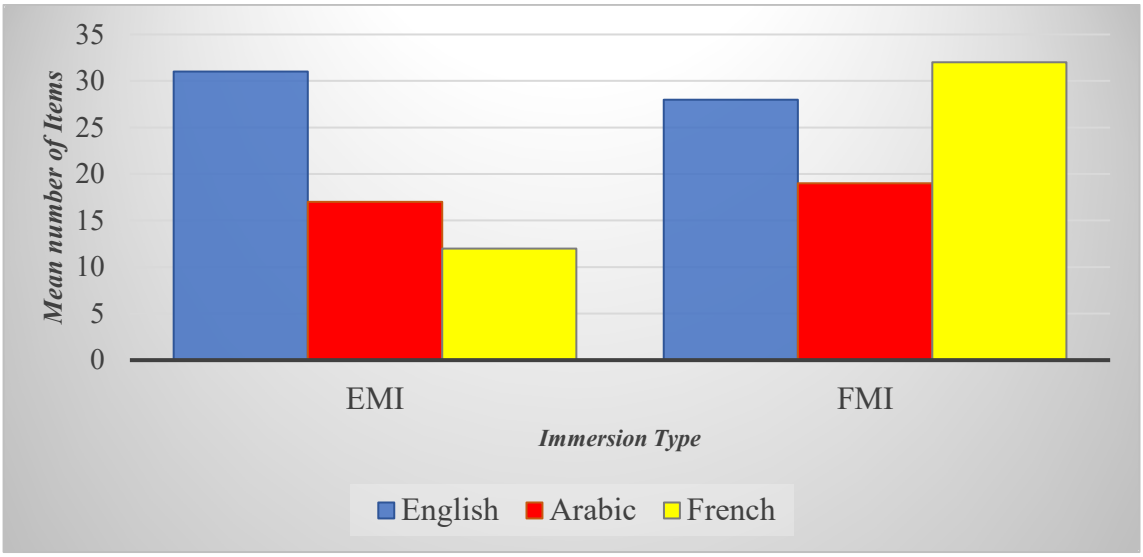


Fig. 1 Mean Number of Item/Language for EMI and FMI Participants in Pilot Study

In the narration task, EMI participants codeswitched mainly to English as opposed to Arabic or French, whereas FMI participants switched between the three languages relatively more equally (Fig. 2, showing switch directionality for each group). These results were expected by both FMI and EMI participants, given the ubiquity of English and extended exposure to and instruction in all three languages. For both the groups, the L2 (English or French) seems to be more dominant than Arabic. Switch costs in these tasks, as well as other tasks, further confirmed this observation and are expected to reveal that the control mechanisms involved in switching are less detectable in more balanced trilinguals (FMI participants), as they are accustomed to switching between the three languages.

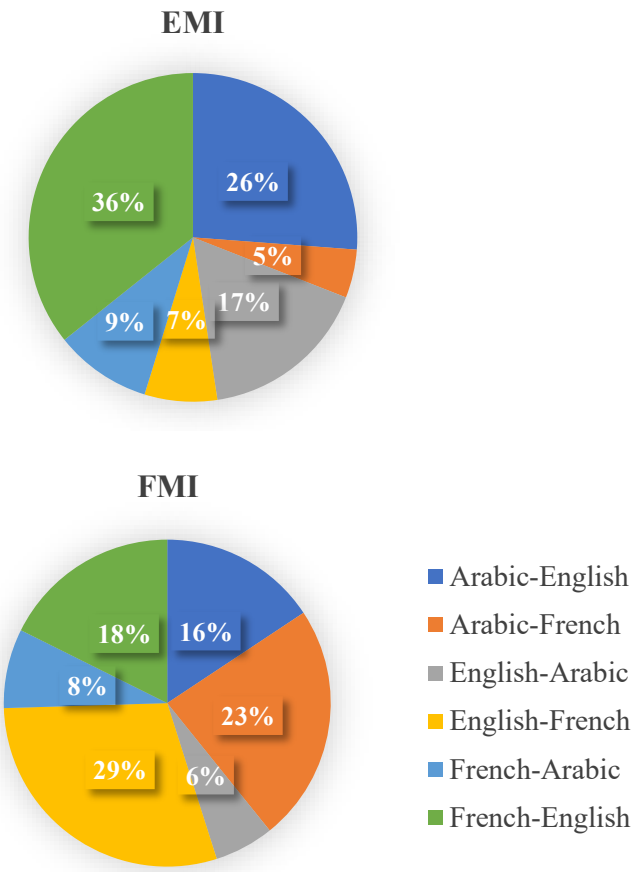


Fig. 2 Frequency of Switch Directionality/Group in Percentage Form

Appendix H

Chapter 4 Supplementary Analyses

This appendix offers additional analyses which were conducted to explore the data more due to an interest in examining results in greater detail. It also offers analyses of exploratory models which were performed when the original models which were adopted in the main thesis revealed no significant results. The exploratory models generally do not differ significantly in fit from the selected models, and they usually zoom in on a predictor's influence of another variable or on correlations between variables.

I. Research Question – Supplementary Analyses

A. Flanker Task

The following section offers analyses of within-group analyses of the Flanker data (i.e., with the data split by group), despite the non-significant main effect of Group and/or non-significant interactions as reported in Chapter 4. These analyses were performed to examine congruence x monitoring contrasts for each group.

Table A4-1 Flanker Task RT Post-hoc Analyses (Congruence-Monitoring Contrasts) for EMI participants

		Low								**Medium**								**High**							
		Congruent				Incongruent				Congruent				Incongruent				Congruent				Incongruent			
		β	SE	T	p	β	SE	t	p	β	SE	T	p	B	SE	T	p	β	SE	T	p	β	SE	t	p
Low																									
	Congruent	-				0.10	0.02	4.59	0.00	0.01	0.02	0.29	0.77	0.11	0.02	4.97	0.00	0.12	0.02	5.46	0.00	0.18	0.02	8.05	0.00
	Incongruent	0.10	0.02	4.59	0.00	-				0.09	0.02	4.30	0.00	0.01	0.02	0.38	0.71	0.02	0.02	0.87	0.39	0.08	0.02	3.46	0.00
Medium																									
	Congruent	0.01	0.02	0.29	0.77	0.09	0.02	4.30	0.00	-				0.10	0.02	4.68	0.00	0.11	0.02	5.17	0.00	0.17	0.02	7.76	0.00
	Incongruent	0.11	0.02	4.97	0.00	0.01	0.02	0.38	0.71	0.10	0.02	4.68	0.00	-				0.01	0.02	0.49	0.63	0.07	0.02	3.08	0.00
High																									
	Congruent	0.12	0.02	5.46	0.00	0.02	0.02	0.87	0.39	0.11	0.02	5.17	0.00	0.01	0.02	0.49	0.63	-				0.06	0.02	2.59	0.01
	Incongruent	0.18	0.02	8.05	0.00	0.08	0.02	3.46	0.00	0.17	0.02	7.76	0.00	0.07	0.02	3.08	0.00	0.06	0.02	2.59	0.01				

