



UNIVERSITAT DE
BARCELONA

Lexical development in very young EFL learners: A longitudinal study on early multilingualism

Cristina Alexandrescu

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

*Lexical development
in very young EFL learners:
A longitudinal study on early
multilingualism*

TESI DOCTORAL

CRISTINA ALEXANDRESCU

Directora: Dra. Immaculada Miralpeix Pujol

Tutora: Dra. Carmen Muñoz Lahoz

Programa de Doctorat: Estudis Lingüístics, Literaris i Culturals

Departament de Llengües i Literatures Modernes i d'Estudis Anglesos

2023



This thesis is dedicated to my son, who was born and raised in a multilingual environment.

“One language sets you in a corridor for life. More languages open every door along the way.”
Frank Smith

ACKNOWLEDGEMENTS

During the preparation of this thesis, there were many people without whose help I could not have managed to finish it.

First and foremost, I would like to thank my coordinator, Dr Imma Miralpeix, for all her support in finding the right topic for my research, for her help with the design of this study, for her advice and for the enormous amount of time she spent on the numerous corrections, as well as for her constant encouragement. Thank you for always being in touch with me, even if I was away, for all the articles and books you sent and for keeping me updated. Thank you for constantly pushing me to improve my academic writing skills and for offering advice and support to help me do so. Last but not least, thank you for the clear and easy-to-follow corrections of my work. I know it has not been easy and that you have invested incalculable time. I can say that after working with you, I am a better professional.

Next, I would like to thank Prof. Carmen Muñoz, the tutor of this thesis, for tracking my progress all these years and for agreeing to let me use the transcripts from the Barcelona Age Factor Project (2006), which she coordinated. This data has helped me in my research and provided useful insights that have helped me interpret my own results. Thanks as well to Dr Elisabet Serrat for her generosity in providing the Catalan translation of the Peabody Picture Vocabulary Test (PPVT) when it was not yet available to the general public. Without it, I would have not been able to carry out my research in both Spanish and Catalan, which would have meant losing invaluable information. In the same vein, I would like to express my thanks and appreciation to Prof. Oliver Strunk for allowing me to take the Statistics Course during my first year, and to Imma Miralpeix (again) for supporting me throughout this entire learning process. The knowledge I have acquired has been of great help in conducting my research and generating results and conclusions.

This research would not have been possible without Ms Esther Herranz, director of European International School in Barcelona; Mr Mike Barret, academic director at the European International School; and Ms Anna Nogue, coordinator of the Infant School Department of European International School, who gave me permission and

supported me as I collected data from 2011 to 2014. Thank you to my colleagues Rosa Tuduri and Sandra Berjillos for helping me with the Spanish and Catalan PPVT and to all the English tutors who assisted me with filling the MacArthur-Bates Communicative Development Inventories (MBCDIs). This research would not have been possible without the availability of the teachers and directors from Santa Isabel School, their willingness to answer my questions and all their help and assistance with filling the MBCIs and the Spanish and Catalan PPVT. The time they have allotted so that this study can be carried out is invaluable.

During these last two years (2021-2023), I have also had the chance to be part of the Maple Bear Central and Eastern Europe teaching staff, which gave me access to relevant information for this dissertation. My appreciation and gratitude go to Ms Jaradat Mihaela, general manager of the Maple Bear Kindergarten in Cluj -Napoca, Romania, for the trust and confidence she put in me, and to the administrative team at Maple Bear Canada for providing relevant materials for this thesis.

Last but not least, I would like to thank my family and friends for always supporting me over these almost eleven years. I would like to thank my mother, my brother, my sister-in-law, my aunt and my uncle for encouraging me to carry on when I felt too tired to continue, and my son Marc, for being the reason I started this process in the first place.

ABSTRACT

Multilingualism is one of the many effects of globalisation. The need to be proficient in several languages has sprung from people's need to expand their universe outside their countries of origin in search for a better life or a more successful career. As a result, more and more children are exposed to languages either from birth or later on, when they start attending kindergarten, and it is now quite common for a child to be acquainted with two or more languages from an early age (Muñoz & Singleton, 2019). English is one of the most highly valued languages across the globe and has the second highest number of speakers in the world. In countries such as Spain, parents encourage their children to learn this language and therefore send them to either private afterschool academies or, to a lesser extent, English immersion schools. In Catalonia, where Spanish and Catalan are spoken, English is learned as a Foreign Language (FL), and nowadays an early introduction to this language is common.

This thesis deals with young Spanish-Catalan bilinguals who learn English in an immersion school. Because lexis is a crucial aspect of language acquisition, especially at very young ages, we measured the lexical development of a group of 51 children in all three languages over three years. Their performance was analysed longitudinally from ages 3 to 5. Their receptive and productive vocabularies in the three languages were compared to a control group consisting of children of the same ages but attending a regular school and following a regular school curriculum. The children's performance at the end of the immersion school was also compared to that of a group of students in seventh grade (1ESO), at the beginning of secondary school in the regular curriculum. To assess lexical development in English, we used the *MacArthur Bates Communicative Development Inventories* (MBCDI; Fenson et al., 1983/2007). The Spanish version of the *Peabody Picture Vocabulary Test* (PPVT; Dunn et al., 2010), together with its Catalan version, were used to assess Spanish and Catalan vocabularies, respectively. At the end of kindergarten (age 5), students also performed an oral storytelling task.

Findings show that the immersion group made steady progress in terms of their receptive and productive vocabularies in English, which were significantly larger than those of the participants in the control group. With regard to Spanish and Catalan, results also showed significant gains over time. When comparing the immersion group to the control group in Spanish and Catalan, results indicate that both groups performed similarly in Spanish, but not in Catalan (in this latter language, participants in the control group scored significantly higher than those in the immersion group). Regarding the productive vocabulary assessed in the storytelling task, results were mixed: some measures show a trend favouring the immersion group, others are in favour of the regular school group and yet others are not significant.

Therefore, the amount of exposure provided in immersion settings seems to quickly boost vocabulary learning in young children, whose performance is, in some aspects, still better than that of students in the regular group at the beginning of secondary school. While lexical competence in Spanish does not seem to be affected by intensive exposure to English (no significant differences were found between groups), the results for Catalan indicate that children in the immersion school lag behind their same-aged peers in the control group. Further analyses taking language dominance into account indicated that there were no significant differences in Spanish and Catalan PPVT scores between regular and immersion groups in Spanish-dominant participants. However, significant differences in favour of the regular curriculum group were found for Catalan-dominant children. It remains to be seen whether this is a temporary effect and whether it disappears in the short run as exposure to the L1/L2 accumulates. Results are discussed in the light of previous research and pedagogical implications are considered in relation to young learners' early introduction to FL.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	I
ABSTRACT.....	III
TABLE OF CONTENTS	V
LIST OF TABLES	IX
LIST OF FIGURES	XIII
LIST OF ABBREVIATIONS	XV
LIST OF APPENDICES	XVII
CHAPTER 1: INTRODUCTION.....	1
CHAPTER 2: THE EARLY ACQUISITION OF TWO AND THREE LANGUAGES: A FOCUS ON VOCABULARY	5
2.1. BILINGUALISM	6
2.1.1. DEFINITIONS AND TYPES.....	6
2.1.2 Simultaneous bilingualism.....	8
2.1.2.1 Parental and school input in simultaneous bilingualism.....	8
2.1.2.2. Language Separation Hypothesis in simultaneous bilingualism in children	11
2.1.2.3. Receptive lexical knowledge in simultaneous bilingualism in children.....	14
2.1.2.4. Productive lexical knowledge in simultaneous bilingualism in children	17
2.1.3 Early sequential bilingualism.....	20
2.1.3.1 Educational settings fostering early L2 acquisition after L1	21
2.1.3.2 Lexical development in sequential bilingualism	24
2.1.3.3 The influence of age of onset (AO) and amount of exposure on linguistic performance	27
2.2. THIRD LANGUAGE ACQUISITION	30
2.2.1. L3 ACQUISITION BY BILINGUAL ADULT LEARNERS	32
2.2.2. L3 ACQUISITION BY YOUNG CHILDREN	34
2.2.2.1. Simultaneous acquisition of L1, L2 and L3.....	34
2.2.2.1.1. Parental and school input.....	34
2.2.2.1.2. The Language Separation Hypothesis	36
2.2.2.1.3. Lexical comprehension and production.....	37
2.2.2.2. Learning an L3 after the simultaneous acquisition of L1 and L2	43
2.2.2.2.1. L3 acquisition in formal settings: Regular curricula and immersion schools.....	43
2.2.2.2.2. L3 acquisition in formal settings: The role of age of onset (AO) and exposure.....	45
2.2.2.2.3. Possible effects of the early introduction of an FL on L1 and L2	48
2.2.3. SOME INSIGHT INTO THE MULTILINGUAL LEXICON	51
2.3. SUMMARY.....	54

CHAPTER 3: LEXICAL ASSESSMENT OF VERY YOUNG CHILDREN IN TWO OR THREE LANGUAGES.....	57
3.1. ASSESSING LANGUAGE DEVELOPMENT IN VERY YOUNG LEARNERS WITH A SPECIAL FOCUS ON VOCABULARY	59
3.1.1. RECEPTIVE AND PRODUCTIVE LEXICAL ASSESSMENT	59
3.1.1.1. Parental diaries	59
3.1.1.2. The MacArthur-Bates Communicative Development Inventories	61
3.1.1.3. Other lexical assessment tools.....	69
3.1.2. RECEPTIVE LEXICAL ASSESSMENT	70
3.1.2.1. The Peabody Picture Vocabulary Test (PPVT-III).....	71
3.1.2.2. Other receptive vocabulary tests.....	76
3.1.3. PRODUCTIVE LEXICAL ASSESSMENT	77
3.1.3.1. Spontaneous speech recordings.....	79
3.1.3.2. Prompted speech recordings.....	83
3.2. THE CHALLENGE OF ASSESSING LEXICAL DEVELOPMENT IN TWO AND THREE LANGUAGES.....	85
3.2.1. USING TRANSLATED AND ADAPTED VERSIONS OF STANDARDISED TESTS.....	86
3.2.2. USING SPEECH RECORDINGS	88
3.3 ASSESSING THE LEXICAL DEVELOPMENT OF VERY YOUNG LEARNERS IN SPANISH, CATALAN, AND ENGLISH	90
3.4. SUMMARY	94
CHAPTER 4: THE STUDY	97
4.1. PRELIMINARY PILOT STUDIES	97
4.1.1. PRELIMINARY REMARKS.....	98
4.1.2. PILOT STUDY 1.....	101
4.1.3. PILOT STUDY 2.....	105
4.2. THE MAIN STUDY	113
4.2.1. OBJECTIVES AND RESEARCH QUESTIONS	114
4.2.2. METHODOLOGY	115
4.2.2.1. Participants	115
4.2.2.2. Instruments and measures.....	119
4.2.2.3. Procedure	122
CHAPTER 5: RESULTS	129
5.1. PRELIMINARY ANALYSES: IG PARTICIPANTS	130
5.2. RQ1: EARLY LEXICAL DEVELOPMENT IN EFL.....	132
5.2.1. RECEPTIVE AND PRODUCTIVE LEXICAL DEVELOPMENT OF CHILDREN ATTENDING THE IMMERSION SCHOOL	133
5.2.1.1. MBCDI total scores	133
5.2.1.2. MBCDI semantic categories.....	136
5.2.1.3. MBCDI word-class measures	140
5.2.2. COMPARATIVE RESULTS OF RECEPTIVE AND PRODUCTIVE VOCABULARY DEVELOPMENT IN IMMERSION AND REGULAR SCHOOL SETTINGS.....	144

5.2.2.1. MBCDI total scores	144
5.2.2.2. The MBCDI semantic categories.....	148
5.2.2.3. MBCDI word-class measures	154
5.3. RQ2: LEXICAL DEVELOPMENT IN SPANISH AND CATALAN OF CHILDREN IN THE IMMERSION AND REGULAR SCHOOLS	159
5.3.1. LEXICAL DEVELOPMENT IN SPANISH AND CATALAN OF CHILDREN IN THE IMMERSION SCHOOL	160
5.3.2. LEXICAL DEVELOPMENT IN SPANISH AND CATALAN OF CHILDREN IN THE IMMERSION SCHOOL AND IN THE REGULAR SCHOOL: A COMPARISON	162
5.3.3. LEXICAL DEVELOPMENT IN SPANISH AND CATALAN LEARNERS IN RELATION TO THEIR LANGUAGE DOMINANCE	165
5.4. RQ3: ENGLISH PRODUCTIVE VOCABULARY OF CHILDREN IN THE IMMERSION SCHOOL AT THE END OF THE EARLY CHILDHOOD EDUCATION PROGRAMME.....	168
5.4.1. ORAL PRODUCTIVE VOCABULARY OF IMMERSION SCHOOL LEARNERS	169
5.4.2. ORAL PRODUCTIVE VOCABULARY OF IMMERSION AND REGULAR SCHOOL LEARNERS: A COMPARISON	170
5.4.2.1. Preliminary analysis: RG2 group.....	170
5.4.2.2. Comparative results of learners' productive vocabulary in the immersion and regular groups.....	171
5.5. SUMMARY	172
CHAPTER 6: DISCUSSION	175
6.1. RQ1: EARLY LEXICAL DEVELOPMENT IN ENGLISH AS AN FL	175
6.1.1. RECEPTIVE AND PRODUCTIVE LEXICAL DEVELOPMENT OF CHILDREN ATTENDING THE IMMERSION SCHOOL.....	176
6.1.1.1. MBCDI total scores	176
6.1.1.2. MBCDI semantic categories and word-class measures.....	181
6.1.2. COMPARATIVE RESULTS OF RECEPTIVE AND PRODUCTIVE VOCABULARY DEVELOPMENT IN IMMERSION AND REGULAR SCHOOLS	186
6.2. RQ2: LEXICAL DEVELOPMENT IN SPANISH AND CATALAN	189
6.2.1. LEXICAL DEVELOPMENT IN SPANISH AND CATALAN OF CHILDREN IN THE IMMERSION GROUP	189
6.2.2. COMPARATIVE RESULTS OF LEXICAL DEVELOPMENT IN SPANISH AND CATALAN OF CHILDREN IN IMMERSION AND REGULAR SCHOOLS	192
6.3. RQ3: PRODUCTIVE VOCABULARY IN ENGLISH OF CHILDREN IN THE IMMERSION SCHOOL AT THE END OF THE EARLY CHILDHOOD EDUCATION PROGRAMME.....	195
6.4. SUMMARY	198
CHAPTER 7: CONCLUSIONS AND FUTURE RESEARCH.....	201
7.1. SUMMARY OF OUR RESEARCH	202
7.2. SOME POSSIBLE PEDAGOGICAL IMPLICATIONS	204
7.3. LIMITATIONS OF THE STUDY.....	206
7.4. ORIGINAL CONTRIBUTION OF THE RESEARCH	208

7.5 FURTHER RESEARCH AND FINAL REMARKS.....	210
REFERENCES:.....	213
APPENDICES.....	231
APPENDIX 1: EUROPEAN INTERNATIONAL SCHOOL REGISTRATION FORM (ATTESTING LANGUAGE DOMINANCE OF THE IG PARTICIPANTS)	231
APPENDIX 2: MBCDI.....	234
APPENDIX 3: PPVT III (SCORING SHEET AND EXAMPLE).....	240
APPENDIX 4: THE DOG STORY	243
APPENDIX 5: DATA-EDITING CONVENTIONS FROM THE ORAL PRODUCTION TASKS AT THE END OF T3	244
APPENDIX 6: EXAMPLE OF ORAL PRODUCTION TASK TRANSCRIPTIONS.....	245
APPENDIX 7: RESULTS OF THE NORMALITY TESTS.....	247
APPENDIX 8: MAPLE BEAR YEARLY SYLLABUSES	252

LIST OF TABLES

Table 1: Freddy’s receptive vocabulary in three languages (Quay, 2001).	38
Table 2: Leandre and Dominique’s MLUm in three languages (Wang, 2008).	40
Table 3: Kathryn’s MLUw in three languages (Montanari, 2010).	41
Table 4: Jenny’s MLUs in three languages (Barnes, 2006).	41
Table 5. Summary of the MBCDI Words and Gestures.	63
Table 6: Summary of the MBCDI Words and Sentences.	64
Table 7: Andreu’s early and late lexical development stages (Perez-Vidal & Juan- Garau, 2002).	80
Table 8: Summary of lexical assessment tools for Spanish and Catalan.	92
Table 8: Cont.	93
Table 9: Summary of lexical assessment tools for young learners in English.	93
Table 10: Detailed description of the MBCDI Words and Gestures.	99
Table 10: Cont.	100
Table 11: Detailed description of the MBCDI Words and Sentences.	100
Table 12: Participants in pilot study 1.	101
Table 13: Descriptive results for the MBCDI Words and Gestures for the participants in P3 and P4 in pilot study 1.	104
Table 14: Descriptive results for the MBCDI Words and Sentences for the participants in P5 in pilot study 1.	104
Table 15: Participants in the second pilot study.	106
Table 16: Cross-sectional descriptive results for the MBCDI Words and Gestures for the participants in P3, P4 and P5 in pilot study 2.	109
Table 17: Cross-sectional descriptive results for the MBCDI Words and Sentences for the participants in P3, P4 and P5 in pilot study 2.	110
Table 18: Longitudinal descriptive results for the MBCDI <i>Words and Gestures</i> for the participants in P4, and P5 in pilot study 2.	110
Table 19: Cross-sectional descriptive results for the Spanish/Catalan PPVT test for the participants in P3, P4, P5 in pilot study 2.	111
Table 20: Cross-sectional descriptive results for the Spanish/Catalan PPVT test for the participants in P3, P4 and P5 in pilot study 2 according to participants’ language dominance.	112
Table 21. Immersion group (IG) participants in the main study according to their AO and hours of exposure to L3.	117
Table 22: Distribution of participants in the main study.	118
Table 23: The MBCDI measures used in the main study.	119

Table 24: The MBCDI semantic categories and word-class measures used in the main study. ...	120
Table 24: Cont.	121
Table 25: Data collection procedure for each test and group involved in the main study of the thesis.	124
Table 26: Measures calculated from the storytelling task.	126
Table 27: Descriptive statistics and one-way between-groups ANOVA of the MBCDI total scores measures at T1 for the participants in the IG (N=51), depending on their AO, with $p \leq .01$	131
Table 28: Descriptive statistics and one-way between-groups ANOVA for the Spanish/Catalan PPVT scores at T1 for the participants in the IG (n= 51), depending on their AO, with $p \leq .01$...	132
Table 29: Descriptive statistics and one-way RM ANOVA results for the MBCDI total scores measures with Time (T) as the within-subjects factor for the participants in the IG (n= 51), with $p \leq .01$	135
Table 30: Descriptive results and Friedman Test results for the MBCDI semantic categories corresponding IG (n=51) with $p \leq .01$. (R: receptive vocabulary, P: Productive vocabulary).....	138
Table 30: Cont.	140
Table 31: Descriptive statistics and Friedman test results of the MBCDI word-class measures for the participants in the IG (n=51), with $p \leq .01$, (R: receptive vocabulary, P: productive vocabulary).142	
Table 31: Cont.	143
Table 32: Summary of the significant differences registered at different data collection times for the semantic categories and word-class measures.	143
Table 33: Descriptive results and the 3x2 RM ANOVA with Time (T) as the within-subjects factor and School (S) as the between-subjects factor for the participants in the IG (n=51) and RG1 (n=15), with $p \leq .01$. Sphericity assumed.	146
Table 34: Descriptive statistics and results of the Mann-Whitney <i>U</i> Tests for the MBCDI semantic categories for IG (n=51) and RG1 (n=15) at T1, T2, and T3 (R: Receptive vocabulary, P: Productive vocabulary).	150
Table 34: Cont.	153
Table 35: Descriptive results and Mann-Whitney <i>U</i> -Test results for the MBCDI word-class measures for the IG (n=51) and RG1(n=15) (R: Receptive vocabulary, P: Productive vocabulary).	157
Table 35: Cont.	158
Table 36: Descriptive results and one-way RM ANOVA for the Spanish/Catalan PPVT scores with Time (T) as the within-subjects factor for the IG (n=51), when $p \leq .01$	161
Table 37: Descriptive results and 3x2 RM ANOVA with Time (T) as the within-subjects factor and School (S) as the between-subjects factor of the Spanish/Catalan PPVT for the IG (n=51) and RG1 (n=15), when $p \leq .01$	164
Table 38: 3x2 RM ANOVA with Time (T) as the within-subjects factor of the Spanish/Catalan PPVT for the Spanish-dominant participants in IG (n=27) and RG1 (n=8), with $p \leq .01$	167
Table 39: 3x2 RM ANOVA with Time (T) as the within-subjects factor and School (S) as the between-subjects factor of the Spanish/Catalan PPVT for the Catalan-dominant participants in IG (n=24) and RG1 (n=7), with $p \leq .01$	167

Table 40: Means and standard deviations of the IG in the oral production task (n=51).	169
Table 41: Descriptive statistics and Kruskal-Wallis results for the oral production measures in the RG2 group participants: school A (n=14), school B (n=10) and school C (n=20).....	170
Table 42: Descriptive results and <i>t</i> -test results for the oral production task for the IG (n=51), RG2 (n=44), with $p \leq .01$	172
Table 43: Comparative results of research conducted on receptive vocabulary using the MBCDI.	179
Table 44: Comparative results of research conducted on productive using the MBCDI.	179
Table 45: Normality tests for the MBCDI total scores for P1 (n= 24), P2 (n=19), P3 (n=8).....	247
Table 45: Cont.	248
Table 46: Normality tests for the Spanish and Catalan PPVT test scores for P1 (n= 24), P2 (n=19), P3 (n=8).	248
Table 46: Cont	249
Table 47: Normality tests for the MBCDI total scores for the IG (n= 51) and RG1 (n=15).	249
Table 48: Normality tests for the Spanish/Catalan PPVT for the IG (n= 51) and RG1 (n=15). ...	250
Table 49: Normality tests for the measures in the oral production task at the end of T3 for the IG (n= 51) and RG2 (n=44).....	250
Table 49: Cont.	251

LIST OF FIGURES

Fig. 1: MBCDI total scores boxplots for each group and time.....	147
Fig. 2: Spanish/Catalan PPVT scores at T1, T2, T3 for both groups	164

LIST OF ABBREVIATIONS

1P/1L	‘1 person /1 language’ approach
1P/2L	‘1 person/ 2 languages’ approach
Adj	Adjectives
Adv	Adverbs
AO	Age of onset
Art	Articles
ASD	Autism spectrum disorder
BAF	Barcelona Age Factor (project)
BFLA	Bilingual first language acquisition
BSLA	Bilingual second language acquisition
BVAT	Bilingual Verbal Ability Test
CCT	Computerised comprehension task (test)
CDI	Communicative Development Inventories
CHAT	Codes for the Human Analysis of Transcripts
CHILDES	Child Language Data Exchange System
CLAN	Computerised Language Analysis (software)
CLI	Cross-linguistic influence
CLIL	Content and Language Integrated Learning
Conj	Conjunctions
CP	Critical period
CPH	Critical Period Hypothesis
Cw	Content words
D	D-value (lexical diversity measure)
DD	Developmental delay
ELI	Test de Evaluación del Lenguaje Infantil
ELLOPA	Early Language Listening and Oral Proficiency Assessment
End 1	Word Endings 1 (measure in the MBCDI test)
End 2	Word Endings 2 (measure in the MBCDI test)
ES	Early starters
EVT	Expressive Vocabulary Test
FL	Foreign language
Fw	Function words
GI	Guiraud’s Index
IG	Immersion Group
Interj	Interjections
L1	First language
L2	Second language
L3	Third language
LS	Late starters
M3L	Mean length of the three longest utterances
MBCDI	MacArthur-Bates Communicative Development Inventories
MIA	Multilingual interactive activation
MLU	Mean length of utterance
MLUm	Mean length of utterance in morphemes
MLUw	Mean length of utterance in words
n	Number
N	Nouns
Num	Numerals

PV	Productive vocabulary
PE	Physical Education
PhsU	Phrases understood
PPVT	Peabody Picture Vocabulary Test
PPVT III	Peabody Picture Vocabulary Test third edition
Prep	Prepositions
Pron	Pronouns
RV	Receptive vocabulary
SC	Sentence complexity
SL	Second language (acquired in naturalistic settings after the first native language)
SLA	Second language acquisition
TCV	Total conceptual vocabulary
TE	Translation equivalents
TLA	Third language acquisition
TTR	Type Token Ratio
TV	Total vocabulary
TROG	Test of Receptive Grammar
V	Verbs
WdsP	Words produced
WdsU	Words understood

LIST OF APPENDICES

Appendix 1	European International School Registration Form (attesting language dominance of IG participants)
Appendix 2	The MBCDI Words and Gestures and MBCDI Words and Sentences
Appendix 3	PPVT (scoring sheet and example)
Appendix 4	The Dog Story
Appendix 5	Data-editing conventions from the oral production task at the end of T3
Appendix 6	Example of oral production task transcriptions
Appendix 7	Results of the normality tests
Appendix 8	Maple Bear Yearly Syllabus

CHAPTER 1

INTRODUCTION

It is quite a common practice for young bilingual children to be exposed to a third additional language from very early ages. This exposure can take place in natural settings (e.g., due to immigration) or formal settings (e.g., through schooling). In Catalonia and the Basque Country, two bilingual communities in Spain, English is taught to children as a foreign language (FL) from very young ages (often as an elective subject or introduced through games and songs), and it becomes compulsory during primary school. The amount of exposure varies according to whether input is received through the regular classroom curriculum (limited amount of exposure) or immersion in special bilingual English schools (extensive amount of exposure). Indeed, immersion schools have become a good option for parents who believe that an early start combined with an extensive amount of exposure may ensure high proficiency in English. At the same time, those who are more sceptical fear that a high amount of exposure at an early age may affect the child's acquisition of their first language. Unfortunately, the data available from researchers on this issue is quite limited. Only a scarce number of studies have been carried out in immersion schools (Genesee, 1983; Montanari, 2014; Bergstörm et al. 2016), and they often deal with older children learning an L2.

Hence, we believe there is a gap in the research targeting young bilingual children attending immersion schools where the third language is used. Aiming to fill this gap, the study we propose in this thesis also attempts to contribute to the work conducted so far in the field of early multilingual lexical acquisition. More precisely, we aim to provide a picture of early lexical development in three languages. Our decision to concentrate on lexis rather than syntax or morphology springs from the fact that words are the first to be learned and produced by children, either in their L1 or in other subsequent languages. Furthermore, lexis is also said to be indicative of language proficiency level: the larger the vocabulary, the higher the proficiency (Miralpeix, 2020). Thus, testing the lexical development of young

learners will offer a picture not only of their lexical knowledge but also of their overall linguistic development.

The importance of lexis has also been acknowledged in the field of teaching methodology. Indeed, many teaching programmes have developed syllabuses based on lexis rather than grammar. These types of syllabuses have been adopted by both immersion and regular schools that work with very young learners. Thus, when FL proficiency is assessed, the number of words students recognise and produce is taken into account.

Acknowledging the importance of lexis in early linguistic development while also being aware of the lack of studies targeting young bilingual learners in immersion schools, this thesis inquiries into three main aspects. Firstly, we question whether a massive amount of input, as received in immersion schools, increases lexical proficiency in the FL. The second aspect we question is whether this massive amount of input in an FL affects the lexical development of bilingual children's first languages. The third question is whether starting to learn an FL through immersion in kindergarten is actually more beneficial in the long run than starting to acquire it later on, following the regular primary school curriculum.

In order to conduct this research, we selected a target group of 51 Spanish-Catalan kindergarten participants learning English as an FL in an immersion school. This group was first compared across time to a control group of 15 children of similar ages but learning English in a regular school, following the regular curriculum, with much less exposure. Then, we selected a comparison group made of 44 Spanish-Catalan secondary school students (aged 14) learning English in a regular school. In order to answer the first two questions, we compared the lexical development in Spanish, Catalan and English of the target group and the first control group. To answer the third question, we compared the oral productions of the target group against those of the second comparison group. The tests and measures used in the assessment were carefully selected on the basis of prior research performed in this field, and after previously conducting two pilot studies.

This thesis is organised into seven chapters. The introduction to the study provides the background for it. It also outlines the organisation of the dissertation together with a brief description of each chapter. Chapter 2 presents findings related to bilingualism and multilingualism which are relevant to our study. We start by discussing the terminology which will be used throughout this thesis. Next, we focus on the research conducted in the field of early bilingualism, both simultaneous and sequential. We finish the chapter by outlining the research conducted with multilingual learners, focusing especially on young bilingual learners acquiring an FL. Chapter 3 focuses on the instruments used to assess early lexical development. We begin by describing three types of tools: those testing both lexical comprehension and production, those assessing receptive lexical knowledge and those focusing on oral production. Next, we examine the challenges of assessing the lexical development of bilingual and multilingual young learners in two and three languages. Finally, we provide possible solutions for assessing the lexical development of very young learners in English, Spanish and Catalan. In Chapter 4 we present the research questions and the methodology used to carry out the studies in this thesis. We also describe two pilot studies conducted before the main study of the thesis with the aim of selecting the most appropriate tools and measures which could be used with young learners in English, Spanish and Catalan across time. Then, Chapter 5 is organised around the research questions already discussed in the previous chapter and reports on the results of the statistical analysis. The data is analysed longitudinally in order to measure development across time in the two groups that we follow. Repeated-measure analyses were performed to compare performance in the three languages of the groups analysed. Cross-sectional comparisons between the immersion school participants and the regular school students were also conducted. The main findings are summarised and discussed in detail in Chapter 6. They are also compared to previous studies carried out thus far in this field. Finally, in Chapter 7 we draw the conclusions of our study, present several limitations of this work and identify possible directions for future research.

CHAPTER 2

THE EARLY ACQUISITION OF TWO AND THREE LANGUAGES: A FOCUS ON VOCABULARY

Nowadays, an increasing number of children are exposed to several languages either from birth or later on during childhood. For this reason, among others, multiple language acquisition is a field that deserves to be studied (Singleton & Aronin, 2019). According to Jessner (2008), multilingualism “covers a wide range of meanings” (p. 17). One of them, also acknowledged in Herdina and Jessner (2002), Cenoz et al. (2003) and Butler (2013), defines multilingualism as the ‘acquisition’ of two or more languages. Other researchers, such as Aronin and O’Laoire (2004) and Kemp (2009), focus on the ‘ability’ to speak more than two languages. This dissertation focuses on bilingual Spanish-Catalan children acquiring a third language.

This chapter presents relevant research conducted in the fields of bilingual (and multilingual) language development, concentrating on vocabulary development. Thus, section 2.1. provides a definition of bilingualism as well as a classification of the different types of bilingual learners. It also explores several topics in the field of simultaneous bilingualism (when the L1 and L2 are simultaneously acquired from birth) and sequential bilingualism (when the L2 is acquired after the L1) in children (see Baker (2006); Butler & Hakuta (2006)). Section 2.2. examines third language acquisition (TLA), called ‘L3 acquisition’ in this thesis. This section consists of three main parts: the first (section 2.2.1.) summarises the most relevant research conducted with bilingual adults who acquire an L3 later in life; the second (section 2.2.2.) focuses on very young bilingual learners who acquire the L3 either simultaneously with their L1 and L2 or after their first languages; and the third (section 2.2.3.) is concerned with general facts about the multilingual lexicon.

2.1. BILINGUALISM

This section presents some of the research conducted in the field of early bilingual lexical development. We begin (section 2.1.1.) by briefly discussing some of the definitions of bilingualism across time and introducing different types of bilingualism such as balanced, simultaneous and sequential. Because simultaneous and sequential bilingualism are two relevant concepts in this thesis, we will pay special attention to them in sections 2.1.1.1. and 2.1.1.2. Thus, section 2.1.1.1. focuses on simultaneous bilingual language acquisition and issues which are pertinent to our study, such as the role of parental and school input and language separation in the lexicon, and it introduces the most relevant research on early lexical comprehension and production in simultaneous bilinguals. Section 2.1.1.2. discusses sequential L2 acquisition and presents schooling models that support L2 acquisition, with a special focus on immersion programmes. It also summarises some interesting studies conducted with sequential bilinguals and offers some considerations in relation to the age factor in second language acquisition (SLA).

2.1.1. DEFINITIONS AND TYPES

One of the first definitions of bilingualism was offered by Bloomfield (1933), who refers to bilinguals as ‘individuals who have native-like control of two languages’ (p. 56). Nevertheless, this definition is very restrictive in that it limits the number of individuals who could be classified as bilinguals. According to Butler and Hakuta (2006), there has been a shift towards broader definitions of bilingualism since then. These definitions have the advantage “that they incorporate the developmental process of SLA into the scope of studies of bilingualism” (p. 115). Consequently, they define bilinguals as “individuals or groups of people who obtain communicative skills, with various degrees of proficiency, in oral and/or written forms in order to interact with speakers of one or more languages in a given society” (p. 115).

In an attempt to classify bilinguals, distinctions have been made between balanced bilinguals (who are equally proficient in both languages) and dominant bilinguals (who are more proficient in one of the languages) (Peal & Lambert, 1962). However, Kroll and Dussias (2006) consider “balanced bilingualism” a construct: they claim

that individuals who speak two languages never have equally developed competences in both, especially if the two languages are used for different purposes in different areas and with different people. Several researchers agree that bilingual children tend to be more dominant in one language than the other and that the dominant language may change. According to Grosjean (2010), for example, “one language weakens if the environment favours the other” (p. 164). One example of this phenomenon is Hildegard (Leopold, 1970) who acquired German and English simultaneously, and while her dominant language was English between the ages two and five as her family lived in England, after four weeks spent in Germany her dominant language became German.

Another distinction has also been made between simultaneous and sequential bilinguals. Padilla and Lindholm (1984) and Baker (2006) use the term ‘simultaneous bilingualism’ to refer to the acquisition of two languages from birth and ‘sequential bilingualism’ to refer to a second language being acquired after the first one. Taeschner (1983) and McLaughlin (1984) also define simultaneous and sequential bilingualism by setting cut-off points. Thus, they use the term ‘simultaneous bilingualism’ to describe the acquisition of two languages from birth until the age of three. They adopt the term ‘successive bilingual acquisition’ to define cases in which the exposure to the L2 takes place after the age of three. Meisel (1989) and later on De Houwer (2009) propose the term ‘bilingual first language acquisition’ (BFLA) to refer to two languages being acquired simultaneously from birth. If the L2 starts being acquired within one month after birth but before the age of two, they use the term ‘bilingual second language acquisition’ (BSLA).

In this dissertation, the term ‘simultaneous bilingualism’ will be used according to Taeschner (1983) and McLaughlin’s (1984) definition, i.e., the acquisition of the L1 and L2 since birth (before the age of three). The term ‘sequential bilingualism’ will be used to describe cases in which the L2 is acquired after the age of three.

2.1.2 SIMULTANEOUS BILINGUALISM

This section summarises relevant research conducted with very young children exposed to two languages simultaneously since birth. During the first years of life, bilingual infants are exposed to L1 and L2 mainly at home. Later on, when these children reach school age, they may be enrolled in different bilingual kindergartens/schools, each with a different language policy. It is obvious that the way L1 and L2 are transmitted since birth both by parents and teachers is relevant to whether or not the child will be able to use both languages. Therefore, section 2.1.2.1. provides information on input transmission models used by parents and later by schools and summarises the research findings on these models. Another issue of interest in simultaneous bilingualism is whether a very young bilingual child (up to the age of two) is actually able to separate the two languages (at the lexical level) since they start talking. In this sense, we attempt to provide an answer in section 2.1.2.2., first by summarising the Language Separation Hypothesis outlined by Volterra and Taeschner (1978), and secondly by presenting evidence from the field of lexical acquisition. Finally, we attempt to shed light on whether a bilingual follows the same milestones of lexical development as a monolingual infant: section 2.1.2.3. describes some of the research conducted in the field of early bilingual lexical comprehension, and section 2.1.2.4. outlines several studies related to oral lexical production.

2.1.2.1 PARENTAL AND SCHOOL INPUT IN SIMULTANEOUS BILINGUALISM

In simultaneous bilingualism, children are exposed to the L1 and L2 both at home and at kindergarten/school, following different patterns. With respect to the input these children receive at home, De Houwer (2009) refers to three models of parental input: the first one is called *the one person-one language (1P/1L) model*, where each parent uses only one language to address the child, one of which is also the community language. In the second model, *one person-two languages (1P/2L)* parents use both languages when speaking to their children. A variation of this pattern is found in Grosjean (2010) whose the ‘language-time’ model points out that the two languages are used at specific times while the child is at home. For example, both parents use the L1 in the morning and the L2 in the afternoon. In the third model presented by De Houwer (2009), *1 parent-1 language, 1 parent-2*

languages (1P/1L and 1P/2L), one parent uses both languages when addressing the child, while the other parent uses only one language.

Regarding the input received at school through both formal instruction and peer interaction, Baker (1988) proposed two schooling models. The first one is *transitional bilingual education*, “where the plan is to phase out one language as the mainstream or the majority language develops” (p. 46). In this type of education, the child’s first language continues to be used, but only as an interim means of communication. Such an approach is characteristic of some schools in Ireland, Wales and Scotland. In Spain, it is used in model D schools in the Basque Country (Lasagabaster, 2005), where the amount of Spanish in instruction decreases in favour of Basque. The second model is *maintenance or enrichment bilingual education*, “where two languages are kept throughout all or most of the schooling” (Baker, 1998; p. 47) in equal proportion. This is found in model B schools in the Basque Country in Spain.

Research on the role of parental and school input in the linguistic development of simultaneous bilingual children has yielded several interesting findings. Regarding parental input, Patterson (1999) found that applying the 1P/1L principle does not prevent the simultaneous bilingual children from mixing the two languages when speaking. He performed a study with mixed utterances in the two languages in which a group of Spanish-English bilingual toddlers who were addressed according to the 1P/1L model was compared against a group that was similar but was addressed to according to the 1P/2L model. The results of the study indicate that bilingual toddlers use mixed sentences irrespective of the language approach their parents used when addressing them.

Another interesting finding is De Houwer’s (2007) conclusion that families where both parents speak both languages have a better chance of their children speaking both languages. Yamamoto (2001) points that if parents use the 1P/1L approach, there is a chance that one of the languages will not be spoken by the child (i.e., the child has acquired the language but s/he does not speak it). More precisely, he conducted a study that showed that out of the 46 children addressed in Japanese by one parent and in English by the other, eight did not actually speak English. The

conclusions of the studies performed by Yamamoto (2001) and De Houwer (2007) on the role of parental language input in the case of bilingual children point out that:

“...language presentation in terms of whether a parent addresses a child in one or two languages is not the most decisive factor in whether children will grow up actually speaking two languages. [...] Rather it most likely matters whether parents offer differing levels of support for either language. In the 1P/1L and 1P/2L situations, both languages have similar levels of parent support: either they are both spoken by just one parent or they are both spoken by both parents. In the 1P/1L and 1P/2L setting, however, one language gets support from both parents and the other only from one. In this more uneven situation, the societal language wins out if that is the language that both parents share.” (De Houwer, 2009; p. 112)

Recently, other studies have linked bilinguals' lexical performance not only with parents' strategies when addressing them but also with their educational level and income. For instance, Gatt et al. (2020) reported the results of a study conducted with 58 Maltese-English bilingual children between 3.11 and 4.3 years old. These children were tested separately on lexical receptive and productive measures in Maltese and English. With respect to lexical comprehension, the educational level of both parents was an indicator of high performance. In the case of lexical production, the scores were higher both when parents had a higher educational level and when the father had a higher occupational status. Similar results were reported by De Cat (2021) on 89 bilingual children between the ages of five and seven whose L1 differed and whose L2 was English. The lexical productive measures in L2 indicated that input quality and parental education as a composite measure were responsible for higher scores in English.

Regarding the input young bilinguals get at school, to our knowledge, the research is quite limited. Dixon (2011) conducted a study on the role of home and school language in predicting the linguistic performance of bilingual children. This study sought to provide an explanation of why Singaporean students perform so well in standardised international evaluations (which they take in English, the language of instruction). The research was performed on 284 bilingual kindergarten children

from Singapore who had been raised in bilingual home environments; 168 participants were Chinese/English bilinguals, 65 spoke Malay and English while 51 spoke Indian and English. The results show that the children who used more English at home or watched more TV in English scored better than those used it just at school. In other words, Dixon attributes the success not to the school programmes but to the family. “Singaporean students perform so well on international comparisons because they start school with a solid foundation of vocabulary in English, which is the language of instruction at school” (Dixon, 2011 p.159).

The studies in this section refer to the role of parental and school input in the acquisition of the general proficiency in the L1 and L2. In line with the objectives of our research, in the following sections we focus exclusively on studies performed in the field of lexical acquisition at very early ages. In this sense, one of the most controversial aspects is whether bilingual infants are able to separate the two languages from the very beginning or whether they start having one system which includes elements from both languages. This debate is important as it also underpins the way bilingual lexical development should be assessed. In the following section, we will examine the research conducted around this topic in an attempt to see what has been found so far regarding lexical development.

2.1.2.2. LANGUAGE SEPARATION HYPOTHESIS IN SIMULTANEOUS BILINGUALISM IN CHILDREN

One of the issues on which there is much debate regarding simultaneous bilingualism is language separation (i.e., “whether the bilingual child develops one or more language systems for his or her languages” Barnes, 2006). According to early research by Leopold (1970), a bilingual child is aware of the two different linguistic systems by the age of two. The example he provides is Hildegard, an English/German bilingual who at the age of 2.2 was found to separate the two languages when addressing different interlocutors. This conclusion was later reinforced by Vihman (1985). Nevertheless, what happens before the age of two is still under debate. On the one hand, Volterra and Taeschner (1978) point to the fact that at the beginning of the speech production, bilingual children start by having a single language system for both languages. On the other hand, researchers like Pearson et al. (1995), Deuchar and Quay (2000), Poulin and Goodz (2001) and

Grosjean (2010) provide evidence that these children develop two separate language systems at the lexical level from the onset of speech. This section discusses the Language Separation Hypothesis posited by Volterra and Taeschner (1978) by first summarising what it proposes and then by providing evidence against it from more recent findings.

In 1978, Volterra and Taeschner defined a three-stage model which described the early language acquisition stages of simultaneous bilingual children. They provided evidence that at the onset of speech, a bilingual child had a single language system. Thus, the initial stage (between the ages of one and two) was characterised by a linguistic system which included elements from both languages. During the second phase (after the age of two to three), the child starts to distinguish the two lexicons but applies the same syntactic rules to both of them. This stage is also characterised by the presence of more Translation Equivalents (TE), also known as ‘cross-linguistic synonyms’ (i.e., words that mean the same thing in different languages). The third stage (from age four onwards) reveals the existence of two lexically and syntactically different codes, but each language is associated with a specific person; that is, the L1 is associated with one caretaker (parents or schoolteachers) and the L2 with another. At this stage, mixing is a process that still occurs, with nouns and verbs being the categories mixed the most frequently.

This theory has prompted academic debate, especially regarding the first stage of the model, which refers to the first two years of life (i.e., between the ages of one and two). As a reminder, during this first stage, the bilingual infant is said to have a unique linguistic system which combines elements from both languages. These conclusions were based on the data Volterra and Taeschner (1978) obtained from the study of Lisa and Giulia, two German-Italian bilinguals who had been found to include elements from both languages in a single linguistic system during the first stage of their linguistic development. For example, Lisa used the Italian word ‘là’ meaning ‘there’ and the German word ‘da’, also meaning ‘there’, to refer to two different concepts irrespective of the interlocutor. She used the Italian ‘là’ for things that are not visible at the moment of speech and the German word ‘da’ to refer to things that are present at the time of speaking. Another similar case was Hildegard (Leopold, 1970) who was found to combine her two languages in one system during

her first two years. The first proof was the fact that the sounds she used were undifferentiated by language. She also mixed English and German words, thus failing to separate the two languages when speaking to a monolingual. Volterra and Taeschner (1978) also found that during this first phase, these children hardly used any TE. Accordingly, in the first stage of Lisa's lexical development, only three out of a total vocabulary of 87 words were TE ('yes' in English and 'ya' in German are translation equivalents, and they were part of Lisa's vocabulary between the ages of one and two). In the same vein, Giulia was found to know six TE out of total vocabulary of 83 words. These findings were also supported by Clark (1987), who proposed the Principle of Contrast with respect to simultaneous bilingual children. According to this principle (whereby every two forms contrast in meaning), these children would reject TEs, implying that they do not separate the two language systems.

A different view, in which simultaneous bilingual children are able to separate the two lexical systems from the onset of speech, is currently supported by modern research. The researchers that defend a new hypothesis that children have two different language systems from the start of speech "are more numerous and have good arguments" (according to Grosjean, 2010; p. 182). In what follows, we outline evidence of the existence of two separate systems at the lexical level based on the research conducted in the field of babbling, language differentiation and TEs. Poulin and Goodz (2001) showed that bilingual babbling was proved to be language-specific, which is a first indication of the separation of the two language systems even from the onset of speech. In the field of lexical differentiation, Bergman (1976) reported the case of her daughter, Mary, who learned English and Spanish at the same time. At the age of 15 months, she distinguished between her two languages by using Spanish to address to her Spanish babysitter and English to address her mother. These results also provide evidence against the Principle of Contrast and therefore Volterra and Taeschner's theory (as pointed out by Pearson et al., 1995; Deuchar & Quay, 2000; De Houwer et al., 2006). Furthermore, the study performed by Pearson et al. (1995) on 27 English-Spanish bilinguals aged 0.8 to 2.6 years old shows that 30% of all words 'coded in the two languages' (p. 346) were TEs. Similarly, Deuchar and Quay (2000) found that TE are produced from the onset of speech and increase with vocabulary growth. Later, De Houwer et al.

(2006) also reported that 13-month-old French-Dutch bilinguals are able to understand TEs.

Research has mainly shown that the evidence of two differentiated systems from very early ages is only clear at the lexical level. Barnes (2006) points out that the morphosyntactic research performed so far has not been able to provide any conclusive results regarding the existence of two separate systems at the morphosyntactic level within the first two years of life. There seems to be no consensus on when the simultaneous bilingual children are able to separate the two morphological systems. Meisel (2001) points that these children are able to separate the two languages “as soon as language-specific word order properties and inflectional morphology emerge” (Meisel 2001; p. 17). Regarding grammar, Deuchar and Quay (2000) suggest that clear data on syntax appears only after the age of two. In this sense, they found that Manuela, an English-Spanish bilingual they researched, was only able to distinguish the two languages syntactically at the age of 2.3 years. Therefore, language separation may depend on what linguistic aspect is being analysed.

Recently, Byers-Heinlein (2020) proposed ‘The Gradual Language Separation Hypothesis’, according to which language separation takes place gradually during the first few years of life. She supports this view by examining studies conducted on phonology. For instance, according to Singh et al. (2014), 18-month-old Mandarin-English bilinguals only start distinguishing the meaningful tones in their two languages at 24 months. This suggests that until the age of 24 months, they had a single phonetic system for both languages.

2.1.2.3. RECEPTIVE LEXICAL KNOWLEDGE IN SIMULTANEOUS BILINGUALISM IN CHILDREN

Another concern is whether simultaneous bilingual children follow the same stages of lexical development in both their languages as their monolingual peers. Of particular relevance is the comparison in vocabulary size. In this sense, researchers have wondered if simultaneous bilingual children lag behind their monolingual peers in terms of the number of words they understand and produce in one of their languages. The following two sections summarise the findings of studies comparing

very young bilingual and monolingual children in terms of stages of lexical development and vocabulary size.

Although language comprehension has received less attention than language production (Poulin et al., 2013), “an understanding of the process through which bilingual children learn to comprehend their two languages is necessary for a deeper understanding of bilingual language acquisition as a whole” (de Houwer et al., 2006; p. 332). Studies in this area include both case studies (e.g., De Houwer, 1990; Perez-Vidal, 1995; Cruz-Ferreira, 2006) and research performed with larger samples of participants (e.g., Pearson et al., 1993; Mahon & Crutchley, 2006; Wreker & Byers, 2008; Bialystok et al., 2010). This section summarises several findings in studies on bilinguals’ lexical developmental stages, vocabulary comprehension and size compared to monolingual and other bilingual peers.

Research points to similarities between the stages of lexical development of monolingual and simultaneous bilingual children. According to Cruz-Ferreira (2006), both monolingual and simultaneous bilingual children start responding to simple commands at around 8 months of age. Poulin et al. (2013) indicate that ten-month-old simultaneous bilingual infants are able to show recognition of familiar words in each of their languages, while Cruz-Ferreira (2006) found that Portuguese-Swedish bilingual children are able to recognise colours when they are between 17 and 19 months old. Likewise, research conducted on TE shows that bilingual children are able to understand TE from very young ages. De Houwer (2006) conducted a longitudinal study on 31 Dutch-French bilinguals five to 20 months old; the results indicate that by the age of 13 months all the children were able to understand both members of at least one TE pair. Other researchers agree that by the end of the second year of life, the comprehension of TE increases by 30% (Pearson et al., 1995; David & Li, 2008).