β= Estimate Coefficient, SE= Standard Error, t= t-value, p= p-value

Table A4-2 Flanker Task RT Post-hoc Analyses (Congruence-Monitoring Contrasts) for FMI participants

		Low								**Medium**								**High**							
		Congruent				Incongruent				Congruent				Incongruent				Congruent				Incongruent			
		β	SE	t	p	B	SE	t	p	B	SE	T	p	B	SE	T	p	β	SE	T	p	β	SE	t	p
Low																									
	Congruent					0.11	0.02	5.10	0.00	0.06	0.02	2.68	0.00	0.11	0.02	4.45	0.00	0.10	0.02	4.45	0.00	0.19	0.02	8.46	
	Incongruent	0.11	0.02	5.10	0.00					0.17	0.02	7.78	0.00	0.11	0.02	5.23	0.00	0.01	0.02	0.65	0.52	0.07	0.02	3.46	0.11
Medium																									
	Congruent	0.06	0.02	2.68	0.00	0.17	0.02	7.78	0.00					0.17	0.02	7.91	0.00	0.16	0.02	7.91	0.00	0.24	0.02	11.14	0.06
	Incongruent	0.11	0.02	5.23	0.00	0.00	0.02	0.13	0.90	0.17	0.02	7.91	0.00					0.02	0.02	0.78	0.44	0.07	0.02	3.23	0.11
High																									
	Congruent	0.10	0.02	4.45	0.00	0.01	0.02	0.65	0.52	0.16	0.02	7.91	0.00	0.02	0.02	0.78	0.44					0.09	0.02	4.01	0.00
	Incongruent	0.19	0.02	8.46	0.00	0.07	0.02	3.36	0.00	0.24	0.02	11.14	0.00	0.07	0.02	3.23	0.00	0.09	0.02	4.01	0.00				

β= Estimate Coefficient, SE= Standard Error, t= t-value, p= p-value

B. Simon Task

Although no significant differences were indicated by the model, it was of interest to obtain further contrasts to gain a better understanding of the results. Furthermore, the use of linear mixed-effects models allows further analyses through refitting and releveling the model to obtain contrasts. These supplementary analyses were simply performed to better acquaint myself with the data.

Table A4-3 Simon Accuracy Between-Group Contrasts

	β	**SE**	**z-value**	**p-value**
Congruent	0.29	0.22	1.30	0.20
Incongruent	0.14	0.15	0.89	0.38
Control	0.37	0.23	1.59	0.11

EMI

As indicated in Table A4-4, EMI participants responded significantly less accurately on incongruent trials than on congruent (β= 0.56, SE= 0.18, t= 3.08, p< 0.01) and control trials (β= 0.61, SE= 0.19, t= 3.28 p< 0.01). All other things being equal, responses on incongruent trials were 1.76 times less likely to be accurate than congruent trials and1.84 times less likely to be accurate than control trials. There was no significant difference in accuracy rates between control and congruent trials (p>0.05).

Table A4-4 EMI Simon Accuracy Congruence Contrasts

	Congruent				*Incongruent*				*Control*			
	B	*SE*	*T*	*P*	*B*	*SE*	*T*	*p*	*B*	*SE*	*T*	*p*
Congruent					0.56	0.18	3.08	0.00	0.37	0.23	1.59	0.12
Incongruent	0.56	0.18	3.08	0.00					0.61	0.19	3.28	0.00
Control	0.37	0.23	1.59	0.12	0.61	0.19	3.28	0.00				

FMI

Like their EMI counterparts, FMI participants performed more accurately on the control trials and congruent trials than the incongruent trials. The incongruent trial performance differed significantly from the control trials (β= 1.12, SE=0.21, z= 5.38, p<0.0001); FMI participants were 3.06 times as likely to be less accurate on incongruent trials relative to control trials. Similarly, the incongruent trial accuracy was significantly lower than that on congruent trials (β= 0.99, SE= 0.30, z= 4.99, p< 0.0001). Accuracy on control and congruent trials were not significantly different (p>0.05). The results are summarised in Table A4-5.

Table A4-5 FMI Simon Accuracy Congruence Contrasts

	Congruent				*Incongruent*				*Control*			
	B	*SE*	*z*	*P*	*B*	*SE*	*Z*	*p*	*B*	*SE*	*Z*	*p*
Congruent					0.99	0.20	4.99	0.00	0.13	0.25	0.53	0.60
Incongruent	0.99	0.20	4.99	0.00					1.12	0.21	5.38	0.00
Control	0.13	0.25	0.53	0.60	1.12	0.21	5.38	0.00				

In terms of Simon Task RT, comparing the two groups, there were no significant differences in RTs in any condition. Table A4-6 summarises the between-group analyses (p>0.05 for all contrasts).

Table A4-6 Simon RT Between-Group Contrasts

	B	***SE***	***t-value***	***p-value***
Congruent	0.29	0.22	1.30	0.20
Incongruent	0.14	0.15	0.89	0.38
Control	0.37	0.23	1.59	0.11

EMI

For the EMI group, response latencies, unlike accuracy rates, were significantly different across the three conditions: Congruent-trial reaction times were significantly smaller than those on the control (β= 0.22, SE=0.04, t= 2.80, p< 0.001); and incongruent trials (β= 0.49, SE=0.04, t= 12.66, p< 0.0001). Furthermore, control-trial RTs were significantly smaller than incongruent-trial RTs (β= 0.38, SE=0.04, t= 9.86, p< 0.0001). The analysis is summarised in Table A4-7.

Table A4-7 EMI Simon RT Scores by Congruence Contrasts

	Congruent				*Incongruent*				*Control*			
	β	*SE*	*t*	*P*	*β*	*SE*	*T*	*p*	*B*	*SE*	*t*	*p*
Congruent									0.22	0.04	2.80	0.00
Incongruent	0.49	0.04	12.66	0.00					0.38	0.04	9.86	0.00
Control	0.22	0.04	2.80	0.00	0.38	0.04	9.86	0.00				

FMI

The FMI group also exhibited significantly different RTs across the three conditions. Reaction times on incongruent trials were significantly higher than those on congruent trials (β= 0.49, SE=0.04, t= 12.84, p< 0.0001) and control trials (β= 0.37, SE=0.04, t= 9.71, p< 0.0001). Furthermore, RTs on control trials were significantly higher than RTs on congruent trials (β= 0.12, SE=0.04, t= 3.13, p< 0.01).). As with the EMI group, therefore, participants were fastest to respond on congruent trials, followed by control trials, and finally incongruent trials. These results are summarised in Table A4-8.

Table A4-8 FMI Simon RT Scores by Congruence Contrasts

	Congruent				*Incongruent*				*Control*			
	β	*SE*	*T*	*p*	*β*	*SE*	*T*	*p*	*β*	*SE*	*t*	*P*
Congruent					0.49	0.04	12.84	0.00	0.12	0.04	3.13	0.00
Incongruent	0.49	0.04	12.84	0.00					0.37	0.04	9.71	0.00
Control	0.12	0.04	3.13	0.00	0.37	0.04	9.71	0.00				

II. Language Experience – Post-Hoc Analyses

Based on the scores from both the language use and language exposure variables, two post-hoc variables were created. Similar to the HPBT variable created for language proficiency, the post-hoc Language Use Variable (LUV) and post-hoc Language Exposure Variable (LEXV) were extracted to offer a clearer description of the two variables. This was based on the notion that the use of and exposure to each language implies a lesser, equal, or more frequent use of another language.

Creating these post-hoc variables was originally performed to reduce multicollinearity; however, the use of separate models in the models which included the variables resolved this issue. Therefore, the purpose of these post-hoc variables was to then provide an overall description of language use and language exposure.

The computation of the post-hoc variables depended on significant differences between each of the three languages for every participant. As there are 13 potential groups to which participants may belong, the post-hoc variable results are combined for both groups in the mixed-effects models. However, for a more detailed description of group differences, Table 1 reveals differences for each group. As observed in the results, EMI participants belonged to only three of 13 potential groups, with the majority (19 of 30) belonging to the categorisation indicating the highest use of English, followed by Arabic, and then French. FMI participants, on the other hand, were more varied in their self-reported language use. Of the 30 FMI participants, 10 participants reported using Arabic most frequently, followed by English, and the French, and 8 participants reported using Arabic and English equally, followed by French. This is an unexpected finding, given that FMI participants were generally expected to use French at the highest or second-highest frequency. The minimal use of French is indicative of the general status to which the French language has been relegated in terms of its utility in social contexts. This is further discussed in Section 6.6.

Table 1. Language Use Post-Hoc Variable (LUV); '+' = highest use; '=' = equal use; '-' = lowest use; '/' = second-highest use (middle).

Description	Arabic	English	French	Label	EMI	FMI
Arabic Highest Use, English and French used equally	+	=	=	AH	0	1
English Highest Use, Arabic and French used equally	=	+	=	EH	0	2
French Highest Use, Arabic and English used equally	=	=	+	FH	0	0
Highest Arabic, then English, then French	+	/	-	A-E-F	4	10
Highest Arabic, then French, then English	+	-	/	A-F-E	0	3
Highest English, then Arabic, then French	/	+	-	E-A-F	19	2

Highest English, then French, then Arabic	-	+	/	E-F-A	0	0
Highest French, then Arabic, then English	/	-	+	F-A-E	0	0
Highest French, then English, then Arabic	-	/	+	F-E-A	0	0
Arabic-English Equal Use, French minimal use	=	=	-	AE-F	7	8
Arabic-French Equal Use, English minimal use	=	-	=	AF-E	0	0
English-French Equal Use, Arabic minimal use	-	=	=	EF-A	0	0
Arabic-English-French Equal Use	=	=	=	AEF	0	4

As some categories did not include any participants, and with some categories including very few participants, it was important to collapse some of the categories to make this variable (statistically) computationally viable. Accordingly, the LUV was collapsed, as presented in Table 2. Despite the variation in numbers for each category, the mixed-effects models could overcome (to a certain extent) the unequal numbers of participants per category.

Table 2. LUV Categorisations EMI and FMI

Description	Arabic	English	French	Label	EMI	FMI	Total N
Arabic Highest Use, English and French used equally	+	=	=	AH	0	1	4
Highest Arabic, then French, then English	+	-	/	A-F-E	0	3	
Highest Arabic, then English, then French	+	/	-	A-E-F	4	10	14
English Highest Use, Arabic and French used equally	=	+	=	EH	0	2	23
Highest English, then Arabic, then French	/	+	-	E-A-F	19	2	
Arabic-English Equal Use, French minimal use	=	=	-	AE-F	7	8	15
Arabic-English-French Equal Use	=	=	=	AEF	0	4	4

As with the ‘Language Use’ construct, ‘Language Exposure’ was further analysed in the same manner of categorising participants in terms of the relative

differences in exposure they each had to each language. The initial results of the categorisation for the extracted variable, LEXV, are summarised in Table 3.

Table 3. Language Exposure Post-Hoc Variable (LEXV); '+' = highest use; '=' = equal use; '-' = lowest use; '/' = second-highest use (middle).

Description	Arabic	English	French	Label	EMI	FMI
Arabic Highest Exposure, equal English and French exposure	+	=	=	AH	1	8
English Highest Exposure, equal Arabic and French exposure	=	+	=	EH	2	1
French Highest Exposure, equal Arabic and English exposure	=	=	+	FH	0	0
Highest Arabic, then English, then French	+	/	-	A-E-F	3	1
Highest Arabic, then French, then English	+	-	/	A-F-E	0	1
Highest English, then Arabic, then French	/	+	-	E-A-F	17	5
Highest English, then French, then Arabic	-	+	/	E-F-A	0	0
Highest French, then Arabic, then English	/	-	+	F-A-E	0	0
Highest French, then English, then Arabic	-	/	+	F-E-A	0	0
Arabic-English Equal Exposure, French minimal exposure	=	=	-	AE-F	7	10
Arabic-French Equal Exposure, English minimal exposure	=	-	=	AF-E	0	0
English-French Equal Exposure, Arabic minimal exposure	-	=	=	EF-A	0	0
Arabic-English-French Equal Exposure	=	=	=	AEF	0	4

As with the reduction of categories of language use, a reduction of categories of language exposure was applied. Categories which had no participants were removed from the variable, and categories which had very few participants were collapsed with other sub-categories as appropriate. This is illustrated in Table 4.