Regarding lexical comprehension, we will first cite the research conducted exclusively with bilingual children. As in the case of monolinguals, it has found that one variable that influences a child’s comprehension is connected to input quantity (De Houwer et al., 2006; David & Li, 2008) and quality (Sun et al., 2020). A study performed by Pearson et al. (1997) on 25 Spanish-English bilingual infants

with different patterns of language exposure shows a strong correlation between the amount of input heard and language comprehension. They also concluded that simultaneous bilingual children may vary in their word comprehension: some understand more words and some fewer depending on the amount of exposure to the respective languages. These findings are supported by De Houwer (2007), who has found that children who understand more in their L1 understand more in their L2, and those who understand less in their L1 understand less in their L2.

With respect to simultaneous bilinguals' total receptive vocabulary size, two types of studies are relevant for our research. The first takes into consideration both L1 and L2 to measure the total receptive vocabulary size of simultaneous bilinguals and then compares it with the total vocabulary size of monolinguals. The second looks at the number of words bilinguals and monolinguals understand, but taking into account their common language only. The results of the studies in the first line of research (i.e., where both L1 and L2 were taken into account in the case of bilingual participants) are inconclusive. Some of them conclude that there are no significant differences in the total receptive vocabulary size of bilinguals and monolinguals. For instance, in a study performed on 25 English-French bilinguals and 12 French monolinguals, Poulin et al. (2013) found no differences between the two groups in terms of total receptive vocabulary size. Other studies revealed that simultaneous bilinguals outperformed their monolingual peers in terms of receptive vocabulary size. For example, Pérez-Pereira (2008) found that 461 Galician-Spanish bilinguals outperformed 275 Galician monolinguals in terms of total receptive vocabulary size in Spanish. In the same vein, De Houwer et al. (2014) conducted a study on 31 French-Dutch bilinguals and a group of 30 Dutch monolinguals. The results indicated that bilinguals understood more words in both their languages than their monolingual peers did in Dutch.

With respect to the studies which took into consideration only the shared language in order to compare bilinguals' and monolinguals' receptive vocabulary sizes, all the outcomes point to monolinguals being superior. In 1978, Doyle et al. conducted a study on 13 French monolinguals' and 13 French-English bilinguals' receptive vocabulary. The results showed that the scores of the bilinguals in their dominant language were lower than the scores of their monolingual peers. Bialystok et al.

(2010) conducted a larger study with big samples of both preschool and school-aged children. They selected 1738 children between the ages of three and ten from different studies previously carried out by Bialystok and assembled them into two groups. The first one had 772 English monolinguals. The other group was made of 966 participants who were educated in English at school. Because this second group was fluent in both their native language and English, they were considered bilinguals. The results of the study showed that the monolinguals understood more words in English than their bilingual peers.

These studies in which monolinguals were found to outperform bilinguals have only taken into account the bilinguals' linguistic performance in one language. Nevertheless, we should also recall that "the bilingual's communicative competence cannot be evaluated through only one language; it must be studied instead through the bilingual's total language repertoire as it is used in his or her everyday life" Grosjean (1989, p. 6).

2.1.2.4. PRODUCTIVE LEXICAL KNOWLEDGE IN SIMULTANEOUS BILINGUALISM IN CHILDREN

Research on productive vocabulary is more extensive than on receptive vocabulary. Therefore, this section relies on more studies than the previous one. We begin by describing the stages of lexical production in young simultaneous bilinguals and then provide information related to differences among simultaneous bilingual children as well as differences between bilinguals and monolinguals in terms of productive vocabulary size.

Simultaneous bilingual children seem to follow the same stages of productive lexical development as their monolingual peers (De Houwer, 2009; Grosjean, 2010). As a result, the existing research places the monolingual and bilingual babbling stage within the same time span. Bosch and Sebastian (2001) situate the monolingual babbling stage between the ages of five and twelve months. According to De Houwer (2009), bilingual babbling starts in the second half of the first year of life and continues even after babies start saying their first words (Leopold 1953, 1970). Comparing canonical babbling in monolinguals and bilinguals, Oller (1980) found the two groups started babbling at the same age.

Between the twelfth and the sixteenth month of life, both monolingual and bilingual infants go through the one-word stage (Clark, 2003; De Houwer, 2009). Caselli et al. (1999) and later David and Li (2005) point to the one-word stage as the reference stage in which simultaneous bilingual infants (just like their monolingual peers) are able to produce their first 100 nouns referring to routines and games. However, researchers agree that sometimes word production does not coincide in both languages in these children (Vila, 1984; Holowka et al., 2002; Cruz-Ferreira, 2006). Children can produce their first words in the L1 first, while the first words in the L2 will be produced slightly later. Moreover, the difference between the times a child produces the first words in the L1 and L2 has nothing to do with any similarities that may exist between the two languages (Aguila et al., 2005; De Houwer et al. 2007). In other words, the fact that the L1 and L2 are typologically similar will not necessarily imply that children will start producing the first words in the L2 immediately after they start producing the first words in the L1.

According to Sinka and Schelletter (1998), bilingual infants in the one-word stage are able to make appropriate language choices for each situation. Bilingual young children can therefore switch from unilingual utterances in L1 to unilingual utterances in L2 (De Houwer, 1990; Genesee et al., 1996; Lanza 1997; Deuchar & Quay, 2000). TEs were reported to be used by bilinguals from the onset of speech (Quay, 1995; Pearson, et al., 1995; Johnson & Lancaster, 1998), but Nicoladis and Secco (2000) found that bilingual children used fewer TEs than their monolingual peers during the one-word stage.

Most monolingual children experience a vocabulary spurt at around 16-18 months (Clark, 2003). Research performed on bilinguals' vocabulary spurt in both languages shows similarities to monolinguals' timing (Vila, 1984; Deuchar & Quay, 2000). During this stage, the production of TEs increases to around 20-25% of the child's total vocabulary (Nicoladis & Secco, 2000). According to Patterson (1998), most simultaneous bilingual children start using two-word combinations between 18 and 23 months, which is similar to the monolingual lexical production milestones cited by Clark (2003). By the end of their second year of life, simultaneous bilingual children are able to form two- or three-word utterances. This

stage is also mentioned in Caselli et al. (1999), David and Li (2005) and Hernandez-Calafat (2021), with the emergence of productive grammar, when grammatical categories are acquired. Children are already able to produce complex sentences by the age of five (Costa, 2006).

In terms of the number of words bilingual children produce, De Houwer (2009) points to differences ‘within each language and across children’ (p. 226). There is variation among bilingual toddlers in how many different words they produce in each language at a given time (Holowka et al., 2002; Junker & Stockman, 2002; David & Li, 2003). Nevertheless, it is agreed that children who say more words in one language say more words in the other (Patterson, 2000 and 2002; Conboy & Thal, 2006), and that this variation is directly related to exposure. For instance, Pearson, et al. (1997) and Patterson (2002) show that the more Spanish children hear, the more Spanish they produce. Other studies performed on bilingual children by Vila (1984), Pearson et al. (1997) and David and Li (2008) conclude that the reason a child produces more words in a language is because (s)he spends more time in the presence of speakers of that language.

Just as with receptive vocabulary size, research comparing the total vocabulary size of simultaneous bilinguals and monolinguals has followed two different lines of inquiry. The first takes both languages spoken by the bilingual child into consideration (Pearson et al., 1993; Hoff et al., 2012; Poulin et al., 2013; McClain et al., 2021), while the second compares the vocabulary sizes only in the shared language (e.g., Hoff & Elledge, 2005; Byers & Wrecker, 2009).

The studies where both languages were considered when calculating the total vocabulary size show no significant differences between simultaneous bilinguals and monolinguals. The results of the research conducted by Pearson et al. (1993) on 25 English-Spanish bilinguals and 35 children growing up either in English or Spanish monolingual homes show no differences between the two groups’ productive vocabulary size. In the same vein, the study performed by Poulin et al. (2013) comparing the expressive vocabulary size of 25 bilinguals and 18 monolinguals shows no significant differences between the two groups. It should be noted that out of the 25 bilinguals, nine spoke English and French, eleven spoke

French and English, three spoke English and Italian, one spoke English and Hebrew and one spoke French and Turkish. In the monolingual group, twelve participants were English while six were French.

When only the shared language was taken into consideration, the results favoured the monolingual children. In this sense, the study performed by Hoff and Elledge (2005) comparing 39 bilingual children (between 16 and 30 months old) and 63 English monolingual children the same age concludes that monolingual children earned better scores in lexical production than bilingual children. The 39 bilinguals had English as their L1 and Spanish, Portuguese, Creole, Hebrew or French as their L2. Similarly, Byers and Wrecker (2009) showed that monolinguals had a larger productive vocabulary than their bilingual peers. Their study compared 16 English monolinguals (17 and 18 months old) and 16 bilinguals of similar ages. The bilingual participants had been simultaneously exposed to English and another language since birth.

As a final remark with respect to receptive and productive lexical development, it is important to stress the fact that bilingual children's lexical knowledge cannot be viewed as the knowledge of two monolinguals in one (Grosjean, 1989). Therefore, it has been suggested that lexical development be measured in both languages by means of conceptual scores (i.e., the number of concepts for which the child has a word in at least one language; see Pearson et al., 1993), as they "offer a more equitable means of measuring lexical performance" (McClain et al., 2021). This has implications in the design and creation of young bilinguals' early lexical assessment tools as well.

2.1.3 EARLY SEQUENTIAL BILINGUALISM

According to Baker (2006), sequential bilingualism refers to the acquisition of a new language (L2) to native-like proficiency after the age of three. This definition is also supported by other researchers such as Taeschner (1983), McLaughlin (1984) and Butler and Hakuta (2006).

There are several ways in which the L2 can be acquired. For early sequential bilinguals, one of the most common ways is as a result of immigration. Often,

children acquire the L2 both informally (e.g., by interacting with the new community members) and formally (e.g., through schooling). Another means of L2 acquisition is exclusively through formal instruction, as nowadays an increasing number of monolingual young children are exposed to an L2 in language immersion schools. Because research on young learners in immersion schools is especially important for our research, we begin by describing the most relevant immersion programmes for young L2 learners. Next, we continue by summarising selected studies conducted with young sequential bilingual learners. Finally, we address the age factor, which has been widely researched with L2 learners: age as an individual variable cannot be ignored when dealing with sequential bilingualism. The section ends with brief presentation of some of the research on lexical comprehension and production in young sequential bilinguals.

2.1.3.1 EDUCATIONAL SETTINGS FOSTERING EARLY L2 ACQUISITION AFTER L1

As mentioned in the previous section, the L2 can be acquired formally in educational settings (see Baker, 2006; Murphy 2014). One of the most frequent forms of L2 acquisition in formal settings is when the community language is taught in schools to immigrant students, that is, minority language students.

Back in 1981, Baker and de Kanter pointed to four L2 instructional alternatives found in the US education system whereby the new community language (English) is taught to very young immigrant children in educational settings. The first way to acquire the L2 is through *Submersion*, in which children are placed in a class with native English speakers without any chance to use their own language to communicate. The second alternative is *English as a Second Language* (ESL), where children are withdrawn from the class for several hours for formal compensatory English classes, where English is taught more explicitly as a foreign language (FL). The third alternative is *Transitional Bilingual Education* (TBE), which is the most common form of bilingual education in the USA. In this kind of classroom, instruction is provided in students' L1, and later (in approximately two years' time) the instruction is conducted exclusively in English. Thus, within two to three years, students are expected to have basic English language proficiency and attend English-only classrooms. *Structured English Immersion* (SEI) is the fourth

alternative in which English is the language of instruction and is used fully with very few L1 clarifications.

The outcomes of TBE and SEI on English proficiency are varied and have been widely discussed in the literature in recent decades. According to Greene (1997) and Slavin and Cheung (2005), there is an increasing number of studies on English language comprehension conducted with sequential bilinguals in the USA that favour the TBE approach. Other research focusing on language production, such as the study conducted by Tong et al. (2008), has found no difference between the two approaches.

Another form which has appeared quite recently is when an L2 other than different from the community language is taught in schools through immersion. Murphy (2014), among others, refers to two types of programmes: one-way immersion and two-way immersion. One-way immersion is exemplified by Lambert and Tucker's experiment in Canada. In 1965, a group of English-speaking parents in Canada succeeded in getting the district to implement a French immersion kindergarten at the request of the Ministry of Education. Prof. Lambert was asked to assess the programme in terms of language skills, mathematics, general intellectual development and attitudinal development. The results of the study carried out by Lambert and Tucker (1972) indicated that with respect to language development, the children studying in the immersion programme were more proficient in the L2 than those who receive regular L2 instruction. Furthermore, they developed similar L1 proficiency as those who received conventional L1 literacy skills. In the case of early total immersion students, there could be a lag in L1 literacy skills in the early years of immersion, but no such lag is recorded when immersion starts later on, such as in the middle of elementary school. With respect to achievement in other domains such as mathematics, science and social studies, this study found no differences between immersion and regular school students. Furthermore, research conducted in this field has also stressed the cognitive advantages of immersion students (see Puric et al., 2017 and Barbu et al., 2019). Genesee (1983) agrees that these outcomes from Canadian studies sustain the integrated teaching approach for language instruction, which is the use of the L2 to teach significant parts of the curriculum.

However, other studies conducted in similar settings have shown a lack of grammatical accuracy or lexical diversity among children studying in immersion programmes compared to those in regular schools (Harley & Swain, 1984; Lyster, 1994). As a result, Lyster (2007), for example, advocated focus-on-form and cross-linguistic or counterbalanced approaches in L2 teaching through immersion. Accordingly, a focus on form in relevant activities for learners can lead to long-term improvements (see Murphy, 2014). In contrast, a cross-linguistic approach would lead students to acquire better L1 and L2 representations. In other words, using children's L1 in immersion settings could have benefits for lexical diversity and accuracy. The research performed by Torregrossa et al. (2023) is a case in point: they conducted a study on 33 Greek-Italian sequential bilinguals between the ages of eight and eleven. These children, who were in a one-way immersion programme in a multilingual school in Greece, were asked to rate Italian sentences for acceptability. The Italian sentences were presented in different ways, one of which included a grammatical/ungrammatical Greek translation. The results indicated that the children were better at noticing and explaining bilingual grammatical errors when presented in the bilingual mode.

In relation to subjects such as science, Turnbull et al. (2012) showed that within a French immersion setting, children who were taught using a counterbalanced approach (i.e., using both L1 and L2) managed to make progress not only in French (L2) but also in science, compared to the students who were taught traditionally, that is, just in the L2.

Two-way bilingual programmes are a mixture of the one-way immersion discussed above and the immersion programmes for minority students mentioned at the beginning of this section. These programmes offer both content language instruction (CLI) and Language and Arts instruction in both languages, for specific periods of time, and they are aimed at both majority and minority language learners. They are quite frequent in countries with a high number of migrants, such as the US. Unlike the mixed results found in one-way immersion programmes, the evaluations of two-way models have concurred on their benefits. More specifically, they have shown that students in these types of programmes perform at or above

the level of their peers attending regular programmes. These results are also comparable or better compared to the L1, maths and reading (see, for example, Murphy 2014).

Since the first implementation of one-way and two-way immersion education programmes in Canada and the USA, a large number of countries all around the world have chosen to give a vote of confidence to this approach towards language learning, the result being that a large number of children are becoming bilingual at very young ages. The types of programmes may differ in terms of ages of onset and amounts of exposure, depending on school policies.

Cases of full one-way immersion have been recorded in the Maple Bear Canadian kindergartens (www.maplebear.ca) all over the world, where children between the ages of three and five are exposed to English or German while at the school. This exposure decreases to approximately 90% in the last year of kindergarten, as the children start preparing for school, and it continues to diminish throughout primary and secondary education. Other kindergartens and schools in Spain introduce immersion at the age of three. For example, at the Europa International School in Sant Cugat, Barcelona (www.eisbarcelona.com), English is introduced at age three, with an amount of exposure equal to 80%, which decreases to 60% in primary school.

These children's lexical development trajectories may be different from those of children acquiring the two languages simultaneously from birth. Under these circumstances, it is important that we also address the lexical development of sequential bilingual learners in terms of comprehension and production.

2.1.3.2 LEXICAL DEVELOPMENT IN SEQUENTIAL BILINGUALISM

In spite of the fact that more and more children acquire an L2 early in life either through immigration, schooling or both, the research targeting this group has been scarcer than with simultaneous bilinguals. Yet two main strands of interest have arisen. Researchers have been concerned first with the relationship between the L1 and the L2 and secondly with comparing the linguistic development of sequential bilinguals to monolinguals and simultaneous bilinguals.

In 1978, researchers' interest was raised by Cummins's Developmental Interdependence Hypothesis (DIH), which has a fair number of supporters and detractors. Just as with simultaneous bilinguals (De Houwer, 2007), this theory highlights the interdependence of a student's L1 and L2 and suggests that proficiency in L1 can have a positive impact on L2 development. For example, the research conducted by Castilla et al. (2009) and Pham (2016) support this theory, proving the existence of a lexical-grammatical association between L1 and L2 and proven that L1 vocabulary size supports L2 lexical growth. More exactly, Castilla et al. (2009) analysed the language interdependence of 49 Spanish-English sequential bilinguals who were 4.7 years old on average. the children were first tested in Spanish by means of the Bilingual English Spanish Assessment, their Mean Length of Utterance (MLU) in words and their lexical diversity. Nine months later they were tested in English using the same measures. Results indicated strong correlations between the languages. In 2016, Pham conducted a similar four-year longitudinal study with 33 seven-year-old Vietnamese-English sequential bilinguals who were exposed to the L2 at the age of four. Data collection started 32 months after first exposure to the L2. The children were tested in both languages with a storytelling and a picture-naming task. Results indicate the existence of grammatical and lexical size associations for the two languages.

Other studies such as the one conducted by Kohnert (2002) point to a shift in dominance, that is, from the L1 being the dominant language to balanced development and finally with L2 being stronger than the L1. In this sense, 28 Spanish English sequential bilinguals between the ages of 6.10 and 15 were tested on a picture-naming task after a five-year exposure to the L2. The study was longitudinal, and the results indicate an increase in accuracy in English from T1 to T2, along with a decrease for Spanish.

At the other end of the continuum, Kohnert et al. (2010) found no evidence of any cross-linguistic relationship between the two languages. The participants of this study were 19 Hmong-English sequential bilinguals between the ages of 2.11 and 5.2. They were assessed on a storytelling and picture-identification task. The results showed weak cross-domain associations between the two languages. Likewise, the

number of different words and the MLU in L1 were significantly higher than in L2. In other words, there was no evidence of a cross-linguistic relationship.

Other studies in the field of sequential bilingualism compare monolinguals and sequential bilinguals in terms of their vocabulary size. To our knowledge, most of the research conducted on lexical comprehension and production compare the linguistic performance of young sequential bilinguals to monolinguals with reference to the shared language (Appel & Verhallen, 1998; Schoonen & Verhallen, 2008; Marianis & Chondrogianni, 2010). For instance, Schoonen and Verhallen (2008) analysed the receptive vocabulary of third and fifth graders from different linguistic backgrounds who were learning Dutch as an L2. These sequential bilinguals were compared to a group of Dutch monolinguals. The outcomes of this study indicate that the older sequential bilinguals earned significantly better results than their younger peers. Nevertheless, both sequential bilingual groups were shown to lag behind the Dutch native participants.

In terms of lexical production, Verhallen and Schoonen (1988) compared the words produced in a word association test of a group of third- and fifth-grade Turkish-Dutch sequential bilinguals against a group of Dutch monolinguals. The results evidenced the superiority of the monolingual participants. Another study, by Appel and Vermeer (1998), compared the lexical development of three groups of secondary school learners: the first were Dutch monolinguals, the second one were Turkish-Dutch bilinguals taking part in an extra vocabulary learning programme, while the third group were Turkish-Dutch sequential bilinguals were not involved in the programme. The study showed that the monolingual learners outperformed sequential bilinguals in lexical comprehension and production. The results indicated that the bilinguals who received extra amount of lexical instruction progressed in all lexical reception measures and production and outperformed their bilingual peers who had not benefited from the extra lexical instruction. Nevertheless, in spite of the vocabulary learning programme, the test results placed them below the monolingual Dutch control group.

Later on, researchers started to question the monolingual norm of testing, that is, comparing monolinguals with bilinguals only in the shared language. As mentioned in section 2.1.2.3. above, McClain et. al (2021) advocated for conceptually scored

vocabulary assessments, alleging that “a monolingual view of bilingualism falsely assumes that bilinguals should have equal proficiency in both languages, that the two languages operate in different spheres” (p. 4).

Finally, one well-known study comparing simultaneous bilinguals with sequential bilinguals of similar ages was performed by Litcofsky et al. (2015). Adult participants in the two groups were compared in terms of their lexical comprehension and production in both languages. The results indicated that the two groups performed similarly in comprehension in both languages. With respect to oral production in L2 (English in this case), the results favoured simultaneous bilinguals. Their better production in English was therefore related to an earlier AO and greater exposure.

In sum, the limited amount of research focusing on receptive and productive vocabularies in sequential bilinguals has yielded different results. When compared to monolinguals (in terms of the shared language), sequential bilinguals seem to be at a disadvantage. On the other hand, when measured against simultaneous bilinguals (in terms of their L1), no significant differences in comprehension and production were found between the two groups. Nevertheless, when the comparison is performed with respect to the L2, simultaneous bilinguals are found to perform significantly better than sequential bilinguals in oral production tasks.

2.1.3.3 THE INFLUENCE OF AGE OF ONSET (AO) AND AMOUNT OF EXPOSURE ON LINGUISTIC PERFORMANCE

In sequential bilinguals, exposure to the new L2 takes place after the age of three. Some of these children are introduced to the L2 during early childhood, while others are later on. Researchers have wondered whether an early start contributes to the achievement of L2 proficiency. In an attempt to summarise the findings on the age factor in L2 acquisition, this section first provides a brief overview of the Critical Period Hypothesis (CPH). Next, it summarises findings with respect to AO defined as ‘time to start learning a language’ and L2 exposure from the research conducted in both naturalistic and formal settings.

The ‘Critical Period Hypothesis’ (CP) (Lenneberg, 1967) is well-known in the field of SLA. It was first posited for L1 acquisition and later on extended to L2. Its strong version states that L2 acquisition is impossible after a certain point, namely puberty. The weak version says that L2 acquisition is still possible after puberty, but it is more difficult and incomplete. At present, there is some reluctance to use the term ‘Critical Period’ as a result of a number of studies which have proven that achieving a native-like competence in L2 is possible after puberty. Among them is the study conducted by Bongaerts et al. (1995) on the acquisition of an English accent by Dutch-English sequential adult bilinguals. Participants in the study were divided into three groups. The first one, the control group, consisted of five English monolinguals approximately 30 years old. The next two groups were made of ten approximately 37-year-old Dutch/English sequential bilinguals and twelve 30-year-old Dutch/English sequential bilinguals. The results of the study indicate that the oldest group of bilinguals outperformed the youngest group in all the tasks that participants were assessed on.

Comparing children and adults, researchers seem to converge on the idea that ‘the earlier, the better’ (Byers & Lew, 2013) based on biological and psychological reasons. The former assumes the existence of biological constraints and innate mechanisms since from a biological standpoint our brains seem to be more receptive to language earlier in life (Johnson & Newport, 1989). Moreover, children are also considered to be equipped with a better system for language learning than adults. The latter is based on the fact that psychologically, children are said to be more successful acquiring an L2 because they are more motivated and less inhibited.

Indeed, other studies conducted in naturalistic settings comparing different groups of sequential bilingual children have also reached the conclusion that an earlier start leads to a higher L2 proficiency. Hyltenstam (1988, 1992) examined the written and oral production of three groups of secondary school students in Sweden. The first group consisted of twelve monolinguals, the other group was made up of twelve sequential bilinguals with Swedish as the L2 and with an AO earlier than six, while the last group was composed of twelve sequential bilinguals with an AO of seven. The conclusions pointed to the fact that all the sequential bilinguals with the AO of seven committed more errors than the monolingual group. Early Starters

(ES), that is, AO= 6, followed two trends: some of the learners were more similar to their monolingual peers, while the other were closer to the older bilingual group. Spadaro (2013) analysed the lexical performance of four groups: the first group consisted of ten monolingual learners. The other three groups were made of sequential bilinguals with different AOs: earlier than six, between seven and twelve and older than 13. The results indicate that the youngest group of sequential bilinguals outperformed the other two bilingual groups in vocabulary use tasks and appeared to sound more native-like in the oral task.

Nevertheless, other studies suggest that *earlier is better in the long run*. The conclusions of studies performed by Snow and Hoefnagel- Höhle (1978) and Singleton (2000) initially favour late starters (LS), who are eventually overtaken by ES. Snow and Hoefnagel- Höhle (1978) conducted a study in the Netherlands with English learners whose L2 was Dutch. The results indicate that the LS initially overtook the ES in terms of L2 vocabulary. However, after a year, the ES began to catch up with the LS in tasks such as storytelling and translation. Singleton (2000) highlighted that there is ‘evidence of more rapid initial learning on the part of adult and adolescent subjects, but also of younger beginners catching up with and beginning to overtake older beginners after about 12 months of exposure’ (Singleton, 2000, p. 22).

It is often true that ES have also received more exposure than LS. Nevertheless, irrespective of the AO, in naturalistic settings, sequential bilinguals need at least 10-25% exposure to the L2 to acquire it successfully (Kessler, 1984; Kastos & Alexopolou ,2011). According to according to Byers and Lew (2013): ‘the take-home messages about bilingual language exposure are clear: more is better, and earlier is better’. As this context is relevant to the study in this thesis, it shall be further discussed in section 2.2.2., which addresses L3 acquisition by young bilingual learners.

2.2. THIRD LANGUAGE ACQUISITION

As we have seen in sections 2.1.2. and 2.1.3., the acquisition of two languages can take place simultaneously from birth (simultaneous bilingualism) or consecutively (sequential bilingualism). In reference to the acquisition of more than two languages, De Angelis (2007) uses the term TLA (Third Language Acquisition) to refer to all languages learned after the L2. As already stated in the introductory section, in this dissertation, TLA will refer specifically the acquisition of an L3. This section will present relevant research conducted in the field of TLA, with a special focus on young learners acquiring an L3 at an early age.

When comparing L2 and L3 acquisition, researchers such as Herdina and Jessner (2000) and Wang (2008) have repeatedly shown that learning two languages is not the same as learning three. First of all, according to Cenoz and Genesee (1989), learning an L3 involves all the factors and processes associated with L2 learning, as well as other complex factors associated with the interaction between the three languages spoken by the same person. Secondly, Cenoz (2013) and many others agree that a person speaking three or more languages has greater cognitive ability and higher metalinguistic awareness than a person speaking two. Thirdly, three languages can be manifested in production in more ways than two languages can: bilinguals may choose to speak language A, language B or a combination of the two. Therefore, they have three linguistic choices. A trilingual has a total of seven choices (language A, language B, language C, and their possible combinations AB, BC, AC and ABC) (Wang, 2008). Finally, the contexts in which a person learns an L3 are also more varied than those in which an L2 is learned.

Thus, as we have seen in section 2.1.3.1., an L2 can be either learned in naturalistic settings as the community language or in FL/L2 in instructional settings. In the case of TLA, Cenoz (2000) described four ways through which an L3 can be acquired. Firstly, L3 can be acquired simultaneously with L1 and L2. Sometimes the three languages are acquired in naturalistic settings, but two of the languages can also be acquired in a naturalistic setting while the third is acquired in a formal setting. A second way could be that when L3 is acquired after the simultaneous acquisition of L1 and L2. This may involve acquiring L3 either in formal or naturalistic settings.

The third way is when all three languages are acquired consecutively (first L1, then L2 and later on L3). L2 and L3 can be acquired either in formal and/or naturalistic settings. The fourth way is when L2 and L3 are being learned simultaneously after the acquisition of L1. In this case as well, L2 and L3 can be acquired in naturalistic settings and formal settings. For instance, this occurs with immigrants to Catalonia and the Basque Country who are exposed to both Spanish and Catalan/Basque both in the community and at school. This scenario is also mentioned by Hufeisen and Jessner (2019).

In line with the study conducted in this thesis, this section references two types of studies. The first is research targeting the simultaneous acquisition of L1, L2 and L3. Such studies, where the three languages are acquired at the same time during early childhood, are rather scarce and have been conducted with small samples of participants (see Hoffmann, 1985; Dewaele, 2000; Barnes, 2006; Montanari, 2006; Wang, 2008). Nevertheless, they are relevant to our study as they target very young children and describe the milestones of early lexical development.

Secondly, we will focus on studies where participants have acquired L3 in formal settings after the simultaneous acquisition of L1 and L2. A large number of such studies target the later acquisition of L3 in formal settings by bilingual adults and school-aged children (Barnes, 2006) and mostly analyse cross-linguistic influence, (i.e., L1/L2 influence on L3) as well as the age factor, where older children, teenagers and adults learning an L3 are compared with respect to their linguistic skills in the L3.

This section is divided into three main parts. With the aim of providing a broader background for our research, the first section (section 2.2.1.) summarises findings from the research on L3 acquisition by bilingual adults. Section 2.2.2. looks into TLA by young learners. We introduce two lines of research: the first addresses the simultaneous acquisition of L1, L2 and L3, while the second focuses on the later acquisition of L3 as an FL by bilingual young learners. Finally, given that this thesis focuses on vocabulary, in section 2.2.3. we provide an overview of the research on the organisation of the multilingual lexicon, as well as possible relationships between the three languages.

2.2.1. L3 ACQUISITION BY BILINGUAL ADULT LEARNERS

Although L3 acquisition by bilingual adult learners does not fall within the scope of this research, the decision to discuss it in this section stems from the fact that the results of the studies conducted with adults are relevant for the study of the language development in younger learners. This section presents several relevant studies carried out in the areas of cross-linguistic influence (CLI) and the age factor because we believe that their results are significant when interpreting research with other populations.

Most of the L3 acquisition studies conducted with adults learning an L3 in formal settings are in the field of CLI. These studies show that when learning an L3, adult learners rely on their L1 and L2 knowledge. According to De Angelis (2007), the way the L1 and L2 influence L3 acquisition is related to aspects such as language distance, proficiency, recency of use, length of residence and order of acquisition (see Cenoz, 2003 and Hufeisen & Jessner, 2019). Most of the research leading to these findings was conducted with sequential bilingual adults.

For instance, a recent study by Pinto (2013) with a group of Arab-French university students learning Portuguese as an L3 showed that the source language was frequently L2 (in this case French), since it is typologically closer to Portuguese. In the same vein, Bartolotti and Marian (2016) showed that language distance had an influence on L3 word spelling and pronunciation. In their study, 20 English-German bilingual adults were learning 48 words in an artificial language similar in type to L1 and L2. The results of the study indicated that these adults relied on word likeness in L1 and L2 to acquire the orthographic structure of the new L3 words.

Research has also been carried out with respect to the age factor. Some of these studies focus on adults alone, while other compare adults with young learners. For example, Pfenninger (2014) analysed the effect of the AO and exposure on the linguistic performance in English as an L3 of four groups of Swiss-German simultaneous bilingual adults. The first group of early starters (Early CLIL) consisted of 50 adults who started to learn English in primary school through a CLIL programme which continued all the way through secondary school. The

second group (Early Mix) consisted of adults who started to learn English in primary school through CLIL but during secondary schools received less hours of instruction in the L3 as a result of shifting to formal instruction English classes. The third group (Late CLIL) was comprised of 50 adults who started to learn English at an older age through CLIL, while the last group contained 50 adults (Late non-CLIL) who began learning English at the same age as the Late CLIL group but through formal instruction. The best results were found for the Early CLIL group, which confirms the link between the AO and massive exposure.

On a different note, Muñoz (2006) conducted a study (the BAF project) which compared the linguistic performance in L3 English of bilingual Catalan-Spanish adults to children and adolescents. In this case, the amount of exposure was controlled. The aim was to check whether there were age-related differences in learning rate (i.e., how fast students progressed in a language) and in ultimate achievement (after 10 years), also taking the exposure factor into account. Three groups of participants were involved. The first one consisted of children who started to learn English at age eight. The second one consisted of older children with AOs of eleven. The third group consisted of adolescents with an AO of 14. The last group was comprised of adults who started to learn English after age 18. All participants were tested on a wide range of linguistic skills at three different points, namely after 200, 416 and 726 hours of exposure. The results point to age-related differences in the learning rate. After 200 hours of exposure, adults and adolescents showed a quicker learning rate than younger children. These differences are explained by Muñoz (2006) in terms of cognitive development: ‘the older learners’ superior cognitive development also allows them to take greater advantage of explicit teaching process in the classroom’ (p. 33). Because the findings related to the age factor are of particular interest for this thesis, they will also be discussed in section 2.2.2.2.2, where we focus on the research conducted with young participants.

2.2.2. L3 ACQUISITION BY YOUNG CHILDREN

This section is comprised of two parts. The first one ([section 2.2.2.1.](#)) deals with the acquisition of the L3 at the same time as the first two languages (L1 and L2). The second part ([section 2.2.2.2.](#)) targets the acquisition of L3 after L1 and L2 have already been acquired.

2.2.2.1. SIMULTANEOUS ACQUISITION OF L1, L2 AND L3

Early or simultaneous trilingualism refers to the simultaneous acquisition of L1, L2 and L3 before the age of three (Cenoz, 2000; Baker, 2006; Barnes, 2006). Because the number of studies targeting early vocabulary acquisition is quite limited, we will also provide data on research carried out in other domains, such as morphosyntax, with the aim of offering a solid background for this thesis' main study. This section will follow the same structure as the one on simultaneous bilingualism. Thus, we begin by offering an overview of parental and school input ([section 2.2.2.1.1.](#)). Then, we address the language separation hypothesis in early trilingualism ([section 2.2.2.1.2.](#)). In [section 2.2.2.1.3.](#), we provide information on receptive and productive vocabularies of early trilingual children by citing several case studies.

2.2.2.1.1. PARENTAL AND SCHOOL INPUT

Early trilingual children can be exposed to the three languages in different ways. Cenoz (2000) refers three input models of how the three languages can be acquired. The first and the most common one is where two languages are acquired at home following the 1P/1L (1 person, 1 language policy) while the third one is learned in the community or at school. This latter model is the case of bilingual children in Catalonia and the Basque Country who are spoken to in Spanish by one parent and in Catalan/Basque by the other parent, while at school (particularly language immersion schools) they are addressed in English or French. In the second model, children learn the local language at home following the 2P/1L (2 people, 1 language policy) principle, another regional language in the community and an L3 at school. In Zaire, for example, a country in which several regional languages are spoken (Kikongo, Tshiluba, Lingala, Kiswahili), parents can speak Kikongo at home while

the community language is Tshiluba. Thus, children would acquire both these languages simultaneously. During their school years, they will learn French, which is the official language of schooling. In the third model, which is not as common, the child is simultaneously exposed to the three languages both at home and within the community. For example, in Luxembourg, children are simultaneously exposed to Luxembourgish, French and German both at home and school.

Research on early trilingualism mainly focuses on the first model. For example, Cristina and Pascual (Hoffmann, 1985) were spoken to in Spanish by their father, German by their mother and English by community members and school teachers. Livia (Dewaele, 2000) was born in England to a French father and a Dutch mother and was spoken to in Dutch by her mother, French by her father, following the 1P/1L principle, and English by the community members. Leandre and Dominique (Wang 2008) were born in the USA and had been acquiring English (the community language), French (the father's native language) and Chinese (the mother's native language) since birth, following the 1P/1L principle. Jenny (Barnes, 2006) was born in the Basque Country. Basque is her father's L1, and he has a B2 level in English, while her mother was born in England but has lived in Spain since 1975. Jenny has been exposed to Basque through her father and English through her mother. When in the presence of Jenny, the parents address each other in English but keep the 'one-parent, one-language policy' when addressing the child. Since birth, Jenny was also spoken to in Spanish by her caregiver. Kathryn (Montanari, 2006) is a Tagalog, Spanish and English trilingual. From the moment she was born, she was spoken to in Tagalog by her mother and grandparents, Spanish by her father and English by her sister and community members.

With respect to the research carried out with simultaneous bilinguals on the role of the parental input, we have already seen in [section 2.1.2.1](#) that applying the 1P/1L principle will not prevent simultaneous bilinguals from mixing the two languages (Patterson, 1999). With respect to the role played by the community (school) in input transmission, the conclusions of the few studies carried out so far (see Dixon 2011) show that the input a simultaneous bilingual child receives from their community members has an important role in their lexical comprehension and production. The studies in this area have mainly focused on the linguistic

development of these young learners. For instance, some research has been carried out on the language separation hypothesis and the lexical production and comprehension of early trilinguals.

2.2.2.1.2. THE LANGUAGE SEPARATION HYPOTHESIS

There is scarce research on the Language Separation Hypothesis in early trilingual young learners. Some researchers (e.g., Mikes, 1990; Turell, 1997; Montanari, 2006) have corroborated the findings in the field of simultaneous bilingualism with respect to the separation hypothesis: their studies show that trilingual children show signs of having three different linguistic systems from the onset of speech. Nevertheless, there is also some evidence against the Language Separation Hypothesis (e.g., Hoffmann, 1985), reporting that language differentiation did not actually occur since the onset of speech in early trilinguals but later on.

In what follows, a brief account of these studies will be provided. We will first refer to the research supporting the hypothesis with young early trilinguals in terms of vocabulary regarding language differentiation and the early existence of TEs but also expand it to the area of morphosyntax. We end by summarising Hoffmann's (1985) contrasting findings, which show that language separation occurs after the onset of speech.

In the field of language differentiation, Murrell (1966) provides evidence that Sandra, a Swedish, English and Finnish trilingual, was able to use all three languages in different contexts with different interlocutors towards the second year of life. Similar evidence is also found in Turell (1997), who describes a Catalan-Spanish-English trilingual child who was able to use the appropriate language with different interlocutors before the age of two.

Montanari (2006) provides more specific data to support the existence of three different linguistic systems from the beginning of the speech. The results of her study, which aims 'to investigate the issue of language differentiation in early multilingual development' (p. 105), reveal that TEs start to be produced before the age of two, which is a sign that the child can differentiate among the three linguistic systems. She states that 'the results of this and previous studies portray a child who

treated her languages in systematically different ways from early on by learning cross-language equivalents, by differentiating her language pragmatically and by combining her words following language specific patterns' (p. 122). Research conducted in morphosyntax has also provided evidence that trilinguals develop all three languages separately. For example, Mikes (1990) describes three children's morphological acquisition of Hungarian, Serbo-Croatian and German. The study suggests that these trilingual children develop separate grammatical systems before the age of two and have three grammatical systems from the beginning. Because of the small number of studies in this field, the results are far from being conclusive. The hypothesis of language separation in early trilinguals remains just a supposition which needs further testing.

There is also some evidence against the Language Separation Hypothesis, since Hoffmann (1985) found evidence that Cristina and Pascual, two Spanish- German-English trilingual children, started to differentiate between the three languages after they started speaking, more precisely, after the age of two. Hoffmann (1985) points out that Cristina used the words indiscriminately when she began speaking, but by 2.6 years old she started to be aware of the different language systems and was clearly associating each person/situation with a language system. Pascual, however, he showed signs of differentiation earlier, (i.e., at the age of two).

The evidence supporting and discarding the Language Separation Hypothesis is mostly based on case studies. In order to generalise these results to all early trilinguals, larger-scale studies involving a more diverse combination of languages are needed.

2.2.2.1.3. LEXICAL COMPREHENSION AND PRODUCTION

There is not a great deal of research performed on trilingual lexical comprehension and production, and what exists is also based on case studies. The aim of these studies has been to analyse vocabulary development during the first years when the languages are being acquired. So far, to our knowledge, no studies have compared large samples of early trilinguals to bilinguals or monolinguals.

The data we rely on for this section is founded in several quantitative studies (Hoffmann ,1985; Dewaele, 2000; Quay, 2001; Barnes, 2006; Montanari 2006; Wang, 2008) following two lines of research. The first targets both the receptive and productive vocabularies, while the second focuses solely on lexical production. After providing a description of these studies, we compare their results with those found with simultaneous bilinguals.

Dewaele (2000) describes the lexical progress in Dutch, English and French in terms of his daughter Livia's reception and production from birth until the age of 48 months, (i.e., four years old). Results indicate that by the age of 14 months, Livia had good passive knowledge of about 150 words in all her languages. She produced her first words in English and the first syllables of words in Dutch at 14 months. Lexical production in French followed one month afterwards. The first multi-word utterances in French and Dutch were produced at the age of 26 months, while more complex utterances were produced between 31 and 35 months. The author recorded the two-word stage in English at the age of 15 months, when Livia told somebody to 'sit down'. Complex utterances in English started to be produced between 31 months and 47 months.

Quay (2001), on the other hand, focuses on the lexical comprehension and production of an English- German- Japanese trilingual named Freddy from the age of 12 to 16 months. In terms of lexical comprehension, the reports from the parents and nursery school teachers suggested that by the age of 16 months Freddy could understand 52 English words (63% of his total vocabulary), 29 German words (35%) and only two Japanese words (2%). The following table summarises the information in Quay (2001) on the child's receptive linguistic development between the age of one and 1.4.

Table 1: Freddy's receptive vocabulary in three languages (Quay, 2001).

Age (in months)	Receptive vocabulary (No. of words)		
	<i>English</i>	<i>German</i>	<i>Japanese</i>
12	15	0	0
13	21	0	1
14	27	0	1
15	40	0	2
16	52	29	2

With respect to lexical production, Quay (2001) records Freddy's vocabulary spurt around the age of 17 months in all the languages. She also reports 88 different vocabulary types from all three languages by the age of 21 months (1.9 years old). Specifically, Freddy produced 50 words in English, 30 in Japanese and eight in German. The production collected from the video recordings shows that Freddy mainly produced Japanese tokens, even if the recordings were done in English and in German. Thus, unlike the receptive level, after five months, 78% of the tokens were in Japanese and only 8% in English, while the rest were in German.

Another two cases are Leandre and Dominique (Wang, 2008), two Chinese-French- English trilinguals. The study describes the children's early lexical development in Chinese and French, which were the parental languages, and English, the language used at school and within the community. In spite of being passively exposed to English since birth, Leandre and Dominique started to receive more input in English only at the age of 48 months (four years old), when they were sent to nursery school. With respect to lexical comprehension in Chinese and French, at the age of four months Leandre showed clear signs of understanding both languages, while Dominique was able to understand only Chinese. Lexical comprehension in French came one month afterwards (i.e., when Dominique was five months old). Specific data on receptive vocabulary in English, however, is not provided in this study. Leandre produced his first words in Chinese at ten months and in French at eleven months, while Dominique produced his words in French and Chinese at twelve months. Leandre went through his vocabulary spurt in Chinese and French between 19 and 21 months, while the vocabulary spurt in English was delayed until Leandre was 35 months old. By 22 months of age (1.10 years old), Leandre could produce 204 Chinese words and 118 French words.

Table 2 summarises the results presented by Wang (2008) of these two children at the productive level. Information on the mean length of utterance of morphemes (MLUm) is provided as an indicator of the mean number of *morphemes* produced in French, Chinese and English.

Table 2: Leandre and Dominique's MLUm in three languages (Wang, 2008).

Age (in months)	Leandre (MLUm)			Dominique (MLUm)		
	<i>French</i>	<i>Chinese</i>	<i>English</i>	<i>French</i>	<i>Chinese</i>	<i>English</i>
12-24	1.5	1.5		1.5	1.5	
24-36	4.5	3.5	3.5	4.5	3.5	3.5
36-48	5.5	5	4.25	5.25	5	5.5

Hoffmann (1985) describes the phonological, grammatical and lexical systems in the oral production by Cristina and Pascual, two Spanish- German- English trilingual children. They were simultaneously exposed to Spanish and German from birth and to English later on, shortly before the age of three, which makes them early trilinguals according to the definition provided in [section 2.2.2.1](#). Thus, at 15 months old, Cristina produced long sentences in Spanish and German. By the age of 36 months, she had acquired an important part of the phonological system as well as most of the morphosyntactic system of Spanish and German. Nevertheless, the two languages did not develop at the same rate. According to Hoffmann (1985), Cristina's German grammar emerged more slowly than her Spanish grammar. With respect to English, Cristina started to utter whole sentences very early (i.e., at the age of 22 months), but her English continued to be telegraphic due to her more limited exposure to it, until she was enrolled school at the age of 57 months, where the input she received was in English. Pascual's language development was different. Although he uttered his first words at the same age as Cristina, his pronunciation sounded less 'adult-like'. He underwent the two-word stage at the age of 24 months. Just like Cristina, his Spanish and German did not develop at the same rate. At the age of 48 months, Pascual had a full control of the German and Spanish phonological systems. Nevertheless, he continued to be grammatically inaccurate in both languages. With respect to English, by the age of 29 months he produced whole utterances, and by 42 months, when input in English increased through formal instruction, English became his strongest language.

Montanari (2006) describes the productive lexical development of Kathryn (a Tagalog- Spanish- English trilingual) between the ages of 16 and 24 months. Kathryn produced her first words in the three languages at around 16 months of age, while her vocabulary spurt was recorded at around 20 months in all the languages. By 23 months, Kathryn had a total vocabulary of 251 words. Montanari (2010) also provides data on Kathryn's lexical development with respect to MLU. Unlike Wang (2008), she performs a different calculation (taking into account words, MLUw, not

morphemes) and therefore reports different findings (with lower results that are not comparable to the previous study). These results provide evidence that Kathryn's one-word stage started at 17 months and ended at about 28 months of age, as shown in the table below.

Table 3: Kathryn's MLUw in three languages (Montanari, 2010).

Age (in months)	Language (MLUw)		
	<i>Tagalog</i>	<i>Spanish</i>	<i>English</i>
17	1	1	1
21	1.08	1.05	1.06
28	1.31	1.27	1.60

Barnes (2006) reports lexical production data focused on English questions from Jenny, a Basque-Spanish-English trilingual. The data collection period was between the ages of 23 months (1.11 years old) and 41 months (3.5 years old). The MLU in words (MLUw) was calculated for English and MLU in morphemes (MLUm) was calculated for Basque and Spanish. The following table shows the results for English, on the one hand, and for Basque and Spanish on the other. Note that the lexical development in English was strictly monitored (with data for all the age milestones), while for Basque and Spanish the data provided is scarcer. However, we do have data for both Spanish and Basque at the age of 38 months, which helps us to compare the lexical development in these two languages at this particular age. At the age of 41 months, we have MLUs for all three languages.

Table 4: Jenny's MLUs in three languages (Barnes, 2006).

Age (in months)	MLUm		MLUw
	Basque	Spanish	English
23			1.52
24			1.90
28			2.83
32	2.24		
34			3.56
36			2.88
37			4.42
38	3.88	2.9	
39			3.54
40			3.13
41	4.48	2.64	2.98

The MLU in English fluctuates and does not show steady progress even though Barnes (2006) reports that their exposure to English was regular and steady. When comparing Jenny's MLU in Basque and Spanish with that of similar developing bilinguals, Barnes (2006) concluded that these results show that Jenny 'is developing in a similar way to other bilingual children' (p. 113).

Overall, the results of the aforementioned studies point to two aspects of lexical comprehension. The first is that in early trilingual children, all three languages do not develop at the same rate. Freddy's receptive lexical development is a clear case in point. His quick English development may be due to the amount of exposure, which totalled 46% between his eleventh and twenty-first months. The second conclusion is that there is a difference in the total receptive vocabulary size from individual to individual at a given age. For example, at the age of 1.2, Livia was reported to understand a total of 150 words in all three languages, whereas Freddy's receptive vocabulary was only 28 words. The same is true regarding oral production. Thus, we found that in trilingual children the three languages develop at distinct rates (see Freddy in Quay, 2001 and Cristina and Pascual in Hoffmann, 1985). At the same time, there is variation among different children's productive lexical development (see Cristina and Pascual in Hoffmann, 1985).

It is also worth noting that when comparing the results of these studies against those from the research conducted with simultaneous bilinguals, we find that early trilingual children follow the same milestones of lexical development as their bilingual peers. According to Wang (2008), the babbling stage of trilingual children is situated at around four months old, which is similar to bilinguals' milestones (Bosch & Sebastian, 2001; De Houwer, 2009). However, in trilinguals, there is no clear evidence showing these children's babbling is language-specific. With respect to TE, Montanari (2006) shows that trilinguals produce TEs early, just when they start speaking, the same as bilingual children. In terms of vocabulary size, just as with bilingual children, there is also variation among trilingual toddlers in how many different words they produce in each language at a given time. The differences in the case of Leandre, Dominique and Jenny are a case in point. When comparing the size of a trilingual's productive vocabulary with that of their monolingual and bilingual peers, Barnes (2006) shows that 'Jenny's development

in English falls within a range similar to that of monolingual and bilingual children' (p. 112).

Research on early trilingual development seems to point to similar conclusions as with simultaneous bilinguals. regarding the babbling stage, knowledge of TEs and total vocabulary size (all the languages taken into consideration), trilingual and bilingual children follow the same patterns. Nevertheless, studies conducted with trilinguals in their early stages are insufficient, and because most of the data presented in this section relies on case studies, we believe that larger-scale research is needed in order to have more solid data regarding early lexical development in three languages.

2.2.2.2. LEARNING AN L3 AFTER THE SIMULTANEOUS ACQUISITION OF L1 AND L2

This section focuses on research on bilingual children learning an L3 as an FL, after the simultaneous acquisition of L1 and L2. It first introduces two types of educational environments which foster L3 acquisition: formal and informal settings, and discusses the importance of AO and exposure in L3 lexical development. It also summarises the possible effects that the early introduction of L3 may have on the first languages, L1 and L2.