Table 4 Language Exposure Post-Hoc Variable (LEXV); '+' = highest use; '=' = equal use; '-' = lowest use; '/' = second-highest use (middle).

Description	Arabic	English	French	Label	EMI	FMI	Total

							N
Arabic Highest Exposure, equal English and French exposure	+	=	=	AH	1	8	9
English Highest Exposure, equal Arabic and French exposure	=	+	=	EH	2	1	3
Highest Arabic, then English, then French	+	/	-	A-E-F	3	1	5
Highest Arabic, then French, then English	+	-	/	A-F-E	0	1	
Highest English, then Arabic, then French	/	+	-	E-A-F	17	5	22
Arabic-English Equal Exposure, French minimal exposure	=	=	-	AE-F	7	10	17
Arabic-English-French Equal Exposure	=	=	=	AEF	0	4	4

It should be noted that while the construction of these variables was aimed at further investigating the influence of language use and language exposure on cognitive and linguistic task performance, the use of the original variables revealed the same findings. As such, the post-hoc variables were not included in the final draft of this thesis. However, the decision to include these post-hoc analyses stems from the relative ease to reveal general trends and facilitate the discussion section in Chapter 6.

III. Research Question 2 – Supplementary Analyses

As indicated in Chapter 4, results from the cognitive tasks were not predicted by language proficiency and/or language experience. Accordingly, exploratory analyses were performed to assess whether significant findings may be obtained by manipulating the models further, generally through a reduction of fixed effects. The results of the analyses are presented here.

A. Conflict Effect

1. Oral Proficiency

Model 1 – Full Model Estimates

Estimated coefficients, standard errors (SE), t values, and p-values from the linear mixed effects models run on log-transformed Conflict Effect scores.
Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	Conflict Effect Scores			
	Estimate	**SE**	**t-value**	**Pr(>|t|)**
Intercept	4888.19	4903.08	1.00	0.32
Group (EMI vs. FMI)	3067.88	2967.03	1.03	0.31
OPT Arabic	455.23	655.73	0.69	0.49
OPT English	564.30	574.79	0.98	0.33
OPT French	862.15	714.43	1.21	0.23
Group*OPT Arabic	259.42	399.82	0.65	0.52
Group*OPT English	-356.17	362.34	0.98	0.33
Group*OPT French	519.14	405.91	1.28	0.21
OPT Arabic*OPT English	52.25	76.85	0.68	0.50
OPT Arabic*OPT French	82.96	93.37	0.89	0.38
OPT English*OPT French	98.82	82.06	1.20	0.24
OPT Arabic*OPT English*OPT French	9.50	10.75	0.88	0.38
Group*OPT Arabic*OPT English	30.27	48.25	0.63	0.53
Group*OPT Arabic*OPT French	47.24	53.32	0.89	0.38
Group*OPT English*OPT French	60.24	47.94	1.26	0.22
Group*OPT Arabic*OPT English*OPT French	5.50	6.26	0.88	0.39

Formula: Conflict Effect ~ Group OPT Scores +(1|Participant)*

a. OPT Scores

Model A:				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	94.91	1	57	<0.0001
OPT Scores	1.11	2	57	0.29

Formula: Conflict Effect ~ OPT scores + (1|Participant)

Model B:				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	103.76	1	57	<0.0001
Group	0.74	1	57	0.38
OPT Scores	1.26	2	57	0.27

Formula: Conflict Effect ~ Group + OPT scores + (1|Participant)

b. VFT Scores

Model C:				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	97.10	1	57	<0.0001
VFT	0.79	2	57	0.47

Formula: Conflict Effect ~ VFT scores + (1|Participant)

Although the simpler models revealed no significant differences, a model of a more complex structure was tested to see whether a level within the VFT variable (and interactions within this variable) as well as interactions with group might reveal a difference. The variable was therefore split into the three levels, each being examined as a separate predictor. No significant results were obtained. A summary of the model's output is presented next.

Model D: Estimated coefficients, standard errors (SE), t values, and p-values from the linear mixed effects models run on Conflict Effect scores.

	Conflict Effect Scores			
	Estimate	**SE**	**t-value**	**Pr(>|t|)**
Intercept	3102.06	3405.78	0.91	0.37
Group	2783.97	2897.88	0.96	0.34
VFT Arabic	29.11	30.60	0.95	0.35
VFT English	26.22	26.55	0.99	0.33
VFT French	28.50	45.64	0.62	0.54
Group*VFT Arabic	27.27	25.36	1.08	0.29
Group*VFT English	24.87	23.90	1.04	0.30
Group*VFT French	22.01	27.17	0.81	0.42
VFT Arabic*VFT English	0.24	0.23	1.05	0.30
VFT Arabic*VFT French	0.25	0.42	0.58	0.56
VFT English*VFT French	0.22	0.31	0.73	0.47
VFT Arabic*VFT English*VFT French	0.00	0.00	0.69	0.49
Group*VFT Arabic*VFT English	0.24	0.21	1.16	0.25
Group*VFT Arabic*VFT French	0.20	0.25	0.82	0.42
Group*VFT English*VFT French	0.19	0.20	0.94	0.35
Group*VFT Arabic*VFT English*VFT French	0.00	0.00	0.97	0.34

Formula: Conflict Effect ~ Group VFT Arabic * VFT English * VFT French + (1|Participant) + (1|Language)*

c. HPBT x OPT

Model E:				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	102.99	1	58	<0.0001
HPBT	1.78	1	58	0.18

Formula: Conflict Effect ~ HPBT + (1|Participant)

Model F:				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	100.89	1	57	<0.0001
HPBT	1.76	1	57	0.19
OPT.A	0.12	2	57	0.73
OPT.E	1.69	2	57	0.19
OPT.F	0.00	2	57	0.99

Formula: Conflict Effect ~ HPBT + OPT Scores + (1|Participant)

d. HPBT x VFT

Estimated coefficients, standard errors (SE), t values, and p-values from the linear mixed effects models run on log-transformed Conflict Effect scores. Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	Conflict Effect Scores			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	2252.24	3521.76	0.64	0.53
HPBT (Bilingual vs. Trilingual)	2069.89	3423.32	0.60	0.55
VFT Arabic	20.20	30.95	0.65	0.52
VFT English	16.75	25.97	0.65	0.52
VFT French	17.61	22.54	0.78	0.44
HPBT*VFT Arabic	18.10	29.88	0.61	0.55
HPBT*VFT English	14.76	25.35	0.58	0.56
HPBT*VFT French	14.84	21.06	0.70	0.48
VFT Arabic*VFT English	0.15	0.23	0.65	0.52
VFT Arabic*VFT French	-0.15	0.20	0.75	0.45
VFT English*VFT French	-0.13	0.17	-0.78	0.44
VFT Arabic*VFT English*VFT French	0.00	0.00	0.75	0.46
HPBT*VFT Arabic*VFT English	0.13	0.22	0.58	0.57
HPBT*VFT Arabic*VFT French	0.13	0.18	0.68	0.50