2.2.2.2.1. L3 ACQUISITION IN FORMAL SETTINGS: REGULAR CURRICULA AND IMMERSION SCHOOLS

Around the world, a growing number of young bilingual children are being exposed to an L3 as an FL in formal settings. Most children start learning it in primary school or even kindergarten. The instruction can take place either in a regular school context where the L3 is taught for one, two or three hours per week (limited exposure) or in an immersion school with a massive amount of exposure from the beginning.

This section describes the regular and immersion school settings which can foster the early acquisition of an FL. In line with the main study of this thesis, we begin by presenting several Spanish regular school models in which simultaneous bilinguals are exposed to English from early ages, with a focus on Catalonia and the Basque Country. We continue by describing different immersion school models,

first focusing on Canada and the US, where this schooling model first emerged, and then on Catalonia, where the study of this thesis was conducted.

Lately, English has been introduced in the regular schools' curricula at early ages in Catalonia and the Basque Country, two bilingual communities in Spain, due to the need to improved children's English language proficiency. In the past, English had been introduced as a compulsory subject in the third year of primary school (at the age of eight) with 2.5 hours/week of instruction (see Cenoz, 2005; Munoz, 2006). Nevertheless, nowadays, children are usually exposed to English before the age of eight and start to learn it at the age of six, when they start primary school. Others start formal instruction in English even earlier, in kindergarten. In the Basque Country in the early 1990s, for example, the 'Early Multilingualism' project introduced the study of English starting at age four for two hours a week in 30-minute sessions (Cenoz, 2005; Lasagabaster, 2005). The situation is similar in most of the state-funded kindergartens in Catalonia, where Spanish-Catalan bilingual children are exposed to English for two hours/week.

Exposure to an FL can also take place in immersion settings through massive input. To summarise some of the main points already mentioned in section 2.1.3.1., it is important to stress that one-way immersion programmes are meant for majority language learners, while two-way immersion programmes target both majority and minority language learners. Immersion can start during kindergarten (early immersion), at nine or ten years old (delayed or middle immersion) or in secondary school (late immersion). With respect to the amount of exposure, Baker (1998) refers to total immersion during the first two or three years, which accounts for 100% of the time and is lowered to 80% during the next three or four years, ending up at 50% in high school. Partial immersion involves exposure to the FL about 50% of the time throughout all schooling.

An important number of language immersion centres can be found in Catalonia, more precisely in the Barcelona metropolitan area. According to <https://www.international-schools-database.com/>, there are 36 language immersion schools in this area. The vast majority of them target students between the ages of three and 18, and the languages of instruction are either English, French or German.

However, majority of the immersion schools in this area use English as the language of instruction. In general, these schools follow the structured immersion model (Baker & de Kanter, 1981) presented in section 2.1.3.1. Thus, despite promoting English, children are allowed to use their Spanish or Catalan to ask questions, and in return, the teacher replies in English. The structural immersion programme starts in kindergarten at the age of three. Between the ages of three and six, daily exposure to English is 75% of the total hours of instruction (six per day). Subjects such as language, maths, art, physical education, music, personal development and environmental studies are taught in English by the native teacher. The classes taught in Spanish and Catalan account for 15% of the total hours of instruction. Between the ages of seven and 16 (primary and secondary school), the exposure to English goes down to 50%. After the age of 16, high school students follow the regular school curricula, in which exposure to English is approximately two hours per week.

The following sections will focus on the research conducted in regular and immersion schools with respect to two topics: first, the importance of AO and exposure for the linguistic proficiency in English as the L3, and then the possible effects that the early introduction English may have on the first developing languages.

2.2.2.2.2. L3 ACQUISITION IN FORMAL SETTINGS: THE ROLE OF AGE OF ONSET (AO) AND EXPOSURE

Research conducted on early lexical comprehension and production with young simultaneous bilinguals (see sections 2.1.2.3. and 2.1.2.4.) shows that the receptive/productive vocabulary sizes in both languages are determined by the amount of exposure in L1 and L2 (Pearson et al., 1997; De Houwer, 2007; David & Li, 2008).

Similar research has also been conducted with simultaneous bilinguals acquiring an L3 in formal settings. Apart from the role of exposure, these studies also investigated the effects of AO in L3 acquisition. Two lines of research were pursued. The first focuses on participants with a different AO who received a similar amount of exposure, while the second targets learners with a similar AO but different amounts of exposure. Because the number of studies targeting vocabulary

acquisition is quite limited, we will present research on different linguistic skills (not just vocabulary knowledge).

With respect to the first line of research, the results in regular curriculum schools in the Basque Country and Catalonia (where the exposure is an average of 2.5 hours per week) show that in almost all situations, the late starters (LS) outperformed the early starters (ES). To our knowledge, no such studies have been conducted in immersion schools with different AOs. In what follows, we will provide a detailed presentation of the studies conducted in regular schools in Spain.

One of the studies in regular curriculum schools in the Basque Country was conducted by Lasagabaster and Doiz (2003). They compared the performance of three groups of Spanish-Basque bilingual children after 700 hours of instruction in English as an FL in different measures of written production. The first group consisted of learners who started English at the age of four and were between 11 and twelve years old at testing time. The second consisted of secondary school students who started learning English at the age of age, while the last group was twelfth-grade students who started English at the age of eleven. The scores earned evidenced the superiority of the first group in all the measures.

The studies conducted in Catalonia led to the same conclusions. Muñoz (2000) conducted a study in which one of the objectives was to explore the influence of age on FL acquisition. To this end, she selected two groups of participants; the first one, ES, consisted of children who started English at the age of eight, while the second one, LS, was made up of learners who started to receive formal instruction in English at the age of eleven. Both groups were tested after 200 hours of exposure on a close test, dictation, a grammar test and a listening comprehension task. The results were significantly in favour of the LS, except for the listening task, where no significant differences were found. In the same vein, Muñoz (2006) performed another study in which two of the five groups tested started learning English at the age of eight (ES) and eleven (LS). These two groups were tested twice, after 200 and 416 hours of instruction, and both times they were administered a dictation, a close test and the listening comprehension task. At both data collection times, the results were in favour of the LS on both the close test and the dictation. The

outcomes of the listening comprehension task became significant in favour of LS only after 416 hours of instruction. Miralpeix (2008) performed a similar study analysing whether an early start has positive effects on productive vocabulary in the long run. To this end, the same groups of participants were chosen and compared after 726 hours of instruction in oral and written production. Oral production was tested through an interview, a picture storytelling task and a role-play. Written production was assessed by means of a composition and a cloze test. After 726 hours of instruction, LS earned significantly higher results than ES in the majority of tasks, although no significant differences were found in the role-play. All these studies clearly indicate that if the amount of exposure is kept constant, LS outperform ES.

Furthermore, when ES receive more exposure, they have been found not to outperform LS. Another case in point is the study conducted by Cenoz (2003b), which investigated whether increased exposure for ES led them to overtake LS. To this end, two groups were compared. The first consisted of children with an AO of eight (ES) and who had been exposed to 800 hours of instruction at testing time. The second group consisted of learners who began to study English at the age of eleven (LS) and who had had 500 hours of instruction. Each group was tested at the age of 15 on a picture storytelling task, an oral comprehension task, a cloze test, a written composition and the Oxford Placement Test (Allan, 2006). The results indicate that in spite of the higher exposure, ES did not manage to outperform LS. Nor were significant differences found between the two groups in the majority of the tasks. However, one exception is the study on writing and vocabulary use conducted by Doiz and Lasagabaster (2004), where exposure was found to be important when similar age groups are compared. Participants were a group of ninth-grade students who started English at the age of eight (ES) and at the age of eleven (LS). ES, who had had more exposure at testing time, outperformed LS in several lexical measures and language use, as well as in average writing scores.

An example of a study highlighting the role of exposure was conducted in the Basque Country by García Mayo (2003), who found that exposure counts in grammar learning (more than age). Participants in the study were divided into four groups. Groups A and C had the same AO, eight years old, but different exposure

(group A had received 396 hours of instruction while group C had received 594). Groups B and D had the same AO, 11 years old, but group B had had 396 hours of exposure, while group D had had 594 hours of instruction. All the participants took a grammatical-judgement task, and the results showed that length of exposure was indicative of the better scores earned by groups C and D, who had received a total of 594 hours of instruction at testing time.

Therefore, it seems that if exposure is kept constant, an early start does not lead to better linguistic proficiency in L3 (sometimes not even with additional exposure time). On the other hand, we see that exposure is an important aspect in achieving better results in an L3 given similar AOs. The studies conducted in Spain with bilingual children learning an L3 in formal settings with a similar AO (but different amount of exposure) have mostly been conducted in these two bilingual communities. No such studies controlling for these variables have been conducted with participants in immersion schools. At this point, we also wonder whether an early start and increased exposure to L3 may affect performance in L1 and L2. In this sense, the following section will focus on the possible effects of the early introduction of an FL on the first languages.

2.2.2.2.3. POSSIBLE EFFECTS OF THE EARLY INTRODUCTION OF AN FL ON L1 AND L2

This section aims to explore what research has found so far regarding the possible effects on the first developing languages when an L3 is introduced as an FL. Because the number of studies in this area focusing on vocabulary is limited, we will refer to the research on different linguistic skills in order to provide some background for our own study.

Research on this topic has yielded mixed results. We will first summarise several studies examining the effects of early L3 introduction in regular schools and then move on to talk about immersion settings. To our knowledge, no such studies have been performed with bilingual children learning an L3 in immersion schools, although there are some studies with older participants, such as Gallardo del Puerto (2007) on phonetics and Kartushina et al. (2019) on phonetics.

With respect to the research carried out with bilingual young learners studying an L3 in regular schools in Spain, the results of the Early Multilingual Project in the Basque Country shows that the introduction of an FL in kindergarten with limited exposure (i.e., two or three hours per week) does not affect linguistic performance in the first language(s) (Cenoz, 2005; Lasagabaster, 2005). This project involved the introduction of English as an L3 for two hours a week in four 30-minute sessions. The experiment was carried out with 195 Spanish/Basque bilingual participants who began learning English at the age of four and another control group of 154 students who began learning English at the age of eight. At testing time, the students were in their second year of compulsory secondary education. The first group had been learning English for ten years and the control group for six. The results of the tests showed no difference between the two groups with respect to their competence in the two official languages. Another study performed by Cenoz (1997) led to similar results when 166 Spanish-Basque bilingual children in Ikastola schools (where instructions is mostly in Basque) were tested on Spanish and Basque linguistic competence when they were 14 years old, that is, three years after being exposed to English as an FL for three hours a week. The results in Spanish and Basque were compared to those of a control group of children the same age who were not exposed to any English. The findings point to no differences in linguistic competence in Spanish and Basque.

Regarding similar studies conducted in immersion settings, so far we have been unable to find any references with simultaneous bilingual young children. However, there is research with young monolingual learners learning an FL through immersion. One of the few studies conducted with kindergarten children learning an FL in an immersion school setting was conducted in Germany by Bergstörn, et al. (2016). Participants were two age-matched groups of children between two and six years old. The first group acquired English as an FL in a partial immersion school, whereas the second one learned English through conventional instruction with less exposure. The results of the study showed that immersion had no detrimental effects on L1.

On the other hand, research performed with school-aged participants led to slightly different results. Studies conducted in French immersion schools in Canada show

that during the first four years of early total immersion in French (between the age of three and seven), children do not progress in English in the same way as their monolingual peers, who learn reading, spelling and punctuation in regular schools. However, after six years (at the age of 13), the early total immersion students catch up with the English monolinguals (Genesee, 1983). The trend is more or less the same in partial early immersion students: they earn similar scores to those of English monolingual children by the end of the elementary school. Similar results have been found in Italian in immersion centres. Montanari (2014) conducted a study comparing the reading skills of 60 American children between the ages of six and eight learning Italian in an immersion school with those of children who were educated only in English in a regular school. The results indicate that children exposed to Italian lagged behind in L1 reading. Nevertheless, by the end of the elementary school, the immersion group was able to perform equally well or even better than the children in the regular school.

Research conducted with young monolingual children learning an L2 (either in natural or formal settings) yields different results. Gibson et al. (2012) performed a study on a group of 123 Spanish-English sequential bilinguals between the ages of five and seven who were exposed to the new community language, English, through schooling. The degrees of language exposure to English were different, and the results showed that Spanish-English bilinguals with more exposure to English earned lower scores on the Spanish lexical production test than those with less exposure to English. No differences were found in the lexical reception scores, and no longitudinal follow-up is available.

In sum, research conducted in regular school settings shows that early L3 introduction with limited exposure does not affect the linguistic development in the first two languages. Nevertheless, studies conducted with young monolingual learners acquiring another language in immersion settings show mixed evidence.

2.2.3. SOME INSIGHT INTO THE MULTILINGUAL LEXICON

The study of the multilingual lexicon is rather new, although there has been an ongoing debate: one of the most researched areas in the multilingual lexicon is the integration of versus separation between languages (Gabrys-Barker, 2019).

The ‘integrated’ view of the lexicon maintains that languages are stored together, and the lexicon contains information from all languages that speakers combine appropriately. In this view, the speaker is able to discriminate between the languages by means of language tags or nodes. The majority of studies in this line have been conducted with bilinguals or monolinguals learning an FL. One paradigmatic study was conducted by Caramazza and Bories (1979) with Spanish-English sequential bilingual adults who were assessed on a word recognition task, in which they were asked to determine whether a string of letters formed a word in one of the languages they knew. The results showed that the reaction time to a word in one language depended on the frequency of the cognate in the other language, which was taken as an indication that both languages were stored together. Other studies that advocated for the integrative approach have focused on lexical comprehension (e.g., de Angelis, 2007) or on cross linguistic influence, language associations or TEs (see Gabrys-Barker, 2019)

On the other hand, researchers such as Hervig (2001) have found evidence of separate lexicons. She examined the production of trilingual speakers who were asked to describe a picture story in their L1 and then translate it in their L2 and L3. A think-aloud protocol was used to identify word association paths when translating. The results showed that ‘lexical selection in situations on non-accessibility of an item in demand involves both automatic and deliberate consultation of several languages’ (p. 128). This was taken as proof that languages were not stored together but rather separately, as the participants had to ‘move’ from one language to the other in order find the right corresponding word to complete the translation task. Other studies supporting this view are related to the modularity hypothesis (which assumes that the information is stored in different modules in the mind, and different languages may be stored in different ones). Other variables such as language typology, language attrition or the ability to select and use the appropriate

language according to the speakers and situation involved seem to suggest that languages are not fully integrated (see Gabrys-Barker, 2019)

Singleton (2003) and De Angelis (2007) advocate for a ‘compromise solution’, which is prevalent nowadays. It posits that languages are neither completely disconnected nor totally integrated. The degree of integration/separation has to do with how words have been acquired, how well they are known or the perceived similarity between them. Several studies may support this view; for instance, Abunuwara (1992) measured the interference effects among L1 speakers of Arabic with Hebrew and English as the two FLs by means of a colour-naming test, a trilingual picture-naming task and a translation test. The results pointed to an independent relation between the two non-native languages, an interdependent relation (word association) between the native language and the weak FL and an intermediate relation (word concept mediation) between the native language and the strong FL. Palma et al. (2022) conducted a study of English- French bilinguals between the ages of 18 and 35, who took part in a novel word-learning task. The findings suggested that similarity to pre-existing lexical representations has an impact on new word lexicalisation. In other words, cognates may be stored together while language-specific words may be stored separately.

Dijkstra’s Multilingual Interactive Activation (MIA) also assumes both the existence of an integrated lexicon and parallel access and activation across languages. MIA occurs at four levels of representation which are hierarchically organised, namely language node (which contains language tags), word, letter and feature. This model proposes that the lexicon is integrated at the word, letter and feature levels, but is separated at the language level. The visual word recognition process is an integrated bottom-up process which first takes place at the feature level, the letter level, the word level and the language node. However, the word production process is top-down, starting at the language node (there are different nodes for different languages). The node of the language that is used would inhibit the words in the other languages via a top-down mechanism.

In the case of trilinguals, Singleton (2003) agrees that the L3 lexicon builds variable connections of different strength with the L1 and L2 lexicons and that the

relationship between the languages is dictated by language proficiency. This view is also supported by De Groot and Hoecks (1995) who conducted a study with Dutch native speakers speaking English (as the strong language) and French (as the weak language) to prove this point. Participants were given two types of translation tasks. In the first one, they needed to translate 40 Dutch words into two FLs, while in the second one, they were given 80 pairs consisting of the Dutch words and their equivalent translation in English and French and then asked to judge if the translations given were correct. The results clearly indicated an association between non-native language proficiency and lexical-semantic organisation. More specifically, the study found that the lexical relation between the native language and a weak FL reflects the word association type (the words of an L1 and weak FL are associated by a direct connection with each other), while the lexical structure between a native language and a strong FL reflects the concept mediation type (the words of an L2/L3 are associated with the words of the L1 by way of a common conceptual store).

Pavicic (2008) also argues that the relationship between languages in the multilingual lexicon is determined by the speakers themselves, which means that “each individual may use the organisational resources in his/her lexicon in a different way, depending on different factors such as the way the word has been acquired or the perception of similarity between the words” (see also Ortega, 2017). In this sense Hufeisen (2005) also states that a learner’s competence can influence their access to a particular lexical item. For instance, beginners would rely on their L1 to access a new word in a different language, while more advanced learners would use conceptual representations when acquiring new words, thus bypassing their L1. This phenomenon may account for code-switching, cross-linguistic influence and even language loss. Other factors that may influence the organisation of learners’ lexicon may be the learning environment, language dominance, the type of linguistic tasks to be performed and the situations in which languages are used (see Gabrys-Barker, 2019).

2.3. SUMMARY

A considerable number of children grow up in bilingual and multilingual environments. Simultaneous bilinguals are exposed to L1 and L2 since birth or to L2 no later than the age of three. Sequential bilinguals are exposed to an L2 later, either in naturalistic or formal settings (usually through immersion). Multilinguals acquire more than two languages (three or more) in either formal or naturalistic settings.

Simultaneous bilingualism is well-documented in the literature. For example, it has been shown that simultaneous bilinguals tend to follow the same lexical development milestones as their monolingual peers. However, they should not be viewed as ‘two monolinguals in one’, given that acquiring two languages also entails other mechanisms and processes. Indeed, at the lexical level, simultaneous bilingual children are able to separate and differentiate the two languages since the beginning of speech, but language mixing still occurs irrespective of how input is received (1P/1L or 1P/2L).

The vocabulary size of young simultaneous bilinguals has been found to be related to parents’ education and income, on the one hand, and to input quantity and quality on the other. Furthermore, children’s receptive and productive vocabularies in both languages tend to be comparable to those of monolinguals (whenever tested on conceptual score measures).

Less research has been conducted with sequential bilingual children focused on the relationship between the two languages, and the results are mixed. On the one hand, some researchers have found interdependence between L1 and L2 (e.g., L1 proficiency is a predictor of L2 proficiency), while others have not.

In line with the main objectives of this thesis, in this chapter we have also focused on different patterns of L2 acquisition through immersion, distinguishing between language minority and majority learners and the school programmes targeting each of the two groups. In the case of language majority immersion programmes, some studies concluded that full immersion programmes may affect grammatical

accuracy and lexical diversity. Therefore, an emphasis on focus-on-form and cross-linguistic issues has been suggested to counterbalance possible adverse effects.

In terms of the lexical comprehension and production of sequential bilinguals, studies have shown that they tend to score lower than monolinguals and are comparable to simultaneous bilinguals, but only in the L1. Regarding the AO, the studies presented in this section indicate that an early AO as well as the quantity of input received are determining factors in L2 proficiency.

Because TLA entails processes different from bilingual acquisition, a section of this chapter was devoted to trilingualism. Cenoz (2013) described four ways in which L1, L2 and L3 can be acquired. The first is when L2 and L3 are acquired after tL1, mostly (but not necessarily) at older ages. Research conducted in this area involves adults and is focused on CLI and the age factor. In terms of CLI, the results of the studies show that typological language difference has an effect on L3 acquisition. With respect to the age factor, the results of these studies showed that an early AO alone does not guarantee successful L3 proficiency.

Another case (when L1, L2 and L2 are acquired at the same time) has been also discussed. The outcome of the research with early trilingual children shows that they seem to be able to separate the three languages since the onset of speech. At the same time, they follow the same lexical acquisition milestones as bilinguals and have comparable vocabulary sizes. Nevertheless, the findings are quite tentative due to the small number of studies conducted in the field of early trilingual development.

A more significant body of studies has been conducted on young bilinguals learning an L3 in instructional settings (i.e., the fourth scenario described by Cenoz, 2013). We focused on two different instructional settings: regular schools and immersion schools. First, we reviewed the research on the role of the age factor and exposure in TLA. The results of the studies indicated that if exposure is kept constant, late starters outperform early starters in the majority of studies. On the other hand, if the AO is similar, the amount of exposure is crucial to L3 performance. These studies were mostly conducted in primary and secondary school settings. To our knowledge,

no such studies have been conducted with immersion school participants in kindergartens thus far. With respect to the influence that early exposure to L3 may have on L1 and L2, the evidence of the research summarised shows that in regular schools, the early introduction of an L3 does not affect lexical development in the first two languages. Given the absence of studies carried out with bilinguals learning an L3 in immersion schools, we reviewed the research with monolinguals learning an L2 in immersion schools. The findings suggest that the early introduction of an L2 with massive exposure seems to affect the linguistic development of the L1 in the short term.

The final part the chapter focused on the multilingual lexicon. Only a limited amount of research has been conducted with multilingual speakers, while more studies have been conducted with bilinguals. In terms of the organisation and relationship between languages, these studies have supported three views: language separation, language integration and the mixed approach (where languages are not fully integrated or separated). Nowadays, researchers seem to agree that the relationship between languages in the mental lexicon can be determined by many factors, including proficiency in each language, learning environment, language dominance, type of linguistic tasks to be performed and situations in which the languages are used.

CHAPTER 3

LEXICAL ASSESSMENT OF VERY YOUNG CHILDREN IN TWO AND THREE LANGUAGES

Assessing the lexical development of very young bilingual and TLA learners between the ages of birth and six is a complex task for several reasons. First of all, these children have very poor (or no) reading and writing skills. Providing them with a written test would be quite difficult. Second, even though they may have acquired many words, such young learners are not always ready to demonstrate their knowledge to teachers when they are asked to. Unlike older learners, who have gained more self-control in managing their emotions, this is harder for young learners. That is why in order to get information on infants and toddlers' lexical acquisition, teachers need to find the most appropriate setting, when the children are more relaxed and are eager to show what they know. Because not all children show this readiness at the same time, and to make the child feel more confident, the assessment needs to be carried out individually. Moreover, in the case of bilingual and TLA children, it needs to be performed in all the languages they speak, which entails a series of challenges with respect to test design (to ensure that it covers all the languages) and score interpretation. Finally, these testing tools should also ideally provide a yardstick against which the lexical performance of all these children could be measured, regardless of the number of languages known.

Currently, researchers find that they need to find such lexical assessment instruments. Parental diaries, oral production tasks and language questionnaires maybe options, as these instruments are able to assess young learners in a variety of situations and in different languages. Nevertheless, very often these tests fail to provide a standard against which these learners can be measured. Standardised vocabulary tests are another possible solution. Unfortunately, most of them (with a few exceptions) are normed only for monolingual young children, and using them with bilinguals and TLA learners faces several constraints.

Studies such as Pearson et al. (1997), Cenoz (2003), Montanari (2010), Gibson et al. (2012) and Poulin et al. (2013) used monolingually normed standardised tests

mostly with simultaneous bilingual participants and in combination with other tools, such as parental diaries, language questionnaires and oral recordings, aiming to provide a more thorough assessment in two languages. For example, standard tests assessing both lexical production and comprehension can usually be combined with language habit questionnaires. For example, standard lexical comprehension tests were combined with questionnaires on language habits (in Dixon, 2011; Gibson et al., 2012) and oral speech recordings (in Pearson et al., 1997), while lexical production tests were mostly combined with questionnaires (in Pearson et al., 1993; Gibson et al., 2012) and sometimes even with other standard language comprehension tests (Gibson et al., 2012; Poulin et al., 2013).

This chapter brings into discussion the early lexical assessment of bilingual and multilingual young children and the challenges this entails. To this end, section 3.1. presents some of the early lexical assessment tools which have been used in research with young bilingual and early TLA participants. Because some of the instruments are used exclusively to assess receptive lexical knowledge, others to evaluate productive knowledge and others to test both, this section was further divided into three subsections according to what is being tested, namely reception, production and reception/production.

Section 3.2. presents the limitations and challenges in using monolingually normed standard tests and oral speech recordings when trying to assess the lexical development of very young bilingual and TLA learners. It also discusses several studies conducted with young children in two and three languages which may offer possible alternatives to these limitations.

Section 3.3. makes a selection of the instruments presented in sections 3.1 and 3.2 that may be used to assess children between the ages of birth and six in Spanish, Catalan and English in two situations: when languages were acquired at the same time (early trilingualism) or when two of the languages (Spanish and Catalan) were simultaneously acquired and English is acquired as an L3 at an early age in formal settings, which is a common situation in Catalonia.

3.1. ASSESSING LANGUAGE DEVELOPMENT IN VERY YOUNG LEARNERS WITH A SPECIAL FOCUS ON VOCABULARY

This section presents some of the tools used the most to assess the lexical comprehension and production of very young bilingual and TLA learners. Section 3.1.1. describes assessment tools which test both vocabulary comprehension and production; section 3.1.2. addresses tools testing only vocabulary comprehension; and section 3.1.3. presents tools used to evaluate productive vocabulary.

3.1.1. RECEPTIVE AND PRODUCTIVE LEXICAL ASSESSMENT

As mentioned in the introductory section, assessing early receptive and productive lexical development in bilingual and TLA contexts is a challenging task, as researchers are usually confronted with the lack of standard lexical assessment tools normed for these participants in different languages. To date, standard tests have been used together with language habits questionnaires or parental diaries, which contribute to getting a more complete picture of early lexical development in two and three languages. This section will focus on standard tools such as parental diaries and the MacArthur-Bates Communicative Development Inventories (Fenson et al., 1993), as well as other less well-known tests.

3.1.1.1. PARENTAL DIARIES

This section introduces parental diaries as lexical assessment tools first by referring to their characteristics and then by summarising some of the studies in which these diaries have been used to assess early lexical development in two and three languages. We also outline the criticisms of this assessment tool by some researchers.

Parental diaries are language records which offer a detailed description of a child's lexical development. As the name indicates, they are kept by either parents or caretakers, given that they are the ones who spend most time with the child and thus are more easily able to track changes in their lexical development. Some of the earliest bilingual and TLA studies of language acquisition were based on parental diaries.

One of the most famous studies using parental diaries is by Leopold (1949), a four-volume account which presents in detail Hildegard's longitudinal lexical development in German and English from the age of eight weeks until her second birthday. More recent studies include Wang (2008), who reported on the Chinese and French lexical acquisition of her trilingual Chinese- French- English children (Dominique and Leandre) from birth until the age of eleven. She described in detail all the stages of her sons' lexical development, including references about babbling and cooing, pronunciation, vocabulary spurts, vocabulary sizes at certain ages, situations of codemixing and codeswitching, as well as issues related to grammar (adjectives, pronouns, irregular verbs, etc.) and early literacy skills. Dewaele (2000) also kept a record of the lexical development of his trilingual English- French- Dutch daughter, Livia, from birth until the age of four. During all this time, Livia was regularly videoed and audio-recorded in the three languages. At the same time, a language diary was also kept when Livia produced 'interesting utterances [...] once the camera has been turned off' (Dewaele, 2000; p. 1). He recorded cases of word production and comprehension with a special focus on vocabulary sizes in the three languages separately. Instances of early lexical mixing were also presented.

Parental diaries can be an important tool for several reasons. First of all, the parents or caretakers spend most of the time with the child and therefore can be more precise in listing all the lexical items he/she produces or understands. Second, all the parents in the aforementioned studies are linguists, and therefore their approach was professional. Finally, these diaries are able to provide accurate longitudinal information relating to children's lexical development. Nevertheless, parental diaries are not free of criticism. First of all, according to Reznick and Schwartz (2001), parents are not always impartial regarding their children's knowledge, and what they report could lead to a distorted picture of the child's lexical development. Secondly, parental diaries provide data from a very small number of subjects (one or two), which raises the doubt regarding whether the participants being studied are indeed representative. Another shortcoming is that sometimes the utterances analysed are used only in the presence of the parent who is collecting the data, and the language sample analysed is very limited. Consequently, the validity of the results from such diaries can be easily questioned. One possible solution to these constraints was provided by Fenson et al. in 1993, who developed a standardised

parental testing tool (an inventory) that can be used with a large number of participants, with a lot of objectives for caretakers or teachers to focus on. These inventories will be described in detail in the following section.

3.1.1.2. THE MACARTHUR-BATES COMMUNICATIVE DEVELOPMENT INVENTORIES

This section provides a detailed description of the MacArthur Bates Communicative Inventories (MBCDI) (Fenson et al., 1993) and then summarises some of the studies on the validity of this tool for monolingual and simultaneous bilingual participants (no such studies have been carried out with early trilinguals). The section ends by commenting on several studies in which the MBCDI was used to assess the lexical development of very young simultaneous bilinguals and early trilinguals across time.

The MBCDI (the full test is presented in Appendix 2) is a standardised parent assessment tool that provides information on the early language skills of children between eight and 36 months old. It relies mostly on recognition rather than recall when assessing infants' and toddlers' vocabulary sizes. It also measures production. According to Pearson et al. (1993), this kind of inventory focuses on “emerging behaviours at times when these behaviours are current and not retrospective” (p. 99). The original version was first created in American English and then adapted for 60 other languages, including Canadian French, Dutch, French, Italian, Spanish, Catalan and more (<http://mb-cdi.stanford.edu/adaptations.html>). The test has only been normed for monolingual populations, but it has extensively been used with bilingual and multilingual populations. In this sense, Weisleder et al. (2023) provide a very thorough review of parental vocabulary assessment tools and according to the data they managed to retrieve from PubMed and PsychInfo, the MBCDI are at the top of the most frequently used tools for early lexical assessment. They also note that very few studies specifically validated these instruments on bilingual/multilingual contexts, so research is needed.

The American MBCDI has three parts. The first one, MBCDI-Words and Gestures (see Appendix 2 and Table 5), is used for children between eight and 18 months old, but it can also be used for children older than 18 months with language delays.

The test measures the comprehension and production of early words, as well as symbolic and communicative gestures. It consists of two main parts. The first is called ‘Early Words’ and the second ‘Actions and Gestures’.

‘Early Words’ evaluates comprehension and production of infants’ early words in four different sections, namely ‘first signs of understanding’, ‘phrases’, ‘starting to talk’ and ‘vocabulary checklist’. The ‘first signs of understanding’ section tests the infant’s response to different commands. The evaluator marks the ‘yes’ or ‘no’ option to show whether the child is able to respond to these commands. The ‘phrases’ section is a checklist of 28 phrases which are marked with a ‘yes’ when the infant shows signs of understanding them. The ‘starting to talk’ section contains two evaluation items. The first refers to frequency of word imitation and concept labelling; the evaluator chooses among ‘never’, ‘sometimes’ or ‘often’ according to infant’s performance. Finally, ‘the vocabulary checklist’ is a list of 396 typical words in child’s vocabulary. This checklist is divided into different lexical categories (see Table 5). The evaluator is asked to mark whether the child is able to just understand the word or both understand and produce it. Separate scores for comprehension and production are calculated out of a total of 396 words.

The second part, ‘Actions and Gestures’, evaluates the infant’s receptive and productive language (through gestures) in five sections. The first section, ‘first communicative gestures’, evaluates the frequency with which infants use certain gestures to communicate. The tester needs to choose among the ‘not yet’, ‘sometimes’ or ‘often’ options according to the child’s performance. In the following four sections, the evaluator needs to choose between the ‘yes’ and ‘no’ options according to whether the child is able to perform these actions. The section on ‘games and routines’ analyses whether the child is able to play different games. The section on ‘actions with objects’ assesses the child performing different actions with the use of objects (e.g., ‘eat with a spoon or fork’ or ‘brush teeth’). The ‘pretending to be a parent’ section analyses the child when he/she takes on the role of a parent while playing, while the fifth one refers to ‘imitating other adult actions. Table 5 summarises the content of the MBCDI Words and Gestures.

Table 5. Summary of the MBCDI Words and Gestures.

Sections		Maximum score	
		Reception	Production
Part I Early Words	First signs of understanding	3	
	Phrases	28	
	Starting to talk		2
	<i>Sound effects and animal sounds</i>	12	12
	<i>Animal Names (real or toy)</i>	36	36
	<i>Vehicles (real or toy)</i>	9	9
	<i>Toys</i>	8	8
	<i>Food and Drink</i>	30	30
	<i>Clothing</i>	19	19
	<i>Body Parts</i>	20	20
	<i>Furniture and Rooms</i>	24	24
	<i>Small Household Items</i>	36	36
	<i>Outside Things and Places to Go</i>	27	27
	<i>People</i>	20	20
	<i>Games and Routines</i>	19	19
	<i>Action Words</i>	55	55
	<i>Words about Time</i>	8	8
	<i>Descriptive Words</i>	37	37
	<i>Pronouns</i>	11	11
	<i>Question Words</i>	6	6
	<i>Prepositions and Locations</i>	11	11
	<i>Quantifiers</i>	8	8
Part II Actions and Gestures	First communicative gestures	18	
	Games and routines	6	
	Actions with objects	17	
	Pretending to be a parent	13	
	Imitating other adult actions	15	

The second part of MBCDI, Words and Sentences (see Appendix 2 and Table 6), is targeted at toddlers between 18 and 30 months old. It evaluates language production and early grammatical constructions. The assessment is performed in two parts. Part I consists of two sections, namely ‘words children use’ and ‘how children use words’. The ‘words children use’ section is structured around a 680-word checklist grouped into categories (presented in Table 6). The section on ‘how children use words’ is a questionnaire where the evaluator needs to choose among the ‘not yet’, ‘sometimes’ or ‘often’ options according to the toddler’s performance.

Part II, ‘Sentences and Grammar’, is divided in five sections. The first one, ‘word endings/part 1’, is a questionnaire which evaluates whether or not the toddler is able to use plural forms ending in ‘-s’, gerunds or regular past-tense endings. The evaluator needs to choose among the ‘not yet’, ‘sometimes’ or ‘often’ options. The second section, called ‘Word Forms’, is a checklist of five irregular plurals and 20

irregular past-tense forms which the toddler is evaluated on. The next section, ‘Word Endings/part 2’, is a checklist of 14 wrong plural noun forms and 31 wrong simple past tense forms which the child produces. The Fourth Section, ‘examples’, contains a list in which the evaluator writes the three longest sentences the child has recently produced. This section is evaluated in terms of the length in morphemes of these three sentences. This metric is also known as M3L (mean length of the three longest sentences). The last section is called ‘Complexity’ and evaluates the child’s use of complex sentences. It consists of a list of 37 pairs of sentences; in each pair, the second sentence is considered to be more complex than the first. One point is given if the toddler is heard to produce the more complex sentence. No points are awarded when the less complex sentence is produced

Table 6: Summary of the MBCDI Words and Sentences.

Sections		Maximum score (lexical production)
Part I Words Children Use	<i>Sound Effects and Animal Sounds</i>	12
	<i>Animals (Real or Toy)</i>	43
	<i>Vehicles (Real or Toy)</i>	14
	<i>Toys</i>	18
	<i>Food and Drink</i>	68
	<i>Clothing</i>	28
	<i>Body Parts</i>	27
	<i>Small Household Items</i>	50
	<i>Furniture and Rooms</i>	33
	<i>Outside Things</i>	31
	<i>Places to Go</i>	22
	<i>People</i>	29
	<i>Games and Routines</i>	25
	<i>Action Words</i>	103
	<i>Descriptive Words</i>	63
	<i>Words About Time</i>	12
	<i>Pronouns</i>	25
	<i>Question Words</i>	7
	<i>Prepositions and Locations</i>	26
	<i>Quantifiers and Articles</i>	17
	<i>Helping Verbs</i>	27
	<i>Connecting Words</i>	6
	How children use words	5
Part II Sentences and Grammar	Word endings part 1	4
	Word forms	25
	Word endings part 2	45
	Examples	mean
	Complexity	37

The third part of the inventories, called MBCDI III, was developed in 2007 and is used for children between 30 and 37 months. It focuses on lexical production, sentence complexity and language use. This test has been used to measure the

linguistic development in L1 for both normally developing infants and toddlers and clinical cases. Because it targets a narrow population sample and is still much less common than the MBCDI I and MBCDI II, we will not refer to it in this thesis.

The validity of the MBCDI has been assessed in a significant number of studies carried out with monolingual participants, both healthy and language- impaired (Heilmann et al., 2005; Galeote et al., 2006; Luyster et al., 2007; Mayor & Plunkett, 2010; Mariscal et al., 2010), and with normally developing simultaneous bilinguals. Out of all these studies, only a few are longitudinal.

In what follows, we will present some of the longitudinal studies testing the validity of the MBCDI with normally developing participants. With respect to the studies carried out with monolinguals, Mayor and Plunkett (2010) investigated the link between estimated total vocabulary size and raw CDI scores. They showed that the CDI scores could be a basis for inferring children's vocabulary sizes. They argued that the CDI is not intended to be an exhaustive list of all the words any infant may know but showed that the CDI can serve as a basis for estimating infants' and toddlers' total vocabulary sizes. To do so, online calculators are available at www.bcbl.eu/cdi and <http://babylab.psy.ox.ac.uk/research/oxford-cdi/>.

One of the studies which tests the validity of the MBCDI on normally developing bilingual children was conducted by Pearson and Fernandez (1994) on 20 English Spanish bilinguals aged ten to 30 months and their vocabulary growth in both languages. The Spanish adaptation of the MacArthur Bates CDI was used together with the original CDI to measure the lexical growth at varying intervals within the same age range. The Spanish adaptation proved to be a valid tool for vocabulary assessment. Similarly, Mancilla-Martinez et al. (2010) conducted a study with 79 Spanish-English bilingual children. The results show that both the original and the Spanish adaptation can be used with this population up to 36 months of age and that parents are accurate reporters of their child's progress.

The validity of the MBCDI was also tested in clinical studies. Luyster et al. (2007) explored whether CDI scores at ages two and three could predict outcomes at age nine in children with autism spectrum disorders (ASD) and developmental delays

(DD). Sixty-two children referred for possible autism at age two and 19 children with DD were tracked up to age nine. Vocabulary, pre-speech and gestures scores on CDIs administered at ages two and three were used to predict follow-up IQ, language development, adaptive skills and scores on several diagnostic measures. The CDI scores at age two and three did not predict outcomes for the DD group. However, the CDI predicted a number of follow-up variables for the ASD group, although scores at age three were more predictive than scores at age two. This led them to conclude that the CDIs may serve as a quick and informative assessment of early verbal and non-verbal skills in children with ASD.

The MBCDI has also been used in developmental studies with both monolingual and bilingual participants. For example, one of the most recent studies was conducted by Tulviste and Schults (2020), who explored individual variability in a group of 100 Estonian children at the age of three. Results showed that children who have initially reported language difficulties scored lower, whereas children with more educated parents or with older siblings or a longer time spent in day-care earned higher scores on the test.

The studies conducted with bilingual participants provide both cross-sectional (Hoff & Elledge, 2005; Serrat et al., 2010; Poulin-Dubois et al., 2013) and longitudinal (Pearson et al., 1993; Pearson & Fernandez, 1994; Pearson et al. 1997, David & Wei, 2005) data.

Two types of studies longitudinally have assessed the early lexical development of simultaneous bilingual children using the MBCDI. The first type has assessed participants' vocabulary size in both languages. The second has compared the vocabulary sizes of simultaneous bilinguals with those of their monolingual peers who were just learning one language.

With respect to the studies focusing on bilinguals' vocabulary sizes, Pearson et al. (1997) separately analysed the lexical development of 25 English-Spanish bilingual children from the ages of eight to 30 months in their two languages. The parents had to fill out the MBCDI Words and Gestures or the MBCDI Words and Sentences according to the child's age in both languages, together with a questionnaire on the

child's language exposure. For a more complete lexical development analysis, the children were also recorded once a month. The results showed that the number of words learned in each language is proportional to the amount of time spent with speakers of that language. David and Li (2005) examined 13 French/English bilinguals between the ages of one and three using the MBCDI Words and Gestures and MBCDI Words and Sentences with respect to the composition of the L1 and L2 lexicons. They found that the lexical categories as measured by this tool evolved at the same rate and in the same proportion.

In the second group of studies, the MBCDI was used to compare the vocabulary size of monolingual and bilingual children. Pearson et al. (1993) analysed and compared the receptive and productive vocabulary sizes of a group of 25 English-Spanish bilingual children to that of a group of 35 English monolingual children. Data were collected by the parents every two or three months. As the age range of the children was between eight and 30 months, the parents had to fill out either the MBCDI Words and Gestures or the MBCDI Words and Sentences together with a questionnaire. The vocabulary size of monolingual participants was calculated directly for the English CDIs from the vocabulary checklists for both comprehension and production. The study found that in lexical reception, bilingual children's ability to understand two languages may be comparable in each language to that of monolingual children. With respect to oral production, it found that bilingual children scored the same as monolinguals if both L1 and L2 were considered. Hoff et al. (2012) tested the productive lexical development of 47 Spanish-English bilinguals and 56 English monolingual children using the MBCDI Words and Sentences. The measures selected from the MBCDI were: total productive vocabulary size in L1 and L2, and M3L and Complexity for English. The M3L and the Complexity measures were only compared for English. The results indicated no differences between the monolingual and bilingual participants with respect to their vocabulary sizes if both English and Spanish were taken into account.

The number of studies in which the MBCDI has been used to assess the lexical development of trilingual children is very limited. With respect to the MBCDI being used to assess longitudinal lexical development, a case in point is the study

conducted by Quay (2001), who reports the lexical development of Freddy in English, German and Japanese. To this end, the English CDI and its German translation were administered every three to four weeks, while the Japanese adaptation (Ogura, 1998) was administered at the ages of 1.1, 1.5 and 1.9. The assessment was also performed with the help of questionnaires and interviews on the child's behaviour and use of gestures, as well as weekly video recordings in all the languages. By the age of 16 months, Freddy's receptive vocabulary consisted of a high proportion of English words (63%), 35% German words and 2% Japanese words. In turn, Freddy was found to produce many more Japanese words (78%) even though the recordings were in English and in German. The English words produced amounted to as much as 8%, while the percentage of German words was 14%.

Byers-Heinlein and Wrecker (2009) conducted a cross-sectional study on word disambiguation with monolingual bilingual and trilingual participants in their shared language, English. To this end, they tested and compared three groups of children between 17 and 18 months old. The first group consisted of 16 English monolingual children, the second of 16 bilinguals and the third of 16 trilingual children. English was one of the languages of the bilingual and trilingual groups. The testing was performed in English with the MBCDI Words and Gestures, and the total comprehension and production scores were calculated. In addition to the MBCDI, further visual and auditory evaluations were conducted. Participants were exposed to brightly coloured visual stimuli and auditory cues in English. It was found that monolinguals were able to decode the meaning of new words by associating them with normal referents while bilinguals showed marginal disambiguation. In contrast, trilinguals were showed no disambiguation at all.

Thus, the MBCDI has been used extensively with monolingual and bilingual populations in their L1 and L2. There are also some accounts of the test being employed with early young trilinguals. However, no studies were found where the tool was used to assess young bilinguals learning an FL.

3.1.1.3. OTHER LEXICAL ASSESSMENT TOOLS

There are other less known tests which have been used with bilinguals in L1 and L2, along with others normed to test children in an FL. This section presents an account of the lexical assessment tools normed for bilingual participants. We begin by describing some of the tests targeting L1 and L2 and then mention others that test bilingual learners in an FL.

The ‘**Test de Evaluación del Lenguaje Infantil**’ (ELI, Saborit & Julian, 2005) assesses the receptive and productive vocabulary and language fluency of children bilingual in Spanish-Valencian between the age of three and five. The comprehension task requires the children to choose the correct item named by the evaluator among five options. The language production test provides children with a list of 30 items to be named, while the fluency task requires them to name as many animals or toys as they can within one minute. One of the studies conducted with the ELI is by Andreu et al. (2011), who used the Catalan translation of the ELI to investigate narrative comprehension and production in a group of Spanish-Catalan BFLA children with language impairment.

The Bilingual Verbal Ability Test (BVAT, Muñoz-Sandoval et al., 2005) is a test designed for a bilingual population from the age of five onwards. The test has been translated into 16 languages. Thus, the language combinations available are more varied than in the ELI and include English along with Arabic, Chinese, French, German, Haitian Creole, Hindi, Italian, Japanese, Korean, Navajo, Polish, Portuguese, Russian, Spanish, Turkish and Vietnamese. The test consists of three parts: the first one is a picture vocabulary task which consists of naming different objects. The second is an oral vocabulary task which consists of finding synonyms and antonyms. The third part is called ‘verbal analogies and comprehension and production’. The order in which they are administered is not random. The English test is always first. The test yields a bilingual verbal ability score, an English proficiency score and scores for each part. So far, we have been unable to locate any studies where the BVAT has been used to test bilinguals younger than five.

Standardised tests have also been designed for bilingual children exposed to an FL in a formal setting. In this sense, the **Woodcock Language Proficiency Battery-**

Revised (Woodcock, 1991) is a set of individually administered tests of English and Spanish that measure language production, reading comprehension and writing. It has been normed for both monolingual and bilingual school-aged children, as well as for adults who are non-native speakers of English or Spanish. The test consists of a test book and a CD. It covers several language areas in 13 sections. Thus, sections one through five test oral language, sections six through nine test reading comprehension while the remaining sections assess writing. In 2012, Gibson et al. used the Woodcock Language Proficiency Battery together with the Peabody Picture Vocabulary Test III (Dunn & Dunn, 1997) to compare the receptive and productive linguistic development of a group of 124 English-Spanish sequential bilinguals and another group of 110 English monolinguals (see [section 2.2.2.2.3.](#) for more details).

Another test battery is the **Early Language Listening and Oral Proficiency Assessment (ELLOPA)** (WIDA Consortium, 2004), an oral interview assessment tool which tests the oral fluency, grammar and lexical comprehension of five-year-old kindergarten children. The children are tested individually, and the teachers need training in sessions organised by ELLOPA to conduct this test and interpret the results. This test has been normed in several languages, including Chinese, English, French, German, Japanese, Russian and Spanish for children learning these languages as an FL in both regular and immersion programmes. Thus far, we have been unable to find published research using ELLOPA.

3.1.2. RECEPTIVE LEXICAL ASSESSMENT

Testing receptive lexical development of young multilingual learners is also a challenging task due to the lack of standard assessment tools for several languages. Given these circumstances, researchers can use either adapted or translated versions of standard lexical assessment tests normed for monolingual use. [Sections 3.1.2.1.](#) and [3.1.2.2](#) present some of these tests, as well as several studies in which they were used in different multilingual contexts. Special attention is given to the Peabody Picture Vocabulary Test, as it is also the test used in our study.

3.1.2.1. THE PEABODY PICTURE VOCABULARY TEST (PPVT-III)

The Peabody Picture Vocabulary Test (PPVT-III) (Dunn & Dunn, 1997) is a widely used standardised tool which assesses receptive language knowledge and has been normed for monolingual populations between the ages of two and 90. The first edition of the PPVT was published in 1965, but it was criticised for being inappropriate for children and adults from ethnic minorities. Therefore, a second edition appeared in 1981, called the PPVT-R, which included changes from the first version to make the test as unbiased as possible. Nevertheless, despite the revisions, Washington and Craig (1992) pointed out that the PPVT-R was still inappropriate to assess the African- American population. The third form PPVT III appeared in 1997 with changes, such as increasing the ethnic minority component of the normative sample.

“The ethnic minority sample in the PPVT-III standardization process was primarily comprised of African-American participants (18.1%) and Hispanic participants (12.9%), the remaining 4.6% was comprised of American Indian, Aleut Eskimos, Asians and Pacific Islanders” (Platt, 2010; p. 29).

The PPVT III consists of two parallel forms (A and B) with 17 sets of twelve items each. Each item consists of a slide with four pictures. The testee has to select the picture that best depicts the meaning of the oral stimulus presented by the examiner. The PPVT III has also been adapted and normed for Spanish (Dunn et al., 2010). This adaptation consists only of one form containing 16 sets of twelve items per set. Throughout the 16 sets, the items are numbered, and there is a total of 192 items. The administration procedure is the same for both the Spanish and the American English versions. At first, the starting set is selected according to the participant's age. This section gives the examiner guidance as to where the testing should start. To determine the precise starting point, the basal (the section where the testee makes no errors) is set. This basal is the point where the testing actually starts. In most cases, the starting set coincides with the basal, although this is not always the case. In situations in which the starting set contains more than one error, the examiner needs to go back to previous sets in order to find a basal. The testing stops when eight or more errors are committed in one set. The ceiling item is the last incorrectly chosen element in the last set. The direct score is calculated as the

difference between the number assigned to the ceiling item and the total number of errors committed from the basal until the ceiling set. The direct score yields to deviation scores such as the CI, percentiles, and rank and development scores, such as mental age, which are calculated according to the standardised norms. The calculation sheet for this test is presented in Appendix 3.