HPBT*VFT English*VFT French	0.11	0.16	0.68	0.50
HPBT*VFT Arabic*VFT English*VFT French	0.00	0.00	0.65	-0.52

Formula: Conflict Effect ~ HPBT VFT Arabic*VFT English*VFT French + (1|Participant) + (1|Language)*

e. Language Exposure and Language Use

Although the models in the main analyses were selected based on their goodness-of-fit, and despite the non-significant results of language use and language exposure as predictors, it was of interest to assess whether 'Group' influenced the findings. Accordingly, the data was analysed without the Group variable for each of these constructs. The results were non-significant (p>0.05) for all main effects and interactions.

Model: Conflict Effect Score ~ Language Use				
	F-statistic	DF1	DF2	p-value
Intercept	104.51	1	56	<0.0001
Arabic Use	0.22	1	56	0.64
English Use	1.67	1	56	0.19
French Use	0.15	1	56	0.71

Model: Conflict Effect Score ~ Language Exposure				
	F-statistic	DF1	DF2	p-value
Intercept	97.10	1	56	<0.0001
Arabic Exposure	0.21	1	56	0.65
English Exposure	0.02	1	56	0.89
French Exposure	0.13	1	56	0.72

B. Simon Effect Scores

For Language Use and Group, the initial model reflected no significant main effects or interactions for/between these two variables, as indicated in the analysis below (Model 1). Accordingly, another model which only examined Language Use as a potential predictor was created. The model (Model 2) also revealed non-significant outcomes.

Model 1: Estimated coefficients, standard errors (SE), t values, and p-values from the linear mixed effects models run on log-transformed Simon Effect scores.
Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	Simon Effect Scores			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	3082.86	1973.13	1.56	0.13
Group (EMI vs. FMI)	171.00	413.02	0.41	0.68

Arabic Use	0.00	0.00	0.50	0.62
English Use	7.19	14.31	1.35	0.18
French Use	21.98	16.27	0.58	0.57
Code-switching Frequency	0.05	0.08	3.44	0.00**
Salient-cue Detection	0.00	0.00	2.42	0.02*
CSF*SalC	0.05	0.02	3.85	0.00**
Group*Arabic Use	1.06	3.10	0.81	0.43
Group*English Use	0.65	2.32	0.50	0.62
Group*French Use	0.03	0.11	0.37	0.71
Group*CSF	10.43	12.09	0.55	0.58
Group*SalC	7.19	14.31	0.13	0.90
Group*CSF*SalC	0.00	0.00	0.35	0.73

Formula: Simon Effect ~Group(Language Use+CSF+SalC) + CSF*SalC +(1+SalC*CSF|Participant)*

Model 2: Simon Effect Scores ~ Language Use				
	F-statistic	DF1	DF2	p-value
Intercept	331.97	1	56	<0.0001
Arabic	0.07	1	56	0.79
English	1.25	1	56	0.27
French	0.14	1	56	0.72

Language exposure also did not significantly influence Simon Effect scores in the more complex models. Removing 'Group' and all other variables and interactions as predictors, Language exposure was assessed as an exploratory variable of inhibitory control. The model revealed no significant main effects:

Model 2: Simon Effect Scores ~ Language Exposure				
	F-statistic	DF1	DF2	p-value
Intercept	310.71	1	56	<0.0001
Arabic	2.13	1	56	0.15
English	0.33	1	56	0.57
French	0.98	1	56	0.33

C. BDST Scores

As the VFT did not predict BDST scores, a model which only examined main effects of each score on the VFT as a predictor of working memory was created. The model was a fitted with BDST scores as the outcome variable and with VFT scores as three separate predictors with no interactions as explanatory variables. Group was not included, and random effects were by participant. The results revealed no significant main effects for any VFT score.

Model: BDST ~ VFT Arabic + VFT English + VFT French				
	F-statistic	DF1	DF2	p-value
Intercept	2080.81	1	57	<0.0001

VFT Arabic	0.57	1	57	0.45
VFT English	3.19	1	57	0.08
VFT French	0.02	1	57	0.90

This model examined the influence of Group, Language Exposure, CSF, and SalC on BDST performance. The model retained the structure of Model 1, with the exception of Language Use being removed and Language Exposure being added to the model. All main effects and interactions between the variables were retained in the structure of this model, and the quality of the model's fit was assessed through the same procedures outlined at the beginning of this section. Analyses, summarised in the table below, revealed that there were no significant main effects for or interactions between any of the variables (all p>0.05).

Estimated coefficients, standard errors (SE), t values, and p-values from the linear mixed effects models run on log-transformed Simon Effect scores. Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	BDST Scores			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	11454.80	7827.83	1.46	0.15
Group (EMI vs. FMI)	4563.83	4394.51	1.04	0.30
Arabic Exposure	114.43	78.31	1.46	0.15
English Exposure	122.56	75.82	1.61	0.11
French Exposure	103.17	66.85	1.54	0.13
Code-switching Frequency	0.51	0.90	0.56	0.58
Salient-cue Detection	0.27	0.40	0.69	0.49
CSF*SalC	0.01	0.02	0.53	0.60
Group*Arabic Exposure	45.61	43.96	1.04	0.31
Group*English Exposure	50.72	42.44	1.19	0.24
Group*French Exposure	47.84	42.56	1.12	0.27
Group*CSF	0.09	0.24	0.35	0.73
Group*SalC	0.12	0.63	0.19	0.85
Group*CSF*SalC	0.00	0.01	0.22	0.82

Formula: BDST ~ Group(Language Exposure + CSF + SalC) + CSF*SalC +(1|Participant) + (1|Language)*

In order to assess whether BDST scores were influenced by codeswitching frequency and salient cue detection, a simpler model which only included CSF and SalC and their interaction was created. Group, Language Use, and Language Exposure, which were non-significant predictors, were not included in the model, as they were noted to be non-significant. However, it was of interest to see whether BDST would be predicted by CSF

and SalC for both theory-driven reasons and because the Simon Effect was influenced by these two variables. The results of the model are presented next:

Model: BDST ~ SalC*CSF				
	F-statistic	DF1	DF2	p-value
Intercept	1995.94	1	56	<0.0001
CSF	0.92	1	56	0.34
SalC	1.27	1	56	0.26
CSF*SalC	0.11	1	56	0.74

Appendix I

Chapter 5 Supplementary Analyses

I. Narration Task Analysis

The transcriptions of the narration task in each condition identified the languages used in each narration and distinguished loans and switches. As indicated in Section 2.3.1.1, loans were considered part of the target language due to their morphosyntactic or phonological adaptation into the cued language. On the other hand, words which retained their original form (from the borrowed language) were considered switches. In the single-language condition, switches constituted 'errors', whereas in the trilingual condition, they signalled the directionality of switches.

Of the total 240 narrations produced by the participants, 36 randomly selected narrations (or 15%) were checked by an external rater who listened to the narrations while reading the transcriptions which included the loan and switch classifications. The examiner, a Lebanese (Arabic-English-French) trilingual lecturer in Linguistics at the American University of Beirut, Lebanon, was provided with a detailed description of the loan vs. switch judgment guidelines. Cohen's κ was run to determine the level of agreement between the researcher and the rater's judgment of switches and loans. Furthermore, any disagreement regarding the transcription was highlighted by the rater; that is, in case of a mismatch between the audio-recordings and transcriptions, the rater informed the researcher.

The rater noted 12 instances in which a discrepancy was observed; eight of these instances were hesitation phenomena which did not influence the outcome of the analysis as these 'meaningless' syllables were classified under 'filled pauses' and not speech which required language assignment. The remaining four instances included two typos and two errors. Accordingly, the transcription was highly accurate. In terms of the loan vs. switch judgments, there was high agreement between the judgments of the researcher and rater, with $\kappa = 0.84$, 95% CI [0.82-0.91], $p < .001$. Instances which were disagreed upon were discussed until a consensus was reached in determining which language the words belonged to. It should be noted that most disagreements stemmed from the Arabic-only condition, as it was difficult to clearly establish whether words were loans or switches, especially given the phonological similarity between Arabic and French in sounds like 'é'.

The source of cross-linguistic influence in the monolingual conditions of the narration task is measured as the number of instances switched into from the non-target languages. These 'instances' may be measured by syllables/utterance. As this was not of direct relevance to the aims of the study, and given time constraints, syllables/utterances were not calculated. Instead, the number of times a switch occurred was recorded as an 'instance', irrespective of the length of the switch (i.e., utterance length as measured by syllables in the non-target language). The number of switches in each of the two non-target languages was recorded and then divided by the total number of switches to get a percentage. The results of the analysis for the two groups are presented below in tabular form.

	Single-Language Condition					
	EMI			FMI		
	Arabic	English	French	Arabic	English	French
Mean number of switches (errors)	3	1	8	4	2	1
Non-Target Language						
Arabic	-	97%	34%	-	41%	38%
English	96%	-	66%	42%	-	62%
French	4%	3%	-	58%	59%	-

Trilingual Language Condition

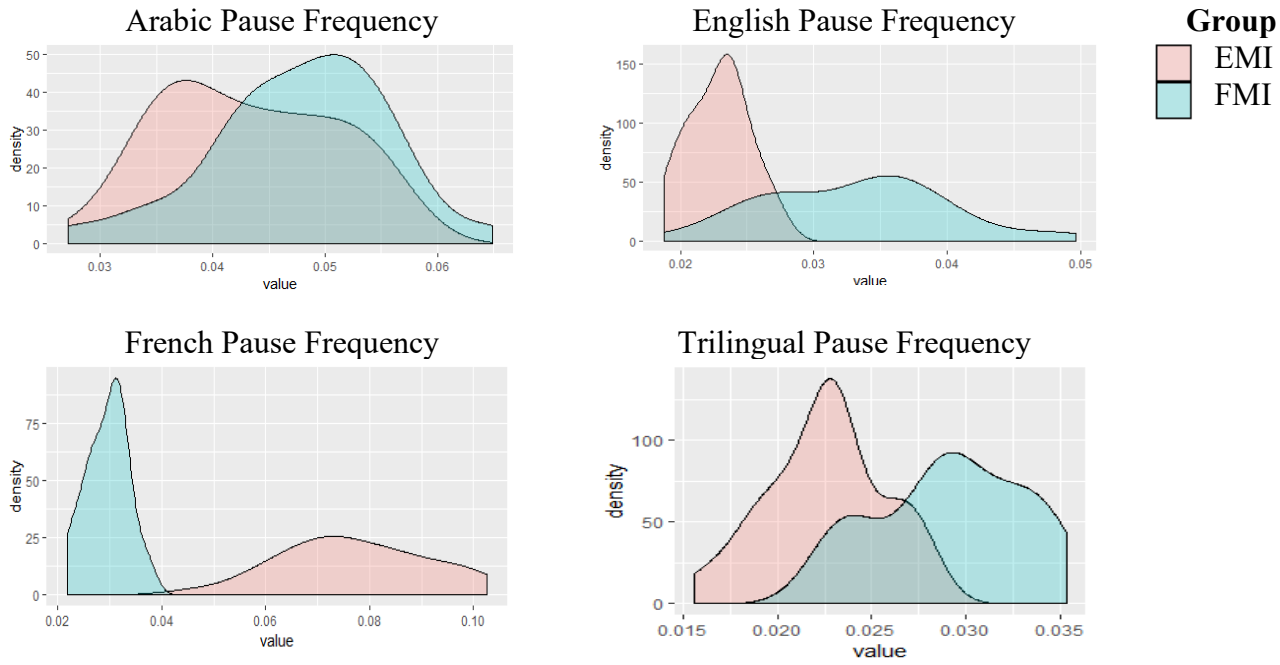
The directionality of switches in the trilingual condition was examined for participants in both groups. There was much variability within and across the two groups, as some participants switched few times, using one language for long stretches (i.e., using Muysken's (2000) typology, alternation typified their speech. Other participants switched very frequently, using a combination of alternations and insertions. Furthermore, some participants produced more congruent lexicalisations than others, with different combination of languages which increased the variation among participants' speech. Some examples are listed below.