These are several cross-sectional studies proving the validity of the PPVT III with normally developing and impaired children. Platt (2010) performed a study on a group of 73 children of different nationalities, such as Korean, Chinese, Pakistani and Persian, who were exposed to English as an L2 in several kindergartens in Ontario, Canada. This group was compared to a group of English monolingual children. The results of the testing showed that the English monolinguals earned better scores and that this difference was not due to test bias. Therefore, the PPVT III could be used with English L2 learners and is also appropriate in a Canadian monolingual context. The validity of the PPVT test has also been proven in clinical studies. For example, Perez et al. (2012) investigated the pseudo-word repetitions in children diagnosed with Down syndrome (DS). They compared a group of 13 normally developing Spanish children with another group of 13 Spanish DS children. The Spanish version of the PPVT III (Dunn et al., 2010) was administered together with an auditory screening test, and it proved to be a valid tool to discriminate between the groups.

Doubt has been cast on the validity of the PPVT III in studies analysing the linguistic development of African-American minority children. Recent findings point to the fact that this instrument should be used cautiously when assessing the receptive language of minority populations. Restrepo et al. (2006) tested a group of 210 African-American (AA) children in Georgia whose mothers reported having low education levels. The PPVT III was used together with the Expressive Vocabulary Test (EVT)¹ (Williams, 1997), as the two tests are co-normed. The results were not consistent from one test to the other. While the EVT placed the

¹ This test is similar to the PPVT in terms of form and testing methodology but is focused on language production.

participants within the typical range of development, the PPVT showed scores below the typical range.

In spite of the fact that the PPVT has not been normed for bilingual populations, it has been used to assess the receptive lexical development of bilingual children in each of their languages, both cross-sectionally and longitudinally. To offer a more precise analysis of children's lexical development in their languages, the test has mostly been used together with other assessment methods, such as questionnaires and oral speech recordings. To our knowledge, there are no studies in which the PPVT was used to assess early trilingual children. Nevertheless, there is one cross-sectional account in which the PPVT was used to test monolinguals' lexical comprehension in an FL.

With respect to the cross-sectional use of the PPVT testing simultaneous bilinguals in two languages, the studies can be classified in two groups. The first group analyses the lexical development of bilingual children in both languages. The second one uses the PPVT to compare simultaneous bilinguals with their monolingual peers in terms of receptive vocabulary. In the first line of research, Dixon (2011) studied the role of parental and community (school) input in the receptive lexical development of BFLA children and reported data from a group of 284 bilingual children who were attending the last year of kindergarten in Singapore. Out of this group, 168 were English-Mandarin bilinguals, 65 were English-Malay bilinguals and 51 were English-Tamil bilinguals. The English PPVT and its translations into Mandarin, Malay and Tamil were used to collect data on the children's receptive linguistic development. The assessment sessions lasted for approximately 45 minutes per child, and the English PPVT was administered first. A questionnaire on parental education, household income, languages used with the child, primary caretaker, TV language and child's age and ethnicity was filled out by all parents before the PPVT testing. The results for English favoured children who also used this language at home.

With respect to the use of the PPVT to compare the bilinguals' and monolinguals' vocabulary sizes, Foroodi and Paradis (2009) conducted a study on CLI transfer in the acquisition of compound nouns. The participants were a group of Persian-

English bilingual children, a group of Persian monolingual children and another group of English monolinguals. At the time of the data collection, the children were approximately four years old. The English version and its Persian translation were used as a tool to determine the relative vocabulary size in each language, as well as linguistic dominance. A parental questionnaire which collected information on the child's background was also used to assist in determining language dominance. For the English test, the raw and the standard scores were calculated. Given that the Persian version has not been standardised, the results were expressed in raw scores only, while a compound word-production task in English and Persian was designed by the researchers and administered to the participants. Results indicated that the bilingual children produced more words in Persian than the Persian monolinguals and more words in English than the English monolinguals.

Several studies have been carried out longitudinally with simultaneous and sequential bilinguals using the PPVT. These studies usually measured the vocabulary comprehension scores of bilinguals against those of monolingual young learners. Doyle et al. (1978) compared the receptive lexical development of a group of 13 English- French simultaneous bilinguals to a group of 13 French monolinguals. To this end, the English version of the PPVT was used together with its French version, which at that time was 'relatively unstandardized and little researched' (p. 14). Apart from the PPVT, the Denver Development Screening Test (Frankenburg et al., 1992) was also used. Results show that bilingual children understood words at the same rate as their monolingual peers and were not as fluent, although they followed the same milestones of lexical acquisition.

Bialystok et al. (2005) used the PPVT to assess and compare the receptive linguistic development of a group of English monolingual Primary 1 children against a similar group of Cantonese- English, Hebrew- English and Spanish- English simultaneous bilingual children. The English and Spanish standardised versions of the PPVT were used together with its translations into Cantonese and Hebrew. The PPVT results were in favour of the English monolingual children.

The PPVT has also been used with sequential bilinguals in L2. In this line of research, Gibson et al. (2012) tested and compared the receptive linguistic

development of a group of 124 English-Spanish sequential bilinguals and another group of 110 English monolinguals. The data were collected on several occasions. For the English tests, the data were collected by an English native specialist. For the Spanish tests, the data were collected differently. The first-year tests were administered by a non-native but fluent Spanish speaker who had previously been trained. In the subsequent years, a native Spanish speaker administered the evaluations, and the PPVT English and Spanish editions were used to measure language comprehension. The scores were interpreted according to the norms, and the means and standard deviations were then calculated for each group. This study also assessed lexical production. The picture vocabulary subset of the Woodcock Language Proficiency Battery (Woodcock, 1991) in English and Spanish was used for this purpose. This subset provides a score which can be compared to the scores on the English and Spanish PPVT.

Studies in which the PPVT was used to test lexical receptive skills in an FL in monolingual children are quite scarce, and the data they provide is cross-sectional only. One of the very few studies was conducted by De Bot (2014) and analysed the receptive vocabulary of a group of 168 Dutch children exposed to English as an FL from young ages, but with different degrees of exposure (i.e., less than 60 minutes a week, between 60 and 120 minutes a week and more than 120 minutes a week). This first group was compared to a control group of 26 Dutch children who were not learning any English at that time. The English PPVT was used to test the receptive vocabulary in English for the first group, while its translation into Dutch was used to test language comprehension in Dutch for both groups. No differences were found between the two groups on the Dutch scores.

Thus, due to the lack of bilingually normed tests which assess lexical reception over time, the PPVT (in spite of being normed for monolinguals) proved usable longitudinally for bilingual participants in their L1 and L2. At the same time, there are also several studies using the PPVT to test receptive vocabulary in an FL.

3.1.2.2. OTHER RECEPTIVE VOCABULARY TESTS

Other standard tests that assess receptive knowledge and were normed for monolingual use are available. Nevertheless, to our knowledge they have not been used in longitudinal research. This following section summarises two of these tools.

The **Computerised Comprehension Task (CCT)**, Friend & Keplinger, 2003) evaluates monolinguals in their L1, while the **Test of Reception of Grammar (TROG 2)**, Bishop, 2003) assesses the receptive skills of monolinguals in an FL. These tests have also been used with bilingual participants as a secondary assessment tool in situations in which other standardised tests such as the MBCDI and the PPVT were not able to provide conclusive results on the lexical development of young learners (see Poulin-Dubois et al., 2013 or De Bot, 2014)

The **CCT** is a laboratory tool which assesses the accuracy of language comprehension in very young English, Spanish and French monolinguals up to the age of 24 months. It relies on the lexical entries listed in the MBCDIs, but the administration time is shorter (ten minutes). The programme presents 41 pairs of images representing nouns (23 pairs), verbs (eleven pairs) and adjectives (seven pairs) in the child's L1. The two images appear simultaneously on a computer touchscreen, one on the left-hand side and the other on the right-hand side. The images are matched for size, colour, brightness, difficulty and word class (noun, adjective and verbs), and accuracy is calculated as the total correct answers for all 41 trials.

Even though it is normed for monolingual children, this test has been used with bilinguals as well. Poulin-Dubois et al. (2013) used the CCT together with the MBCDI Words and Sentences to analyse the receptive and productive lexical development of 75 English- French bilingual children and compare it to that of a group of 18 monolingual children. Out of the monolingual group, twelve children were English monolinguals and six were French. No significant differences were found between the two groups.

TROG 2 (Bishop, 2003) assesses comprehension with a focus on lexicon and grammar and has been normed for students from the ages of four onwards who study English as an FL in formal settings. The presentation is similar to that of The Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 1997). Thus, each test stimulus is presented in a four-picture multiple-choice format with lexical and grammatical foils. The difficulty range has been increased in order to be adequate for secondary students and young adults. This test was used in the study conducted by De Bot (2014) presented in the previous section (in combination with the PPVT test) to provide a more complete picture of students' linguistic development.

One of the few tests normed for bilingual use is the **Receptive One-Word Picture Vocabulary Test-4, Spanish Bilingual Edition** (ROWPVT-4: SBE, Martin, 2013). It is the PPVT's counterpart, although it is only normed for English-Spanish bilinguals. In a review of this test, McClain et al. (2021) highlight the fact that it can be used for populations from age three onwards, as the item difficulty increases progressively. Furthermore, many of the items assessed are specific to the academic content students normally find at school.

3.1.3. PRODUCTIVE LEXICAL ASSESSMENT

There are several ways of assessing productive lexical development during the early stages of language learning. One way entails the use of standardised tools such as the **Expressive Vocabulary Test (EVT)** (Williams, 1997). This test has been designed as a complementary tool to the PPVT and is used to test lexical production in populations between the ages of two and 90 in English L1. Another way of testing lexical production is by means of audio/video recordings. In some cases, researchers audio- or videotape their participants in naturalistic situations (i.e., at home during daily activities). These recordings are thus based on spontaneous language production. In other situations, participants are recorded in formal settings while performing a prompted language production task, usually describing pictures or telling a story.

Both the spontaneous and the prompted recordings are then usually transcribed with the help of the CHAT transcription format (MacWhinney & Snow, 1990). CHAT is an acronym of Codes for the Human Analysis of Transcripts and is the standard

transcription system for Chiles (Child Language Data Exchange System) (MacWhinney & Snow, 1990), the largest database of conversational interactions. The CHAT system includes conventions for both transcription and coding. All files transcribed in CHAT can be analysed by the CLAN programmes, which can track different language structures, calculate automatic indices and analyse morphosyntax. Another possible way of analysing data from recorded transcriptions is with vocabulary profilers. For example, a popular one is available on Lextutor by Tom Cobb (www.lextutor.ca). The vocabulary profiler breaks the text down and categorises words in frequency bands according to how often they appear in real language.

Both the vocabulary profilers and the CLAN programmes are able to calculate several measures of fluency, lexical richness and complexity. Thus, the number of word tokens (total number of words produced by children in each language) and word types (the number of different words produced in a language) are used to describe lexical competence. Lexical richness has traditionally been calculated by the type-token ratio (TTR), which represents the number of different words as a ratio of the total number of running words in a text. This measure shows how likely it is for a student to repeat the same words. However, one of the main problems of this measure is that it is sensitive to text length. To solve this problem, a more appropriate measure is *D* (Malvern et al. 2004) which can be calculated by CLAN or D-Tools (Meara & Miralpeix, 2004). *D* shows how the TTR varies over a range of token sizes for each speaker and is more reliable than the TTR because it is not affected by text length. Lexical complexity is measured through the mean length of utterance (MLU), which can be expressed in morphemes (MLUm) or in words (MLUw). The MLUm indicates the mean number of morphemes produced per utterance, while the MLUw indicates the mean number of words produced per utterance. Importantly, MLUw is said to become less reliable when it surpasses a value of 4.0 (Wang, 2008). Nevertheless, it is a frequent measure used to describe children's language development. The MLU can be calculated using CLAN but not from the vocabulary profile.

The following two sections present several case studies based on spontaneous speech recordings ([section 3.1.3.1.](#)) and prompted recordings ([section 3.1.3.2.](#)) in

bilingual and trilingual learners. These recordings provide longitudinal data and are generally used together with other measures and tools, such as questionnaires or language tests, to provide a more complete picture of the vocabulary knowledge in different languages.

3.1.3.1. SPONTANEOUS SPEECH RECORDINGS

In the 1960s, it became very popular to record (and later to videotape) and transcribe children's early everyday lexical productions. One of the earliest studies doing so was Brown (1973), which assessed the English language development of three monolingual children during their preschool years. The children were periodically recorded by three different researchers when they interacted with their respective mothers. The data were transcribed and the syntax was analysed.

Similar studies were also conducted with bilingual and TLA children whose lexical development was assessed through video or tape recordings which were transcribed and then analysed. As we shall see below, one of the most common measures used in such studies are MLUm, and MLUw, number of types and number of tokens. In what follows, we present a number of case studies which used video or tape recordings to assess the lexical development of simultaneous bilingual and TLA participants.

With respect to studies conducted with simultaneous bilinguals, Volterra and Taeschner (1978) analysed the lexical development across time of two German-Italian bilinguals, Lisa (between the ages of 1.5 and 3.6) and Giulia (between the ages of 1.2 and 2.6). The girls were recorded monthly in 30-minute sessions and in alternate language environments. The recordings were carefully transcribed together with notes on children's behaviour that the researchers considered relevant for the study. A detailed lexical and syntactical analysis was performed afterwards, which showed the number of words each girl produced at a given stage, as well as the context in which these words were produced.

De Houwer (1990) described the morphosyntactic development of Kate, an English-Dutch simultaneous bilingual from the ages of 2.7 to 3.4. Kate was recorded for one hour every two weeks. All sessions included both English and

Dutch interactions in a naturalistic setting (Kate's home). The recorded data were transcribed, and separate MLUs were calculated and reported for each language.

Pérez-Vidal and Juan-Garau (2002) performed a study on the morphological and lexical development of Andreu, an English-Catalan bilingual child. To this end, a corpus made up of 87 recordings in the two languages between the ages of 1.3 and 4.2 was used. The MLUm per utterance was calculated for the early and later development stages. Table 7 shows the periods of late and early development expressed in years for both languages, along with the MLUm corresponding to each period. The results show that the two languages were acquired in similar ways, but their progress was not balanced. It is worth noting that Andreu's vocabulary in English was richer than in Catalan even though he was Catalan-dominant.

Table 7: Andreu's early and late lexical development stages (Perez-Vidal & Juan-Garau, 2002).

Development stage	Language	Age (in years)	MLUm
Early	English	3.2	2.24
	Catalan	2.5	
Late	English	3.4	3.21
	Catalan	2.8	

Zwanziger, et al. (2005) conducted a study on the presence of CLI in six English-Inuktitut simultaneous bilingual children. The ages of the children ranged from 1.8 to 3.9 years old. The participants were videotaped between four and six times over the course of one year, with each recording session lasting two hours. The recordings were transcribed in CHAT format (MacWhinney & Snow, 1990), and each utterance was coded for language, while the data analysis was carried out using the CLAN programmes (MacWhinney, 2000).

Cantone (2007) investigated the early mixing of five German-Italian simultaneous bilingual children. All the children had been recorded every two weeks for 30 minutes in each language as they played freely. Apart from the recordings, a diary was kept as well in order to observe relevant changes. Recordings were transcribed with intonations, hesitations, interruptions, movement and noise. To get a picture of the children's production in each language, during the study, the MLU/recording was calculated.

Comeau et al. (2003) conducted a study on the code mixing of six English- French BFLA children with the average age of 2.4 years. Each child was visited five times; the first two visits were used to observe the child's linguistic interaction with each parent. Thus, each child was video- and audiotaped with each parent for approximately 30 minutes. Most of the sessions were filmed at the children's homes. These recordings provided data on the children's linguistic development in the two languages. To this end, the following measures were used: MLUm, MLUw, word types in each language, word tokens in each language and multiword utterances (the number of utterances produced by each child when interacting with each parent). The word types and multiword utterances were transformed into percentages based on the total number of words and the total number of multiword utterances used in both languages. After the parental sessions, the children were audio- and videotaped in the presence of researchers in three separate sessions lasting approximately 30-40 minutes each. The researchers were bilingual in English and French. Twenty minutes of each session was transcribed using the CHAT transcription system, lexical mixing was analysed and the mixing rates were calculated.

Recordings and video sessions have also been used for the lexical assessment of very young early trilingual participants. Barnes (2006) recorded Jenny, a Basque-Spanish-English trilingual, between the ages of 1.11 and 3.6. The data were collected approximately every two weeks in conversations with her mother both on audio and video. The recording sessions lasted between 45 and 60 minutes, but only the first 30 minutes were transcribed. The recordings took place in Jenny's bedroom before the evening meal and included activities such as 'playing together with bricks or beakers, talking about toys, family and clothes looking at storybooks and photos, colouring, bathing and toilet' (p. 95). The transcriptions followed conventions of the CHILDES CHAT system (MacWhinney, 2000). To get a general picture of Jenny's development in English, the MLU and the TTR were calculated for all the scripts using the CLAN programme from CHILDES.

Hoffman and Stavans (2007) analysed the trilingual code switching of two English-Spanish- Hebrew trilingual children. The children started being recorded when the youngest was three and the oldest was six years old. The recordings included the children's interactions when playing with monolingual friends and when interacting

with other bilingual relatives and friends. The second set of data were recorded three years later, but this time the children were asked to tell a story. After the data were transcribed, the number of types and the percentages of codeswitching and code mixings were calculated.

Quay (2001) conducted a case study on the lexical development of Freddy, an English- German- Japanese trilingual. This study was performed with several data collection methods. One of them was video recordings, in which Freddy was recorded on a weekly basis in both German and English at home and in Japanese at the daycare centre. They were then transcribed in CHAT and classified according to language.

Montanari (2010) studied the lexical development of Kathryn, a Tagalog- Spanish- English trilingual. From the ages of 1.4 until 2.2, Kathryn's spontaneous speech was audio recorded by the researcher once a month until the child was 1.9. As she became more productive, these recordings were made on weekly or biweekly basis while she was performing activities such as drawing, reading books, eating and free playing. All family members were present so that Kathryn was able to produce utterances in all three languages. Then the recordings were transcribed and the transcriptions were doublechecked by her mother, a Tagalog speaking assistant and a Spanish-English bilingual. Apart from the recordings, a diary of Kathryn's lexical production was kept by her Spanish-speaking grandmother, a Spanish-English bilingual linguist and a trained phonetician. The notes were taken on a weekly or biweekly basis when Kathryn was interacting with her grandmother or other family members. This data contains information on the utterances, pronunciation and context. A cumulative lexicon was reconstructed from the recording transcriptions and diary notes, taking the first occurrence of each lexical item and placing it in order of appearance in a table containing the phonetic transcription of the word, the age at which the form appeared, the English translation and the context in which it appeared, which was defined by the language of the adult with whom Kathryn was interacting. TEs were identified on the basis of her interchangeable use of one word for another, while the words common to all two or three language were counted only once.

Most of these studies in which samples of oral production were recorded are case studies or were conducted with a small number of participants. Although the results of case studies cannot be generalised, ‘they may produce data that can always refute an existing generalization’ (Barnes, 2006; p. 94), and they are always a reliable source of hypotheses. More general results can be obtained from research conducted with larger amounts of participants.

3.1.3.2. PROMPTED SPEECH RECORDINGS

This section introduces prompted speech tasks as a means of assessing lexical production in young children. Of these tasks, the most frequent is storytelling through pictures. Among its advantages, we can highlight that they

“allow for a closer inspection of linguistic richness that may not be captured when utilizing standard assessments and this is particularly important for culturally, socioeconomically and linguistically diverse children” (Cureton & Lucas, 2007; p. 403).

Another advantage is that these tasks are not as subject to socioeconomic bias as standard tests may be. Research examining narrative production and comprehension in low-income children shows that these children are able to perform well using these methods. According to Craig and Washington (1994) and Benson (1997), narrative assessments place fewer constraints on the child’s language than the structured question format of standardised tests.

Many studies in this field have been conducted with cross-sectional designs. In the following sections, we will provide several examples of well-known studies on storytelling used to test oral production in simultaneous bilinguals in their L1 and L2, as well as bilingual young learners who were tested through storytelling tasks in an FL. Finally, one example of a study with young early trilinguals is provided in which storytelling tasks in all languages were used.

With respect to oral production through storytelling in bilingual young learners, Fiestas and Pena (2004) report a study with twelve Spanish-English bilingual children between the ages of four and six who were asked to tell a story in both languages using a wordless picture book. The stories were analysed in terms of the

number of total words, number of C-units and MLU. Minami (2005) conducted a study with 40 English- Japanese bilinguals between six and twelve years old who were asked to tell the story '*Frog where are you?*' in English and Japanese. To avoid the order of presentation effect, half the children were asked to tell the story first in English and then in Japanese, while the other half was asked to do it in the opposite order. The recordings were transcribed in the CHAT format of the CHILDES and then the number of verbs in past tense, the number of verbs in present tense and the number of nominals both in English and Japanese were analysed.

Research using oral speech recordings has also been conducted with bilingual participants in an FL. For example, Cenoz (2003) conducted a study on a group of 135 Spanish-Basque BFLA primary and secondary school children learning English as an FL in a regular school. Some of the 135 students had started to learn English in kindergarten at the age of four, others in primary school at the age of eight and others in secondary school at the age of eleven. At the time of the study, all participants had been exposed to 600 hours of instruction. Participants were asked to tell the story '*Frog where are you?*' in English, and the recordings were then transcribed. The aim of the assessment was to identify whether the learners who started to study English in kindergarten at the age of four codeswitched more than those who started to study English at the ages of eight or eleven.

Munoz (2006) used the *Dog Story* (see Appendix 4) as a production task to assess vocabulary richness, textual cohesion and narrative development in English as an FL. Participants were two groups of Spanish-Catalan simultaneous bilinguals learning English as an FL in regular schools. The first group had started to learn English at the age of eight, while the second one was exposed to English at the age of eleven. The amount of exposure, 273 hours, was the same for both groups. The recordings were transcribed and doublechecked following the CHAT conventions in CHILDES. The number of types, tokens and word families were obtained from the vocabulary profile. To analyse the lexical richness, the *D* measure was used, which was calculated with the help of D-Tools (Meara & Miralpeix, 2004).

The number of similar studies with early trilingual participants is limited. Stavans and Muchnik (2007) conducted a case study on two Spanish-Hebrew-English trilingual children where they analysed the instances of lexical mixing. To this aim, the children were asked to tell the story ‘*Frog where are you?*’, and the data were collected in two phases. In the first, the children were asked to tell the story in the three languages within a three- week intervals, while the second phase took place three months later. The children were given the audio recordings of the story told in the three languages by three native speakers and after listening to the stories, they were asked to retell it in all the languages they knew. The stories were transcribed and then divided into C-units consisting of argument and predicate clusters rather than lexical items or morphosyntactic units.

3.2. THE CHALLENGE OF ASSESSING LEXICAL DEVELOPMENT IN TWO AND THREE LANGUAGES

Assessing the lexical development of bilingual and TLA children has proved to be a challenge, as conducting such assessments implies evaluating the lexical development in all the languages spoken, ideally with the same normed tool. So far, only a few tests have been especially designed and normed for young bilingual children, but they do not cover a vast array of language combinations (to date, no tests have been created for children speaking more than two languages). As mentioned above, one example is the **Test de Lenguaje Infantil ELI** (Saborit and Julian, 2005), which was designed exclusively for Spanish-Valencian bilingual children under the age of six. The **BVAT**, on the other hand (Muñoz-Sandoval et al., 2005), uses more language combinations, but again they are limited.

Given the lack of tests, other ways of assessing bilingual and trilingual young children’s lexical development with respect to receptive and productive vocabulary needed to be found. For instance, as mentioned in Section 2, to asses vocabularies in more than one language, researchers have used either standard tests (together with their adaptations and translations) normed for monolinguals or oral speech recordings in each of the languages tested. Nevertheless, these methods have their limitations. This section presents some of the challenges of these tools and methods to test lexical knowledge of young bilingual and trilingual children.

3.2.1. USING TRANSLATED AND ADAPTED VERSIONS OF STANDARDISED TESTS

As seen in section 3.1.1.2. the MBCDI is one of the most comprehensive tools for the lexical assessment of young monolingual children in their L1. Not only does it test both lexical comprehension and production, but it also has broad linguistic coverage (adapted to more than 60 languages). In spite of being normed for monolingual use, these inventories have also been used to assess young bilingual and early trilingual children in their first languages (L1, L2 and L3). So far, this test has not been used with bilingual children learning an L3 in formal settings.

Using a tool such as the MBCDI to assess bilingual and TLA children has several limitations. According to García (2009), when bilinguals and TLA children are assessed in all their languages according to monolingual norms, their total vocabulary size and their abilities such as codeswitching, code mixing or transfer are not assessed. Under these circumstances, researchers using the MBCDI to compare bilinguals and monolinguals have come up with different solutions by adding language-combined measures to test the bilingual participants. These scoring solutions have been proposed and implemented with young bilinguals. However, thus far there are no accounts of how these adaptations can be made with early trilingual young children in L1, L2 and L3 or with young bilinguals learning an L3 in school. In what follows, we will provide two examples of language-combined measures to assess the total vocabulary size of young bilingual learners.

Pearson et al. (2003) analysed and compared the receptive and productive vocabulary sizes of a group of 25 English-Spanish bilingual children to that of a group of 35 English monolingual children, taking both languages into account. The vocabulary size of the monolingual participants was calculated directly for the English CDIs from the vocabulary checklists both for comprehension and production. In order to analyse the vocabulary size of the English-Spanish bilingual infants and toddlers, Pearson et al. (2003) took four measures into consideration. The first two, the English and the Spanish receptive/productive vocabulary size, were already part of the MBCDI. They then added two more language-combined measures: Total Vocabulary (TV) (the sum of all the words understood/produced in one language and all the words understood/produced in the other minus the

cognates understood/produced) and the Total Conceptual Vocabulary (TCV) (the sum of all words understood/produced in one language plus all the words understood/produced only in the other language).

The same method was applied by Poulin-Dubois et al. (2013), who analysed the productive and receptive lexical development of 75 English- French bilingual children by comparing it to that of a group of 18 monolingual children. For the bilingual participants, the TV (total vocabulary) size was calculated as outlined above. The proportion of TE was also calculated: the number of TE was multiplied by two and the result was divided by the total vocabulary size minus the cognates, semi-cognates (a pair of words that sound similar but have a slightly different spelling) and non-equivalents (a word that does not have a translation in the CDI in the child's other language).

A different type of calculation of bilinguals' vocabulary size was performed by Hoff et al. (2012), who tested the productive lexical development of 47 Spanish-English bilinguals and 56 English monolingual children. The TV size for the bilingual group was calculated as the sum of the total vocabulary size in English and the total vocabulary size in Spanish. Other researchers who analysed the vocabulary size of bilingual participants without comparing it to that of monolinguals continued to use the MBCDI original measures. The research conducted by Pearson and Fernandez (1994) and David and Li (2005), summarised in section 3.1.1.2., are cases in point.

The PPVT, which is also normed for monolingual use, is quite different from the MBCDI. Unlike the MBCDI, it only tests lexical reception and it has very limited linguistic coverage. So far, this test has assessed the lexical development of young bilingual learners in L1 and L2. However, there are no accounts of the PPVT being employed with young trilinguals. As in the case of the MBCDI, using the PPVT with young bilingual children is not free of challenges. Some of the limitations of using the PPVT with these learners are quite similar to those of using the MBCDI, and which were described above; for example, the PPVT is unable to yield a 'language-combined score' for all languages involved in the testing. Consequently, researchers continue to test participants in each of the languages involved and

obtain total scores for each language separately. The studies already cited in [section 2.2.1](#) are cases in point.

As mentioned before, another limitation of the PPVT is its narrow linguist coverage. Thus, in order to complete their lexical assessment in different languages, researchers need to rely on translations or adaptations (see Bialystok et al., 2005; Nejad & Paradis, 2005; Dixon, 2011). According to Kester and Pena (2002) when translating tests from one language to another, the order of difficulty of the listed items is not taken into consideration. In the original language, the test may start with the least difficult items, which may not correspond to the least difficult items in the target language. Nor is the frequency of word occurrence, given that the contextual variations of a certain words are not taken into consideration either. The grammatical structure may also affect the validity of a translation. For example, the category of gender is marked in Spanish but not in English. If a test in English is translated into Spanish, it will miss the gender aspect. Nevertheless, given the absence of standardised tests adapted to all languages, researchers continue to use translated versions of the PPVT to assess young bilingual and multilingual learners. Both the MBCDI and the PPVT are being extensively used in research to measure lexical development in two languages. Researchers have adapted some of the MBCDI scores to test young bilinguals. With respect to the PPVT, no language-combined scores are used. So far, there is not enough evidence of how these tests could work with young trilinguals.

3.2.2. USING SPEECH RECORDINGS

This section presents the difficulties that may arise when measuring children's lexical development through recordings of both spontaneous and prompted oral speech.

With respect to spontaneous speech recordings, evaluating these learners is quite challenging for several reasons. First of all, different languages are involved, and the data collection methodology in more than one language is more complex than in the case of monolingual speech. Thus, several interlocutors and researchers may be needed in order to collect and interpret the data in each of the languages involved. For example, Comeau et al. (2003) used both parents (one for each language) and

an English- French bilingual researcher to interact with the participants and collect and interpret the data. Two interlocutors, one for each language, were used to elicit data from Carlota, a German-Italian bilingual. Several German interlocutors were needed for data elicitation in the case of Lukas, another German-Italian bilingual (Cantone, 2007). Next, recording multilingual oral data involves more careful planning of data collection in the different languages, as more than one interlocutor needs to be available at the same time. Furthermore, several specialists for each of the languages are also needed to transcribe the oral speech samples. In the case of Kathryn, an English- Spanish- Tagalog trilingual (Montanari, 2010), the recordings for each language had to be programmed in such a way as to have all the interlocutors for each language present in order for the child to produce utterances in the three languages. To ensure reliability, the transcriptions of the recordings needed to be doublechecked by her mother, a Tagalog speaking assistant and a trained Spanish-English bilingual phonetician.

As far as prompted oral speech is concerned, narrative elicitation is one of the procedures frequently used. This methodology has proved challenging in the case of bilingual and early trilingual children for several reasons. First of all, the same story needs to be elicited several times in the different languages, which means that a productive effect may arise (Pavlenko, 2008). Thus, the last narrative may be less detailed than the previous ones and, therefore, will contain fewer tokens. At the same time, the last recording may also be more elaborate, as “having seen the ending during the first telling, the narrator can attribute more significance to previously unnoticed elements of the plot” (Pavlenko, 2008; p. 313). To overcome these effects, in most cases the elicitations are spread weeks apart, or the order is counterbalanced in the different languages. Thus, in the study performed by Stavans and Muchnik (2007), the trilingual participants were asked to tell the *Frog Story* in Spanish, Hebrew and English within a three-week interval. If the speaker’s languages are balanced, the language order is randomised and the narratives in a non-dominant language tend to be elicited first (Lanza, 2001). Another method that is also used is asking storytellers to switch languages at some point in the story.

Secondly, according to Chon (2012), the measures used to interpret oral speech production, such as number of types, tokens or MLU, are monolingually normed.

If used with speakers of more than one language, they can only offer a partial picture of participants' lexical development. In terms of vocabulary size measures (such as the number of types and tokens), one possible solution would be to use compound language measures such as TV and TCV. As far as complexity measures, such as MLU, Chon (2012) suggests that MLU should also be calculated in mixed utterances to get a complete picture of a child's language proficiency.

Therefore, researchers have adopted different solutions to overcome these limitations. For instance, oral speech transcriptions have been used together with questionnaires of language habits or diaries of vocabulary production in all the languages (Barnes, 2006; Montanari, 2010). Those who used multilingual language diaries together with oral speech recordings have also created cumulative scores for each language separately. As we have seen, in the case of Kathryn (Montanari, 2010), for example, the cumulative lexicon in all three languages was reconstructed by taking the first occurrence of each lexical item both from the diary and recordings and placing it in order of appearance in a table containing the phonetic transcription of the word, the age of appearance and the English translation, as well as the context in which it appeared.

Thus, in spite of the challenges entailed in assessing comprehension and production in very young bilingual and trilingual learners, different assessment solutions have been found that provide a reliable picture of early vocabulary development in two and three languages.

3.3 ASSESSING THE LEXICAL DEVELOPMENT OF VERY YOUNG LEARNERS IN SPANISH, CATALAN, AND ENGLISH

Focusing specifically on the languages being learned by the participants in our study, this section concentrates on the tools assessing lexical competence in Spanish, Catalan and English. We will refer to the possible evaluation tools meant to test Spanish-Catalan simultaneous bilinguals between the ages of three and six who are learning English in formal settings from very young ages.

As already mentioned, to date, no standardised assessment tools have been designed and normed for the lexical assessment of young trilingual learners. Therefore, in

this section, we will propose several possible options introduced in the previous sections for the lexical assessment of young Spanish-Catalan children immersed into English at school from young ages.

Even though it is not normed for multilingual use, the MBCDI could be used to assess early lexical development of both simultaneous bilinguals and early trilingual children in different language combinations. Because the Catalan and Spanish adapted versions of the original American test are available, the MBCDI could be used for the lexical assessment of early trilingual infants and toddlers between the age of eight and 30 months in languages such as Spanish, Catalan and English. So far, there is no evidence from research in which the MBCDI has been used to test young trilinguals in Spanish, Catalan and English. Nevertheless, there are accounts of research being conducted with Spanish-English bilinguals, as outlined in [section 3.1.1.2](#). There is also evidence of the MBCDI being used with Catalan children, such as the study performed by Serrat et al. (2010). The participants were 517 Spanish-Catalan BFLA children from 18 to 30 months old. The Catalan adaptation of MBCDI was used to study the relationship between children's lexical and morphosyntactic development. Furthermore, Serrat et al. (2022) have published the MBCDI Manual for the Catalan version.

Assessing Spanish-Catalan-English early trilingual children between the ages of three and six using standardised tools is also challenging because tests like the PPVT, which targets this age group, do not have a version adapted to Catalan. One possible solution would be to use the translated version together with further tests or questionnaires for a more thorough assessment. To our knowledge, there are no published studies conducted with the Catalan translated version of the PPVT, but we do have accounts of other studies which used translated versions of the PPVT, such as those in Farsi (Foorodi & Paradis, 2005) and Mandarin, Malay and Tamil (Dixon 2011), together with other tests.

In order to assess linguistic development in Spanish, Catalan and English, researchers have also created and used their own tests, such as a study conducted by Muñoz (2000) with school-aged children aimed at examining the Spanish, Catalan and English linguistic competence of three groups of students aged ten,

twelve and 17. The materials used for the evaluation were a cloze test and a dictation. All the cloze tests followed the same format, that is, 30 blanks, one every ten words. The Spanish and Catalan versions were adaptations of newspaper articles by the same writer, while the English cloze test was an adaptation of a well-known story. The dictations had all the same 50-word lengths and were adaptations of literary texts. Although these tests have been validated, they cannot be used with younger children because they would be too difficult for them.

Because we are analysing Spanish-Catalan bilinguals learning English as an FL in the age range of three to six years, we will begin by looking at tools assessing young learners' receptive and productive vocabulary in Spanish and Catalan. Next, we will focus on tests able to evaluate the receptive and productive vocabularies of these young learners in English.

Table 8 summarises the tools which have been used in research to measure the lexical development of young learners in Spanish and Catalan. On closer examination, we can see the pros and cons of the instruments that could be used in our own research. For example, the BVAT applies to children between the ages of five and nine but does not have a Catalan version. The MBCDI is normed and comprehensive, but it targets children below the age of three. The ELI could be appropriate, but it does not target children older than five. Regarding the PPVT, even though it targets a wide age group and has Spanish and Catalan versions, it only assesses lexical comprehension. Its counterpart, the EVT, which tests oral production, does not have a Catalan version available.

Table 8: Summary of lexical assessment tools for Spanish and Catalan.

	Test	Age (in years)	What it tests	Spanish- Catalan
<i>Bilingually normed</i>	BVAT	5-9	Comprehension/Production	No
	MBCDI	< 3	Comprehension/Production	Yes
<i>Monolingually normed</i>	ELI	3-5	Comprehension/Production	Yes*
	CCT	2	Comprehension	No
	PPVT	2-90	Comprehension	Yes
	EVT	2-90	Production	No

Test	Age (in years)	What it tests	Spanish- Catalan
<i>Others</i>	Oral speech recordings	Production	

**There is an adapted version in Valencian*

Table 8: Cont.

If we take into consideration the age range to be tested, as well as the availability of a Spanish-Catalan version, the most appropriate tool for testing these learners would be the PPVT. Because this test focuses on receptive vocabulary only, oral speech recordings could be used in these two languages to get information on the children's productive vocabularies between the ages of three and six.

Table 9 summarises the tools measuring lexical comprehension and production in FL. It shows that ELLOPA is targeted at five-year-old kindergarten children only, while the Woodcock Language Proficiency Battery has been designed for school aged children older than six.

Table 9: Summary of lexical assessment tools for young learners in English.

Test	Ages (in years)	What it tests
Woodcock Language Proficiency Battery	>6	Comprehension/production
ELLOPA	5	Comprehension/ production

Given that these tests are not compatible with the age group of our study, we believe that using standard tools designed for younger speakers of English as an L1 may be an option. As Table 8 shows, the MBCDI, which targets infants and toddlers between eight and 30 months of age, could be used to test the lexical comprehension and production of older children in English as an FL. To our knowledge, there are no studies in which the MBCDI has been used for this purpose. However, the tests may offer insight into these children's lexical development in the FL, although some data from speech recordings may be needed in order to obtain further information on lexical development in an FL at such young ages. However, we should recall that at early ages, it is difficult that children are elicited samples of oral speech in different languages.

3.4. SUMMARY

Assessing the receptive and productive lexical development of very young bilinguals and trilinguals is complex, because there are no readily available standard assessment tools especially designed for the purpose. To overcome this limitation and to test lexical competence in more than one language, researchers can use standard lexical assessment tools normed for monolingual young children, as well as oral speech recordings adapted for the purpose of the study.

The MBCDI is one of the most frequently used tools which aims to test both the receptive and productive vocabulary of infants and toddlers between the ages of eight to 36 months in more than 60 languages. Versions adapted to different languages keep appearing, such as Frota et al. (2016) in Portuguese and Serrat et al. (2022) in Catalan. This test has also been employed with simultaneous bilinguals in their L1 and L2. There are also several accounts of trilingual children tested with this tool in all three languages. No studies testing bilinguals in an FL have been conducted with the MBCDI.

The PPVT is also widely used to test the receptive vocabularies of populations of different ages (two to 90, as stated in the manual). There is also a considerable number of studies using this tool to test simultaneous bilinguals in their L1 and L2. Furthermore, there are accounts where the test has been used to assess FL receptive vocabulary. To our knowledge, the PPVT has not been used to evaluate the receptive vocabularies of early trilingual speakers. Oral speech recordings and especially storytelling have frequently been used as a means of testing oral production with different types of learners and language combinations.

The use of these tools with bilinguals in their L1 and L2 and trilinguals in L1, L2 and L3 is not free of constraints. For example, the limitations of standardised tests such as the MBCDI and the PPVT are due to the fact that they are not able to convey a complete picture of the lexical development of bilingual and trilingual children because, given that they are normed for use in one language, they do not take into account these learners' linguistic particularities (i.e., the existence of TE, or instances of mixing and transfer, in individuals speaking more than one language). In order to counterbalance these limitations, researchers use complementary

language assessment methods such as questionnaires and oral speech recordings. Using free speech (sometimes elicited by prompts) has also proved challenging in terms of the data collection methodology (the number of testers needed to collect and transcribe the data, scheduling the recordings, etc.) and the measures used to analyse the data, which are sometimes normed for monolingual use.

Regarding the evaluation of the early lexical development in Spanish, Catalan and English, two situations need to be taken into account. The first refers to very young early trilinguals in Spanish, Catalan and English (between the ages of birth and six), while the second describes Spanish-Catalan bilinguals learning English as an FL between the ages of three and six. As indicated in section 3.1.1.2., the MBCDI is used to assess lexical reception and production of Spanish, Catalan and English in early trilingual children between eight and 30 months old in all languages. No standardised tests meant for these children are available to cover the age range between three and six, which is the age targeted in this study. For Spanish-Catalan bilinguals between the ages of three and six who are learning English as an FL, one possible solution would be to use a test to evaluate lexical development in L1 and L2 and another test for the FL. Regarding lexical evaluation in Spanish and Catalan, to our knowledge there are no tests targeted at this age group that assess both lexical comprehension and production. Therefore, in order to have a complete lexical evaluation in the two languages, one of the options may be using the PPVT for lexical reception and storytelling to test oral production. With respect to FL assessment tools, no tests targeting this age group are available. However, other standardised tools used for much younger participants in their L1 can be used to test the lexical development in English of learners between the ages of three and six. Because the MBCDI assesses both receptive and productive lexical development, and because of the large number of studies conducted with this tool showing its reliability in first languages, we believe it could be used to test young children who are being introduced to an FL at an early age. Chapter 4 provides a detailed description of how the MBCDI, PPVT and storytelling were piloted and integrated as testing tools in the main study of the thesis, which is also presented in the next chapter.

CHAPTER 4

THE STUDY

While the process of language acquisition in simultaneous bilingualism has often been the focus of research, studies on the acquisition of an L3 as an FL by simultaneous bilingual children are very much needed, as the topic has hardly been explored, despite the fact that learning different languages from early ages is becoming quite common in our globalised society. In Chapter 2 section 2.2.2.2., we have summarised some of the research conducted with bilingual primary school children in Spain learning English in regular schools in Catalonia and the Basque Country. Nevertheless, to the best of our knowledge, no research has been conducted with simultaneous Spanish-Catalan bilingual kindergarten children learning English in immersion schools. One of the possible explanations may be the fact that evaluating the lexical development in L3 of such young learners is a difficult task, primarily due to the lack of standardised lexical assessment tools (see Chapter 3).

Taking this situation into account, our study expands on the research performed with simultaneous bilingual children learning an L3 in formal settings. More precisely, it aims to provide information related to the lexical acquisition of L1, L2 and L3 in young bilingual learners attending immersion schools. This chapter is divided in two sections. Section 4.1. describes two pilot studies which were performed before conducting the main study of the thesis. Taking into account the results of these previous studies, section 4.2. describes the objectives, the research questions and the methodology and design adopted in the main study of the thesis.

4.1. PRELIMINARY PILOT STUDIES

The two pilot studies were planned to provide insight for the data collection of the main study and to choose appropriate tools for the lexical assessment of the participants. Before describing the results of these studies, we believe it is necessary to justify why they were conducted. The rationale for these studies is presented in

section 4.1.1 (Preliminary remarks). Next, in section 4.1.2., we describe the first pilot study: its objectives, design, methodology, results and methodological decisions taken. Section 4.1.2. presents the second pilot study, following the same format as in the first study.

4.1.1. PRELIMINARY REMARKS

Although normed to test monolingual young children in their L1, the MBCDI has also been used for the lexical assessment of bilingual infants (eight to 18 months) and toddlers (18 to 30 months) in their L1 and L2 (see Chapter 3). To our knowledge, there are no standard tools available which assess the lexical development of preschool bilinguals of different ages in an L3, and we wondered whether using the MBCDIs with preschool bilingual participants in L3 could be useful, valid and informative. Thus, pilot studies were needed to assess the behaviour of this test.

The participants in the studies were chosen to have the same characteristics as the participants in the main study of the thesis. Thus, they were Spanish-Catalan simultaneous bilingual children between the ages of three and five who had been learning English as an L3 through immersion in the same school where the main study was going to be conducted. Given that at the time of the data collection the three-year-olds were attending the third year of the early childhood education programme, the group will be referred to as P3. Similarly, the four- and five-year-old groups will be referred to as P4 and P5, respectively.

These students were part of the one-way immersion programme described in section 2.1.3.1. This programme is meant to instruct majority language learners in an FL with a massive amount of input. Thus, the amount of exposure to the L3 in the early childhood education programme accounts for 62.5% of the entire time spent in school. It is an immersion model that is different from Baker's (1988) total immersion model (100% of the time, narrowed down to 80% after four years) or partial immersion models (50% of the time). Within the early childhood education programme, subjects such as language, maths, art, physical education, music, personal development and environmental knowledge are taught in the L3 by the class's native teacher. The children receive 20% exposure to Spanish through both formal teaching (three hours per week) and peer interaction on the playground. In

international schools, the exposure to Catalan is slightly lower than Spanish, accounting for 17.5% of the time the children spend in school (counting both formal instruction and peer interaction). Nevertheless, these children were allowed to use Spanish or Catalan (L1 or L2) to ask questions. In return, the teacher would reply in English.

The instruments used to assess lexical development in English were the MBCDIs parental report questionnaires (Fenson et al., 1993, 2007). We decided to use the revised version of the MBCDI reprinted in 2007. As previously outlined in Chapter 3, the MBCDI consists of two forms. The first one, called MBCDI Words and Gestures (see Appendix 2), is a language habit questionnaire used to assess the lexical development of children between eight and 18 months old in their L1. It measures comprehension and production of early words, as well as symbolic and communicative gestures. The second form, MBCDI Words and Sentences (see Appendix 2), targets children between 18 and 30 months old in their L1. It evaluates language production and early grammatical constructions. Because this is particularly relevant for the assessment, Tables 10 and 11 provide a more detailed description of the two tests, including the organisation of the sections, the linguistic content assessed, the measures (and labels) used by the researcher to report the results in the pilot tests and the maximum score that can be earned in each.

Table 10: Detailed description of the MBCDI Words and Gestures.

MBCDI Words and Gestures	Sections	Content	Measures (labels)	Maximum score
Part I Early Words	First signs of understanding	Checklist of commands	Commands (<i>Commands</i>)	3
	Phrases	Checklist of phrases that assess language comprehension	Phrases understood (<i>PhsU</i>)	28
	Starting to talk	Checklist of situations where the child imitates situations or labels words	Imitation and labelling (<i>Imitation</i>)	2
	Vocabulary checklist	Checklist of words grouped in 19 thematic sections	Words understood (<i>WdsU</i>)	396
		Checklist of words grouped in 19 thematic sections	Words produced (<i>WdsP</i>)	396
Part II Actions and	First communicative gestures	Checklist of gestures	Communicative gestures (<i>Gestures</i>)	18

MBCDI Words and Gestures	Sections	Content	Measures (labels)	Maximum score
Gestures	Games and routines	Checklist of games	Games and routines (<i>Games</i>)	6
	Actions with objects	Checklist of actions	Actions with objects (<i>Actions</i>)	17
	Pretending to be a parent	Checklist of actions	Parental role-play (<i>Role play</i>)	13
	Imitating other adult actions	Checklist of actions	Imitating adults (<i>Imitation</i>)	15

Table 10: Cont.

Table 11: Detailed description of the MBCDI Words and Sentences.

MBCDI Words and Sentences	Sections	Content	Measures (labels)	Maximum score
Part I Words Children Use	Vocabulary checklist	Checklist of words grouped in 22 thematic sections	Words produced (<i>WdsP</i>)	680
	How children use words	Situations in which a child refers to past or future events, or the absence of an object or owner	Words use (<i>Use</i>)	5
	Word endings, part 1	Situations where different word endings are used (plural in –s, possessive in –‘s, progressive forms, regular past tense)	Word endings 1 (<i>End1</i>)	4
Part II Sentences and Grammar	Word forms	Checklist of irregular verbs and plurals	Word forms (<i>Forms</i>)	25
	Word endings, part 2	Checklist of mistakes children make when using regular/irregular plurals and past tense	Word endings 2 (<i>End2</i>)	45
	Examples	The mean of the three longest sentences	Mean of the three longest utterances (<i>M3L</i>)	mean
	Complexity	The most grammatically correct sentence	Sentence complexity (<i>SC</i>)	37

Because the children’s parents did not have the linguistic competence in English needed to fill in the questionnaires, at this point the MBCDIs were completed by the researcher, who had been the children’s English teacher since their enrolment in the school and was thoroughly familiar with their linguistic profile in L3. To ensure reliability, the participants’ native English teachers, who were spending an average of six hours per day with the children, were asked to double-check the completed questionnaires and make any pertinent remarks in case of disagreement.

If the researcher and the teachers did not concur in their evaluation, cumulative scores were calculated (De Houwer et al., 2006). This cumulative score ‘counts the best score for an item of the CDI that any one rater gives’ (p. 336). In this sense, ‘a word was counted as comprehended if at least one of the raters (but not necessarily all of them) indicated that the child understood an item’ (p. 337).

Given that the number of participants involved in the two pilot studies was small (N=12 for each pilot study), the results are presented at the descriptive level and were not analysed in terms of statistical significance. The results are thus expressed only in terms of means and standard deviations.

4.1.2. PILOT STUDY 1

Research objectives

The first pilot study (cross-sectional) provides a description of participants’ vocabularies in L3. The objective was to test the adequacy of the MBCDI Words and Gestures and MBCDI Words and Sentences as L3 lexical assessment tools for preschool Spanish-Catalan bilingual participants, who are learning English through immersion following the syllabus described in [section 2.1.3.1](#) In order to see if this test could be used with this population across a three-year span, we examined possible ceiling effects, as well as this test’s ability to indicate progress from one year to the next, that is, from P3 to P5.