Type	Example
Alternation	They're running after him but they can't catch up fa raħo min maħal teneh ta yil'atoo 'They're running after him but they can't catch up so they went in a different direction to catch him'
Insertion	Hon bissoura fi bint zgheeri shakla pauvre w badda masareh yimkin 'Here in this picture there is a small girl who looks poor and wants money maybe'
Congruent Lexicalisation	La femme ħatta lunetteita w lebseh chapeauyitta la anno she's at the beach 'The woman is wearing (or putting on) her sunglasses and is wearing her hat because she's at the beach'
Arabic → red font English → blue font French → green font	

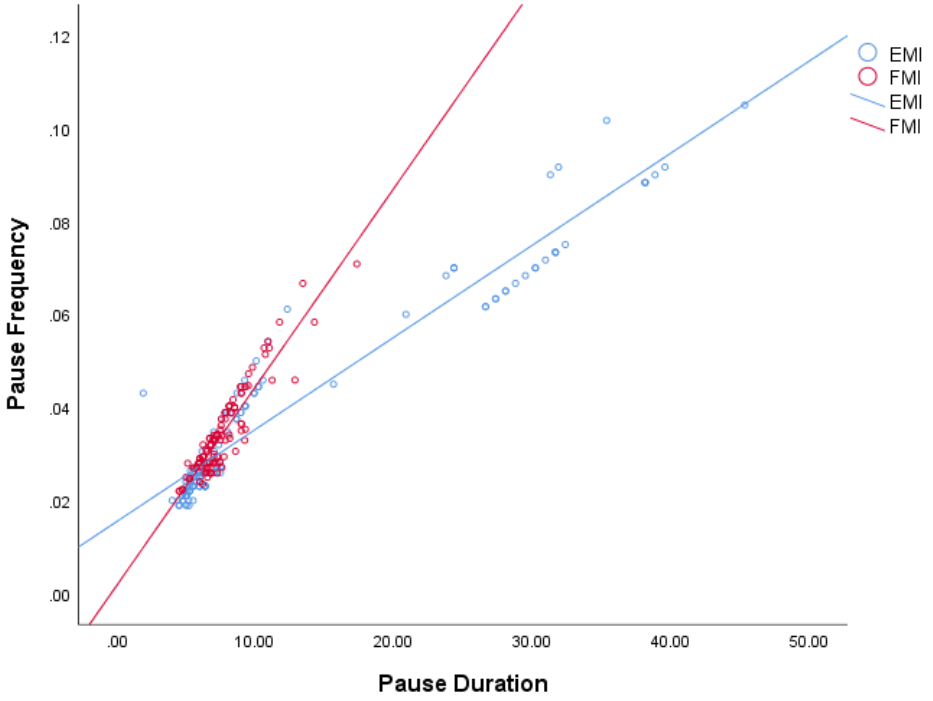
As suggested in Chapter 7 (Recommendations for Future Research), as well as limitations of the analyses, data from the narration task is very rich and may be analysed in great detail. The study does not attribute one type of switching to a particular context,

as the data from this task revealed that participants were not limited by the type of switches in the trilingual condition. However, subdividing and analysing participants' speech according to the type of switches produced may provide further information about individual differences in language production and allow investigating whether participants differ in executive functions given how frequently they produce each type of switch.

Figure A5-1 Pause Frequency Analyses- Group distributions per Language

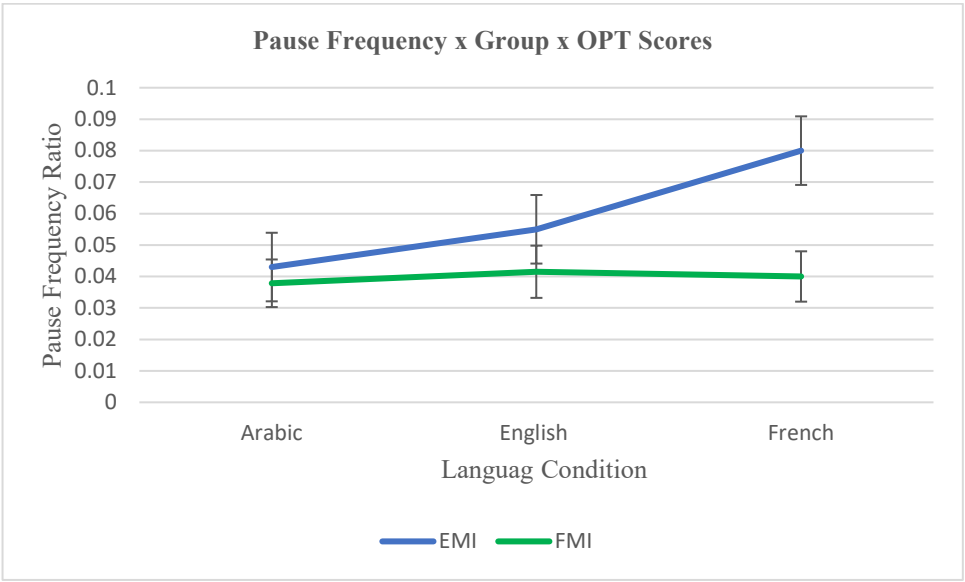


Correlation Graph: Pause Frequency and Pause Duration by Group



II. OPT scores as Predictors of Linguistic Performance

An alternative visualisation of the magnitude of oral proficiency scores in Arabic, English, and French in predicting pause frequency ratios for each group. It offers the same results, indicating that for all three languages, language proficiency scores predicted pause frequencies of EMI participants to a larger extent than they did for FMI participants.



III. VFT Scores as Predictors of Linguistic Performance – Exploratory Models

The following are summaries of different general linear models, linear models, generalised linear models, and linear mix-effects models which were performed to examine VFT as a predictor, given the non-significant results of the best-fit model. Although this is not an exhaustive list of all the models which were tested, this section provides further information about ways in which the data was analysed, even if these analyses were not included in the main thesis.