Participants

We tested a total of twelve students: four average P3, P4 and P5 Spanish-Catalan simultaneous bilingual children learning English as an L3 in the immersion school. The amount of exposure to English was 25 hours/week, that is, 62.5% of the time they spent in school. Students who attended extracurricular English classes or had native au pairs were excluded from the sample. Table 12 shows the main features of each group of students.

Table 12: Participants in pilot study 1.

Group	P3	P4	P5
<i>n</i>	4	4	4
<i>Age</i>	3-4-year-olds	4-5-year-olds	5-6-year-olds

Instruments

The MBCDI Words and Gestures form, targeting mostly comprehension, was believed to be more adequate for the P3 and P4 groups, who showed clear signs of understanding at the time the data were collected. After a careful analysis of all the sections, the decision was taken to use the ‘phrases’ and ‘vocabulary checklist’ sections, which assessed lexical development. As indicated in Table 10, the other sections examined infants’ behaviour during playtime and were therefore omitted. As a result, receptive vocabulary was measured according to the total number of *Phrases Understood (PhsU)*, which is found by adding all the phrases from the ‘phrases’ section that the learners understand, and the total number of *Words Understood (WdsU)*, which is found by the total of all the words the participants understand from the ‘vocabulary checklist’ section. Given that these children were also able to produce some words in English, the decision was taken to measure productive vocabulary as well. As a result, productive vocabulary was calculated according to the total number of *Words Produced (WdsP)*, which was found by counting the words from the vocabulary checklist section that children could produce, in addition to the ones they understood, just like the *WdsU*. That is, according to the MBCDI guidelines, *WdsU* indicates the words children just understand while *WdsP* indicates the words children can understand and produce. Thus, lexical comprehension is tested both explicitly through the total number of *WdsU* and implicitly through the total number of *WdsP*.

Given that the P5 children were already able to produce more than their P3 and P4 peers and were at a more advanced level, the decision was taken to use the second section, MBCDI Words and Sentences, which focuses only on lexical production. Productive vocabulary was measured according to the following measures: *Words Produced (WdsP)*, calculated by adding all the words the children produced from the ‘vocabulary checklist’ section; *Word Forms (Forms)*, found by totalling all the correct irregular past tense and plural forms the children produced from the ‘word forms’ section; *Word Endings 2 (End2)*, found by adding the wrong endings (viewed as a sign of progress at these stages) that the children made when using regular/irregular plurals and past tense (from the checklist in the ‘word endings, part 2’ section); *Sentence Complexity (SC)*, calculated by adding all the grammatically complex sentences the child produced which are listed in the

‘complexity’ section; and *the mean of the three longest utterances (M3L)*, found by calculating the mean number of words in the three longest utterances produced.

The reason for choosing *WdsP* is that it specifically targets lexis (i.e., it measures how many words these children were able to produce at a given time). The other measures selected (*Forms*, *End 2*, *SC* and *M3L*), included in the second part of the test (‘Sentences and grammar’, see Table 11) are more grammatically oriented. Given that the results of research conducted in the field of early grammatical development show that there is a close relationship between grammatical and lexical competence (see, for example, Serrat et al., 2010, among others, proving the existence of a linear relationship between morphological and lexical development), the decision was taken to include them in the pilot study as well. Other measures assess situations in which children produced certain types of words which were omitted.

Procedure

Data were collected in February 2011. First, the questionnaires were filled in by the researcher. Next, the children’s English native teachers (who were different for P3, P4 and P5) were asked to double-check the researcher’s assessment and in case of disagreement make pertinent comments in a different coloured pen. Each teacher verified only the questionnaires pertaining to the children in their classes. The researcher took about 30 minutes to complete each form, while the teachers needed an average of 30 minutes for possible corrections. The sample was small (twelve questionnaires), and both raters concurred in their evaluations. As a result, there was no need to calculate cumulative scores.

Results

In spite of the fact that inferential statistical analyses were not conducted with the data, due to the low number of participants, the results of MBCDI Words and Gestures test show that the P4 group earned better scores than the P3 learners in all lexical reception measures and production (see Table 13). This result suggests that this form can capture progress from P3 to P4. The findings also point out that when the data were collected, participants in both the P3 and P4 groups had higher scores in receptive than productive vocabulary, as could be expected. A closer look at the

data reveals no ceiling effect in any of the measures, thus indicating that this test may be used with larger populations at these levels.

Table 13: Descriptive results for the MBCDI Words and Gestures for the participants in P3 and P4 in pilot study 1.

Measures	Max. score	Group (n=4)	Mean	Sd.	Range	
					Min.	Max.
<i>Receptive</i>	PhsU	P3	10.75	2.87	9	15
		P4	17.25	1.70	15	19
	WdsU	P3	74.00	13.37	62	93
		P4	115.75	20.85	90	139
<i>Productive</i>	WdsP	P3	50.25	24.30	22	72
		P4	107.25	12.84	98	126

The results of the MBCDI *Words and Sentences* for P5 are summarised in Table 14. Note that the ‘maximum score’ cell corresponding to the *M3L* measure has been left blank, given that the score depends on each participant’s production. Looking at the findings, we see that they suggest there is no ceiling effect with the participants in the P5 group. This outcome indicates that these sections could be used for English as an L3 with similar populations.

Table 14: Descriptive results for the MBCDI Words and Sentences for the participants in P5 in pilot study 1.

Measures	Max. score	Group (n=4)	Mean	Sd.	Range	
					Min.	Max.
<i>Productive</i>	WdsP	P5	268.25	25.19	258	305
	Forms		7.00	1.41	5	8
	End2		2.25	1.89	1	2
	SC		19.00	1.15	18	20
	M3L		4.15	0.59	5	4

Summary of results and decisions reached

The results from the MBCDI Words and Gestures showed that this test could measure the lexical development in the L3 of the P3 and P4 participants. Because the P5 participants were tested with MBCDI Words and Sentences, which used different measures, a comparison between the P3-P5 and P4-P5 groups was not possible (the *WdsP* measure, which was used in both tests, did not assess the same words). In order to measure P3 and P4 against P5, a unique set of measures would be necessary. Furthermore, this pilot study reported only cross-sectional results, and lexical proficiency only for L3 (English). Longitudinal data on English was

necessary to complete the picture. Similarly, assessing the behaviour of the L1/L2 vocabulary test was crucial.

4.1.3. PILOT STUDY 2

Research objectives

In light of the results of the first pilot study, the decision was taken to conduct a second one. The second pilot study was performed in February 2012. The aims of the study were as follows:

Objective 1: To choose the most appropriate set of measures from the MBCDIs that would allow the P3, P4 and P5 groups to be compared cross-sectionally and longitudinally.

Objective 2: To find an adequate standard lexical assessment tool of participants' L1 and L2 that had already been used in similar contexts (i.e., preschool bilingual children being tested in their L1 and L2 on several occasions).

Objective 3: To detect any possible problems that may arise during data collection from the use of the same test in L1 and L2 (Catalan and Spanish), in order to avoid them in the main study of the thesis with larger number of participants.

Participants

The participants were selected from the same school programme as the ones in the first pilot study. They were chosen according to the same parameters as in the first study, that is, average Spanish-Catalan simultaneous bilingual students who are exposed to L3 in the formal school setting. The first group (P3) consisted of children who were between three and four years old. The participants in the second group (P4) were between four and five and the ones in the third group (P5) were between five and six. Three of the four students in P5 had been tested in the previous academic year (pilot study 1) with the MBCDI Words and Gestures, which meant that they were longitudinal. The P4 students could not be tested longitudinally, as they had not been tested on the same measures the previous year (when they were in P3). Table 15 shows the participants in the second pilot study.

Table 15: Participants in the second pilot study.

		Group		
		P3	P4	P5
n		4 (all cross-sectional)	4 (all cross-sectional)	4 (3 longitudinal)
Age		3-4-year-olds	4-5-year-olds	5-6-year-olds
Amount of exposure per week	English	25h (62.5%)	25h (62.5%)	25h (62.5%)
	Spanish	8h (20%)	8h (20%)	8h (20%)
	Catalan	7h (17.5%)	7h (17.5%)	7h (17.5%)

Instruments

To measure lexical development in L3, all the groups were administered the same sections of the MBCDI as in the first pilot study with the aim of confirming the findings of the previous study and choosing the most appropriate sections from the two tests for the longitudinal study of this thesis. To this end, all the participants were assessed on the total *PhsU*, *WdsU*, and *WdsP* from the MBCDI Words and Gestures and on *WdsP*, *Forms*, *End 2*, *M3L* and *SC* from the MBCDI Words and Sentences.

Assessing lexical development in Spanish and Catalan also entailed looking for the means to assess vocabulary in these two languages. Seeing that the learners were already too proficient to be tested with the MBCDI in their L1 and L2, a tool suitable for our purpose had to be found. After conducting extensive research on other possible validated tests (see Chapter 3), the *Peabody Picture Vocabulary Test* (PPVT, Dunn et al., 2010) or the *Test de Evaluación del Lenguaje Infantil* (ELI, Saborit & Julian, 2005) stood out as the most adequate for the study. To obtain reliable results in all three languages, it was important that the test assessed similar dimensions of lexical knowledge as what the MBCDI assess. The ELI proved to be similar to the MBCDI in that it assesses both receptive and productive language; however, it was recommended for use with children up to age five, while our participants at Time 3 would actually be five or a bit older, and this may have led to a ceiling effect. The PPVT, in contrast, could be used with a wider range of participants and has also been more extensively used in previous research, both linguistic and clinical studies (see Chapter 3). Therefore, the Spanish and Catalan versions of this test were used. Because Catalan and Spanish are typologically close languages (and no cultural difference was found for the items to be tested), we

determined that the Catalan translation of the Spanish PPVT² could also be used to test the participants in Catalan.

Procedure

Data were collected in February 2012. The testing in L3 was done following the same procedure as in the first pilot study: forms were first completed by the researcher. Next the participants' English native teachers were asked to double-check the researcher's evaluations and add comments in a different coloured pen on the same form in case of disagreement. In contrast to the first pilot test, the time allotted for the assessment was longer because the testing included more measures. More specifically, the researcher spent one hour assessing, each participant while the teacher needed an extra hour for possible corrections. As in the previous pilot study, the researcher and the teacher concurred in their evaluations, and the cumulative scores did not need to be calculated.

The Spanish and Catalan testing was performed by the researcher on a one-on-one basis. According to the instructions in the Spanish PPVT manual (Dunn et. al. 2010), the data collection was conducted in a quiet room, free of distractions and under appropriate visibility conditions (i.e., light from both the ceiling and behind the person being evaluated). Due to the fact that the PPVT is not a test with a time limit, the participants were not rushed in their answers. Nevertheless, if our subjects took more than 15 seconds to decide, the researcher would encourage them to answer. Irrespective of their linguistic profile (Spanish- or Catalan-dominant), the participants were first tested in Spanish and then in Catalan three days later. An average of 15 minutes per participant and test was needed to complete the evaluation. The scoring sheet is provided in Appendix 3

Results and discussion

Because the number of students participating in this pilot was small (n=12), the data were not analysed statistically. Therefore, we start by presenting the descriptive results of the MBCDI in English, and next we summarise the results of the PPVT in Spanish and Catalan.

² We thank Dr Elisabet Serrat for the translated version of the test.

Results in English (MBCDI)

Because this section presents both cross-sectional and longitudinal results, it has been divided into two parts. The first presents the cross-sectional descriptive results of the participants at each different level (MBCDI Words and Gestures and MBCDI Words and Sentences sections). Afterwards, we examine the lexical development over time of the three longitudinal participants, who were first assessed in P4 (T1/pilot 1) through the MBCDI Words and Gestures set of measures and then in P5 (T2/pilot 2).

Cross-sectional descriptive results

The results for the MBCDI Words and Gestures in the pilot study 2 provide mixed findings. Regarding *PhsU*, the participants in the P3 group scored the lowest, followed by the P4 learners, who were, in turn, surpassed by the older P5 group, which got the best results. Nevertheless, this same pattern was not found for *WdsU*: the P3 learners earned the lowest scores, followed by the P5 participants, who unexpectedly scored lower than their P4 peers. With respect to the oral production scores (*WdsP*), the results were not surprising: the P5 learners produced more words than their younger peers in P4, while the P3 learners were found to produce the smallest number of words. In spite of these mixed outcomes, we noticed no ceiling effects in any of the measures.

Because *WdsU* and *WdsP* are both part of the same section ('vocabulary checklist'), where the same word is evaluated in terms of both comprehension and production, we were able to compare receptive and productive vocabularies within the same group. Thus, we found that for the P3 and P4 children, the total number of words understood exceeds the total amount of words produced, as also found in the first pilot. This result was in line with our expectations, as we believed that during initial stages young learners are usually able to understand more than they are able to produce. However, this assumption did not prove true for the P5 participants, who surprisingly earned a better score for the total number of *WdsP* than for *WdsU*. These results are summarised in Table 16.

Table 16: Cross-sectional descriptive results for the MBCDI Words and Gestures for the participants in P3, P4 and P5 in pilot study 2.

Measures		Max. score	Group (n=4)	Mean	Sd.	Range	
						Min.	Max.
<i>Receptive</i>	PhsU	28	P3	10.75	2.87	9	15
			P4	17.25	1.70	15	19
			P5	18.00	3.46	13	21
	WdsU	396	P3	74.00	13.34	62	93
			P4	115.75	20.85	90	139
			P5	100.50	15.35	78	112
<i>Productive</i>	WdsP	396	P3	50.25	24.30	22	69
			P4	107.25	12.84	98	126
			P5	194.00	17.64	175	216

Note: The unexpected results are marked in bold letters

However, the unexpected results (marked in bold in Table16) in the second pilot test can be explained taking into account the format of the test. As already stated in the previous chapter and again in [section 4.1.1.](#) of this one, lexical comprehension is tested both explicitly through *WdsU* and implicitly through *WdsP*. Therefore, a total comprehension score can be found with a sum of the results in *WdsU* and *WdsP*. In our case, this score for P5 would be 294.50 which is higher than the total scores calculated for P4 (223.00) and P3 (124.25). At the same time, this score is also higher than the total number of *WdsP*. Therefore, it is in line with our expectations: P5 learners produced more than the P3 and P4 students, and their lexical comprehension is also better than their oral production.

On examining the three groups' scores in the MBCDI Words and Sentences (see Table 17), we did not find any ceiling effects in any of the measures. As expected, the P5 group was also found to have the best results. With respect to the P3 and P4 children, their performance in measures such as for *Forms* and *End2* was very low and reflects poor oral production skills. In the same vein, the P3 group did not earn high results in the *SC* and *M3L*, probably for the same reason. Nevertheless, these scores are higher in P4 and could therefore be compared against those in P5.

Table 17: Cross-sectional descriptive results for the MBCDI Words and Sentences for the participants in P3, P4 and P5 in pilot study 2.

Measures		Max. score	Group (n=4)	Mean	Sd.	Range Min.	Max.
<i>Productive</i>	WdsP	680	P3	36.00	12.90	24	47
			P4	124.25	16.54	112	147
			P5	268.25	25.19	248	305
	Forms	25	P3	0.25	0.50	0	1
			P4	0.00	0.00	0	0
			P5	7.00	1.41	5	8
	End2	45	P3	0.00	0.00	0	0
			P4	0.00	0.00	0	0
			P5	2.25	1.89	1	5
	SC	37	P3	0.00	0.00	0	0
			P4	14.50	4.04	11	18
			P5	19.00	1.15	18	20
	M3L		P3	0.00	0.00	0	0
			P4	4.35	0.66	3.8	5.3
			P5	5.15	0.59	3.6	4

Longitudinal descriptive results

This section presents the longitudinal results for the participants who were involved in both pilot tests (N=3), first while attending P4 and now P5. The results could only be reported for the MBCDI Words and Gestures, because at T1 the learners only took this test. As seen in Table 18, the test shows better results for *PhsU* and *WdsP* one year later. As in the cross-sectional results, the outcomes for *WdsU* (see the bold scores in Table 18) are unexpectedly in favour of the P4 group. At the same time, the P5 group's scores are higher for lexical production than lexical comprehension. These findings can again be accounted for by the test format. On calculating the total comprehension score (*WdsU* + *WdsP*), as we did in the previous section, we found a lower score for P4 (225.33) and a higher one for P5 (294.00). Furthermore, the score earned by the P5 group in comprehension is higher than in production.

Table 18: Longitudinal descriptive results for the MBCDI Words and Gestures for the participants in P4, and P5 in pilot study 2.

Measures		P4				P5			
		Mean	Sd.	Range		Mean	Sd.	Range	
				Min.	Max.			Min.	Max.
<i>Receptive</i>	PhsU	17.00	2.00	15	19	19.66	1.15	19	21
	WdsU	124.33	14.50	110	139	98.00	17.77	78	112
<i>Productive</i>	WdsP	101.00	3.60	98	105	196.00	20.59	175	216

Note: Unexpected results are marked in bold

Results in Spanish and Catalan (PPVT)

As illustrated in Table 19, the scores for Catalan are always higher than for Spanish. At the same time, we can also notice that the P4 group earned the lowest scores and the P5 children the highest.

Table 19: Cross-sectional descriptive results for the Spanish/Catalan PPVT test for the participants in P3, P4, P5 in pilot study 2.

Group	PPVT	Mean	Sd.	Rank	
				Min.	Max.
P3	Spanish	50.75	12.41	33	62
	Catalan	61.25	36.94	38	116
P4	Spanish	38.50	10.23	21	47
	Catalan	51.50	7.93	44	61
P5	Spanish	61.50	11.56	47	75
	Catalan	83.25	12.86	64	91

P3 would have been expected to score lower than P4; however, there were also high standard deviations (Sd) in P3. This unexpected result led us conduct another analysis taking into account the participants' dominant language. As Table 20 shows, whenever it was possible to have children with a different dominant language, the results of the PPVT show that the children tended to earn higher scores in their dominant language (either Spanish or Catalan). However, there are also a few exceptions (children who scored higher in their non-dominant language), which may have to do with the short time span (two to three days) within which the second test was administered; specifically, given the similarities between both languages, they may have remembered the correct answers. There seems to be no noticeable improvement between P3 and P4, which will need to be further explored with a larger group of learners, where statistical analyses can be conducted.

Table 20: Cross-sectional descriptive results for the Spanish/Catalan PPVT test for the participants in P3, P4 and P5 in pilot study 2 according to participants' language dominance.

Group	PPVT	Language dominance	Mean	Sd.
P3	Spanish	Spanish	58.00	5.65
		Catalan	44.50	9.19
	Catalan	Spanish	42.00	12.72
		Catalan	78.00	53.74
P4	Spanish	Spanish		
		Catalan	38.50	11.81
	Catalan	Spanish		
		Catalan	51.50	7.93
P5	Spanish	Spanish	53.50	9.19
		Catalan	76.50	17.67
	Catalan	Spanish	69.50	7.77
		Catalan	90.00	1.41

Note: Shaded cells indicate lack of participants for that measure.

Decisions reached based on the results

When examining the results of both the English and Spanish-Catalan tests, several decisions were taken with the aim of using an adequate methodological design for the main study of the thesis.

In spite of the results in the P5 group in *WdsU* that seem unexpected at first sight, the decision was taken to continue to use this measure according to the MBCDI standards, seeing that it affected the scores of one group out of three and that these scores could easily be interpreted by performing further calculations recommended in the test manual. Therefore, the measures to be used in the main study of the thesis were to be: *PhsU*, *WdsU* and *WdsP* from the MBCDI Words and Gestures and *SC* and *M3L* from the MBCDI Words and Sentences. Our decision was grounded on three findings. First of all, when participants were tested with these measures, no ceiling effect was found. At the same time, the results proved that this set of indices may more easily capture between-group differences. Furthermore, we also found these measures to be indicative of development with longitudinal participants. The *WdsP* from the MBCDI Words and Sentences was omitted to avoid redundancy with the *WdsP* in Words and Gestures, which could only be used in the more proficient groups, while the former could be used across the board. Nevertheless, it was decided not to use *Forms* and *End 2* in the main study of the thesis because the scores earned by the P3 and P4 groups were practically null.

Regarding Spanish and Catalan, the results showed that PPVT could be a useful tool for testing samples of participants at these stages of L1 and L2 development. Given that the PPVT has been fully standardised for this population, the lack of progress between P3 and P4 could be attributed to: 1) the low number of participants, which in any case a study with larger number of testees will help to elucidate; or 2) real backsliding due to the amount of exposure in English at school, which would definitely be worth exploring further.

In addition, several shortcomings were detected in the data collection process that could have affected the results. For instance, we believe that a three-day span between the two tests may have been too short, as some students could have remembered some of the answers and thus earn better scores at the second testing. This is the case of some of the participants in the P5 group, who scored higher in the Catalan test even though they are Spanish-dominant. Consequently, a change in the administration procedure was needed. It was clear that the order of the tests should be counterbalanced: half of the children should be tested first in their dominant language, while the other half should be tested first in their non-dominant language. This data collection procedure will be thoroughly explained in section 4.2.2.3.

4.2. THE MAIN STUDY

The research presented in this thesis is a three-year longitudinal study that assesses the lexical development in three languages in a group of 51 Spanish-Catalan simultaneous bilingual children who are learning English as an L3 in an immersion setting (thus, Immersion Group: IG). The study also compares their lexical performance with that of a control group, consisting of 15 Spanish-Catalan bilinguals of the same ages who are learning English as an L3 in a school with a regular curriculum in Barcelona (Regular School: RG1). Data were collected from both groups at three different testing times (towards the end of each academic year: in P3 (T1), P4 (T2) and P5(T3)). Furthermore, the oral lexical performance of the main group (IG) in P5, the end of kindergarten, will be compared to that of a control group of 44 older Spanish-Catalan children learning English as an L3 in regular curriculum schools in Barcelona (RG2). This comparison group was only tested

once and was added in the study because at the end of kindergarten children attending regular schools cannot produce much in English apart from several isolated words, while the group in the immersion school was able to produce longer utterances. Therefore, we thought it would be very informative to compare the attainment of those in the immersion group to the performance of more advanced students in regular curriculum schools.

This thesis addresses three main research questions, which are presented in section 4.2.1., following the main objectives. Next, in section 4.2.2. we describe the design and methodology of the main study of the thesis. This includes a presentation of the three groups of participants (IG, RG1 and RG2), the assessment tools and measures chosen to assess their lexical development in English and Spanish-Catalan, and information on the procedure and the analyses conducted.

4.2.1. OBJECTIVES AND RESEARCH QUESTIONS

Three main objectives were set for this study. Firstly, we aimed to assess the lexical development in English (L3) of a group of very young learners in an immersion school (IG) over the course of three years and compare it to that of a control group (RG1) consisting of children of the same ages and linguistic profile (Spanish-Catalan bilinguals) who were learning English in a regular school. Secondly, we wanted to examine the lexical development in L1 and L2 (Spanish and Catalan) of the IG learners and compare it longitudinally to that of the RG1 control group. By conducting this comparison, we aimed to see whether the early introduction of English as an L3 with a massive amount of formal exposure, as in immersion schools, can possibly affect L1-L2 lexical development. Finally, the third objective of this study was to analyse and describe the lexical production in English of the IG children at the end of kindergarten (P5). Thus, we compared the free productive vocabulary used by the IG group in an oral task to that of a comparison group (RG2), the latter consisting of older Spanish-Catalan simultaneous bilingual learners studying in regular schools. In order to guide our research, we formulated the three research questions (RQs) below and broke them down in to different sub-questions so as to address each RQ more thoroughly:

RQ1. What are the characteristics of the receptive and productive vocabularies in English as an L3 of very young Spanish-Catalan bilingual children learning English in an immersion school setting, as measured by the MBCDIs?

RQ1a. How does the English vocabulary of children learning L3 in an immersion school compare with that of children attending a regular school?

RQ2: What is the lexical development in Spanish and Catalan of very young Spanish-Catalan bilingual children learning English in an immersion school setting, as measured by the PPVT?

RQ2a. How does the Spanish-Catalan vocabulary of children learning an L3 in an immersion school compare to that of children attending a regular school?

RQ2b. Does language dominance (Spanish or Catalan) play a role in the children's lexical development in these languages?

RQ3: What are the characteristics of the lexical production in English in terms of fluency, diversity and complexity of Spanish-Catalan bilingual children learning English in an immersion school?

RQ3a. How does it compare to that of children at the beginning of secondary education attending a regular school?

4.2.2. METHODOLOGY

This section presents the methodology employed in the main study of the thesis. In section 4.2.2.1., we provide a detailed description of the groups of participants in relation to their AO (Age of Onset), AT (Age at Testing), language dominance, school they attend and type of syllabus followed. Section 4.2.2.2. presents the assessment tools used to test our participants in English, Spanish and Catalan, while section 4.2.2.3 presents the procedure and section 4.2.2.4 the data analysis.

4.2.2.1. PARTICIPANTS

The participants in the main group (immersion group: IG) were initially 60 Spanish-Catalan bilingual children who were learning English as an L3 through formal immersion at the same school where the pilot studies had been conducted (for a complete description of the school, see section 4.1.1.) Because school populations

can be unstable, especially in private schools, three students left at T2 (n=57) and six at T3 (n=51). In the light of these changes, the final size of this group was 51 students, who were assessed longitudinally in three consecutive years.

The 51 participants were chosen according to the same criteria as in the two pilot studies (i.e., students with no English native au pairs and no extracurricular English lessons). This time, a more careful selection was made in terms of language dominance, as we wanted to have a relatively equal number of Spanish- and Catalan-dominant students. It is worth noting that when registering their children at the school, parents needed to fill out a form (see Appendix1) where they chose the L1 (in this case, the dominant language), as well as the language (Spanish, Catalan or English) in which they would like to receive messages or important documents (schedules, marks, menus, etc.) from the school. Nevertheless, to make sure the information received from the parents was accurate, after collecting their answers to these questions, we double-checked them with the early childhood coordinator, who knows each family very well, with the children's main teachers and with the Catalan and Spanish teachers. After re-checking (to ensure we had correct information and an equal number of participants in each group), we were able to include 27 Spanish-dominant and 24 Catalan-dominant participants.

When children start P3, some of them have already been exposed to L3 in the same school even before the age of 3, albeit for few hours every week. Therefore, the participants were divided into three subgroups according to their age of onset (AO1, AO2 and AO3) to see if there were any significant differences between them that would prevent us from considering the IG one group and force us to treat these three subgroups separately instead (the results in the 'Preliminary Results' section in Chapter 5 indicated that there were no significant differences between groups). The first two subgroups (AO1 and AO2) consisted of children who had been exposed to English since the age of one and two, respectively, for approximately 2.5 hours a week, that is, 6.25% of their time at school. The rest of the time they received balanced input in Catalan and Spanish following the language-time model (Grosjean, 2010) described in Chapter 2. The last sub-group (AO3) consisted of children who started the immersion programme at the age of three, in P3, with

approximately 25 hours of exposure to English per week. The following table shows the distribution of the IG participants in relation to their AO and exposure.

Table 21. Immersion group (IG) participants in the main study according to their AO and hours of exposure to L3.

Group	Data collection time	n	Hours of exposure to English (according to the age of onset)		
			AO1 (N= 22)	AO2 (N=21)	AO3 (N=8)
<i>Immersion (IG)</i>	T1 (P3)	51	1170	1072	975
	T2 (P4)	51	2145	2047	1950
	T3 (P5)	51	3120	3022	2925

From the first year of nursery school (P3) onwards, the input received during the day was mainly in English both as means of communication and for the target vocabulary they were taught in class. The targeted vocabulary in the L3 consisted of words describing routines, as well as words related to different thematic units. Even though these units are the same from P3 to P5, the amount of targeted vocabulary increased from year to year. Vocabulary acquisition also occurred in Spanish and Catalan, although they were less exposed to these languages at school: the participants were exposed to Spanish eight hours a week and Catalan seven hours a week, both through formal instruction and peer interaction during playground time.

The participants in the longitudinal control group (RG1) were 15 Spanish-Catalan bilinguals of the same age as the children in the first group, but who had regular school curricula where English is taught twice a week. These participants were initially exposed to the L3 at the age of three, mostly through explicit instruction.

This school follows the official state school curriculum, which has a communicative approach with a strong focus on form and grammar. Thus, the students have fewer opportunities for production. From primary education onwards, students also have a textbook. Our participants followed this curriculum, but sometimes they used handouts created by the teacher instead of the textbook. Given that they were very young, the focus was on vocabulary related to the thematic units proposed by the teacher (e.g., Welcome to School, Autumn and Autumn Festivities, Christmas, Winter, Spring, Easter, Gardening and Summer). Instruction was mostly in English, but from time to time, when they were difficult to understand, the teacher also used

Catalan. The rest of the time these learners were addressed by their teacher mainly in Catalan. The exposure to Catalan totalled 30 hours per week, including formal instruction and peer interaction during playground time. These children were also spoken to in Spanish during both their language lessons and free play or playground time. As a result, their weekly exposure to Spanish totalled eight hours. Care was taken in the selection of RG1 participants to ensure that they were evenly distributed regarding language dominance. Thus, eight participants were Spanish-dominant and seven were Catalan-dominant. Given that the population studying at regular schools is more stable, we managed to have the same 15 longitudinal participants at T2 and T3.

The participants in the second comparison group (RG2) had already been tested in the Barcelona Age Factor Project (BAF) (Muñoz, 2006), which took place over several years. They were 44 Spanish-Catalan bilingual children who studied in three different secondary schools in Barcelona where English was taught twice a week (two hours a week). Just like with the RG1, these children followed the official curriculum and used a textbook as support (i.e., the Burlington textbook for IESO). Participants in this group were first exposed to L3 at the age of eight (i.e., when they were in the third year of primary school), and at testing time they had already been exposed to a total of 416 hours of English. Table 22 presents the distribution of the participants in this study in terms of age and exposure to L1, L2, L3 at testing time. In the case of the RG2 participants, lexical development in Spanish and Catalan was not considered in this thesis. Therefore, the cells corresponding to the amount of exposure to these languages are shaded.

Table 22: Distribution of participants in the main study.

					Hours of exposure at testing time		
Group	Data collection time	Grade	Age in years at testing time	n	English	Spanish	Catalan
<i>Immersion (IG)</i>	T1	P3	3-4	51	975	312	237
	T2	P4	4-5	51	1950	624	546
	T3	P5	5-6	51	2925	936	819
<i>Regular (RG1)</i>	T1	P3	3-4	15	78	312	1170
	T2	P4	4-5	15	156	624	2340
	T3	P5	5-6	15	234	936	3510
<i>Regular (RG2)</i>	T3	Secondary1	15	44	416		

Note: The shaded cells indicate information not taken into account in this thesis.

4.2.2.2. INSTRUMENTS AND MEASURES

This section presents the instruments used in the main study of the thesis in order to answer the three main research questions. We begin by presenting the tools and measures used to assess the longitudinal lexical development in English in terms comprehension and production (MBCDI). We continue by describing the tools employed to test lexical comprehension in Spanish and Catalan across a three-year time span (PPVT). As these two tests have been already described in Chapter 3, we will concentrate here on the specific points that were not previously mentioned. Finally, we introduce the task and the measures used to assess vocabulary in the participants' oral production at T3, at the end of kindergarten, and compare it with that of older students in regular schools.

The MBCDI (English)

Even though the MBCDI is used to assess lexical development of children between birth and 36 months in their L1, the conclusions of the previous pilot studies show that several sections of this test could be used with older participants to evaluate lexical comprehension and production in English as an L3. Taking into account the results from the two pilot studies, the set of MBCDI measures used in this study is summarised in Table 23.

Table 23: The MBCDI measures used in the main study.

Test	Section	Measures
MBCDI Words and Gestures	Phrases	PhsU
	Vocabulary checklist	WdsU
		WdsP
MBCDI Words and Sentences	Examples	M3L
	Complexity	SC

In addition to the measures displayed in Table 23, we decided to add an extra one, as we believed it could provide information on children's lexical development in L3. The new measure (word-class measures) was created by reorganising the vocabulary checklist (i.e., semantic categories) of the MBCDI Words and Gestures according to the nine different grammatical categories (total number of content words, total number of function words, total number of nouns, total number of verbs, total number of adjectives, total number of adverbs, total number of pronouns, total number of prepositions, total number of interjections). Table 24 provides a detailed

presentation of all the semantic measures (already part of the MBCDI Words and Gestures) and word-class measures used in the main study of the thesis. It should be noted that each measure will be assessed in terms of both reception and production.

Table 24: The MBCDI semantic categories and word-class measures used in the main study.

MBCDI measures	Maximum score	Thematic sections (labels)
Semantic categories	12	Sound Effects and Animal Sounds (<i>Sounds</i>)
	36	Animal Names (<i>Animals</i>)
	9	Vehicles (<i>Vehicles</i>)
	8	Toys (<i>Toys</i>)
	30	Food and Drink (<i>Food</i>)
	19	Clothing (<i>Clothes</i>)
	20	Body Parts (<i>Body</i>)
	24	Furniture and Rooms (<i>Furniture</i>)
	36	Small Household Items (<i>Household</i>)
	27	Outside Things and Places to Go (<i>Places</i>)
	20	People (<i>People</i>)
	19	Games and Routines (<i>Games</i>)
	55	Action Words (<i>Actions</i>)
	8	Words about Time (<i>Time</i>)
	37	Descriptive Words (<i>Descriptives</i>)
	11	Pronouns (<i>Pronouns</i>)
	6	Question Words (<i>Questions</i>)
	11	Prepositions and Locations (<i>Locations</i>)
	8	Quantifiers (<i>Quantity</i>)
Word-class measures	350	Total number of content words (<i>Cw</i>)
	46	Total number of function words (<i>Fw</i>)
	237	Total number of nouns (<i>N</i>)

MBCDI measures	Maximum score	Thematic sections (labels)
	59	Total number of verbs (<i>V</i>)
	37	Total number of adjectives (<i>Adj</i>)
	17	Total number of adverbs (<i>Adv</i>)
	17	Total number of pronouns (<i>Pron</i>)
	11	Total number of prepositions (<i>Perp</i>)
	18	Total number of interjections (<i>Interj</i>)

Table 24: Cont.

The PPVT (Spanish and Catalan)

To assess the lexical development in L1 and the L2 across time and thus provide an answer to RQ2, we maintained the decision taken after the second pilot test, namely to use the adapted Spanish version and the Catalan translation of the PPVT test. As previously presented in Chapter 3, the Spanish adaptation of the PPVT (Dunn & Arribas, 2010) contains 192 slides with four black-and-white pictures each. The participants' task is to select the picture that best shows the meaning of the oral stimulus presented by the examiner. The stimuli are divided in 16 sets, and the test was administered following the PPVT guidelines. At first, the starting set is selected according to the participant's age. This section gives the examiner a possible point of departure for the testing. Then the basal, which represents the first set with one or no errors, is determined. This is the point where the testing actually starts. In most cases the starting set dovetails with the basal, but not always. In situations in which the starting set contains more than one error, the examiner needs to go back to previous sets in order to find a basal. The testing stops when eight or more errors are committed in one set. The 'ceiling element' is the last incorrectly chosen element from the last set. Appendix 3 shows an example of a PPVT filled in at T2, when the participant in question was four years and nine months old. First, the starting set was chosen according to the age (e.g. set 2). However, as there was more than one error in this set, the basal (the actual starting point) was considered set 1, where our participant made only one mistake. Set 4, where eight mistakes were made, was considered the 'ceiling set', where the assessment finished.

Storytelling

According to Cureton and Lucas (2007) storytelling tasks evaluate linguistic variety and diversity better than standardised tests. Assessing L3 oral production skills at the end of kindergarten, which is the time immersion students start to produce sentences, entailed finding a proper story to suit our aim. The ‘Dog Story’ (Heaton, 2006) was chosen for participants to tell. The story consists of a set of six pictures showing a boy and a girl preparing food and going to a picnic with their dog, who is hiding in the basket that children are carrying. At the end, they discover that the dog has eaten all the food (see Appendix 4). ‘The Dog Story’ was appropriate for the students’ age and has been used with participants of similar English competence before (e.g., Miralpeix, 2006)

4.2.2.3. PROCEDURE

The data for the MBCDIs was collected towards the end of each academic year for three consecutive years (May-June 2012, 2013, 2014). For the IG group, the data were collected by the researcher (who was already familiar with the participants’ L3 linguistic profile) with the assistance of their native teachers, who spent approximately six hours per day with the children. Just as in the pilot studies, the researcher filled out each form and then it was handed in to the participants’ respective teachers, who were asked to double-check the questionnaires and make pertinent remarks with a different coloured pen in case of disagreement. Approximately 30 minutes were needed by the researcher to assess one child. In order to verify the researcher’s assessment, the teachers needed another 30 minutes per child. Because the number of participants was much higher than in the two pilot studies, the researcher and the teachers did not always concur in their assessments. When the lexical items were assessed differently by either rater (researcher/tutor), the cumulative scores ‘count the best score for an item of the CDI that any one rater gives’ (De Houwer et al., 2006; p. 336), as explained above. For example, if a word was marked as understood/produced only by one rater, it was accepted as such by the other. If one rater marked a word as understood and the other rater marked it as produced, the word was counted as produced. Appendices 1a and 1b contain samples of the MBCDI Tests.

In the regular school group (RG1), the class's English native teacher did the assessment after being trained by the researcher. The researcher also found several English language classes in the school. The same form was used for the assessment, where the teacher wrote the evaluation and the researcher made observations in case of disagreement. Because the two raters concurred in their evaluations, no cumulative scores needed to be calculated.

The Spanish and the Catalan Versions of the PPVT were collected by the researcher one by one at both schools, towards the end of the academic year in May-June, 2012, 2013 and 2014. To avoid possible misinterpretations of test results, as the second pilot study suggested, the data collection was performed in two stages but with a larger time span between the two tests. Half of the participants were first tested in their dominant language and the other half in the non-dominant one, and after a one-month interval they were tested in the other language. This testing procedure was also applied by Gibson et al. (2012), who used the PPVT in a longitudinal study with English-Spanish sequential bilingual young learners. The testing was performed according to the PPVT criteria (i.e., in a quiet room, free of distractions and under appropriate visibility conditions). A total of 15 minutes for each test was needed to evaluate each participant at each data collection time.

Oral speech samples from the IG group were collected at the end of T3, in June 2014, using the Dog Story. The participants were given several minutes to look at the drawings and were then asked to tell the story following the picture clues while being recorded. The researcher recorded their retellings and tried to intervene in the storytelling task as little as possible, offering encouragement only when participants showed signs of timidity but without providing any words. The comparison group RG2's oral production samples had already been collected by the researchers in the BAF project (Muñoz, 2006). Table 25 below summarises the data collection procedure for each group and test.

Table 25: Data collection procedure for each test and group involved in the main study of the thesis.

Test	Group	Data collection time	Rater/child (n)	Assessment	
				Time/Rater	Length of recordings
<i>MBCDI</i> s	IG	May-June 2012, 2013, 2014	2	30 min	
	RG1		2	30 min	
<i>Spanish/Catalan PPVT</i>	IG	May-June 2012, 2013, 2014	1	15 min	
	RG1		1	15 min	
<i>Oral production task: Dog Story</i>	IG	June 2014 1998	1	5 min	5 min
	RG2		1		

Note: Shaded cells indicate that the information does not apply.

4.2.2.4. Data analysis

This section summarises the analysis of the data. It discusses the data processing of the MBCDI and PPVT tests and of the transcriptions of the oral speech samples of the IG participants. We also outline several minor methodological decisions and editing issues and present the computer tools used to analyse both groups' oral data.

After the MBCDI forms were filled out, using cumulative scores when necessary, we proceeded to calculate total scores for each measure. Thus, for *PhsU*, we counted the total number of phrases marked as understood from the Phrases section. For *WdsU* and *WdsP*, we counted the total number of words marked as understood/produced from the vocabulary checklist. For *SC*, we calculated all the correct answers, while the *M3L* was found by calculating the mean of the three longest utterances produced. We also took into account the semantic and word-class categories of the words understood and produced. The semantic categories matched those already present in the MBCDI vocabulary checklist, so calculating them was quite straightforward. Therefore, the scores for each *semantic category* were found by adding first all the words understood and then all the words produced which were part of that category. Calculating the *word-class measures* scores was more time-consuming, as it first meant regrouping all the words in the vocabulary checklist according to different grammatical categories such as nouns, verbs, adjectives, adverbs, pronouns, interjections, content words and function words. Next, the scores for each category were calculated by adding all the words understood and all the words produced.

The data from the PPVT test was processed following the instructions in the test manual. Thus, the raw scores were found by calculating the difference between the ceiling element of the test (the last incorrectly chosen element) and the total number of incorrect answers.

The oral production samples of the 51 participants in the IG group were transcribed by the researcher and formatted according to the requirements of the IT tools used for the analysis of the data. First, the oral data were transcribed following the CLAN format and then they were double-checked by the researcher while listening to the recordings again.

In order to ensure that the number of types, tokens and grammatical categories used was reported consistently for each participant, the corpus transcribed (storytelling of the immersion and regular groups) was trimmed. Thus, false-starts and hesitation markers were eliminated, and some editing of the data were needed (see Appendix 6). L1 and L2 words, though infrequent, were also omitted. After performing these operations, the files were saved as a plain text (.txt) so they could be easily analysed with the computer tools described below.

After the oral speech samples were edited, we calculated the total number of tokens, types and word families produced by each learner. To do, the online version of VocabProfile 'VPkids' by Tom Cobb was used. Thus, each plain text sample was inserted into the programme and processed to find these results. The different word-class categories were found by means of the CLAWS part-of-speech tagger developed at Lancaster University and retrieved from www.ucrel.lancs.ac.uk/claws/trial.html. The programme takes the plain text and assigns each word a grammatical part-of-speech tag. The output was checked for any possible inconsistencies using the vertical output display, which was more suitable for manual post-tagging.

The storytelling task was analysed according to three dimensions (fluency, diversity and complexity) using different measures, which are summarised in Table 26. Lexical fluency was assessed by the total number of types, tokens, families and other different class measures (total number of content words, nouns, verbs,

adjectives, adverbs, function words, pronouns, numerals, prepositions, articles and conjunctions). For lexical diversity, we used Giraud's index (*GI*), as this measure is not length-dependent. The formula to calculate this measure is $\frac{types}{\sqrt{tokens}}$. The *GI* was used because at their level, not all participants would be able to produce the minimum of 50 words needed to calculate this measure (Meara & Miralpeix, 2017). Complexity was measured using the mean of the three longest utterances produced (*M3L*) and the mean length of utterance (*MLU*).

Table 26: Measures calculated from the storytelling task.

Blocks	Measures	Labels used
<i>Lexical fluency</i>	Total number of tokens	Tokens
	Total number of types	Types
	Total number of families	Families
	Total number of content words	Cw.
	Total number of nouns	N.
	Total number of verbs	V.
	Total number of adjectives	Adj.
	Total number of adverbs	Adv.
	Total number of function words	Fw.
	Total number of pronouns	Pron.
	Total number of numerals	Num.
	Total number of prepositions	Prep.
	Total number of articles	Art.
	Total number of conjunctions	Conj.
<i>Lexical diversity</i>	Giraud Index	GI
<i>Complexity</i>	Mean of the three longest utterances	M3L
	Mean length of utterances	MLU

The data were analysed with the SPSS Statistical Programme, version 25. Normality tests were conducted for each data collection period (T1, T2 and T3) and for each test used. Next, statistical analyses were conducted in order to answer all three research questions.

To answer RQ1 (*What are the characteristics of the receptive and productive vocabularies in English as an L3 of very young Spanish-Catalan bilingual children*

learning English in an immersion school setting, as measured by the MBCDIs?), we performed a one-way RM ANOVA with *Time* (*T1*, *T2*, *T3*) as the within-subjects factor for all three sets of measures (i.e., total scores, semantic measures and word-class measures). We also applied the Greenhouse-Geisser correction when necessary. For variables which did not meet the normality assumption, we conducted a Friedman test. In order to detect whether significant changes occurred between *T1/T2*, *T2/T3* and/or *T1/T3*, we used Bonferroni pairwise comparisons for variables showing a normal distribution and the Wilcoxon signed-rank test for those which did not. In order to answer RQ1a (*How does the English vocabulary of children learning L3 in an immersion school compare with that of children attending a regular school?*), we conducted a 3x2 RM ANOVA with *Time* (*T*) as a within-subjects factor and *School* (*S*) as the between-subjects factor (or a Mann-Whitney *U*-Test for each data collection time, if normality was not met) for all three sets of measures involved.

To answer RQ2 (*What is the lexical development in Spanish and Catalan of very young Spanish-Catalan bilingual children learning English in an immersion school setting, as measured by the PPVT?*), we conducted a one-way RM ANOVA with *Time* (*T*) as the within-subjects factor, as well as Bonferroni pairwise comparisons to detect significant differences between *T1/T2*, *T2/T3* and *T1/T3*. Sometimes the Greenhouse-Geisser correction needed to be applied. Answering RQ2a (*How does the Spanish-Catalan vocabulary of children learning an L3 in an immersion school compare to that of children attending a regular school?*) and RQ2b (*Does language dominance [Spanish or Catalan] play a role in the children's lexical development in these languages?*) entailed conducting a 3X2 RM ANOVA with *Time* (*T*) as a within-subjects factor and *School* (*S*) as the between-subjects factor, as well as applying the Greenhouse Geiser correction when necessary.

To answer RQ3 (*What are the characteristics of the lexical production in English terms of fluency, diversity and complexity of Spanish-Catalan bilingual children learning English in an immersion school?*), we provide the means and standard deviations of all measures used. Next, we conducted a *t*-test (or the Mann-Whitney *U*-test for variables which did not meet the normality assumption) to answer RQ3a (*How does it compare to that of children at the beginning of secondary education*

attending a regular school?). It should be noted that because the secondary school participants went to different schools in Barcelona, a Kurskal Wallis test was conducted to make sure their storytelling could be taken as a group. The next chapter presents the results summarised above.

CHAPTER 5

RESULTS

This chapter presents the results of the analysis carried out to answer our research questions. As indicated in Chapter 4, this research consists of a three-year longitudinal study that assesses the lexical development in three languages (English, Spanish, Catalan) of a group of 51 Spanish-Catalan bilingual children who were learning English as an L3 in an immersion setting. At the beginning of our study (T1), the participants were attending early childhood school, P3 (i.e., they were three years old). Some of these children had been exposed to L3 in the same school before the age of three but for just a few hours per week. Therefore, the participants in the IG were initially divided into three subgroups (P1, P2 and P3) according to the age at which they were first exposed to English at school (AO of one, two or three years old). The reason for doing so was to assess whether the different amounts of exposure to English could have led to any significant differences in lexical competence among children who enrolled in the programme in P1, P2 or P3.

Therefore, section 5.1, following this introduction, presents the results of the analysis that aimed to assess the lexical performance of the P1, P2 and P3 subgroups, both in English and in Spanish/Catalan. Sections 3, 4 and 5 provide the results for the research questions (RQs) proposed in Chapter 4. In section 5.2, which addresses RQ1, we first analyse the longitudinal lexical development in L3 of the participants in the IG in terms of reception and production and based on the MBCDI's three sets of measures. Then, in response to RQ1a, this group's performance is compared across time against that of the control group (RG1), which consists of 15 Spanish-Catalan bilinguals of the same age as the IG children and who were attending P3 in a regular school where English was taught twice a week (two hours weekly) at the beginning of data collection. The two groups are measured and compared at T1, T2 and T3 in terms of their lexical receptive/productive vocabulary in the FL using the same measures in the MBCDI.

The results for second research question (RQ2) are presented in [section 5.3](#). We begin by describing the receptive vocabulary development in Spanish and Catalan of the participants in the IG. Then, in response to the RQ2a, their results are compared to those of the participants in RG1 with the aim of detecting whether early, massive exposure to the L3 has any effect on the lexical competence in the L1/L2. Given that the results of the L1/L2 may be influenced by participants' language dominance, we performed three different analyses. First, we compared the receptive vocabulary in L1 and L2 of all participants in the two groups. Next, we analysed the Spanish-dominant participants in the two groups, and finally, we performed the same analysis for the Catalan-dominant participants.

The results for the third research question (RQ3) are presented in [section 5.4](#), where we show the results of the IG in the oral production task in terms of lexical fluency, diversity and complexity at T3. These outcomes were compared against those of a comparison group (RG2) consisting of 44 Spanish-Catalan bilingual secondary school children in 1ESO who were learning EFL in three different regular secondary schools in Barcelona (namely, schools A, B, and C).

5.1. PRELIMINARY ANALYSES: IG PARTICIPANTS

This section presents the results of the preliminary analyses conducted with the P1, P2 and P3 participants with respect to their receptive/productive vocabularies in L3 and their receptive vocabularies in L1/L2. As mentioned in the introductory section, these analyses aimed to make sure that participants in the IG were all at the same level despite the fact that some had received some more exposure to the FL in years previous to P3.

The section provides the results of the P1, P2 and P3 subgroups on the MBCDI **total score** measures at T1; that is, when our study started and the participants were three years old (from then onwards, the amount of exposure was kept constant). The scores on the MBCDI lexical and grammatical categories are not presented separately, as they were actually part of the total score of the test.

Before presenting the results, we consider it necessary to restate the terminology used to refer to each of the MBCDI total measures. As already indicated in Chapter

4, the measure *phrases understood* is abbreviated as *PhsU*, *Words Understood* as *WdsU*, *Words Produced* as *WdsP*, *Sentence Complexity* as *SC* and *Mean of the three Longest Utterances* as *M3L*.

The results of the Saphiro-Wilk normality tests (Appendix 7, Table 45) showed that most of the variables in the MBCDI total score measures had a normal distribution. Therefore, a one-way between-groups ANOVA test was performed in order to compare the subgroups' performance. For the *M3L* measure, which did not meet the normality assumption (marked with * in Table 27), a non-parametric Kruskal-Wallis test was also conducted. As summarised in Table 27, the scores on the receptive vocabulary measures showed no significant differences among the three groups. Regarding the productive vocabulary measures, the findings confirm the same tendency. The Kruskal-Wallis test results indicate a non-significant difference for *M3L* ($\chi^2(2)=71.152$, $p=.213$).

Table 27: Descriptive statistics and one-way between-groups ANOVA of the MBCDI total scores measures at T1 for the participants in the IG (N=51), depending on their AO, with $p \leq .01$.

Measures	Group	N	Mean	Sd.	F	p
Receptive vocabulary	PhsU	P1	24	11.83	6.34	.966 .387
		P2	19	9.68	4.88	
		P3	8	9.38	5.82	
	WdsU	P1	24	93.79	25.08	.713 .495
		P2	19	87.59	18.52	
		P3	8	85.13	12.40	
Productive vocabulary	WdsP	P1	24	33.79	17.64	.311 .734
		P2	19	29.79	16.24	
		P3	8	31.50	13.88	
	M3L*	P1	24	1.67	.83	2.880 .066
		P2	19	1.09	1.35	
		P3	8	.85	1.05	

Note: *Normality assumption violated.

As in the case of the MBCDI **total scores**, the analysis conducted for these subgroups was also performed with the PPVT scores. The normality analysis (Appendix 7, Table 46) revealed a normal distribution for all the PPVT measures. Therefore, a one-way between-groups ANOVA was conducted to see whether there were significant differences between P1, P2 and P3 regarding their lexical knowledge in L1/L2. As summarised in Table 28, the results show no statistical differences in any of the measures.

Table 28: Descriptive statistics and one-way between-groups ANOVA for the Spanish/Catalan PPVT scores at T1 for the participants in the IG (n= 51), depending on their AO, with $p \leq .01$.

PPVT Measures	Group	n	Mean	Sd.	F	p
Spanish	P1	24	33.21	2.68	.146	.865
	P2	19	31.74	11.17		
	P3	8	30.88	10.04		
Catalan	P1	24	32.67	11.46	.964	.389
	P2	19	28.26	9.80		
	P3	8	29.88	8.32		

None of the results of the ANOVAs and the non-parametric Kurskall-Wallis test carried out for the IG participants showed significant differences between the children who enrolled in the programme at the ages of one, two or three. These findings allow us treat these participants as the Immersion Group (IG).

5.2. RQ1: EARLY LEXICAL DEVELOPMENT IN EFL

This section addresses the first research question and its sub-question, which was formulated in the previous chapter as follows:

RQ1. What are the characteristics of the receptive and productive vocabularies in English as an L3 of very young Spanish-Catalan bilingual children learning English in an immersion school setting, as measured by the MBCDIs?

RQ1a. How does the English vocabulary of children learning L3 in an immersion school compare with that of children attending a regular school?

For RQ1, results are presented organised in three sets of measures (i.e., the MBCDI **total scores**, the MBCDI **semantic categories** and the MBCDI **word-class categories**). One-way RM ANOVAs and Friedman tests were conducted to assess development with these measures.

In order to answer RQ1a, we also examined our data with respect to the same three sets of MBCDI measures. We conducted a 3x2 RM ANOVA (or a Mann-Whitney *U*-Test if normality was not met). These analyses sought to compare the lexical development of the IG and the RG1 over a three-year period.

5.2.1. RECEPTIVE AND PRODUCTIVE LEXICAL DEVELOPMENT OF CHILDREN ATTENDING THE IMMERSION SCHOOL

This section addresses RQ1 by describing and analysing the longitudinal lexical development of the participants in the IG with respect to their receptive and productive vocabularies in English. Three different types of measure are used. Section 5.2.1.1. presents the results of the MBCDI **total scores**, while sections 5.2.1.2. and 5.2.1.3. present descriptive information on the MBCDI **semantic** and **word-class categories**, respectively.

5.2.1.1. MBCDI TOTAL SCORES

The scores for the lexical receptive measures summarised in Table 29 revealed that the means for *PhsU* and *WdsU* increased steadily from T1 to T2 and from T2 to T3. For the vocabulary production measures (*WdsP* and *SC*), we found considerable growth from T1 to T2 and from T2 to T3. Regarding the *M3L* scores, we detected an important increase from T1 to T2 but little change from T2 to T3.

Because most of the variables proved to have a normal distribution (except *M3L*), we performed a one-way RM ANOVA with *Time* (T1, T2, T3) as the within-subjects factor. The Greenhouse-Geisser correction was applied for all the variables which did not meet the sphericity condition (the ones in shaded cells in Table 29). Bonferroni pairwise comparisons were used to detect whether significant changes had occurred between T1/T2, T2/T3 and/or T1/T3. For the variable that did not meet the normality assumption (*M3L*), we also conducted the non-parametric Friedman test, followed by Wilcoxon signed-rank test in order to detect whether significant changes occurred between T1/T2, T2/T3 and/or T1/T3.

The results showed an overall significant effect of *Time* on the lexical reception measures. The largest main effect was found for *PhsU* ($\eta^2p = .58$). As summarised in Table 29, the Bonferroni pairwise comparisons revealed significant differences between T1/T2, T2/T3, and T1/T3 in all the measures related to receptive lexical knowledge. With respect to lexical production, the one-way RM ANOVA also showed an overall effect for *Time* in all the measures. The largest effect size was found for *WdsP* ($\eta^2p = .91$) and *SC* ($\eta^2p = .90$). Similar to reception, the pairwise

comparisons showed significant differences between T1/T2, T2/T3 and T1/T3 for all measures. For the variable *M3L*, which presented a non-normal distribution, the Friedman Test confirmed the parametric results. Thus, an overall *Time* effect was found: *M3L* [$\chi^2(2) = .87$, $p = .000$]. The Wilcoxon signed-rank test also confirmed the significant differences. The results were as follows: *M3L*: [T1/T2 ($Z = -6.181$, $p = .000$); T1/T3 ($Z = -6.255$, $p = .000$); T2/T3 ($Z = -4.284$, $p = .000$).

Table 29: Descriptive statistics and one-way RM ANOVA results for the MBCDI total scores measures with Time (T) as the within-subjects factor for the participants in the IG (n= 51), with $p \leq .01$.

Measures		Mean (<i>Sd.</i>)			One-way RM ANOVA						
		T1	T2	T3	F	<i>p</i>	η^2p	Power	T1/T2	T2/T3	T1/T3
<i>Receptive vocabulary</i>	PhsU	10.65 (5.78)	12.92 (5.21)	17.61 (3.69)	T: F(1.15-57.8)=70.370	.000	.58	1	.000	.000	.000
	WdsU	90.12 (21.15)	106.41 (31.98)	123.71 (17.60)	T: F(2-100)=26.635	.000	.34	1	.000	.000	.000
<i>Productive vocabulary</i>	WdsP	31.94 (16.39)	75.61 (18.34)	146.35 (33.94)	T: F(1.23-61.80)=543.887	.000	.91	1	.000	.000	.000
	M3L*	1.32 (1.02)	3.12 (.73)	3.83 (1.06)	T: F(2-100)=177.951	.000	.78	1	.000	.000	.000
	SC	.00 (.00)	10.31 (2.42)	16.59 (4.74)	T: F(1.41-11.17)=452.214	.000	.90	1	.000	.000	.000

Note: *Normality assumption violated. Greenhouse-Geisser correction applied for variables in shaded cells.

5.2.1.2. MBCDI SEMANTIC CATEGORIES

In this section, we focus on the lexical development of the learners in the IG in relation to the MBCDI semantic categories across time. We start by offering descriptive results and then proceed to describe any significant changes across years.

As regards receptive knowledge, out of the 19 semantic categories summarised in Table 30, we found a steady increase of the mean scores earned from T1 to T3 for twelve of them (*Sounds, Toys, Clothes, Household, People, Games, Actions, Qualities, Pronouns, Questions, Locations, Quantity*). For the measures that did not follow this tendency, five different trends can be found. First, we found no changes between T1 and T2 for the lexical category of *Animals*. Nevertheless, the scores did increase from T2 to T3. Secondly, we saw a descending trend from T1 to T2 followed by an improvement from T2 to T3 for words related to *Furniture*. Note that for this category the score earned at T3 is lower than the score earned at T1. Thirdly, we noticed a decrease from T1 to T3 for the semantic category of *Places*. Fourthly, we found an improvement from T1 to T2 followed by a decline from T2 to T3 for the categories of *Vehicles, Body* and *Games*. The values found at T3 were lower than those at T2 for the first two categories (*Vehicles* and *Body*), while for the *Games* category the T3 score was higher than the T1 score. Finally, the means improved from T1 to T2 but showed no changes from T2 to T3 for the semantic category of *Time*. Regarding the results for productive use, the means displayed in Table 30 show a unique T1-T3 upward trend for all 19 measures. It was also found that in all data collection times, some of the standard deviations are high. We will try to account for this finding in the next chapter, where we discuss the results.

When comparing the receptive and productive vocabularies at the different data collection times (Table 30), we noticed that at T1 the results for receptive knowledge were higher than productive knowledge. The same was found at T2, as the means of the receptive vocabulary measures continued to be higher for all categories except for *Games*. Changes occurred at T3, where the means on production were higher than those of reception for all measures except the lexical categories of *Clothes, Descriptives, Questions* and *Quantity*. However, this does not mean that the children's receptive vocabulary at T3 was inferior to their productive vocabulary, as

explained above. It should be noted that because of the format of the MBCDI test (explained in Chapter 4, above), the receptive knowledge score is also part of the lexical production score.

Next we will report on the results from the statistical analysis. Given that all the variables on semantic categories presented a non-normal distribution in at least one data collection time, Friedman non-parametric tests were conducted. We also performed the Wilcoxon signed-rank tests in order to check where specifically the statistically significant differences were found. The results are summarised in Table 30.

Regarding receptive knowledge, the Friedman test indicated an overall significance for the majority of measures except the semantic categories of *Furniture*, *Places*, *Games* and *Time*. The Wilcoxon signed-rank test indicated three distinct significance patterns: the first showed progress between all data collection times for the lexical fields of *Clothes*, *People*, *Toys* and *Quantity*. The second pattern showed that the IG improved significantly between T1/T2 and T1/T3 (but not T2/T3) in words referring to *Sounds*, *Food*, *Actions* and *Questions*. Finally, the third pattern showed statistical differences between T2/T3 and T1/T3 (but not T1/T2) for *Animals*, *Vehicles*, *Body*, *Household*, *Qualities*, *Pronouns* and *Locations*.

When we examine lexical production, the Friedman test indicated an overall significant effect for *Time* in all the measures. The pairwise comparisons according to the Wilcoxon signed-rank test showed that the results were statistically significant for almost all the measures between all the data collection times, thus indicating that in the IG, students' productive lexical knowledge progressively improved. Nevertheless, for the lexical categories of *Questions* and *Quantity*, the participants showed signs of relevant improvement especially between T1/T2 and T1/T3 (not between T2/T3).

Table 30: Descriptive results and Friedman Test results for the MBCDI semantic categories corresponding IG (n=51) with $p \leq .01$. (R: receptive vocabulary, P: Productive vocabulary).

Measures		Mean (<i>Sd.</i>)			Friedman Test				
		T1	T2	T3	Chi-square	<i>p</i>	T1/T2	T2/T3	T1/T3
Sounds	R.	3.84 (1.71)	4.67 (1.47)	5.02 (2.59)	14.319	.001	.001	.315	.017
	P.	1.80 (1.56)	3.39 (1.20)	5.61 (3.35)	57.330	.000	.000	.000	.000
Animals	R.	9.49 (2.82)	9.49 (3.48)	12.63 (2.40)	37.905	.000	.784	.000	.000
	P.	3.82 (3.33)	7.08 (3.64)	13.71 (3.12)	86.616	.000	.000	.000	.000
Vehicles	R.	2.49 (1.12)	2.73 (1.20)	1.63 (1.18)	14.217	.001	.143	.000	.001
	P.	.82 (.88)	2.43 (1.13)	3.43 (1.41)	79.774	.000	.000	.000	.000
Toys	R.	1.67 (1.46)	2.27 (1.53)	3.12 (1.79)	20.772	.000	.005	.023	.000
	P.	.80 (1.29)	2.25 (1.32)	3.35 (2.57)	67.185	.000	.000	.000	.000
Food	R.	5.75 (2.46)	7.51 (3.17)	7.94 (2.52)	14.663	.001	.002	.267	.000
	P.	4.31 (2.25)	6.92 (2.62)	11.45 (2.59)	88.294	.000	.000	.000	.000
Clothes	R.	2.47 (1.78)	3.51 (2.00)	6.69 (1.65)	65.978	.000	.001	.000	.000
	P.	.53 (.83)	2.51 (1.48)	4.75 (1.77)	88.294	.000	.000	.000	.000
Body	U.	6.57 (2.69)	7.57 (3.31)	5.33 (1.21)	11.191	.004	.056	.000	.010
	P.	2.53 (1.98)	5.04 (2.34)	11.47 (2.60)	100.567	.000	.000	.000	.000
Furniture	R.	4.80 (1.51)	4.45 (1.51)	4.61 (2.43)	1.401	.496	.139	.679	.622
	P.	.59 (.94)	3.49 (1.71)	5.57 (2.35)	85.672	.000	.000	.000	.000

Measures		Mean (<i>Sd.</i>)			Friedman Test				
		T1	T2	T3	Chi-square	<i>p</i>	T1/T2	T2/T3	T1/T3
Household	R.	7.59 (2.87)	8.33 (3.24)	9.57 (1.60)	12.734	.002	.276	.010	.000
	P.	1.18 (1.58)	4.25 (1.89)	10.43 (2.96)	101.507	.000	.000	.000	.000
Places	R.	8.33 (3.39)	8.04 (3.00)	7.18 (1.59)	2.221	.329	.277	.136	.108
	P.	3.92 (2.49)	6.84 (2.69)	11.94 (3.22)	87.364	.000	.000	.000	.000
People	R.	3.78 (1.46)	6.43 (3.42)	7.96 (1.01)	47.409	.000	.000	.003	.000
	P.	2.35 (1.05)	5.00 (1.88)	9.25 (2.04)	98.517	.000	.000	.000	.000
Games	R.	4.12 (1.72)	4.43 (1.25)	4.33 (1.62)	2.341	.310	.174	.402	.493
	P.	3.96 (2.09)	5.37 (1.83)	8.94 (2.04)	84.746	.000	.000	.000	.006
Actions	R.	8.75 (4.32)	15.61 (6.26)	15.90 (3.75)	50.643	.000	.000	.972	.000
	P.	1.69 (1.47)	4.65 (1.74)	18.92 (5.39)	100.567	.000	.000	.000	.006
Time	R.	2.65 (.84)	2.73 (1.07)	2.73 (.87)	.357	.837	.600	1.000	.712
	P.	.35 (.62)	2.16 (.80)	3.75 (1.18)	93.536	.000	.000	.000	.000
Qualities	R.	6.59 (3.11)	7.43 (4.09)	14.04 (2.75)	54.031	.000	.307	.000	.000
	P.	3.20 (1.86)	5.37 (1.84)	10.84 (3.80)	90.163	.000	.000	.000	.000
Pronouns	R.	2.51 (1.30)	2.86 (1.37)	3.88 (1.16)	28.048	.000	.079	.000	.000
	P.	.16 (.50)	2.22 (1.02)	4.63 (1.75)	97.283	.000	.000	.000	.000
Questions	R.	2.45 (.73)	2.86 (1.11)	3.06 (.78)	12.413	.002	.013	.164	.000

Measures					Friedman Test				
Mean (<i>Sd.</i>)									
T1 T2 T3					Chi-square	<i>p</i>	T1/T2	T2/T3	T1/T3
<i>Locations</i>	P.	.16 (.50)	1.88 (.88)	2.06 (1.06)	76.837	.000	.000	.481	.000
	R.	3.24 (1.01)	3.29 (1.46)	4.06 (1.10)	18.335	.000	.971	.004	.000
	P.	.45 (.82)	2.37 (.95)	4.73 (2.08)	88.096	.000	.000	.000	.000
<i>Quantity</i>	R.	1.59 (.98)	2.63 (1.66)	4.00 (1.09)	70.486	.000	.000	.000	.000
	P.	.00 (.00)	1.75 (.93)	1.78 (.57)	84.746	.000	.000	.913	.000

Table 30: Cont.

5.2.1.3. MBCDI WORD-CLASS MEASURES

This section addresses the lexical development of the IG with respect to the MBCDI **word-class measures** understood and produced over time (T1, T2 and T3). As in the previous section, the results for the receptive and productive measures are first presented descriptively, and then the results from statistical analyses are reported.

As summarised in Table 31, the mean scores earned by the IG in all the receptive measures showed an increase from T1 to T2 and from T2 to T3. However, this upward trend shows different variations. More precisely, *Fw*, *N*, *V*, *Adv*, *Pron* and *Interj* increased steadily across time, while *Cw*, *Adj* and *Prep* experienced a higher increase from T2 to T3. In the case of production, an upswing across time was seen for all measures, but different patterns were found. Thus, the totals of *Fw*, *N*, *Adv*, and *Prep* increased steadily over time, while the mean scores of *Cw*, *V*, *Adj* and *Interj* increased more from T2 to T3. Comparing receptive and productive performance, we found that the scores earned on all measures of receptive vocabulary at T1 and T2 were higher than those on the measures of lexical production. Nevertheless, at T3, the oral production scores were higher than the receptive vocabulary scores with respect to the totals of *Cw*, *Fw*, *N*, *Adj*, *Pron*, *Prep* and *Interj*. As mentioned above,

it is important to recall that receptive vocabulary knowledge is also part of the productive score on the test. With respect to the standard deviations, some are particularly high for oral productive measures at T1 and T2.

As most of the variables violated the normality assumption, Friedman tests were carried out. We also conducted Wilcoxon signed-rank tests with pairwise comparisons to further assess where significant differences were found. For the variable that presented a normal distribution at all data collection times (marked with * in Table 31), a one-way ANOVA was also conducted to further check the findings.

Regarding the receptive vocabulary, the Friedman test indicated an overall time effect in all cases. Exploring this outcome more in depth, the Wilcoxon signed-rank test pointed to three distinct patterns of significance. Firstly, we noticed a statistical difference between T1/T2, T2/T3 and T1/T3 for *Cw*, *Fw*, *Adv*, and *Pron*. Secondly, we found significant differences between T1/T2 and T1/T3 (but not T2/T3) for *V* and *Interj*. Thirdly, significant variations for *N*, *Adj* and *Prep* were found between T2/T3 and T1/T3 (but not T1/T2).

Regarding production, the Friedman test also showed an overall effect for time in all measures. Next, the Wilcoxon signed-rank test indicated significant differences between T1/T2, T2/T3 and T1/T3 for all the measures. Regarding *Cw*, which showed a normal distribution at all three data collection times, the results of the Paired-Samples T-Test conducted showed borderline significant differences between T1/T2 and T2/T3: T1/T2: [$t(65) = -18.24$, $p = .000$, $\eta^2 p = 0.42$], T2/T3: [$t(65) = -22.63$, $p = .000$, $\eta^2 p = 0.47$], T1/T3 [$t(65) = -27.86$, $p = .000$, $\eta^2 p = 0.52$]

Table 31: Descriptive statistics and Friedman test results of the MBCDI word-class measures for the participants in the IG (n=51), with $p \leq .01$, (R: receptive vocabulary, P: productive vocabulary).

Measures					Friedman Test				
		Mean (<i>Sd.</i>)			Chi-square	<i>p</i>	T1/T2	T2/T3	T1/T3
		T1	T2	T3					
Cw	R.	73.25	83.43	100.08	27.517	.000	.007	.003	.000
		(19.06)	(27.56)	(14.03)					
	P.*	27.06	60.27	121.00					
		(14.61)	(15.63)	(26.78)					
Fw	R.	16.71	19.96	23.63	25.270	.000	.000	.001	.000
		(3.83)	(5.23)	(4.67)					
	P.	5.39	14.94	25.35					
		(3.92)	(4.59)	(7.90)					
N	R.	58.35	64.12	70.63	13.488	.001	.172	.032	.000
		(14.95)	(20.73)	(10.10)					
	P.	21.45	50.57	91.92					
		(13.46)	(13.65)	(20.09)					
V	R.	9.84	16.51	17.06	47.370	.000	.000	.857	.000
		(4.31)	(6.30)	(3.36)					
	P.	2.59	5.71	20.55					
		(1.51)	(1.95)	(5.36)					
Adj	R.	6.59	7.43	14.04	54.031	.000	.307	.000	.000
		(3.11)	(4.09)	(2.75)					
	P.	3.20	5.37	10.84					
		(1.86)	(1.84)	(3.80)					
Adv	R.	2.24	3.57	4.78	53.929	.000	.000	.000	.000
		(.55)	(1.25)	(1.31)					
	P.	.55	2.67	3.45					
		(.61)	(1.16)	(.90)					
Pron	R	4.98	5.73	6.94	26.793	.000	.023	.003	.000
		(1.58)	(2.13)	(1.69)					
	P.	.31	4.10	6.61					
		(.99)	(1.64)	(2.30)					
Prep	R.	3.24	3.29	4.06	18.335	.000	.971	.004	.000
		(1.01)	(1.46)	(1.10)					
	P.	.45	2.37	4.73					
		(.82)	(.95)	(2.08)					

Measures					Friedman Test				
		Mean (<i>Sd.</i>)							
		T1	T2	T3	Chi-square	<i>p</i>	T1/T2	T2/T3	T1/T3
Interj	R.	4.76	5.76	6.20	14.489	.001	.001	.270	.009
		(1.92)	(1.40)	(2.84)					
	P.	3.14	4.82	8.25	71.947	.000	.000	.000	.000
		(2.13)	(1.57)	(3.84)					

Note: *Normality assumption violated.

Table 31: Cont.

Table 32: Summary of the significant differences registered at different data collection times for the semantic categories and word-class measures.

Significant differences	Semantic categories	Word-class measures
T1-T2-T3	Furniture Places Games Time	
T1-T2 < T3	Animals Vehicles Body Household Qualities Pronouns Locations	
T1 < T2-T3	Sounds Food Actions Questions	
T1 < T2 < T3	Clothes People Toys	Cw Fw Pron
T1 < T2-T3 > T1		V Interj
T1-T2 < T3 > T1		N Adj Prep

5.2.2. COMPARATIVE RESULTS OF RECEPTIVE AND PRODUCTIVE VOCABULARY DEVELOPMENT IN IMMERSION AND REGULAR SCHOOL SETTINGS

In this section we address RQ1a, which was formulated as follows:

RQ1a. How does the lexical development in English of children learning an FL in an immersion school compare to that of children attending a regular school?

This section is divided into three parts. Section 5.2.2.1. presents the results from the IG and RG1 at T1, T2 and T3 with respect to the MBCDI **total scores**. In section 5.2.2.2., we provide a comparison of the lexical performance of the two groups across time regarding the MBCDI **semantic categories**. Similarly, section 5.2.2.3. focuses on the longitudinal results of both groups with respect to the MBCDI **word-class measures**.

5.2.2.1. MBCDI TOTAL SCORES

This section examines the lexical development of the IG by comparing it to that of the RG1 group across the three data collection times with respect to the MBCDI **total scores**. As in the previous sections, we start by providing the descriptive results of our data and then present the results of the statistical analyses. As summarised in Table 33, the mean scores for the vocabulary receptive measures show that the IG scored above the RG1 control group at all data collection times. Results for productive knowledge were similar.

The normality analysis conducted (see Appendix 7, Table 47) showed that most of the variables followed a normal distribution. Therefore, a 3x2 Repeated Measures ANOVA was conducted with *Time (T)* as the within-subjects factor and *School (S)* as the between-subjects factor. Sphericity was also assumed. For the only variable which did not meet the normality assumption (marked with * in Table 33), in the absence of the corresponding 3x2 Repeated Measures ANOVA non-parametric counterpart, we also performed a Mann-Whitney *U*-Test at each data collection time.

Regarding receptive vocabulary, results indicate an overall time effect for all measures, with the largest effect size found for *PhsU* ($\eta^2p = .55$). A large school effect in favour of the IG was also found for all measures, with the largest effect size being registered for *WdsU* ($\eta^2p = .60$). The results for the vocabulary production measures follow the same pattern. Thus, a large overall time effect was found for all measures, with the largest size effect for *WdsP* ($\eta^2p = .87$). There is also a school effect in all measures of lexical production in favour of the IG (the largest effect size was found in *SC*: $\eta^2p = .63$). These differences between groups are summarised in Table 33 and graphically illustrated in the boxplots in Figure 1. As for the *M3L*, which showed a non-normal distribution, the results of the Mann-Whitney *U*-Test confirmed the previous outcomes. The mean ranks of the IG were significantly higher than those of the RG1. Therefore, the statistical results for all three data collection times are summarised as follows: T1 ($Z = -4.200, p = .000$), T2 ($Z = -5.871, p = .000$), T3 ($Z = -4.867, p = .000$).

Table 33: Descriptive results and the 3x2 RM ANOVA with Time (T) as the within-subjects factor and School (S) as the between-subjects factor for the participants in the IG (n=51) and RG1 (n=15), with $p \leq .01$. Sphericity assumed.

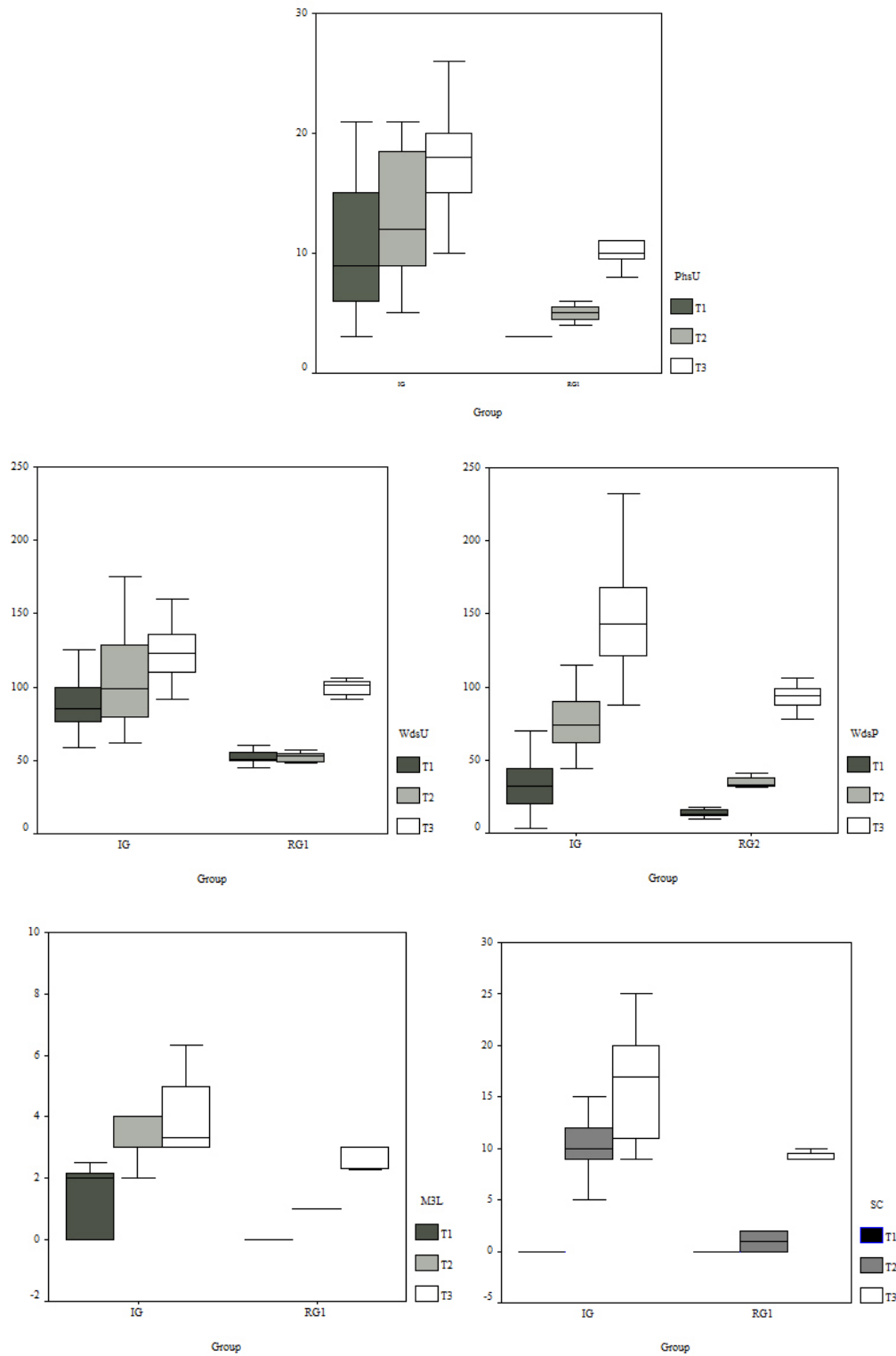
Measures		Mean (<i>Sd.</i>)						3x2 RM ANOVA			
		T1		T2		T3		F	<i>p</i>	$\eta^2 p$	Power
		IG	RG1	IG	RG1	IG	RG1				
Receptive vocabulary	PhsU	10.65	3.13	12.92	4.93	17.61	10.00	T: F (2-128) = 80.538	.000	.55	1
		(5.78)	(.51)	(5.21)	(.88)	(3.69)	(.92)	S: F (1-64) = 47.196	.000	.42	1
	WdsU	90.12	51.13	106.41	52.53	123.71	99.47	T: F (2-128) = 49.172	.000	.43	1
		(21.15)	(13.26)	(31.98)	(4.24)	(17.60)	(5.02)	S: F (1-64) = 94.152	.000	.60	1
Productive vocabulary	WdsP	31.94	14.13	75.61	34.60	146.35	92.80	T: F (2-128) = 446.800	.000	.87	1
		(16.39)	(3.88)	(18.34)	(3.39)	(33.94)	(.44)	S: F (1-64) = 54.643	.000	.46	1
	M3L*	1.32	.00	3.11	1.00	3.82	2.67	T: F (2-128) = 194.044	.000	.75	1
		(1.02)	(.00)	(.73)	(.53)	(1.06)	(.35)	S: F (1-64) = 57.507	.000	.48	1
	SC	.00	.00	10.31	.87	16.59	9.13	T: F (2-128) = 306.712	.000	.82	1
		(.00)	(.00)	(2.42)	(.91)	(4.74)	(.83)	S: F(1-64) = 111.039	.000	.63	1

Note: *Normality assumption violated.

Fig. 1: MBCDI total scores boxplots for each group and time.

From top to bottom and left to right:

- 1) Words Understood, 2) Phrases Understood, 3) Words Produced, 4) M3L, 5) SC.



5.2.2.2. THE MBCDI SEMANTIC CATEGORIES

In this section, the two groups (immersion group, IG, and regular school group, RG1) are compared at all three data collection times with respect to the MBCDI **semantic categories**. First, we present the descriptive results at each testing time and then the outcomes of the statistical tests.

The mean scores regarding the semantic categories understood (receptive knowledge) show that the IG scored higher than the RG1 on almost all measures and at all data collection times. Exceptions are the category of *Games* at T1 and *Places* and *Quantity* at T3. A closer look at the means at T1, T2 and T3 show that at a descriptive level, in some cases the differences between the two groups were small. Thus, at T1 the IG was slightly higher than the RG1 group in the lexical categories of *Toys*, *People* and *Games*. At T2, small differences between groups were found for the lexical categories of *Sounds* and *Toys*, with the IG performing slightly above the RG1 group. At T3, however, we noted minor variations for the lexical category of *Quantity*, but this time in favour of the RG1 group.

Regarding the results of the semantic categories produced (productive knowledge), the means show that overall the IG performed better than the RG1 group. A closer look at T1, T2 and T3 reveals that for some measures the differences between the groups were small, as well. In fact, at T1 these minor variations (in favour of the IG) were generalised to eleven out of the total of 19 measures in the section. At T2, the IG's scores improved considerably compared to those of the RG1 control group, and a small difference between the groups was noticed solely for the lexical category of *Vehicles*. At T3, small distinctions between groups were found for words for *Quantity*, also in favour of the IG.

The Kolmogorov-Smirnov and Shapiro-Wilk normality tests showed a violation of the normality assumption for all variables. As a result, non-parametric Mann-Whitney *U*-Tests were performed at each data collection times. As shown in Table 34, the results for the measures of vocabulary reception at T1 show significant differences in favour of the IG in almost all measures except for *Toys*, *Clothes*, *People* and *Games*. In terms of oral production, statistical differences were also found in favour of the IG. Therefore, the immersion learners produced significantly

more words in the category of *Vehicles*, *Toys*, *Food*, *Places* *People*, *Games*, *Actions*, *Qualities* and *Locations*. No further significant differences between the two groups were found. At T2, differences were also found to be significant in favour of the IG in all lexical reception measures except for *Sounds* and *Toys*. The scores for all measures of lexical production were significantly higher in the IG as well. The analysis performed at T3 showed that the IG earned significantly better results in almost all lexical reception measures except for *Sounds*, *Toys*, *Food*, *Furniture*, *Games* and *Quantity*. With respect to the oral production measures, we found statistical differences favouring the IG in all measures except for *Sounds*, *Toys*, *Questions* and *Locations*.

Therefore, significant differences were always found to be in favour of the IG. Differences at all data collection times were found in the following semantic classes with respect to receptive vocabulary: *Animals*, *Vehicles*, *Body*, *Household*, *Places*, *Actions*, *Time*, *Qualities*, *Pronouns*, *Questions* and *Locations*. For the lexical categories of *Food*, *Furniture* and *Quantity*, the IG performed better only at T1 and T2 (but not at T3). Differences at T2 and T3 (but not T1) were noted for the semantic category of *Clothes* and *People*. Statistical variation only at T1 (but not T2 and T3) was found for the category of *Sounds*, while variation at T2 (but not T1 and T3) was found for *Games*. No statistical differences at T1, T2 and T3 were found for *Toys*.

Regarding productive use, the significant differences between the groups were always found at T1, T2 and T3 for the following variables: *Vehicles*, *Food*, *Places*, *People*, *Games*, *Actions*, *Descriptives* and *Locations*. At T1 and T2 (but not T3), the IG only significantly outperformed the RG1 group in the category of *Toys*. Differences were found at T2 and T3 (but not T1) for *Animals*, *Clothes*, *Body*, *Furniture*, *Household*, *Time*, *Pronouns* and *Quantity*. At T2 (but not T1 and T3) significant differences were found for the categories *Sounds* and *Questions*.

Table 34: Descriptive statistics and results of the Mann-Whitney *U* Tests for the MBCDI semantic categories for IG (n=51) and RG1 (n=15) at T1, T2, and T3 (R: Receptive vocabulary, P: Productive vocabulary).

Measures		T1				T2				T3			
		IG	RG1	Z	<i>p</i>	IG	RG1	Z	<i>p</i>	IG	RG1	Z	<i>p</i>
		M (<i>Sd.</i>)	M (<i>Sd.</i>)			M (<i>Sd.</i>)	M (<i>Sd.</i>)			M (<i>Sd.</i>)	M (<i>Sd.</i>)		
Sounds	R	3.84 (1.71)	2.53 (.64)	-3.282	.001	4.67 (1.47)	4.53 (.99)	-.143	.887	5.02 (2.59)	5.40 (.91)	-.078	.938
	P	1.80 (1.56)	1.40 (.82)	-.850	.396	3.39 (1.20)	2.53 (.64)	-2.722	.006	5.61 (3.33)	4.73 (.79)	-1.337	.181
Animals	R	9.49 (2.82)	4.73 (1.22)	-4.746	.000	9.49 (3.48)	3.87 (1.45)	-5.160	.000	12.63 (2.40)	10.67 (1.17)	-2.823	.005
	P	3.82 (3.33)	2.20 (1.04)	-1.731	.083	7.08 (3.64)	3.27 (1.03)	-4.267	.000	13.71 (3.12)	10.80 (1.47)	-3.493	.000
Vehicles	R	2.49 (1.12)	1.33 (.72)	-3.699	.000	2.73 (1.20)	1.87 (.64)	-2.572	.010	1.63 (1.18)	.87 (.51)	-2.350	.019
	P	.82 (.88)	.13 (.51)	-2.940	.003	2.43 (1.13)	1.07 (.25)	-4.586	.000	3.43 (1.41)	2.53 (.83)	-2.395	.017
Toys	R	1.67 (1.46)	1.07 (.70)	-1.342	.180	2.27 (1.53)	2.07 (.70)	-.345	.730	3.12 (1.79)	3.33 (.90)	-1.125	.261
	P	.80 (1.29)	.00 (.00)	-3.123	.002	2.25 (1.32)	1.27 (.79)	-2.740	.006	3.35 (2.57)	2.47 (.74)	-.845	.398
Food	R	5.75 (2.46)	4.60 (2.46)	-2.381	.017	7.51 (3.17)	4.20 (.94)	-3.909	.000	7.94 (2.52)	8.40 (.73)	-1.220	.223

Measures		T1				T2				T3			
		IG	RG1	Z	p	IG	RG1	Z	p	IG	RG1	Z	p
		M (Sd.)	M (Sd.)			M (Sd.)	M (Sd.)			M (Sd.)	M (Sd.)		
	P	4.31 (2.25)	1.13 (.64)	-4.346	.000	6.92 (2.62)	2.07 (.79)	-5.363	.000	11.45 (2.50)	9.53 (1.35)	-2.667	.008
Clothes	R	2.47 (1.78)	1.93 (.88)	-1.103	.270	3.51 (2.04)	1.60 (.73)	-3.701	.000	6.69 (1.65)	4.20 (.86)	-4.540	.000
	P	.53 (.83)	.27 (.59)	-1.136	.256	2.51 (1.48)	1.40 (.63)	-2.854	.004	4.75 (1.77)	3.07 (.79)	-3.645	.000
Body	R	6.57 (2.69)	3.93 (.70)	-4.491	.000	7.57 (3.31)	2.53 (.51)	-5.702	.000	5.33 (1.21)	7.73 (.59)	-5.437	.000
	P	2.53 (1.98)	1.73 (.88)	-1.408	.159	5.04 (2.39)	3.40 (1.12)	-2.490	.013	11.47 (2.60)	6.27 (.96)	-5.740	.000
Furniture	R	4.80 (1.51)	3.20 (1.37)	-3.481	.001	4.45 (1.51)	2.87 (1.12)	-3.648	.000	4.61 (2.43)	3.33 (1.34)	-.877	.381
	P	.59 (.92)	.20 (.41)	-1.313	.189	3.49 (1.71)	1.67 (.58)	-4.227	.000	5.57 (2.35)	4.20 (1.52)	-1.972	.049
Household	R	7.59 (2.87)	3.87 (.64)	-4.818	.000	8.33 (3.24)	3.87 (.74)	-5.104	.000	9.57 (1.60)	8.40 (.91)	-3.479	.001
	P	1.18 (1.58)	.47 (.51)	-.899	.368	4.25 (1.89)	1.60 (.50)	-4.822	.000	10.43 (2.96)	6.40 (1.18)	-4.890	.000
Places	R	8.33 (3.39)	4.60 (.98)	-4.369	.000	8.04 (3.00)	4.20 (.67)	-4.833	.000	7.18 (1.59)	8.93 (1.22)	-3.718	.000

Measures		T1				T2				T3			
		IG	RG1	Z	p	IG	RG1	Z	p	IG	RG1	Z	p
		M (Sd.)	M (Sd.)			M (Sd.)	M (Sd.)			M (Sd.)	M (Sd.)		
People	P	3.92 (2.49)	1.60 (1.18)	-3.693	.000	6.84 (2.69)	2.33 (1.54)	-5.130	.000	11.94 (3.22)	8.27 (.88)	-4.928	.000
	R	3.78 (1.46)	3.53 (.99)	-.695	.487	6.43 (3.44)	2.47 (1.06)	-4.669	.000	7.96 (1.01)	5.60 (.73)	-5.648	.000
	P	2.35 (1.05)	1.27 (.70)	-3.528	.000	5.00 (1.88)	2.13 (1.06)	-4.949	.000	9.25 (2.04)	5.07 (1.03)	-5.733	.000
Games	R	4.12 (1.72)	4.33 (.97)	-.836	.403	4.43 (1.25)	3.53 (1.18)	-2.460	.014	4.33 (1.62)	3.93 (.25)	-.148	.883
	P	3.96 (2.09)	1.00 (.75)	-4.760	.000	5.37 (1.83)	2.27 (.88)	-5.185	.000	8.94 (2.04)	6.33 (1.71)	-4.163	.000
Actions	R	8.75 (4.32)	4.67 (.61)	-4.409	.000	15.61 (6.26)	6.53 (1.40)	-4.755	.000	15.90 (3.75)	11.87 (3.88)	-3.615	.000
	P	1.69 (1.47)	.00 (.00)	-4.268	.000	4.65 (1.74)	2.20 (1.26)	-4.375	.000	18.92 (5.39)	7.20 (1.65)	-5.625	.000
Time	R	2.65 (.84)	1.33 (.61)	-4.658	.000	2.73 (1.07)	1.60 (.50)	-3.741	.000	2.73 (.87)	1.93 (.88)	-3.126	.003
	P	.35 (.62)	.07 (.25)	-1.707	.088	2.16 (.80)	1.27 (.45)	-3.778	.000	3.75 (1.18)	.60 (.73)	-5.863	.000
Qualities	R	6.59 (3.11)	3.33 (2.07)	-4.366	.000	7.43 (4.09)	3.20 (.86)	-4.239	.000	14.40 (2.75)	6.40 (.82)	-5.887	.000

Measures		T1				T2				T3			
		IG	RG1	Z	p	IG	RG1	Z	p	IG	RG1	Z	p
		M (Sd.)	M (Sd.)			M (Sd.)	M (Sd.)			M (Sd.)	M (Sd.)		
	P	3.20 (1.86)	2.07 (.25)	-2.499	.012	5.37 (1.84)	2.60 (.73)	-4.769	.000	10.84 (3.80)	5.73 (1.16)	-4.950	.000
Pronouns	R	2.51 (1.30)	.80 (.56)	-4.391	.000	2.86 (1.37)	1.47 (.51)	-3.809	.000	3.88 (1.16)	1.93 (.25)	-5.453	.000
	P	.16 (.50)	.00 (.00)	-1.251	.211	2.22 (1.02)	.87 (.35)	-5.035	.000	4.63 (1.75)	3.40 (1.45)	-2.173	.030
Questions	R	2.45 (.73)	1.53 (.64)	-4.070	.000	2.86 (1.11)	1.40 (.50)	-4.546	.000	3.06 (.78)	1.47 (.74)	-5.372	.000
	P	.16 (.50)	.00 (.00)	-1.251	.211	1.88 (.88)	.93 (.59)	-3.534	.000	2.06 (1.06)	1.67 (.72)	-1.006	.314
Locations	R	3.24 (1.01)	.80 (.41)	-5.735	.000	3.29 (1.46)	.67 (.61)	-5.523	.000	4.06 (1.10)	2.00 (.65)	-5.189	.000
	P	3.24 (1.01)	.80 (.41)	-5.735	.000	3.29 (1.46)	.67 (.61)	-5.523	.000	4.06 (1.10)	2.00 (.65)	-5.189	.000
Quantity	R	1.58 (.98)	.00 (.00)	-5.693	.000	2.63 (1.16)	.47 (.51)	-5.491	.000	4.00 (1.09)	4.27 (.79)	-1.248	.212
	P	.00 (.00)	.00 (.00)	.000	1.000	1.75 (.93)	.87 (.35)	-3.537	.000	1.78 (.57)	1.13 (.83)	-3.072	.002

Table 34: Cont.

5.2.2.3. MBCDI WORD-CLASS MEASURES

In this section, we start by examining the differences between the two groups in terms of grammatical categories from a descriptive perspective. Next, we proceed to present the results of the statistical tests. As indicated in Table 34, at T1 the descriptive results are more favourable for the IG in both receptive and productive measures. More specifically, with respect to receptive vocabulary at T1, the differences between the groups were large for *Cw*, *Fw*, *N* and *V*. For the remaining word-class measures, the differences were minor. A similar situation was recorded for lexical production, where large variations were noticed for *Cw*, *Fw* and *N*, while the differences were small for *Adj*, *Adv*, *Pron*, *Prep* and *Interj*. At T2, the results continued to be in favour of the IG; the differences were large for the majority of the lexical reception measures, with the exception of *Interj*. In terms of oral production, we also found minor differences for *Interj*, while the differences continued to be substantial for the remaining variables. At T3, at the descriptive level, we found important differences between the two groups in lexical reception (always in favour of the IG), except for the category of *Adv*, where the variations in scores were minor. With respect to the oral production measures, the differences between groups were large with the exception of *Adv*, *Pron*, *Prep* and *Interj*.

Because the Kolmogorov-Smirnov and the Saphiro-Wilk analyses showed that the data were not normally distributed for the majority of the variables (at least in one of the data collection times), they were treated non-parametrically. To this end, Mann-Whitney *U*-Tests were conducted at each data collection time in order to analyse the groups' performance regarding the lexical reception and production measures. The data which were shown to present a normal distribution (marked with an asterisk * in Table 35) were also analysed parametrically by means of independent samples *t*-tests.

With respect to the lexical reception measures, we found significant differences at T1 in favour of the IG in all measures except for *Interj*, where the difference between the two groups was not significant. The *t*-test conducted for *Fw*, which showed a normal distribution at T1 (marked with an * in Table 35), confirms the results of the non-parametric test. More precisely, the results [$t(64) = 6.99, p = .000$]

point to a significant difference favouring the IG. In the case of the oral production measures, the statistical results were in favour of the IG in all cases, except for *Pron* and *Interj*, where the differences were not significant. The parametric *t*-test carried out for *Cw* and *N*, which showed a normal distribution (marked with * in Table 34), confirmed the results of the non-parametric test. The results are as follows: *Cw*: [$t(63.42)=6.76, p=.000, \eta^2p = 0.41$], *N*: [$t(63.65)=5.92, p=.000, \eta^2p = 0.35$].

Looking at the T2 data for the reception measures, the situation is similar to what was found at T1: the IG earned significantly better results in all measures except for *Interj*. The *t*-test was also performed for *F.*, which showed a normal distribution (marked with * in Table 35), and the results are in line with the non-parametric test [$t(40.30)=7.26, p=.000, \eta^2p = 0.45$]. Regarding lexical production, all results were statistically in favour of the IG. For *Fw* and *V*, which showed a normal distribution (and are therefore marked with * in Table 35), the *t*-test showed significant differences between the two groups in favour of the IG, thus confirming the results of the non-parametric test: *Fw*: [$t(42.83)= 8.88, p=.000, \eta^2p = 0.55$], *V*: [$t(64)= 3.98, p=.000, \eta^2p = 0.19$]

The outcomes for the lexical reception measures at T3 also prove the superiority of the IG in almost all measures, except for *Adv* and *Interj*, where the results were not significant. For the variables which showed a normal distribution at T3 (marked with * in Table 35), the parametric results confirm the non-parametric findings as follows: *Cw*: [$t(63.98)= 7.64, p=.000, \eta^2p = 0.47$], *Fw*: [$t(63.98)= 6.97, p=.000, \eta^2p = 0.43$], *N*: [$t(51.38)= 4.23, p=.00, \eta^2p = 0.21$], *V*: [$t(64)= 4.62, p=.000, \eta^2p = 0.25$]. With respect to lexical production, we found statistical variations in all measures except for *Fw*, *Prep* and *Interj*. For *Cw*, which was normally distributed at T3 (marked with * in Table 35), the results of the parametric analysis were in line with those of the non-parametric test: *Cw* [$t(62.80)= 11.78, p=.000, \eta^2p = 0.68$].

A closer look at the scores for reception indicates that the statistical results are all in favour of the IG, with several emerging patterns. For instance, at T1, T2 and T3, significant differences between the groups were registered for *Cw*, *Fw*, *N*, *V*, *Adj*, *Pron* and *Prep*. At T1 and T2, but not T3, the contrasts between the groups were

meaningful with respect to the number of *Adv*. No significant differences were found for the category of *Interj* at any data collection time.

As for the oral production results, the IG significantly outperformed the RG1 in the majority of measures. However, just as in lexical reception, between-group significance was not always found in all data collection times. For *Cw*, *N*, *V*, *Adj* and *Adv*, the differences were maintained from T1 through T3. Nevertheless, for other measures, the contrasts between groups occurred at one or two data collection times. Thus, at T1 and T2 there are differences in *Fw* and *Prep*. At T2 and T3 the two groups differed significantly with respect to the number of *Pron* produced. For the category of *Interj*, differences between the two groups were only found at T2.

Table 35: Descriptive results and Mann-Whitney U-Test results for the MBCDI word-class measures for the IG (n=51) and RG1(n=15) (R: Receptive vocabulary, P: Productive vocabulary).