Model A: Arabic Switch Costs ~ Group*Arabic VFT Scores				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	26814.29	1	56	<0.0001
Group	4.82	1	56	0.03
VFT Arabic	1.82	1	56	0.18
Group*VFT Arabic	0.19	1	56	0.67

Model B: Arabic Switch Costs ~ Arabic VFT Scores				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	24829.25	1	58	<0.0001
VFT Arabic	0.18	1	58	0.67

Model C: English Switch Costs ~ Group*English VFT Scores				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	41959.24	1	56	<0.0001
Group	76.57	1	56	<0.0001
VFT English	0.35	1	56	0.56
Group*VFT English	0.96	1	56	0.33

Model D: Trilingual Switch Costs ~ Group*VFT English				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	73932.04	1	56	<0.0001
Group	55.25	1	56	<0.0001
VFT English	0.82	1	56	0.37
Group*VFT English	1.32	1	56	0.26

Model E: Trilingual Switch Costs ~ Group*VFT English				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	75114.96	1	56	<0.0001
Group	56.51	1	56	<0.0001
VFT English	2.49	1	56	0.12

Group*VFT English	2.63	1	56	0.11

Model E: Trilingual Switch Costs ~ Group*VFT Arabic				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	73020.09	1	56	<0.0001
Group	54.57	1	56	<0.0001
VFT Arabic	1.39	1	56	0.24
Group*VFT Arabic	0.03	1	56	0.86

Model F: English Pause Durations ~ Group*VFT English				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	34205.19	1	56	<0.0001
Group	37.46	1	56	<0.0001
VFT English	0.07	1	56	0.79
Group*VFT English	1.88	1	56	0.17

Model G: Arabic Pause Durations ~ Group*VFT Arabic				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	114413.72	1	56	<0.0001
Group	3.21	1	56	0.08
VFT Arabic	0.63	1	56	0.43
Group*VFT Arabic	3.15	1	56	0.08

Model G: French Pause Durations ~ Group*VFT French				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	15032.86	1	56	<0.0001
Group	2.18	1	56	0.15
VFT Arabic	0.38	1	56	0.54
Group*VFT Arabic	0.72	1	56	0.40

Model H: Pause Duration ~ Group*Language Condition*VFT Scores (two-way interactions only.

Estimated coefficients, standard errors (SE), t values, and p-values from the linear mixed effects models run on log-transformed Pause Frequency. Asterisks (∗) indicate: p< 0.05 (∗), p< 0.01 (∗∗), p< 0.001(∗∗∗) and p< 0.0001 (∗∗∗∗).

	Pause Duration			
Predictor	**Coefficient**	**SE**	**t-value**	**p-value**
Fixed Effects				
Intercept	1.79	0.12	15.09	0.00****
Group	0.18	0.13	1.32	0.19
VFT Arabic	0.05	0.05	0.91	0.37
VFT French	0.02	0.05	0.35	0.73
VFT English	0.05	0.05	0.96	0.34
Arabic-only	0.04	0.08	0.57	0.56
English-only	0.05	0.08	0.62	0.54
French-only	1.03	0.08	12.79	0.00*****
Trilingual	0.09	0.08	1.18	0.24
Group*VFT Arabic	0.02	0.01	1.98	0.05
Group*VFT French	0.02	0.01	1.23	0.22
Group*VFT English	0.02	0.01	1.30	0.19
Group*Arabic-only	0.01	0.03	0.28	0.77
Group*English-only	0.03	0.03	1.06	0.29
Group*French-only	0.00	0.03	0.02	0.98
Group*Trilingual	0.04	0.03	1.25	0.21
VFT Arabic*Arabic-only	0.00	0.00	0.39	0.69
VFT English*English only	0.00	0.00	0.00	0.99
VFT French*French-only	0.00	0.00	1.12	0.26

*Formula: Pause Duration ~ Group + VFT scores + Language Condition + VFT*Group + VFT*Language Condition + Group*Language Condition +(1|Participant)*

Model I: Arabic Pause Durations ~ Group*VFT Arabic*VFT French				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	15410.320	1	52	<0.0001
Group	2.24	1	52	0.14
VFT Arabic	0.38	1	52	0.54
VFT French	1.35	1	52	0.25
Group*VFT Arabic	0.78	1	52	0.38
Group*VFT French	1.92	1	52	0.17
VFT French*VFT Arabic	0.15	1	52	0.70
Group*VFT French*VFT Arabic	1.89	1	52	0.17

Model J: French Pause Durations ~ Group*VFT French				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	45049.73	1	57	<0.0001
Group	14.86	1	57	0.0003
VFT French	0.81	1	57	0.37
Group*VFT French	1.33	1	57	0.25

IV. Exploring Language Experience as a Predictor of RTs in TPNT

As Language Exposure did not reveal significant findings in the full model, a series of exploratory models which included each level of Language Exposure and Group as predictors in three separate models, with the outcome variable being switch costs in the relevant language. Random effects were included in the models for participants. The results of these models revealed that the inclusion of 'Group' nullified the significance of Language Exposure.

MEM ANOVA (Type III) Summary of Main Effects and Interactions of Switch Costs and Language Exposure

	F-statistic	DF1	DF2	p-value
Intercept	287.03	1	56	<.0001
Group	2.63	1	56	0.11
Arabic Exposure	0.09	1	56	0.75
Group*Arabic Exposure	0.21	1	56	0. 65

MEM ANOVA (Type III) Summary of Main Effects and Interactions of Switch Costs and Language Exposure

	F-statistic	DF1	DF2	p-value
Intercept	256.39	1	56	<.0001
Group	11.64	1	56	0.00***
English Exposure	0.01	1	56	0.90
Group*English Exposure	0.07	1	56	0.79

MEM ANOVA (Type III) Summary of Main Effects and Interactions of Switch Costs and Language Exposure

	F-statistic	DF1	DF2	p-value
Intercept	597.28	1	56	<.0001
Group	127.65	1	56	<.0001
French Exposure	0.59	1	56	0.44
Group*French Exposure	0.07	1	56	0.78

The analysis of Language Experience revealed that CSF and SalC were significant predictors of switch costs, where the more frequent codeswitching and cue detection are reported (no interaction), the smaller the switch costs. This was true for all languages. It was of interest to see whether frequent codeswitchers and those who have high monitoring skills (SalC) were also faster in their responses compared to participants who

provided lower self-reported frequencies for both variables. Accordingly, a model was fitted with mean RTs (for both switch and nonswitch) trials as an outcome variable and with CSF and SalC (and their interaction) as predictor variables. The results of the analysis, as summarised in Model K, revealed no significant results.

Model J: TPNT Reaction Times ~ Group*CSF*SalC				
	F-Statistic	DF1	DF2	p-value
Predictors				
Intercept	148.43	1	56	<0.0001
CSF	1.91	1	56	0.17
SalC	1.53	1	56	0.22
CSF*SalC	1.03	1	56	0.31

MEM ANOVA (Type III) Summary of Main Effects of Pause Frequency, Codeswitching Frequency, and Salient Cue Detection

	Pause Frequency			
	F-statistic	DF1	DF2	p-value
Intercept	22565.548	1	180	<0.0001
CSF	0.50	1	57	0.48
SalC	0.73	1	57	0.39

Formula: Pause Frequency~CSF+SalC + (1|Participant)