		T1				T2				T3			
Measures		IG	RG1	Z	p	IG	RG1	Z	p	IG	RG1	Z	p
		M (<i>Sd.</i>)	M (<i>Sd.</i>)			M (<i>Sd.</i>)	M (<i>Sd.</i>)			M (<i>Sd.</i>)	M (<i>Sd.</i>)		
Cw	R	73.25 (19.06)	44.80 (6.46)	-5.504	.000	86.40 (27.56)	41.60 (3.04)	-5.824	.000	100.08 (14.03)	81.47 (5.05)	-4.716*	.000
	P	27.06 (14.61)	12.13 (3.79)	-3.814*	.000	60.27 (15.63)	26.73 (3.80)	-5.809	.000	121.00 (26.77)	74.33 (5.92)	-5.549*	.000
Fw	R	16.71 (3.83)	8.93 (2.31)	-5.272*	.000	19.96 (5.23)	10.93 (2.25)	-5.522*	.000	23.63 (4.67)	18.00 (2.17)	-4.345*	.002
	P	5.39 (3.92)	2.00 (.92)	-3.270	.001	14.94 (4.59)	7.87 (1.95)	-5.107*	.000	25.35 (7.90)	18.47 (3.31)	-3.172	.080
N	R	58.35 (14.95)	37.67 (6.96)	-4.885	.000	64.12 (20.75)	32.00 (3.11)	-5.824	.000	70.63 (10.10)	63.73 (4.69)	-2.888*	.004
	P	21.45 (13.46)	9.07 (3.51)	-3.340*	.001	50.57 (13.56)	20.27 (5.92)	-5.786	.000	91.92 (20.09)	59.93 (4.94)	-5.457	.000
V	R	9.84 (4.31)	5.07 (.70)	-4.690	.000	16.51 (6.30)	7.13 (1.64)	-4.770	.001	17.06 (3.36)	12.67 (2.69)	-4.094*	.000
	P	2.59 (1.15)	1.00 (.65)	-3.738	.000	5.71 (1.95)	3.53 (1.59)	-3.559*	.000	20.55 (5.36)	9.07 (1.62)	-5.625	.000
Adj	R	6.59 (3.11)	3.33 (.90)	-4.366	.000	7.43 (4.09)	3.20 (.86)	-4.239	.000	14.40 (2.75)	6.40 (.82)	-5.887	.000

		T1				T2				T3			
Measures		IG	RG1	Z	p	IG	RG1	Z	p	IG	RG1	Z	p
		M (<i>Sd.</i>)	M (<i>Sd.</i>)			M (<i>Sd.</i>)	M (<i>Sd.</i>)			M (<i>Sd.</i>)	M (<i>Sd.</i>)		
	P	3.20 (1.86)	2.07 (.25)	-2.499	.012	5.37 (1.84)	2.60 (.73)	-4.769	.000	10.84 (3.80)	5.73 (1.16)	-4.950	.000
Adv	R	2.24 (1.03)	.67 (.48)	-4.899	.000	3.57 (1.25)	1.00 (.75)	-5.555	.000	4.78 (1.31)	5.20 (.86)	-1.014	.311
	P	.55 (.61)	.00 (.00)	-3.373	.001	2.67 (1.16)	1.00 (.53)	-4.930	.000	3.45 (.90)	2.60 (1.24)	-2.705	.007
Pron	R	4.98 (1.58)	2.33 (1.04)	-4.818	.000	5.73 (2.13)	2.87 (.83)	-4.818	.000	6.94 (1.69)	3.33 (.72)	-5.558	.000
	P	.31 (.99)	.00 (.00)	-1.381	.167	4.10 (1.64)	1.80 (.77)	-5.059	.000	6.61 (2.03)	5.13 (1.76)	-2.092	.036
Prep	R	3.24 (1.01)	.80 (.41)	-5.735	.000	3.29 (1.46)	.67 (.61)	-5.523	.000	4.06 (1.10)	2.00 (.65)	-5.189	.000
	P	.45 (.83)	.00 (.00)	-2.058	.040	2.37 (.95)	.93 (.45)	-4.485	.000	4.73 (2.08)	3.87 (1.40)	-1.465	.143
Interj	R	4.76 (1.92)	3.93 (1.28)	-1.907	.057	5.76 (1.40)	5.67 (1.29)	-.450	.653	6.20 (2.84)	6.40 (.91)	-.240	.810
	P	3.14 (2.13)	2.00 (.92)	-1.922	.055	4.82 (1.57)	3.27 (.79)	-3.276	.000	8.25 (3.84)	6.47 (.83)	-.389	.697

Note: *Normally distributed variable.

Table 35: Cont.

5.3. RQ2: LEXICAL DEVELOPMENT IN SPANISH AND CATALAN OF CHILDREN IN THE IMMERSION AND REGULAR SCHOOLS

In this section, we address the second research question formulated in the previous chapter as follows:

RQ2: What is the lexical development in Spanish and Catalan of very young Spanish-Catalan bilingual children learning English in an immersion school setting, as measured by the PPVT?

RQ2a: How does the Spanish-Catalan vocabulary of children learning an L3 in an immersion school compare to that of children attending a regular school?

RQ2b: Does language dominance (Spanish or Catalan) play a role in the children's lexical development in these languages?

In order to answer this question and its sub-questions, the data were analysed by means of several ANOVAS. First, to explore the longitudinal development of the IG participants in terms of their L1 and L2, a one-way RM ANOVA with *Time(T)* as the within-subjects factor was conducted. The results of this analysis are presented in [section 5.3.1](#).

In [section 5.3.2](#), we address RQ2a by comparing the receptive vocabulary in Spanish and Catalan of the IG with that of the RG1. The aim of this analysis was to see whether the early introduction of English with a massive amount of exposure, as with the IG participants, may have an effect on the lexical development of their first languages compared to Spanish-Catalan children attending non-immersion schools. Therefore, a 3x2 RM ANOVA was conducted with *Time (T)* as the within-subject factor and *School (S)* as the between-subjects factor.

The data collected in terms of the participants' language dominance is further explored in [section 5.3.3](#), in order to answer RQ2b. We wanted to observe whether competence in Spanish and Catalan while learning a third language in any of the two programmes was related to the child's language dominance. Thus, two 3x2 RM ANOVAS were conducted with *Time (T)* as the within-subjects factor and *School (S)* as the between-subjects factor: the first ANOVA compared the Spanish-

dominant participants, while the second one compared the Catalan-dominant students.

5.3.1. LEXICAL DEVELOPMENT IN SPANISH AND CATALAN OF CHILDREN IN THE IMMERSION SCHOOL

This section addresses RQ2 by analysing the receptive lexical development in Spanish and Catalan over time of the learners in the IG as measured by the PPVT. As shown in Table 36, the mean scores on the Spanish test indicate a steady increase from T1 to T3. The Catalan test results follow the same upward trend. At a descriptive level, when comparing the scores earned by the IG at both tests, we find that at each of the testing times our participants earned slightly better results in the Spanish test.

Given that the PPVT scores showed a normal distribution (see Appendix 7, Table 48), we conducted a one-way RM ANOVA with *Time (T)* as the within-subjects factor. We also used Bonferroni T1/T2, T2/T3 and T1/T3 pairwise comparisons to further assess where exactly the significant changes occurred. Given that the sphericity condition was not fulfilled by any of the variables, the Greenhouse-Geisser correction was applied.

We found an overall effect for *Time* in both tests, with the largest effect size for the *Spanish PPVT* ($\eta^2p = .70$). A closer look at the pairwise comparisons showed significant differences between T1/T2, T2/T3 and T1/T3 on oth tests, thus showing that learners' receptive vocabulary in Spanish and Catalan significantly improved over time.

Table 36: Descriptive results and one-way RM ANOVA for the Spanish/Catalan PPVT scores with Time (T) as the within-subjects factor for the IG (n=51), when $p \leq .01$.

PPVT	Mean (Sd.)			One-way RM ANOVA						
	T1	T2	T3	F	<i>p</i>	$\eta^2 p$	Power	T1/T2	T2/T3	T1/T3
Spanish	32.29 (11.81)	45.73 (14.70)	58.53 (16.39)	T:F(1.35-67.96)=119.313	.000	.70	1	.00	.00	.00
Catalan	30.59 (10.44)	42.69 (13.11)	54.35 (15.53)	T:F(1.39-69.80)=96.877	.000	.66	1	.00	.00	.00

Note: Greenhouse-Geisser correction applied for both tests.

5.3.2. LEXICAL DEVELOPMENT IN SPANISH AND CATALAN OF CHILDREN IN THE IMMERSION SCHOOL AND IN THE REGULAR SCHOOL: A COMPARISON

This section addresses RQ2a and explores any possible differences between the IG and the RG1 group across time in terms of receptive vocabulary development in Spanish and Catalan, as measured by the PPVT in both languages. We aimed to see whether the massive exposure to English from a very young age had any an effect on the lexical development of children's first languages or whether their performance in Spanish and Catalan was comparable to that of their peers attending a regular school.

The descriptive results in terms of means and standard deviations of both groups on the Spanish and Catalan PPVT test are summarised in Table 37. The results are also visually displayed in boxplots (see Figure 2). The mean scores of the two groups at all three data collection times show that the RG1 group earned better results than the IG at both tests over time. In order to see if these differences were statistically significant, we conducted a 3x2 RM ANOVA (the data showed a normal distribution) with *Time (T)* as the within-subjects factor and *School (S)* as the between-subjects factor. Because the sphericity assumption was violated, the Greenhouse-Geisser correction was applied for all variables.

As shown in Table 36, the Spanish PPVT scores revealed significant differences across time, with a large effect size ($\eta^2 p = .64$). However, no significant differences between the groups were found, indicating that the early introduction of English in an immersion context does not affect the receptive vocabulary size in Spanish, as measured by the PPVT. However, note that the level of significance was borderline (.051). In addition, no interaction effects were found.

The PPVT results for Catalan also show significant differences across time, with no interaction effects and with a large effect size as well ($\eta^2 p = .62$). In this case, there is a school effect, with a very small effect size ($\eta^2 p = .08$), showing that the children attending the regular school earned better results on the Catalan PPVT.

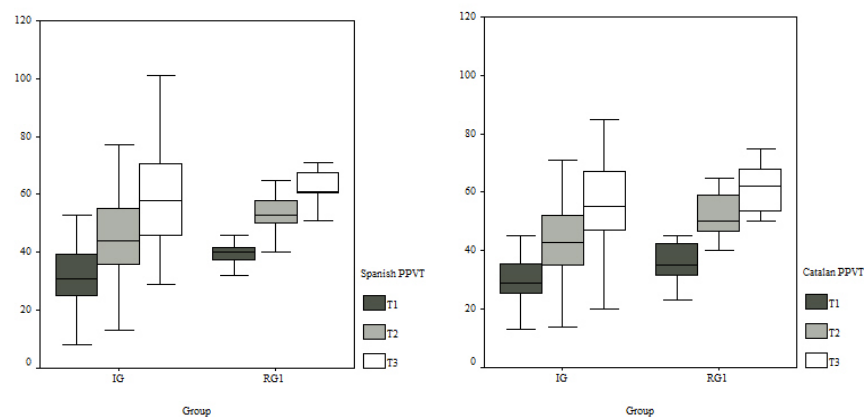
The summary of these results can also be found in Table 37. In the following section, we further explore the results in Spanish and Catalan in relation to these young learners' dominant language.

Table 37: Descriptive results and 3x2 RM ANOVA with Time (T) as the within-subjects factor and School (S) as the between-subjects factor of the Spanish/Catalan PPVT for the IG (n=51) and RG1 (n=15), when $p \leq .01$.

PPVT	Mean (<i>Sd.</i>)						3 x 2 RM ANOVA			
	T1		T2		T3		F	Sig	$\eta^2 p$	Power
	IG	RG1	IG	RG1	IG	RG1				
Spanish	32.29	38.67	45.73	54.40	58.53	63.80	T: F (1.35-86.93) = 118.365	.000	.64	1
	(11.81)	(9.10)	(14.70)	(8.66)	(16.39)	(6.91)	S: F (1-64) = 3.967	.051	.06	1
Catalan	30.59	36.27	42.69	52.27	54.35	60.93	T: F (1.36-87.64) = 108.116	.000	.62	1
	(10.44)	(7.45)	(13.11)	(7.62)	(15.53)	(8.44)	S: F (1-64) = 5.728	.000	.08	1

Note: *Greenhouse-Geisser correction applied for both tests*

Fig. 2: Spanish/Catalan PPVT scores at T1, T2, T3 for both groups



5.3.3. LEXICAL DEVELOPMENT IN SPANISH AND CATALAN LEARNERS IN RELATION TO THEIR LANGUAGE DOMINANCE

This section answers RQ2b by analysing the possible influence of language dominance in L1 and L2 lexical development. To this end, the participants in both groups were divided according to their language dominance. In the IG, 27 participants were Spanish-dominant while 24 were Catalan-dominant. In the RG1, there were eight Spanish-dominant and seven Catalan-dominant participants. We wanted to see whether the massive amount of exposure to the FL in the immersion school had an effect on the L1 and L2 of our Spanish- and Catalan-dominant participants and compare the results to those in the regular school. Therefore, two separate analyses were conducted: the first compared the Spanish-dominant children in both schools, while the second compared the Catalan-dominant children.

With respect to the Spanish-dominant learners, at a descriptive level, the scores earned by the RG1 on the Spanish test were superior to those earned by the IG at all data collection times. In the Catalan test, although the RG1 was superior to the IG at T1 and T2, at T3 the mean scores of the IG were slightly higher (see Table 38).

A 3x2 RM ANOVA with *Time (T)* as the within-subjects factor and *School (S)* as the between-subjects factor was also conducted. Because sphericity could not be assumed for the variables under analysis, the Greenhouse-Geisser correction was applied. The results summarised in Table 38 revealed that for the Spanish-dominant students, there was an overall large *Time* effect in both tests, indicating that both receptive vocabularies, as measured by the Spanish/Catalan PPVT, improved across time. However, in the between-subjects analysis (i.e., role of *School*), no significant differences were found between the groups in either test. A small effect size of .01 was found for both groups, and no interaction effects were found. Therefore, school did not significantly affect the results on any of these tests.

Regarding the Catalan-dominant participants, the mean results on both tests and at all three data collection times were always in favour of the RG1. To analyse the data statistically, we conducted a 3x2 RM ANOVA with *Time (T)* as the within-subjects factor and *School (S)* as the between-subjects factor, as we did with Spanish-dominant learners. Furthermore, the Greenhouse-Geisser correction also

needed to be applied, as the variables analysed violated the sphericity assumption. The results indicate an overall large effect for *Time* on both Spanish and Catalan scores, as well as the L1 and L2 receptive vocabularies improving significantly over time. A significant effect for *School* was found in favour of the RG1 in both tests, although the effect sizes were smaller ($\eta^2p = .23$ for Catalan and $.22$ for Spanish). No interaction effects were noted. The results for the Catalan-dominant participants are reported in Table 39, below.

Table 38: 3x2 RM ANOVA with Time (T) as the within-subjects factor of the Spanish/Catalan PPVT for the Spanish-dominant participants in IG (n=27) and RG1 (n=8), with $p \leq .01$.

PPVT	T1		T2		T3		3x2 RM ANOVA			
	IG	RG1	IG	RG1	IG	RG1	F	p	$\eta^2 p$	Power
Spanish	38.33	41.00	49.22	56.38	63.85	65.25	T: F(1.31-43.42)= 46.634	.000	.58	1
	(10.99)	(8.31)	(17.39)	(10.51)	(17.85)	(8.73)	S: F(1-33)=.545	.466	.01	.11
Catalan	29.63	34.13	43.30	46.62	56.11	55.00	T: F(1.6-52.08)=52.427	.000	.61	1
	(11.88)	(5.96)	(12.83)	(4.50)	(15.36)	(6.02)	S: F(1-33)=.295	.591	.01	.08

Note: Greenhouse-Geisser correction applied for both tests.

Table 39: 3x2 RM ANOVA with Time (T) as the within-subjects factor and School (S) as the between-subjects factor of the Spanish/Catalan PPVT for the Catalan-dominant participants in IG (n=24) and RG1 (n=7), with $p \leq .01$.

PPVT	T1		T2		T3		3 x 2 RM ANOVA			
	IG	RG1	IG	RG1	IG	RG1	F	p	$\eta^2 p$	Power
Spanish	25.50	36.00	41.79	52.14	52.54	62.14	T:F(1.2737.05)=83.605	.000	.74	1
	(8.69)	(9.84)	(9.87)	(5.92)	(12.28)	(4.05)	S: F(1-29)=8.610	.006	.22	.80
Catalan	31.6	38.71	42.00	58.71	52.38	67.71	T:F(1.1232.55)=58.411	.000	.61	1
	(8.66)	(8.67)	(13.65)	(4.57)	(15.81)	(4.85)	S:F(1-29)=8.791	.000	.23	.08

Note: Greenhouse-Geisser correction applied for both tests.

5.4. RQ3: ENGLISH PRODUCTIVE VOCABULARY OF CHILDREN IN THE IMMERSION SCHOOL AT THE END OF THE INFANT EDUCATION PROGRAMME

This section addresses RQ3, which has been formulated as follows:

RQ3: What are the characteristics of the lexical production in English in terms of fluency, diversity and complexity of Spanish-Catalan bilingual children learning English in an immersion school?

RQ3a: How does it compare to that of children at the beginning of secondary education attending a regular school?

In order to answer RQ3, in section 5.4.1. we present the descriptive results of the IG with the aim of ascertaining what these children are able to produce at the end of kindergarten after receiving massive input in the FL (2,925 hours of exposure). Their oral production was analysed using lexical fluency, diversity and complexity measures. We wanted also to see how their productions compared to those of learners in regular schools who have followed a traditional regular curriculum throughout primary education (RQ3a) and students who are some years older (given that students in regular schools are unable to speak English at the end of early childhood education). Therefore, we selected another comparison group (RG2) consisting of 44 older Spanish-Catalan bilinguals—1ESO students—who were studying English at three schools in Barcelona (A, B, and C) with a regular curriculum. The participants in this control group were first exposed to the L3 at the age of eight, that is, when they were in the third year of primary school, and at testing time they had received a total of 416 hours of instruction. These participants were tested in the Barcelona Age Factor Project (BAF) (Muñoz, 2006). To answer RQ3a, we compare the oral lexical performance of the IG against that of RG2 in terms of lexical fluency, diversity and complexity in section 5.4.2.

5.4.1. ORAL PRODUCTIVE VOCABULARY OF IMMERSION SCHOOL LEARNERS

In this section, we present the results of the IG participants in the storytelling task at the end of kindergarten. As shown in Table 40, with respect to the lexical fluency measures, our group produced an average of 44.18 tokens and 22.88 types. In relation to the different word-class measures, they produced similar numbers of *Fw* and *Cw*. They were also found to primarily use *N*, *Art* and *V*, as well as some conjunctions and prepositions, while they produced very few *Adj* and *Adv*. Their mean lexical diversity, according to Guiraud, is 3.46. With respect to the lexical complexity measures, the scores from the *M3L* and *MLU* indicate that at 70 months old (5.8 years) and after a massive amount of exposure at early ages our participants produce utterances in the FL up to 6 or 8 words.

Table 40: Means and standard deviations of the IG in the oral production task (n=51).

Measures		<i>Mean</i>	<i>Sd.</i>
<i>Fluency</i>	Tokens	44.18	13.06
	Types	22.88	5.16
	Families	21.88	4.84
	Cw	20.82	6.28
	Fw	23.22	6.42
	N	13.57	3.76
	V	7.20	2.78
	Adj	.22	.46
	Adv	1.63	.91
	Pron	1.78	1.33
	Num	.00	.00
	Prep	4.37	1.90
	Art	11.80	3.88
	Conj	3.14	1.54
<i>Diversity</i>	GI	3.46	.42
	M3L	6.84	1.94
	MLU	8.78	2.28

5.4.2. ORAL PRODUCTIVE VOCABULARY OF IMMERSION AND REGULAR SCHOOL LEARNERS: A COMPARISON

Before providing the results of the oral production of RG2, recall that because the participants in the RG2 went to three different schools, a preliminary analysis was conducted to make sure these learners could all be included in the same group. The results of this analysis are presented in [section 5.4.2.1](#). [Section 5.4.2.2](#) addresses RQ3a by comparing the oral performance of the IG and the RG2 with respect to lexical fluency, diversity and complexity measures. These results are first examined descriptively and then explored by means of *t*-tests and Mann Whitney *U*-tests.

5.4.2.1. PRELIMINARY ANALYSIS: RG2 GROUP

Participants in the RG2 went to three different schools in the Barcelona area (14 participants were selected from school A, ten from school B and 20 from school C). A one-way between-groups ANOVA was needed to check if there were any statistical differences in three basic measures, namely the number of tokens, types and word families, between the children studying at schools A, B and C.

After performing the normality analysis, a Kruskal-Wallis test was conducted given that the data were not normally distributed (see Appendix 7, Table 49). As summarised in Table 40, the results revealed no significant differences among the three subgroups in any of the measures, indicating that these participants produced texts with similar lengths and characteristics and therefore could be taken as a comparison group (RG2) in our study.

Table 41: Descriptive statistics and Kruskal-Wallis results for the oral production measures in the RG2 group participants: school A (n=14), school B (n=10) and school C (n=20).

Measures	Group	n	Mean	Sd.	Kruskal-Wallis		
					Chi-square	F	p
Tokens	A	14	52.93	22.94	3.78	2	.15
	B	10	57.50	36.92			
	C	20	40.30	22.62			
Types	A	14	20.50	7.09	3.60	2	.16
	B	10	20.70	9.49			
	C	20	16.40	5.38			
Families	A	14	19.07	6.15	2.46	2	.29
	B	10	19.40	8.54			
	C	20	15.85	5.12			

5.4.2.2. COMPARATIVE RESULTS OF LEARNERS' PRODUCTIVE VOCABULARY IN THE IMMERSION AND REGULAR GROUPS

In this section, we compare the results of the IG participants with those of the RG2 group. We begin by providing a descriptive summary of the outcomes and then present the results of the *t*-tests. It is worth mentioning that the means are expressed in raw numbers (see Table 42). Percentages were not used because there were no significant differences between the groups regarding the number of tokens produced. Table 42 shows that the IG earned higher results in the diversity and complexity measures than the RG2. For the fluency measures, the results favour either one group or the other.

The next step consisted of exploring the differences statistically by means of a *t*-test, as most of the data showed a normal distribution. Nevertheless, the normality assumption was violated for some of the variables (marked with * in Table 42). As a result, these variables were also treated non-parametrically by means of Mann-Whitney *U*-tests.

Mixed results were found in the lexical fluency measures: significant differences in favour of the IG were found for the number of *types* and *families*. Within the word-class measures, statistical differences in favour of the IG were found for *V*, *Adv* and *Prep*. On the other hand, the RG2 group earned significantly higher scores for the total number of *Cw*, *N*, *Adj*, *Num*, and *Conj*. However, there were no statistical differences in the number of *Tokens*, *Fw*, *Pron*, and *Art*. For the variables which did not meet the normality criteria, the Mann-Whitney *U*-test confirmed the previous findings for the majority of the measures: *families* ($Z=-950$, $p=.000$), *Adj* ($Z=-5.298$, $p=.000$), *Adv* ($Z=-6.461$, $p=.000$), *Num* ($Z=-3.756$, $p=.000$), *Prep* ($Z=-4.259$, $p=.000$) and *Conj* ($Z=-2.177$, $p=.002$). Nevertheless, the Mann-Whitney test did not confirm the results of the *t*-test for *Pron* ($Z=-2.791$, $p=.000$). With respect to diversity, the Guiraud Index was significantly higher for the IG. Regarding the lexical complexity measures, we noticed significant differences in favour of the IG both in *M3L* and *MLU*, thus indicating that the immersion school participants produced longer and more complex sentences than the RG2.

Table 42: Descriptive results and *t*-test results for the oral production task for the IG (n=51), RG2 (n=44), with $p \leq .01$.

		IG		RG2			
Measures	Mean	<i>Sd.</i>	Mean	<i>Sd.</i>	<i>T</i>	<i>p</i>	
<i>Fluency</i>							
Tokens	44.18	13.06	48.32	26.96	60.06	.36	
Types	22.88	5.16	18.68	7.16	93.00	.00	
Families*	21.88	4.84	17.68	6.14	93.00	.00	
Cw	20.82	6.28	25.09	12.26	62.08	.04	
Fw	23.22	6.42	21.52	14.76	56.91	.48	
N	13.57	3.76	17.25	6.97	63.83	.00	
V	7.20	2.78	7.18	7.17	54.14	.00	
Adj*	.22	.46	1.25	1.86	47.51	.00	
Adv*	1.63	.91	.32	.63	89.28	.00	
Pron*	1.78	1.33	1.23	1.95	93.00	.10	
Num*	.00	.00	.45	1.02	43.00	.00	
Prep*	4.37	1.90	2.41	2.48	93	.00	
Art	11.80	3.88	11.70	9.46	55.43	.94	
Conj*	3.14	1.54	5.14	3.73	55.67	.00	
<i>Diversity</i>							
GI	3.46	.42	2.76	.57	93.00	.00	
<i>Complexity</i>							
M3L	6.84	1.94	3.59	1.47	93.00	.00	
MLU	8.78	2.28	6.45	2.64	93.00	.00	

Note: *Normality assumption violated.

5.5. SUMMARY

In this section, we briefly summarise the results of each of the three RQs presented in this chapter. First, to answer RQ1, the RM ANOVA conducted for the MBCDI total scores of the children in the immersion programme shows a significant improvement from T1 to T2 and from T2 to T3 (significant differences were found in both receptive and productive vocabularies). With respect to the MBCDI semantic categories, the Friedman test and the Wilcoxon signed-rank pairwise comparisons indicated three distinct significant patterns for the lexical reception measures: for some variables, significant changes were found between T1/T2 and T2/T3. For other variables, we found statistical differences only between T1/T2 but not between T2/T3, and for others only between T2/T3. In lexical production,

there was a significant increase in the majority of the measures in T1/T2 and T2/T3. However, for a few variables significant differences were detected only between T1/T2. The analysis conducted for the MBCDI word-class measures displayed various patterns of significance both for the lexical reception measures and oral production: for vocabulary comprehension, the Friedman test and the Wilcoxon signed-rank test revealed significant differences between T1/T2 and T2/T3 for some of the measures. In other cases, significant changes were only found between T1/T2, and sometimes just between T2/T3. Regarding oral production, the statistical tests revealed that significant changes occurred for all measures between T1/T2 and T2/T3.

In RQ1a, we conducted 3x2 RM ANOVA and Mann-Whitney *U*-tests to compare the lexical performance of the IG and RG1 in English. As indicated by the ANOVA, the results on MBCDI total scores were significant and favourable for the IG. Likewise, the Mann-Whitney *U*-test conducted for the MBCDI semantic categories and word-class measures showed that significant results were always in favour of the IG in all the measures in both comprehension and production.

Regarding RQ2 on the performance of immersion learners in Spanish and Catalan, an RM ANOVA with *Time* as the within-subjects factor showed that there were significant differences between T1/T2 and T2/T3 in both the Spanish and Catalan tests, evidencing that the IG steadily and significantly progressed over time in all these measures. In response to RQ2a (group comparison), we conducted a 3x2 RM ANOVA with *Time* as the within-subjects effect and *School* as the between-subjects effect; the comparison with the RG1 group showed no significant differences between the two groups in Spanish test and a significant difference in favour of the RG1 group in the Catalan test, with a small effect size ($\eta^2 p = .06$). We also conducted further analyses in terms of language dominance: 3x2 RM ANOVAs, one for the Spanish-dominant and another for the Catalan-dominant participants. The results showed that there were no differences in performance in the Spanish and Catalan PPVT for Spanish-dominant students in both schools. On the contrary, differences in favour of the RG1 students were found in Spanish and Catalan PPVT tests when comparing the Catalan-dominant participants.

Finally, in order to explore the productive vocabularies of the IG at the end of kindergarten, we analysed these children's storytelling at T3 with respect to lexical fluency, diversity and complexity measures. The participants were found to produce similar numbers of *Fw* and *Cw*, and *N*, *V*, and *Art* were the grammatical categories that the IG participants produced most often when telling the story in English.

When comparing the productive vocabulary results of the IG at the end of kindergarten with those of the RG2 (1 ESO), mixed results were found. The *t*-tests identified significant differences in almost all the measures, but not all of them were always in favour of the same group. With respect to the lexical fluency measures, we found significant differences favouring the IG (e.g., number of *types*, *families*, *Adv* and *Prep*), while the RG2 was significantly superior in the total amount of *Cw*, *N*, *Adj* and *Conj*. No statistical differences were found in the number of tokens, indicating that both groups produced texts of similar length. Regarding the results of the lexical diversity and complexity measures, they were all in favour of the IG. Thus, the young immersion participants were found to produce significantly longer and more complex utterances in English than their older peers. All these results will be thoroughly discussed in the next chapter.

CHAPTER 6

DISCUSSION

In the previous chapter, we presented the results on receptive and productive vocabularies in English (L3), Spanish and Catalan (L1 and L2) and we have compared them to those of two control groups (RG1 and RG2) learning English in regular curriculum schools. This chapter discusses the results of our research following the RQs outlined in Chapter 4. Thus, in an attempt to answer RQ1 and RQ1a, section 6.1, provides an interpretation of the findings in English for the IG and then discusses the comparative outcomes of the IG and RG1. Section 6.2, focuses on RQ2, RQ2a and RQ2b and the results in Spanish and Catalan of the participants in the IG. It offers an interpretation of the comparative outcomes of the IG and RG1 on the Spanish and Catalan PPVT, while also taking into account the participants' language dominance. The discussion ends with section 6.3, which focuses on RQ3 and RQ3a, specifically the results of the IG in an oral production task at the end of pre-primary school. These results are then interpreted compared to those of another control group, RG2, who had been learning English at different regular schools in Barcelona since the age of eight and were 14 years old at testing time.

6.1. RQ1: EARLY LEXICAL DEVELOPMENT IN ENGLISH AS AN FL

This section begins by restating RQ1 and RQ1a, which were initially formulated in Chapter 4 as follows:

RQ1. What are the characteristics of the receptive and productive vocabularies in English as an L3 of very young Spanish-Catalan bilingual children learning English in an immersion school setting, as measured by the MBCDIs?

RQ1a. How does the English vocabulary of children learning L3 in an immersion school compare with that of children attending a regular school?

To answer RQ1, we conducted a longitudinal analysis of the IG participants' lexical comprehension and production in English. The analysis was carried out with three

sets of measures: MBCDI **total scores**, MBCDI **semantic categories** and MBCDI **word-class measures**, and the results are discussed in [section 6.1.1](#). To answer RQ1a, we compared the results of the IG across time with those of the RG1 in terms of comprehension and production and with the same sets of measures mentioned above. The discussion of these outcomes is presented in [section 6.1.2](#).

6.1.1. RECEPTIVE AND PRODUCTIVE LEXICAL DEVELOPMENT OF CHILDREN ATTENDING THE IMMERSION SCHOOL

In the following sections, we discuss the results of the IG in MBCDI **total scores**, MBCDI **semantic categories** and MBCDI **word-class measures**. More specifically, [section 6.1.1.1](#) discusses the results of the MBCDI total scores, but here we start by examining the longitudinal outcomes. After that, we cite some relevant cross-sectional data with the aim of presenting a picture of the English vocabulary the IG acquired. Finally, we end the section by examining the suitability of the MBCDI as a testing tool for an FL. The outcomes of the MBCDI semantic categories and word-class measures are discussed together in [section 6.1.1.2](#), as they follow a similar pattern (i.e., the Friedman tests carried out for both categories led to mixed results). In doing so, we will follow a structure similar to that of [section 6.1.1.1](#).

6.1.1.1. MBCDI TOTAL SCORES

The longitudinal results show significant differences in all lexical measures between all three data collection times (T1/T2, T2/T3 and T1/T3), indicating that the participants' receptive and productive vocabularies in English improved. This progress is most likely due to the fact that the learners accumulated more hours of exposure at each data collection time (by T1 there were 975 hours, T2 1950 and T3 2925). These findings are corroborated by the results of other longitudinal studies conducted in L1 and L2 with learners who were younger than the participants in this study (i.e., 18 months old). More specifically, the results indicate improvement in receptive and productive lexical knowledge as the hours of exposure increased (Vila, 1984; Pearson et al., 1997; Patterson, 2002; David & Li, 2008). Similar outcomes are also found in instructional settings where learners were acquiring an FL even with less exposure. For instance, the study carried out by García Mayo

(2002) with Spanish-Catalan bilinguals learning English as an FL in regular schools shows that more hours of exposure to the FL lead to better performance. Comparable outcomes in formal immersion settings were found by Bergstörm et al. (2016), who carried out a study with a group of German monolingual children learning English as an FL.

Looking at the cross-sectional results at T1, we can provide a picture of these children's lexical development in English after one school year of exposure. At this point, we should remark that we have been unable to compare the results of our study against those of other similar research due to the fact that we have not found comparable evidence of learners' performance in this particular context (i.e., young bilingual children acquiring English as an L3 in formal immersion settings). Instead, we will rely on the findings with monolingual and bilingual children in their L1 and L2. Needless to say, the process and factors involved in TLA are different than those related to first and second language acquisition (Cenoz & Genesee, 1989; Cenoz, 2013), and therefore the comparative results discussed in the next paragraphs are just orientative.

With respect to lexical receptive knowledge, after nine months of studying English (i.e., 975 hours of exposure), the participants in our study were found to be able to understand and respond to simple commands (they earned a mean of 10.65 in *PhsU*) and understand less than one-quarter (22%) of the number of words measured by the MBCDI. As mentioned in Chapter 2, the studies conducted with very young monolingual and bilingual children in their first languages found that both monolingual and simultaneous bilingual young learners start responding to simple commands nine months after they were first exposed to the language (Cruz & Ferreira, 2006). Furthermore, Poulin et al. (2013) showed that ten months after the first exposure to L1 and L2, simultaneous bilinguals are able to recognise familiar words in each of their languages. These studies do not specify the number of hours of exposure to L1 and L2, but we assume it is much greater than the IG, who learned English as an FL at school for eight hours a day, five days a week. In an attempt to make a comparison in terms of early lexical comprehension and the data from monolingual and bilingual learners, we can say that at the receptive level, the lexical knowledge of children in immersion school settings is quite similar to these populations.

The scores earned for *WdsP* and *M3L* at T1 show that after nine months of massive exposure (975 hours), learners were able to produce a few familiar words in English and one-word utterances. Nevertheless, they were not able to produce complex sentences ($SC=0.00$), which reveals their lack of grammar at this stage. At T2, after 20 months (1950 hours) of exposure, the IG learners were found to produce two- to three-word sentences. Furthermore, they were found to produce complex sentences, which is an indicator of productive grammar (Caselli et al., 1990; David & Li, 2005). Comparatively, monolingual and bilingual young learners in other studies were reported to undergo the one-word stage twelve to 16 months after their first exposure to the language (De Houwer, 2009; Grosjean, 2010) and the two-word stage after 18-20 months (Patterson, 1998; Clark, 2003). Thus, our participants seem to follow the same milestones in vocabulary production as their monolingual and bilingual peers, acquiring vocabulary in L3 at about the same age, possibly with less exposure. Again, we should remark that broad comparisons across students are just tentative.

In terms of vocabulary size, the results of our study provide further information on young learners' receptive and productive vocabulary sizes as measured by the MBCDI. In the absence of research assessing lexical comprehension and production in an FL with the MBCDI, we will rely on other MBCDI studies conducted with very young bilingual and trilingual children acquiring English as one of their first languages, also bearing in mind that the TLA process is different than first and second language acquisition, as already pointed out. In order to offer a more reliable comparison of the results, we performed proportional calculations. Therefore, with respect to *WdsU* (see Table 43), we found that the scores of the IG in English are lower than those of other much younger bilingual and trilingual children. These results could be expected, as we assumed that our participants would earn poorer scores in English than bilingual and trilingual learners who were exposed to English since birth. This may be an indicator of the fact that the *WdsU* measure of the MBCDI provides consistent results with FL kindergarten.

Table 43: Comparative results of research conducted on receptive vocabulary using the MBCDL.

Study	Participants	Age in months (at testing time)	WdsU in English
Alexandrescu (2023)	Bilinguals learning English as an L3	46	90.12
		58	106.41
		70	123.71
Byers & Wrecker (2011)	Bilinguals	17	156
	Early trilinguals	17	202
Quay (2001)	Early trilinguals	16	52

With respect to *WdsP*, when comparing the results in English on the same test by the IG and other bilingual children, we found that first, the IG scores in English at the ages of 46 and 58 months (T1 and T2, respectively) were lower than those of other much younger bilingual and trilingual children who acquired English as one of their first languages (see Table 44), and secondly, at the age of 70 months (T3), the IG scores were comparable to those found by Byers and Wrecker (2011) (see shaded cells in Table 44). All these results indicate that with a few exceptions, our participants understood and produced fewer words in English as an FL than much younger bilinguals and trilinguals do. This suggests that the *WdsP* measure can provide valuable information for the population we are analysing.

Table 44: Comparative results of research conducted on productive using the MBCDL.

Study	Participants	Age in months (at testing time)	WdsP in English
Alexandrescu (2023)	Bilinguals learning English as an L3	46	31.94
		58	75.61
		70	146.35
Byers & Wrecker (2011)	Bilinguals	17	37
	Early trilinguals	17	75
Quay (2001)	Early trilinguals	21	50

The outcomes of *PhsU*, *M3L* and *SC* could not be compared with those from other studies, as their results are expressed in percentiles, which we were unable to use, as the MBCDL has only been normed to test young learners in their first languages but not in an FL. Nevertheless, after carefully examining our results by comparing them against those of other studies conducted with monolinguals and bilinguals, we

see that the MBCDI results for the L3 in immersion children can also give coherent indications when testing lexical acquisition in an FL.

First of all, the longitudinal outcomes of this research point to a significant improvement in receptive and productive vocabulary as the number of the hours of exposure to L3 increased. Similar results have been found in studies using the MBCDI to test young monolingual and bilingual infants in their L1 and L2 (Pearson et al., 1997; Patterson, 2002; David & Li, 2008). Secondly, the test also provides reasonable cross-sectional results in terms of lexical acquisition milestones and vocabulary size. We have shown that our participants follow the same steps in learning vocabulary as their monolingual and bilingual peers, as measured with this test (see the results found using the MBCDI in Caselli et al., 1990; Patterson, 1998; David & Li, 2005; De Houwer, 2009; Poulin et al., 2013). The proportional calculations performed to compare the vocabulary sizes in English for our learners with monolinguals, bilinguals and trilinguals (see the research conducted by Quay, 2001; Byers & Wrecker, 2011) has yielded the expected results; that is, our students understand and produce fewer words in English than monolingual and bilingual children exposed to English since birth.

Analysing these results in connection to the setting where participants' FL instruction has taken place shows that learning an FL through one-way immersion at very young ages does indeed have beneficial results on vocabulary acquisition. The progress these children have made from T1 to T3 shows that this type of immersive education in which L3 accounts for 65% instruction is effective. Indeed, these results give a vote of confidence to immersion schools and advocate for an early start in FL acquisition.

Furthermore, these children seem to pass the same milestones in lexical development as their bilingual and monolingual peers. We can therefore see them as early trilinguals, even if they seemed to understand and produce less at testing time.

6.1.1.2. MBCDI SEMANTIC CATEGORIES AND WORD-CLASS MEASURES

Turning to the semantic categories and word-class measures, we would like to point out that because the results showed similar patterns, they will be discussed together in this section. Thus, we start with preliminary remarks on the MBCDI vocabulary checklist to frame the discussion. Next, we examine the longitudinal results first in terms of comprehension and then in terms of production, by discussing both the semantic and word-class measures. We continue by commenting on the cross-sectional results, and the section ends by focusing on the reliability of these measures in the lexical assessment of young learners acquiring English in an immersion school setting.

As stated in Chapter 4, the MBCDI vocabulary checklist consists of 19 semantic categories with a total of 396 lexical entries. Our participants' scores were first calculated for each semantic category, and then these categories were regrouped into nine word-class measures to provide greater insight into what learners' lexicon looks like at different stages. Each lexical entry was tested either for comprehension ('the child understands') or production ('the child understands and says'). Note that when assessing vocabulary production, there is also the implicit assessment of comprehension.

With respect to the **semantic categories** and **word-class measures** understood, it is necessary to discuss several findings. First of all, as the results indicate, we found high standard deviations for some of the measures, implying that some participants scored much higher or lower than the mean. This shows the high degree of variability among participants, probably due to the fact that lexical performance at this age is not very homogeneous. This type of finding has also been reported by De Houwer (2007) in the case of simultaneous bilinguals assessed in their L1 and L2. Secondly, from a longitudinal perspective, we noticed an overall significant effect of time in the majority of semantic categories and word-class measures understood. These results once more provide evidence that improvement is due to the accumulation of hours of exposure (which has also been corroborated in the previous section discussing the MBCDI **total scores**).

Nevertheless, looking at the results more in depth, we noticed that, surprisingly, unlike the total scores measures (discussed in the previous section), the differences between certain data collection times were not always significant for some of the semantic and word-class measures. Thus, for the semantic categories understood, the Wilcoxon signed-rank test indicated three distinct significance patterns, as follows. For the categories of *Sounds*, *Food*, *Actions* and *Questions*, the differences were not significant at T2/T3, whereas for the categories of *Animals*, *Vehicles*, *Body*, *Household*, *Qualities*, *Pronouns* and *Locations*, the differences were not significant at T1/T2. The categories of *Clothes*, *People* and *Toys*, were the only ones where significant differences were found between all data collection times (T1-T2-T3). For the semantic categories of *Furniture*, *Places*, *Games* and *Time*, no significant differences at T1-T2-T3 were found. A similar pattern was noticed for the word-class measures understood. Thus, differences at T1/T2 and T1/T3 were found for the class of *V* and *Interj*; T2/T3 and T1/T3 differences were found for *N*, *Adj* and *Prep*; while significant T1-T2-T3 differences were found for the categories *Cw*, *Fw* and *Pron*.

These results need to be interpreted with caution. Given that both the semantic categories and the word-class measures were found from the total amount of *WdsU*, we thought it was plausible that the outcomes would follow the same trend (i.e., significant differences at all data collection times, presumably due to the increased exposure). Moreover, it seems that when taking into consideration the scores earned for specific categories (*Furniture*, *Places*, *Games* and *Time*), unlike the total score (the total amount of *WdsU*), the outcomes are a bit different and do not show a significant increase from one data collection time to another. Under these circumstances, and in the absence of other studies which analyse the type of words known by learners by means of the MBCDI **semantic categories**, we believe we should offer some reflections on the structure of the MBCDI, with the aim of providing a possible explanation for these results. According to the format of the test, lexical comprehension is also tested through production, as the words produced are defined as the words the child understands and says. Therefore, a more reliable score could be found by totalling the words understood (those that the child understands) and the words produced (those that the child understands and says)

for each semantic category involved, even if a distinction between reception and production cannot be made.

Another possible explanation may have to do with what researchers such as Gairns and Redman (1986), Laufer (1990) and Meara and Milton (2003) suggest: lexical performance may also vary according to individual styles, degree of recycling and other aspects of word learnability such as word length and cognates. It is important to remember that at each data collection time, our IG had different English native teachers with different teaching styles, which may have also had an impact on the results. As other studies point out (Cote et al., 2022; De Cat, 2020; Sun et al., 2020), lexical development of young learners is also a function of input quality. Different teachers may have placed more emphasis on certain semantic categories than others.

Regarding the non-significant differences at T1-T2-T3 found for the semantic categories of *Furniture*, *Places*, *Games* and *Time*, the results were predictable. These categories all include high-frequency words that our participants were exposed to on a daily basis through their morning or afternoon routines and were expected to use daily. Indeed, most of these children readily produce these words rather than just recognise them. Therefore, these categories were not indicative of progress and may not be used to assess children in these contexts.

With respect to lexical production, the results were quite different. For all measures, the differences recorded were significant at all data collection times, following the pattern indicated by the total amount of *WdsP*. This may be an indicator that the participants in the IG were turning their receptive knowledge into productive knowledge as they grew older, and that the immersion setting was facilitating this development.

We found high standard deviations for some lexical reception and production measures. Comparing the results of the semantic categories to those of other studies conducted with young learners in their first language, we found that during the initial stages of FLA, our participants were exposed to and learned similar word categories as those acquired by very young monolinguals in their L1. For instance,

research conducted by Harris and Chasin (1999) with English monolingual children between the ages of six and 18 months showed that in terms of vocabulary comprehension, these young learners were able to recognise names of objects, actions, words and people. The semantic category labelled ‘names of objects’ may correspond to several of the semantic measures in the MBCDI, such as *Vehicles*, *Toys*, *Furniture* and *Household*, which, as our results indicate, are semantic categories our participants had been exposed to and acquired. Another study conducted by Kauschke and Hofmeister (2002) with 32 German monolingual children at different ages (13 months, 14 months, 21 months and 36 months) showed that at the onset of speech production, children acquired relational words, social words and onomatopoeic terms. In our data, relational words are included in the semantic category of *Family* and social words in the category *Games* and *Routines*. As expected, the participants in our study were shown to acquire the same semantic categories as monolinguals do in their L1, but their FL vocabulary size is generally much smaller than that of much younger monolinguals or bilinguals in their first languages.

In the same vein, the proportion of word-class measures acquired by the participants in our study showed to be similar to those acquired by very young learners in their first languages. Thus, Bates et al. (1994) found that the word-class categories acquired during the first stages of lexical development in children’s L1 are basically nouns, verbs and closed-class items, and that the consistency of the lexis varies from one stage to another. Thus, in the first stage (when the vocabulary totals nearly 100 words), the children were found to learn mostly nouns; in the second stage (100-400 words), there is an increase in the number of verbs; and in the last stage (more than 400 words), the number of closed-class items increases. Similarly, research conducted by Wei and Li (2008) with bilingual young children in both their languages showed a similar trend. Taking a look at the receptive and productive vocabularies of the participants in our study, as measured by the MBCDI, we noticed that the composition of their incipient lexis in an FL is quite similar to that of the participants in the aforementioned studies. It is also worth noting that even though the test assessed more nouns and verbs than any other grammatical category, even though all of them are well represented and word-class proportions

can be calculated, these outcomes may also suggest that formal instructional settings provide similar output to natural immersion contexts.

Therefore, regarding the ability of the MBCDI semantic and word-class measures to assess receptive and productive vocabularies in an FL, we can say that our results may be comparable to those of bilingual populations (who earned higher scores). Therefore, these measures are also able to offer relevant information on what young learners of English as an FL acquire during the first three years of intensive exposure to the language.

This study builds on research targeting early lexical acquisition in an FL. First, it attempts to offer a picture of the early lexical development of bilingual young learners who are learning English in an immersion school setting. According to our data, the IG participants in the study acquired similar semantic categories and proportions of different word categories as their monolingual and bilingual peers, even though they had less exposure to English than native speakers. Nevertheless, we found that their vocabulary size (number of words known and produced) was smaller than monolingual and bilingual infants. From a longitudinal perspective, vocabulary sizes increased significantly after each data collection time due to increased exposure. At the same time, when looking at different semantic and word-class measures, we noticed that young learners' lexical performance is quite varied and that individual variability is high. This may be attributed to the large amount of input they receive at school, teachers' focus on different target vocabulary or even different learning styles. Secondly, in relation to the adequacy of the testing tool for lexical assessment in an L3, the results of all three categories of the MBCDI measures did not disagree with outcomes of studies in children's L1 and L2 that were carried out with the same tool, thus showing that the MBCDI can provide insightful longitudinal and cross-sectional information in an FL as well. Finally, the results show that teaching English through immersion brings benefits to the young Spanish-Catalan learners in that not only do they increase their receptive and productive vocabularies in the FL, but they are also exposed to similar semantic and word-class categories as their native peers, which may be an indication that if all the other factors are equal and exposure is kept constant, in due time they may achieve high proficiency in English.

6.1.2. COMPARATIVE RESULTS OF RECEPTIVE AND PRODUCTIVE VOCABULARY DEVELOPMENT IN IMMERSION AND REGULAR SCHOOLS

This section discusses the results aiming to provide an answer to RQ1a, which is formulated as follows:

RQ1a. How does the English vocabulary of children learning L3 in an immersion school compare with that of children attending a regular school?

The MBCDI **total scores** showed that the IG was superior to the RG1 in all lexical reception and production measures. In other words, the IG learners understood and produced significantly more words than those following a regular curriculum. Furthermore, they were also able to construct longer and more complex sentences. This predictable outcome is accounted for by the fact that the IG received considerably more exposure to the L3 than the RG1. In the literature, these findings have also been confirmed by EFLA studies conducted in regular schools (García Mayo, 2002) and immersion school settings (Bergström, 2016). More specifically, they showed that when comparing students of similar ages at testing time and with similar ages of onset, the group with more hours of instruction scored higher than those who had received less instruction. Nevertheless, different results were found by Cenoz (2003b), who focused on two groups of Spanish-Basque bilingual learners with different amounts of exposure (500 vs 300 hours) and with similar ages at testing time. In this case, the results revealed no significant differences between the groups. The explanation provided by Cenoz has to do with the quality of input, in that the younger participants benefitted from oral input while the older ones from grammar-based input, which helped them catch up. We should also note the difference in exposure between this study (200 hours) and our study (1950 hours on average)

Regarding our interpretation to the outcomes found for the semantic and word-class measures both understood and produced, several points should be considered. Firstly, we noticed that whenever results were significant, they were always in favour of the IG. Secondly, we found that in some cases, no significant differences were found between the two groups despite the difference in the number of hours of instruction (T1: 898, T2: 1,794, T3: 2,691). For instance, in the semantic

categories, at T1 the difference between groups was more striking with receptive vocabulary, as the IG scored significantly better than the RG1 in 15 semantic categories. The difference in production proved statistically significant for only nine categories, which indicates that the IG group's advantage is not large at the beginning, even if they had had more hours in contact with English than the RG1. The situation changes at T2, when the results indicated a significant difference between the two groups in 17 reception measures and for all oral production measures, a trend which holds at T3. The outcomes follow the same pattern in the nine word-class measures. The results at T1 were significant for eight reception measures and seven production measures. At T2, differences were significant for eight reception measures and nine production measures, while at T3 the IG was superior in seven and six measures, respectively.

These results need to be interpreted with caution. First of all, the amount of exposure can indeed account for the IG's better performance. Similar conclusions have been reached by Bergström et al. (2016), who also focused on receptive and productive vocabularies in immersion and regular school settings. Other studies conducted in regular schools but focusing on grammar (García Mayo, 2002) and written production (Lasagabaster & Doiz, 2004) have also found similar outcomes. Apart from the amount of exposure that may account for these results, input quality and the syllabus used by the teacher should also be taken into account. More specifically, the IG had the chance to learn implicitly more often than the RG1 given the amount of input provided (Muñoz, 2006). On the other hand, the RG1 was taught explicitly while implicit learning resources used by children when learning languages could not be exploited at the fullest in this case. We should note that the RG1 learned English as an FL, whereas English was also the language of instruction for the IG.

Nevertheless, in our situation, in spite of a large difference in the number of hours of exposure, no between-group differences were found in some of the measures. Clearly, other factors need to be considered. The fact that the number of lexical entries in each category was not very high may have influenced the statistical results. At the same time, the very young age of our participants may have also played a role, as we have already seen that performance at these young ages is not very

homogeneous, and the content taught in the regular school in any explicit way may have favoured the more effective acquisition of these words by the participants in RG1. Thus, we can interpret the measures for which non-significant results were found by the fact that less exposure for the RG1 may have been counterbalanced by focused instruction, and this may be the reason why significant differences between the groups were not always found either at the receptive or the productive levels.

Indeed, the results of these two sets of measures also shed light on other factors (apart from exposure) that need to be taken into account when analysing the lexical performance of young children in two different settings. For instance, we have seen that the teaching style may also contribute to the results of these two blocks of measures. On the other hand, children benefit more from the language immersion approach. However, explicit teaching, with the teacher using the children's native language (as sometimes happens in regular schools) can 'bridge the gap' between the lexical performance of learners in an immersion setting and those in a regular school by helping the latter catch up in terms of vocabulary size. This idea can be supported by findings in studies such as Torregossa et al. (2023), who showed the benefits of bilingual pedagogy in children's literacy skills. To this end, they tested 33 Greek-Italian bilingual children between the ages of eight and eleven who were studying Italian through immersion in a school in Greece. When presented with Italian sentences that they needed to assess for acceptability, the participants proved to be better at noticing the mistakes and accounting for them if the sentences were presented bilingually (i.e., in both Italian and Greek).

With regard to the reliability of the MBCDI when longitudinally comparing two groups, we have seen that the tool is also used not only to measure progress but also to capture the between-group differences. Therefore, it may be a good option when contrasting different groups of young learners in terms of their lexical development in an FL.

Finally, to sum up, we can say that young learners who are acquiring an FL in an immersion setting have better receptive and productive vocabulary knowledge than same-age learners who are learning English in a regular school, with less exposure.

The implications these results have in teaching English as an FL are quite straightforward: the more English is used in teacher-student interaction, the better. Yet, there may be some instances in which using the L1 could possibly accelerate the acquisition of new words.

6.2. RQ2: LEXICAL DEVELOPMENT IN SPANISH AND CATALAN

This section analyses the results of the Spanish and Catalan PPVT test. We begin our discussion by sequentially addressing RQ2, RQ2a, and RQ2b, stated in the Methodology chapter as follows.

RQ2: What is the lexical development in Spanish and Catalan of very young Spanish-Catalan bilingual children learning English in an immersion school setting, as measured by the PPVT?

RQ2a: How does the Spanish-Catalan vocabulary of children learning an L3 in an immersion school compare to that of children attending a regular school?

RQ2b: Does language dominance (Spanish or Catalan) play a role in the children's lexical development in these languages?

Thus, section 6.2.1. examines the results of the IG in Spanish and Catalan and discusses RQ2. Next, in section 6.2.2. we discuss the comparative results in Spanish and Catalan of the IG and RG1 (RQ2a), also focusing on the participants' language dominance (RQ2b).

6.2.1. LEXICAL DEVELOPMENT IN SPANISH AND CATALAN OF CHILDREN IN THE IMMERSION GROUP

To answer RQ2, we conducted both descriptive analyses at each data collection time and a longitudinal analysis of the IG participants' lexical comprehension in their L1 and L2. The descriptive analyses showed that the scores in the Spanish receptive vocabularies were higher than those in Catalan. This may be due to the fact that our learners had accumulated more hours of exposure at each data collection time in Spanish (312 vs 237 at T1, 624 vs 546 at T2, and 936 vs 819 at T3). A further look at the results from the cross-sectional perspective shows high

standard deviations at all data collection times for both Spanish and Catalan. This indicates, as it did with the L3, that performance at this age is quite uneven for both the L1 and L2. It appears that some students understand more and others understand less, confirming the fact that, as De Houwer (2007) asserted, simultaneous bilingual children may vary in their word comprehension abilities. Some children may have been exposed to more Spanish and therefore their lexical comprehension in Spanish is better than in Catalan, and vice-versa.

From a longitudinal point of view, the findings of the one-way RM ANOVA showed an overall effect for time in both tests. Furthermore, the Bonferroni pairwise comparisons indicated significant differences between T1/T2, T2/T3 (and T1/T3) in both tests. These results show that the IG learners made significant progress in their lexical development, as measured by the PPVT. This progress, found between all data collection times in both languages, is due to both exposure and cognitive maturity (the children were older at each data collection time). Note that for Spanish, the hours of instruction increased from 312 at T1 to 625 at T2 and 936 at T3. With respect to Catalan, we also found increased yearly exposure from 237 hours at T1 to 546 hours at T2 and 819 hours at T3. Similar outcomes have been found in studies carried out with young simultaneous bilingual learners in their L1 and L2 (Pearson et al., 1993; Pearson et al., 1997; De Houwer, 2006, 2007; David & Li, 2008), which provides evidence that exposure influences lexical abilities. These studies have also shown that simultaneous bilingual children who understand more in L1 also understand more in their L2, and vice-versa. This is further confirmed by our study: there is steady progression in both languages, which is not surprising due to the typological similarities between the two.

It seems that the L1 and L2 instruction model proposed by this school (the enrichment bilingual education model which was initially defined by Baker in 1998 and referred to in Chapter 2) has also led to an increase in participants' receptive vocabulary sizes in both Spanish and in Catalan (recall that these are the languages these children speak at home). Furthermore, it also seems that in spite of the early introduction of English as L3, with a high number of hours of instruction (975 at T1, 1950 at T2, and 2925 at T3), our participants' receptive vocabularies in L1 and L2 continued to increase significantly throughout all data collection times.

In relation to the reliability of the PPVT as a testing tool for bilingual populations, we saw in Chapter 3 that this test has not actually been normed for bilingual use. Nevertheless, it has been used to test the receptive vocabularies of bilingual populations in each of their languages (e.g., Nejad & Paradis, 2009; Dixon, 2011). In this situation, researchers need to bear in mind the fact that a bilingual should not be perceived as ‘the sum of two monolinguals’ (Grosjean, 1989), and therefore the outcomes of these tests need to be interpreted with caution. Furthermore, care should be taken when interpreting results of tests which have been translated and not adapted to a certain language. In this sense, there are testimonies of researchers who used translations of this test, in the absence of versions adapted for a particular language (Bialystok, 2005; Nejad & Paradis, 2005; Dixon, 2011), always bearing in mind the limitations that this type of translated tests entails. In our case as well, we used the Spanish adaptation and the Catalan translation of the PPVT. All in all, aware of the limitations of this type of assessment, our results seem coherent when compared to the outcomes of other studies (Pearson et al., 1993; Pearson et al., 1997; De Houwer, 2006, 2007; David & Li, 2008). Therefore, in our context as well, the findings with the PPVT add evidence to the body of results with bilingual young learners in each of their languages.

Our study attempted to build on the existent research with simultaneous bilingual young learners, but adding a third language to the equation and testing across three consecutive years. Thus, we can answer RQ2 by stating that our participants continued to acquire and significantly increase their receptive lexical knowledge in their L1 and L2 in spite of receiving massive input in L3 from a very young age. Given that oral production in their first languages was not assessed in this thesis, it shall not be brought into discussion, but it would definitely make a topic for future research. It is also important to say that although the results indicated predictable lexical development in L1 and L2, they did not (up to this point) indicate whether participants’ development in L1/L2 was hindered by the amount of exposure to L3. The next section discusses this by comparing the results in Spanish and Catalan of participants in the IG and those in the RG1.

6.2.2. COMPARATIVE RESULTS OF LEXICAL DEVELOPMENT IN SPANISH AND CATALAN OF CHILDREN IN IMMERSION AND REGULAR SCHOOLS

Answering RQ2a entailed comparing the results of the IG across time with those of the RG1 in terms of receptive vocabulary, as measured by the PPVT. The results of the Spanish test indicated no significant differences between the two groups, thus showing that early massive exposure to L3 did not affect the IG's receptive lexical development in Spanish. Note also that both groups had received similar hours of exposure to Spanish at testing time. Similar results were found in the research conducted by Bergstörm et al. (2016) and in studies in regular school settings by Cenoz (1997, 2005) and Lasagabaster (2005), who found that the exposure to an FL through a massive amount of input did not affect lexical development in the children's first language. In the immersion schools, Spanish is extensively used in teaching along with English (in our case) and Catalan, which may account for the results in the IG. In the case of the regular schools, Spanish is not the language of instruction (except for in Spanish lessons). However, students use Spanish at home and during playground time to communicate with their immigrant peers who may not be as comfortable using Catalan, for example.

The results of the Catalan test, on the other hand, pointed to significant differences in favour of the RG1 with a medium effect size (.051). This outcome can be accounted for by the greater exposure to Catalan in the regular school, as Catalan is the official language of education and the means of instruction in Catalonia. Note that there is a mean difference in favour of the RG1 of 1,806 hours of exposure in this language. This result is partially confirmed by other studies with bilingual learners in L2 (Genesee, 1983; Gibson et al., 2012; Montanari, 2014), and contradicted by Bergstörm et al. (2016), who focused on bilingual students learning an L3. Note that in Genesee and Montanari, the effect found is in the short term, while in Gibson (2012) the results are confirmed for oral production only. More specifically, Genesee (1983) shows that during the first four years (between the age of three and seven) of early total immersion in French in French immersion schools in Canada, children did not progress in English (L1) reading, spelling and punctuation the same way as their monolingual peers who were studying in regular schools. However, after six years (at the age of 13), the early total immersion students caught up with the English monolinguals, perhaps because the amount of

exposure to French decreased in secondary school, as stipulated by the school's policy. Similar results were found by Montanari (2014), who compared the reading skills of 60 American children between the ages of six and eight studying Italian through immersion with children who were educated only in English in a regular school. The results indicate that the children exposed to Italian showed an initial lag in reading. Nevertheless, by the end of elementary school, the immersion group was able to perform equally well or better than the children in the regular school, also probably because the exposure to Italian had decreased.

The fact the outcomes in the Catalan test were different from those in the Spanish test may have to do with the number of hours of exposure, as indicated above, as well as the participants' language dominance. Therefore, below the outcomes of our study will be discussed in relation to our learners' language dominance. For the Spanish-dominant students, no significant differences were found in either test, meaning that the participants performed at the same level in both languages regardless of the school they attended. We can therefore say that the results of L1 and L2 receptive vocabulary task were not affected by the participants' school. With respect to the Catalan-dominant students, significant differences were found in favour of the RG1 participants in both tests. Firstly, the Catalan-dominant students in immersion settings performed worse in Catalan than their regular school peers. This can be explained by the fact that the Catalan-dominant learners in the IG had received less formal exposure to Catalan. Secondly, the Catalan-dominant students in immersion settings performed worse in Spanish than their regular school peers, even though the number of hours of instruction in Spanish was the same in both schools. This result was not expected, given that the number of hours of exposure to Spanish was the same. Nevertheless, based on the evidence provided by Genesee (1983) and Montanari (2014), this may be a short-term effect, which could be confirmed (or not) only after further research is conducted. It may also be the case that extensive exposure to Catalan (in the regular curriculum students, who may also use more Catalan outside school) may have a positive effect on performance in Spanish (given that the two languages are typologically similar and share a large number of cognates). A tentative explanation could be that a common underlying proficiency (Cummins, 1979) may help transfer between the two languages, so they reinforce one another. For the time being, however, it seems that

in the case of the Catalan-dominant students, the results of receptive vocabulary in L1 and L2 need to be interpreted with caution: these results indicate that learning English from young ages with massive exposure to this language only seems to affect the L1/L2 vocabularies of Catalan-dominant children. Further research on this topic and more longitudinal studies should be conducted with larger samples of participants. To our knowledge, there are no similar studies targeting language dominance and L3 learning at young ages upon which we could rely to interpret the results. More research is needed to see if these findings relating to language dominance are temporary.

Regarding the suitability of the PPVT when comparing different groups, earlier research has used this test to compare the receptive lexical development of both simultaneous bilinguals and monolingual learners (see, for example, Doyle et al., 1978; Bialystock et al., 2005) and sequential bilinguals and monolinguals (Gibson et al., 2012). The results of our study regarding development of L1/L2 over time are confirmed by the studies mentioned above, which used the PPVT as an assessment tool. Furthermore, our results are also along the lines of studies such as Genesee (1983) and Montanari (2016), which used other assessment tools. Therefore, we could say that the PPVT can be used to compare the development of Spanish and Catalan in bilingual children.

With respect to the implications of these results for teaching, they suggest that immersing young learners in an FL at a young age can be beneficial for their lexical development in the FL, as outlined in the previous section. It does not seem to affect the performance in their L1 and L2 either, in the case of Spanish-dominant learners. However, our results for Catalan-dominant learners suggest that their peers following the regular school curriculum have better receptive vocabularies in L1/L2 in pre-primary education. Further research will help to determine the extent and persistence of these possible differences.

6.3. RQ3: PRODUCTIVE VOCABULARY IN ENGLISH OF CHILDREN IN THE IMMERSION SCHOOL AT THE END OF THE EARLY CHILDHOOD EDUCATION PROGRAMME

In this section, we examine the results in the oral production task undertaken by our participants at T3 when they were between five and six years old. The discussion addresses RQ3 and RQ3a, initially formulated in the Chapter 4 as follows:

RQ3: What are the characteristics of the lexical production in English in terms of fluency, diversity and complexity of Spanish-Catalan bilingual children learning English in an immersion school?

RQ3a: How does it compare to that of children at the beginning of secondary education attending a regular school?

The first part of this section addresses RQ3. To do so, we first conducted a cross-sectional analysis at the end of T3 with results of the oral production task done by the IG participants at the end of pre-primary education. The results revealed high standard deviations, thus showing that some participants were able to produce more than others. These variations were also found in other studies and are discussed by researchers such as De Houwer (2007), who researched the lexical development of young bilingual learners in L1 and L2.

The scores of the lexical fluency measures indicate that by the end of kindergarten, our participants produced similar numbers of Cw and Fw . This suggests they are able to properly link the parts of their discourse. Hyltenstam (1988) suggests that a very low number of Fw means the inability to construct a coherent text, which is not the case here. N , Art , and V are the grammatical categories that these learners produced the most, due to the fact that during their years of schooling these children have probably been explicitly taught more nouns, articles and verbs than adjectives, prepositions or conjunctions. Similar results were found by researchers targeting lexical development in first languages (Bates et al., 1994; David & Li, 2005). It appears that in the initial stages of lexical development in an FL, our group thus produced words which pertain to the same word-class categories as those produced by monolinguals and bilinguals in their L1 and L2, given the fact that they are exposed to English as a means of communication. It seems logical that the results

from the storytelling are in line with those found with the MBCDI and are somehow predictable, given that these children study English in an immersion school, where the language acquisition process is rather similar to that in naturalistic settings. Nevertheless, we are unable to determine the proportion of each word-class measure in relation to all FL lexis, because the story elicits just a sample of the productive vocabulary that students may know.

With respect to the lexical complexity measures, we noticed quite a high mean length of utterance ($MLU = 8.78$, $M3L = 6.25$), showing thus that by age six the IG participants were able to produce quite long sentences in the FL. It should be noted, though, that the $M3L$ score earned in the oral production task is much higher than the score earned in the MBCDI at T3 (6.25 vs 3.83), indicating that our learners were able to produce longer sentences when assessed through the storytelling task. This outcome may have several explanations. The first is linked to the MBCDI testing procedure (i.e., upon recall: meaning that the teacher writes the three longest sentences that he/she remembers the child saying, not what the child has actually just said). Another possible explanation is related to the fact that generally speaking, children tend to be more creative when telling stories. The task may have motivated them to say longer sentences than those recorded in the MBCDI, which they may say more often in class. Therefore, this tool seems to have been more appropriate at testing oral production at these ages (five to six years old), as it offered the participants the opportunity to show more of what they could actually do in terms of oral production.

At this point, several considerations regarding the reliability of the MBCDI are in order. First, our learners' lexical fluency results in the oral task are correlated with those of the MBCDI measures at T3. Nevertheless, the $M3L$, tested through the MBCDI, is a measure that at T3 seems not to have yielded an accurate score. At the end of kindergarten, our participants are able to produce longer sentences than those measured by the MBCDI. Therefore, the $M3L$ needs to be used with caution when assessing older children in English.

To answer RQ3b, we compared the results of the IG in the oral production task against those of another comparison group RG2 at older ages (14) who had been

learning English in different regular schools in Barcelona since the age of eight. The results showed no significant differences in some measures, while in others we found mixed results. First of all, the IG scored higher in *the number of types*, *word families*, and the number of *Adv* and *Prep*. Second of all, the RG2 earned significantly better results for *Cw*, *N*, *Adj*, *Num* and *Conj*.

The results in favour of the IG with respect to *types*, *word families* and *complexity* indicates that this group was able to produce longer and more complex sentences in English. It seems that in this case the implicit learning in immersion contexts, where oral interaction is promoted, is beneficial. Therefore, for this particular case we may advocate for acquiring an FL ‘the earlier, the more, the better’. It seems that an early start (age three vs 15) combined with an extensive amount of exposure (2925 vs 416 hours of instruction) account for the significant differences between the two groups. These results are contradicted by Harley and Swain (1984) and Lyster (1994), who found that children studying in immersion schools scored significantly lower than regular school children in measures of grammatical accuracy and lexical diversity, thus advocating for a cross-linguistic approach in immersion schools as well.

Regarding the results for the number of *tokens*, *V*, *Pron*, *Art* and *Fw*, where no significant differences were found, we attribute it to the cognitive maturity of the participants in the RG2. Even though they started later and received less exposure, they are able to be as fluent as the IG participants.

Finally, for the other measures for which mixed results were found (i.e., the IG scored better in *Adv* and *Prep* while the RG2 scored better in *Cw*, *N*, *Adj*, *Num* and *Conj*), it seems that the IG’s massive amount of exposure is offset by the cognitive maturity of the participants in the RG2. As a general rule, the productive vocabularies of immersion students at the end of pre-primary education (AO=3, with average exposure of 2,925 hours) is similar to that of IESO students (AO=8,416 hours). Cognitive maturity and more explicit teaching is needed to achieve comparable results to those of children who need more exposure to learn implicitly.

These findings show that the results of children in immersion schools, where they are massively exposed to English since a very early age, can be compared at the end of kindergarten to those earned by children in regular secondary schools (14 years old). The results show that either there are no differences in productive vocabulary or when there are they mostly in favour of the IG, mainly in terms of diversity and complexity. It seems that children who learn English through immersion in formal settings are able to produce a richer and more diverse discourse than those who learn English in a regular school.

6.4. SUMMARY

This chapter discusses the results of this research with the aim of providing insight into the lexical development in three languages of young bilingual learners studying in immersion school settings and how they compare to learners following regular school programmes. Thus, we can answer RQ1 by saying that the receptive and productive vocabularies of very young Spanish-Catalan bilingual children learning English in an immersion school setting improve as the number of hours of exposure to the language increases. Furthermore, they seem to follow the same lexical development process as monolingual and bilingual children do in their first language: they naturally learn the same types of semantic and grammatical categories, but they have smaller vocabulary sizes. When comparing them to another group of children of the same age and characteristics but who are learning English in a regular school, we found that the immersion school learners have larger vocabulary sizes in the majority of the measures tested by the MBCDI, which has been attributed to exposure. Nevertheless, we also conclude that other factors apart from exposure need to be taken into consideration, especially when analysing the semantic and word-class measures understood and produced. In some situations, the larger amount of explicit teaching in the regular school settings and greater cognitive maturity may ‘bridge the gap’ between the immersion school learners (who benefit more from implicit learning) and their regular school peers.

In order to answer RQ2, we examined the IG learners’ results in the Spanish and Catalan PPVT test. We noticed that despite the massive amount of exposure to L3, the vocabulary sizes in L1 and L2 continued to increase significantly at each data

collection time. Therefore, massive exposure to another language did not stop the children from acquiring more words in their first languages. Nevertheless, the children who attend immersion schools were found to score lower in Catalan than their regular school peers (although the effect size was small), while their Spanish scores were quite similar. This slight disadvantage in Catalan when compared to their regular school peers can be explained by the fact that Catalan is not used as a means of communication and language of instruction as in regular schools. It is also worth mentioning that L1 and L2 in Spanish-dominant learners were not found to be affected by the type of school they study in, while in Catalan-dominant students there was a significant difference between the two school types in favour of the RG1 in terms of vocabulary size. The immersive system seems to affect their receptive vocabularies in L1 and L2, which are slightly smaller than those of the participants in the regular school. We have cited other studies conducted on the effect of the early introduction of an FL on L1 and L2 lexical development, which have also provided mixed results. When the results suggest an effect of the massive amount of input in an FL on the development of the first language, the effect was found to be short-term. In this sense, we may also believe that in Catalan-Spanish the effect of the early introduction of English with a massive amount of input may be momentary and disappear in primary school, although further research is needed.

Regarding productive vocabulary, we saw that those of young learners who acquire English from very young ages through massive exposure are comparable to those of secondary school students in the first year of ESO who started later (at the age of eight) and study in regular school with a normal curriculum where English is taught as just another subject.

Even though we have to be cautious about the results we found from the three instruments used (MBCDI, PPVT and the storytelling task), they can provide information on the lexical development and attainment of young bilingual who are learning English in immersion and regular schools in Catalonia. Until now, there was no evidence in this context of learners' performance in all three languages, and what is presented in this project can constitute preliminary evidence to conduct further research.

CHAPTER 7

CONCLUSIONS AND FUTURE RESEARCH

Nowadays, more and more children are growing up in bilingual and multilingual communities. Therefore, it is of the utmost importance to study how learning more than one language at early ages may affect their overall linguistic development and the ways multilingualism can be enhanced.

Due as well to international mobility and many parents' desire for their children to learn as many languages as early as possible (especially English, as it is considered a lingua franca today), the number of immersion schools has risen in recent decades. They focus on teaching the target language using a communicative approach: the target language is also used as a means of instruction, and the amount of exposure to this language varies depending on the school (although compared to regular school curricula in Spain, for instance, it totals at least 70% of the instruction time in the first years, that is, kindergarten). It is also worth noting that in order to prevent language loss, some multilingual schools are also striving to maintain the children's native languages and therefore have stressed the importance of following the enrichment bilingual education model (Baker, 1998) as well.

Measuring these young children's lexical development is quite a challenge, given that for the time being there is an extremely limited number of testing tools especially designed for bilinguals and multilinguals. Not all tests devised to test a particular language are valid for testing others, especially when the testees have different proficiency levels in each and are children (therefore, not all test formats are valid). Care must be taken to choose tools that are reliable or that, at least in the past, have been shown to be valid for similar situations or for testing bilingual children in different languages. In this respect, researchers have advocated the need for instruments especially tailored for young bilingual and multilingual learners. Chapter 3 presented several monolingually normed tests that assess young children's receptive and productive vocabularies, which have also been used by

researchers for bilingual and trilingual learners. They seem to be able to provide some insight into their lexical development. To conduct our research, we chose the MBCDIs and the PPVT III. They were first piloted twice, which proved that could yield reliable results under the conditions that we have set for our study (Chapter 4) and for our group of participants.

In what follows, we will summarise our study and its results, while offering remarks on the implications our research has for teaching and testing. We will also mention several drawbacks of our study and offer recommendations for future research.

7.1. SUMMARY OF OUR RESEARCH

Three objectives were set for this study. Firstly, we aimed to see how the lexis of Spanish-Catalan bilingual young children studying English as FL in an immersion school develops across time. Additionally, we wanted to compare the results of the IG with those of a regular school group (RG1) the same age (end of pre-primary education), but who were learning EFL with much less exposure. The second objective was related to students' L1 and L2 (Spanish and Catalan). More precisely, we wanted to see how the L1 and L2 of IG develop across time, in terms of receptive vocabularies, and whether the massive exposure to the FL affects lexical development in L1 and L2. Finally, the third objective was to analyse and describe the lexical production in the FL of IG children at the end of kindergarten by comparing their vocabulary against that of a group of 15-year-old students (RG2) who were studying English in a regular school but had received much less exposure.

In relation to the first objective, the results show that participants in the IG improved their receptive and productive vocabularies across time. After analysing their FL lexicon in terms of MBCDI total scores in *PhU*, *WdU*, *WdP* and *M3L*, we came to the conclusion that their lexicons develop similarly to those of English L1 children, even if their receptive and productive vocabulary sizes are smaller. Looking at the semantic categories and word-class measures in the MBCDI, two aspects stood out. The first was that generally speaking, the IG's scores increased significantly across time, meaning that our participants made progress from year to year in terms of their receptive and productive vocabularies. However, no significant difference was found for some of the measures. The second aspect we noticed were the similarities

between our participants' receptive and productive vocabularies and those of their monolingual and bilingual peers. This provides evidence that the kind of vocabulary these children are learning is very close to native English-speaking children. This could be considered as an indication that these very young children acquiring EFL through immersion can actually be considered trilingual, even if they started the immersion programme after the age of three.

When comparing the IG against the RG1 in the total scores of MBCDI, the IG was significantly superior across time, which indicates that input quality and quantity plays a role in learners' vocabulary sizes. For the semantic categories and word-class measures, the results favoured the IG, with a few exceptions, in which no significant differences were found between groups. These exceptions can be accounted for by teaching style. As already indicated by Lyster (1994), apparently sometimes the more grammatical and cross-linguistic approach, which is generally used in regular curriculum schools, may facilitate acquisition, especially when time is a constraint. A formal focus on certain lexical aspects may enhance vocabulary learning in children as well.

With respect to the second objective of this thesis, the results in Spanish and Catalan show that the IG significantly improved their receptive vocabularies in L1 and L2 across time. When compared to the control group RG1, no significant differences were found for Spanish, indicating that massive exposure to the FL does not affect the IG's Spanish receptive vocabularies. On the contrary, there was a small effect size for Catalan and Spanish PPVT scores, revealing that the L1/L2 receptive vocabularies of the IG participants may lag behind. This may be a short-term effect, as previous findings in other studies have pointed out, but it is definitely worth investigating more in depth. These outcomes were also discussed in relation to participants' language dominance. They showed that Spanish-dominant immersion students performed equally well in L1/L2, and no significant difference were found between the immersion and the regular school groups in these languages. The results also showed that the Catalan-dominant students in the IG performed worse in Spanish and Catalan than their regular school peers. The poor scores in Catalan can be justified by the smaller amount of Catalan input the IG received in school. The results for Spanish are surprising, as both groups benefited from the same

amount of input. Again, this result may be a short-term effect only and should be further explored. With a few exceptions, these results show that immersion in formal settings where school policy advocates for the preservation of the bilingual education model can indeed shape early trilingual learners' proficiency in all languages.

Regarding the third objective, the results for the participants in the IG showed that by the end of kindergarten, they produced similar numbers of *Cw* and *Fw*. *N Art* and *V* were the categories that these children produced the most. These results match those of other studies conducted with monolinguals (Bates et al., 1994; David & Li, 2005), confirming again that the IG learners have similar lexicons in terms of content as the lexicons of their monolingual peers. Furthermore, these children are able to produce quite long and complex sentences in English (MLU 8.78). This is another clear indicator that language immersion schools can be appropriate environments for language learning and the acquisition of productive vocabularies that are similar to those of teenagers already in the first year of secondary education following regular instruction. The IG learners were found to produce richer, more creative, precise, diverse and complex structures than the RG2 participants. These results favouring the IG may support the hypothesis that 'the earlier the better', proving that an early start in FL learning (with considerable amount of exposure to the target language) has benefits at the lexical level. However, it should be noted that cognitive maturity is crucial for language learning, as the teenagers achieved in four years with 416 hours of exposure what the children achieved in three years with nearly 3,000 hours.

7.2. SOME POSSIBLE PEDAGOGICAL IMPLICATIONS

With respect to EFL, the results indicate, with a few exceptions, that language immersion schools are adequate places for young learners to develop their lexical competence in an FL. Yet we all know that the vast majority of these schools are extremely costly for parents, and that not everybody can afford them. On the other hand, regular schools are much less costly (some of them are even free), but the time children spend learning an FL is considerably lower, and even if the authorities decide to increase it, it would still not yield the same amount as the instruction time as in immersion schools. Under these circumstances, in order to bridge this gap

between immersion and regular schools, we believe that either CLIL approaches and adopting extracurricular practices where English is used could be feasible solutions to the problem of a lack of input in the target language. Larger amounts of exposure could help yield results similar to those in immersion schools earlier than secondary education, as seen in this study.

At the same time, the results of our research also have implications for kindergarten curriculum design and lesson-planning strategies and highlight the importance of approaching diverse themes of interest for children, which can help enrich their receptive and productive vocabularies. For instance, the Maple Bear Canadian Curriculum proposes a list of interesting topics which children between the ages of three and six study in their language immersion schools. Looking at the Catalan FL linguistic model for kindergarteners (<https://educacio.gencat.cat/web/.content/home/departament/publicacions/monografies/model-linguistic/model-linguistic-Catalunya-CAS.pdf>), we notice that it leaves a lot of room for teachers' decisions regarding topics and approaches. In this sense, English teachers can be encouraged to study the EFL curricula from immersion schools (see, for instance, Appendix 8) and adapt their content to fit their needs. Another option for kindergarten English teachers would be to use the MBCDI's content (see Appendix 2), which was especially created for English native young learners, to create their own particular thematic units. It is worthwhile to start teaching vocabulary that is frequent and may be relevant for the children at these levels.

With respect to testing these young learners in English, the MBCDI as a testing tool in schools has its advantages, as well as some drawbacks. One of its main advantages is that the final scoring can be done by a computer programme today, which saves teachers a lot of time. However, the teachers still need to fill out the questionnaires first, and this requires time and in-depth knowledge of the participants they are evaluating (this may be stressful when it is required on top of the usual teaching and assessment load). The MBCDI can also be used longitudinally, thus helping kindergarten teachers to get a comprehensive, coherent picture of students' lexical development throughout pre-primary education .

With respect to Spanish and Catalan, we believe that it is important for language immersion schools to preserve the bilingual education model, even if they provide a massive amount of EFL. The PPTV can be a good instrument to use at the beginning and the end of each education cycle to check that L1 and L2 learning are taking place equally and growing at the recommended levels. The results of our study indicate that this may usually be so, although poorer knowledge of L1/L2 has been noticed among the Catalan-dominant bilinguals who are learning the L2 in immersion schools.

We believe that kindergarten teachers can use this tool to test their pupils' receptive vocabularies in Spanish and Catalan for a series of reasons. First of all, testing can be easily done and is not time consuming. Second of all, score interpretation is transparent, and clear instructions are provided in the test manual. Third, given that the Spanish version has already been normed for Spanish-speaking populations, this test can also serve as a kind of yardstick against which young children's lexical reception can be measured. In addition, there is now a Catalan version (Serrat et al., 2022), which was not available at the time the data were collected. Finally, this test can be used longitudinally at different junctures, which can provide teachers with a more complete and reliable picture of students' lexical development in Spanish and Catalan.

7.3. LIMITATIONS OF THE STUDY

It should be also acknowledged that there are several different limitations to this study. First of all, there is a dearth of studies assessing vocabulary development in multilingual young children over an extended period of time (and there is a lot of variability in terms of ages, tests, languages, proficiency levels or contexts in which languages were acquired). Most of them are case studies or studies involving a limited number of children. Therefore, it was difficult to establish comparisons between our participants and data already published. It was also hard to make inferences from previous findings, as highlighted by previous research. In relation to the design, the fact that this research was longitudinal and set in a classroom context entailed challenges that do not normally occur in experimental settings. The selection of testing tools was also problematic due to the initial lack of instruments measuring what we intended to measure.

Our study was designed longitudinally across three years, and at the beginning, we selected 60 participants in order to get results from a considerable sample of participants. Selecting them was not easy: it is common practice for parents whose children attend private immersion schools to also sign their children up for extracurricular English lessons and hire English-speaking nannies to spend time with the children in the afternoons and evenings while they are at work. Because we wanted to make sure that the only input in English these children received was from the school, we needed to ask their parents for confirmation. We assumed that the parents provided honest answers.

As the population attending private schools is quite unstable, we found ourselves in the situation that nine of the students initially selected left the school during the three years our study took place. Therefore, we ended up with 51 longitudinal participants. Nevertheless, we believe that the data collected from them provided reliable results for our study, as they constitute a representative sample. The fact that these children attended only one school is also a limitation. However, we wanted them to share exact conditions, and had students from different immersion schools been selected, further variability would have been introduced into the sample.

Choosing the appropriate testing tools was also a challenge. Given the lack of standardised assessment tools that test young children's comprehension and production in EFL, we relied on the MBCDI, which has been normed for monolingual and some bilingual (English- Spanish) infants and toddlers. The two pilot tests conducted before the main study showed us that this test could provide reliable results for this age group in the FL. Yet we believe that there is always a small margin of error with this type of assessment, which is based upon recall. For this reason, two assessors were used. However, even if data were collected by the researcher who knew the linguistic profile of the children and then double-checked by the native teacher, it is still possible that the child knows more than what s/he demonstrated in class and that his/her knowledge has been underestimated.

With respect to the testing in Spanish and Catalan, we had to rely on the Spanish version of the PPVT III and its Catalan translation, as this test is not normed for

these bilingual populations. Furthermore, the assessment in both Spanish and Catalan was conducted face-to-face on a one-to-one basis, which may also have caused some distress to the children, who may have felt intimidated and may not have shown everything they knew. Another limitation comes from the fact that some children in the immersion school had started being exposed to English before P3 (even though with fewer hours of exposure), and that the participants from the regular schools (RG2) came from three different high schools in Barcelona (as we looked for participants with specific conditions, such as not having attended extracurricular lessons or been abroad to English-speaking countries). For this reason, we conducted further statistical analyses to confirm that these students could be grouped together. Of course, we would have liked to have bigger samples, as some of the groups are small (especially if they are broken up into subgroups, as they are for language dominance).

Despite these limitations, we think that the research was worth conducting due to the lack of studies with early multilinguals. Even if the conclusions cannot be taken merely at face value, they can serve as starting point to conduct further research with these populations.

7.4. ORIGINAL CONTRIBUTION OF THE RESEARCH

In spite of the limitations mentioned in the previous section, this study provides original and new information on the lexical acquisition of very young bilingual EFL learners in immersion schools, in terms of both language development and testing tools. To our knowledge, there are no longitudinal studies following kindergarten groups of bilingual children acquiring English as an FL in immersion schools. Therefore, our study not only provides a picture of vocabulary learning in these young learners across time, but it also looks into how the three languages develop in these specific circumstances, which, in turn, are a reality for a growing number of children nowadays.

Our study has shown that in terms of FLA, children who study an FL in immersion schools follow similar patterns as those of monolingual and bilingual children and learn similar vocabulary. In general, they produce significantly more words than children of the same age learning English following the state school regular

curriculum. Furthermore, at the end of the three kindergarten years, their speech in English tends to be richer and more varied than that of older children who learn English in a regular curriculum school.

With respect to the findings in Spanish and Catalan, our study is, to our knowledge, among the very few conducted with young bilingual participants looking at the impact that the early introduction of an FL may have on the development of other languages. It should be noted that most studies doing so are case studies, and that the sample in this study is much bigger (though limited to 51 students).

In relation to the structure of the multilingual lexicon of young bilingual learners who acquire an FL in immersion schools, this study only outlines several basic factors regarding the relationship between the three languages, contributing to the research already conducted in this field (using other inventories or parental diaries) from a different perspective. More precisely, we have shown that despite the early introduction of an FL through massive exposure, the lexicons of the first two languages continue to grow significantly across time.

The testing tools have been a secondary but important aim in this study, as they had to be carefully selected, and the results provided had to be cautiously compared to those from previous research. At the time this study was set up, there were no standard lexical assessment tools normed for multilingual young children. Therefore, we tried to find a solution to this problem by using reliable instruments that would give us an indication of development in all three languages, even if the same tool could not be used for all three of them. For English, the longitudinal use of the MBCDI yielded reliable results for this educational period. Furthermore, complementing it with an oral production task in the last year of kindergarten has provided us with a more complete picture of what immersion children are able to do at the end of pre-primary education. The PPVT III helped to describe the children's receptive vocabularies in Spanish and Catalan, even though it has not been normed for bilingual populations and the Catalan version is just a translation from the Spanish more than an adaptation. The combination of these three testing tools (previously piloted with similar populations to the main study) helped us conduct the study with some confidence. Therefore, given the results of this thesis,

it can be stated that given the lack of multilingually normed lexical assessment tools, these instruments can be safely be used both by researchers and teachers alike to assess these young multilingual populations.

7.5 FURTHER RESEARCH AND FINAL REMARKS

There are still many different aspects that this thesis has not managed to cover. Indeed, the main focus here was the receptive and productive lexical knowledge in English of bilingual learners who are acquiring the FL in immersion schools. Yet to date, there are very few studies focusing on the FLA of kindergarten bilinguals in different educational contexts. Here, we have compared two very different contexts (immersion vs non-immersion). However, introducing other contexts ‘between’ these two could be an interesting way to see how vocabularies develop when different amounts of exposure are provided in formal contexts: for example: immersion, semi-immersion and CLIL.

With respect to the Spanish and Catalan lexicons and the consequences of the early introduction of a massive amount of English input, it would be interesting to see whether the results from this study would hold or differ in primary education (i.e., whether L1/L2 vocabularies of Catalan-dominant children have recovered, as in other studies, or whether, to the contrary, the early introduction of the FL causes a more profound effect on these languages). It would also be worthwhile to check whether Spanish-dominant children experience the same phase as their Catalan-dominant peers, or whether the results found here were just due to the small size of the two groups, while language dominance does not actually play a role in L1/L2 vocabulary development when attending immersion schools. This additional research could expand the view presented in this doctoral thesis and add to its value. At the same time, in order to complete the picture of the lexical development in our participants’ L1 and L2, having an insight into L1 and L2 lexical production may also be helpful, perhaps by using storytelling (the same task here) but in the first languages.

Furthermore, exploring the relationship between these children’s three languages more in depth could also be of interest. This could be done by taking language errors into account (and paying attention to cross-linguistic influence) or by interviewing

the children's parents in relation to their children's linguistic development. Hopefully, research along these lines will be conducted in the near future to help further account for the complex process of acquiring three languages at early ages.

REFERENCES

- Abunuwara, E. (1992). The structure of the trilingual lexicon. *European Journal of Cognitive Psychology*, 4(4), 311-322.
- Aguila, E., Ramon, M., Pons, F., & Bosch, L. (2005). Efecto de la exposición bilingüe sobre el desarrollo léxico inicial. In A. Mayor, Cinca, B., Zubiauz de Pedro & E. Diez-Itra (Eds.), *Estudios sobre la adquisición del Lenguaje* (pp. 676-692). Universidad de Salamanca.
- Allan, D. (2006). *Oxford placement tests*. OUP.
- Andreu, L., Sanz-Torrent, M., Guardia Olmos, J., & MacWhinney, B. (2011). Narrative comprehension and production in children with SLI: an eye movement study. *Child Linguist Phonology*, 25(9), 767-783.
- Aronin, L., & O'Laoire, M. (2004). Exploring multilingualism in cultural contexts: towards a nation of multilinguality. In C. Hoffmann & Y. Jehannes (Eds.), *Trilingualism in family, school and community* (pp. 13-29). Multilingual Matters.
- Appel, R., & Vermeer, A. (1998). Speeding up second language vocabulary acquisition of minority children. *Language and Education*, 12 (3), 159–173.
- Baker, C. (1988). *Key issues in bilingualism and bilingual education*. Multilingual Matters.
- Baker, C (2006). *Foundations of bilingual education and bilingualism*. Multilingual Matters.
- Baker, C., & de Kanter, A. A. (1981). *Effectiveness of bilingual education: A review of literature*. Washington D.C.: Office of Planning, Budget and Evaluation, U.S. Department of Education.
- Barbu, C., González, A., Gillet, S., & Poncelet, M. (2019). Cognitive advantages in children enrolled in a second language immersion elementary school program for one year. *Psychologica Belgica*, 59(1), 416-435.
- Barnes, J. D. (2006). *Early trilingualism. A focus on questions*. Multilingual Matters.
- Bartolotti, J., & Marian, V. (2016). Bilinguals' existing language benefit vocabulary learning in a third language. *Language Learning*, 67(1), 1-31.

References

- Benson, M. S. (1997). Psychological causation and global-based episodes. Low income children's emerging narrative skills. *Early Child Research Quarterly*, 12(4), 439-457.
- Bergman, C. (1976). Interference vs. independent development in infant bilingualism. In G. Keller, R. Taschner & S. Viera (Eds.), *Bilingualism in the bicentennial and beyond* (pp. 86-96). Bilingual Press.
- Bergstörm, K., Klatte, M., & Steinbrink, C. (2016). Effects of language immersion in kindergarten. *Language Learning*, 66(2), 386-418.
- Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading*, 9(1), 43-61.
- Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13, 525-531.
- Bishop, D. V. M. (2003). *Test for reception of grammar TRGO2*. Pearson Assessment.
- Bloomfield, L. (1933). *Language*. Allen and Unwin.
- Brown, R. (1973). *A first language: The early stages*. Harvard University Press.
- Bongaerts, T., Planken, B., & Schils, E. (1995). Can late starters attain a native accent in a foreign language? A test of the Critical Period Hypothesis. In D. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition* (pp. 30-50). Multilingual Matters.
- Bosch, L., & Sebastian-Galles, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, 2(1), 29-49.
- Butler, Y. G. (2013). Bilingualism/ multilingualism and second language acquisition. In K. T. Bathia & W. C. Ritchie (Eds.), *The handbook of bilingualism and multilingualism* (pp. 109-136). Blackwell Publishing.
- Butler, Y. G., & Hakuta, K. (2006). Bilingualism and second language acquisition. In K. T. Bathia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 114-144). Blackwell Publishing.
- Byers-Heinlein, K. (2020). Challenges of infant language acquisition in bilingual environments. <https://doi.org/10.31234/osf.io/fuhjb>
- Byers-Heinlein, K., & Wreker, J. F. (2009). Monolingual, bilingual, trilingual: infants' language experience influences the development of a word-learning heuristic. *Developmental Science*, 12(5), 815-823.

- Byers-Heinlein, K., & Lew-Williams, C. (2013). Bilingualism in the early years: what the science says. *Learning Landscapes*, 7(1), 95-112.
- Cantone, K. F. (2007). *Code switching in bilingual children*. Springer.
- Caramazza, A., & Brones, I. (1979). Lexical access in bilinguals. *Bulletin of the Psychonomic Society*, 13(1), 212-214.
- Caselli, C., Cassadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language*, 26, 69-111.
- Castilla, A. P., Restrepo, M. A., & Perez-Leroux, A. T. (2009). Individual differences and language interdependence: a study of sequential bilingual development in Spanish-English preschool children. *International Journal of Bilingual Education and Bilingualism*, 12(5), 565-580.
- Cenoz, J. (1997). L'acquisition de la troisième langue: Bilinguisme et plurilinguisme au Pays Basque. *AILE*, 10 (1), 159-180.
- Cenoz, J. (2000). Research on multilingual acquisition. In J. Cenoz & U. Jessner, (Eds.), *English in Europe: The acquisition of a third language* (pp. 39-53). Multilingual Matters.
- Cenoz, J. (2003). The influence of age on the acquisition of English: general proficiency, attitudes and code mixing. In M. P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language: Theoretical issues and field work* (pp. 77-93). Multilingual Matters.
- Cenoz, J. (2003b). Facteurs déterminants l'acquisition d'une L3: Age, développement cognitif et milieu. *AILE*, 18, 38-51.
- Cenoz, J. (2005) English in Bilingual Programs in the Basque Country. *International Journal of the Sociology of Language*, 171, 41-56.
- Cenoz, J. (2013). The influence of bilingualism on third language acquisition: Focus on Multilingualism. *Language Teaching*, 46 (1), 71-86.
- Cenoz, J., & Genesee, F. (1998). Psycholinguistic perspectives on multilingual education. In J. Cenoz & F. Genesee (Eds.), *Beyond bilingualism: Multilingualism and multilingual education* (pp. 16-32). Multilingual Matters.
- Cenoz, J., Hufeisen, B., & Jessner, U. (2003). Why investigate the multilingual lexicon? In J. Cenoz, B. Hufeisen & U. Jessner (Eds.), *The multilingual lexicon* (pp. 1-9). Kluwer.
- Chon, H.C., Sawyer, J., & Ambrose, N.G. (2012) Differences of articulation rate and utterance length in fluent and disfluent utterances of preschool children who stutter. *Journal of Communication Disorders*, 45 (6), 455-467.

- Costa, A. (2006). Speech production in bilinguals. In T. Bhatia & W.C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 201-223). Blackwell.
- Cote, L. S., Gonzalez- Barrero, A.M., & Byers-Heinlein, K. (2022). Multilingual toddlers' vocabulary development in two languages: Comparing bilinguals and trilinguals. *Journal of Child Language*, 49, 114-130.
- Clark, E. (1987). The principle of contrast: A constraint of language acquisition. In B. MacWhinney (Eds.), *Mechanisms of language acquisition* (pp. 1-33). Lawrence Erlbaum Associates.
- Clark, E. (2003). *First language acquisition*, CUP.
- Comeau, L., Genesee, F., & Lapaquette, L. (2003). The Modelling Hypothesis and child bilingual codemixing. *International Journal of Bilingualism*, 7, 113-126.
- Conboy, B. T., & Thal, D. (2006). Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Development*, 77, 712-735.
- Craig, H., & Washington, J. A. (1994). The complex syntax skills of poor urban African American pre-schoolers at school entry. *Language Speech and Hearing Services in Schools*, 25(3), 171-180.
- Cruz-Ferreira, M. (2006). *Three is a crowd? Acquiring Portuguese in a trilingual environment*. Multilingual Matters.
- Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, 19, 121-129.
- Cureton, S. M., & Lucas, T. D. (2007). Assessing young children's oral narrative skills- The story pyramid framework, In K. L. Pen (Eds.), *Assessment in emergent literacy* (pp. 377-432). Plural Publishing.
- David, A., & Li, W. (2003). To what extent is code-switching dependent on a bilingual child's lexical development? *Sociolinguistica*, 18, 1-12.
- David, A & Li, W. (2005). The composition of the bilingual lexicon. Retrieved from <http://www.cascadilla.com/isb4.html>.
- David, A., & Li, W. (2008). Individual differences in the lexical development of French-English bilingual children. *International Journal of Bilingual Education and Bilingualism*, 11, 598-618.
- De Angelis, G. (2007). *Third or additional language acquisition*. Multilingual Matters.
- De Cat, C. (2021). Socioeconomic status as a proxy for input quality in bilingual children? *Applied Psycholinguistics*, 42, 301-324.

- De Bot, K. (2014). The effectiveness of early foreign language learning in the Netherlands. *Studies in the Second Language Learning and Teaching*, 4(3), 409-418.
- De Groot, A., & Hoecks, J. (1995). The development of bilingual memory: evidence from word translation by trilinguals. *Language Learning*, 45(4), 683-724.
- De Houwer, A. (1990). *The acquisition of two languages from birth: A case study*. Cambridge University Press.
- De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics*, 28(3), 411-424.
- De Houwer, A. (2009). *Bilingual First Language Acquisition*. Multilingual Matters.
- De Houwer, A., Bornstein, M. H., & De Coster, S. (2006). Early understanding of two words for the same thing: A CDI study of lexical comprehension in infant bilinguals. *International Journal of Bilingualism*, 10(3), 331-347.
- De Houwer, A., Bornstein, M. H., & Putnick, D. L (2014). A bilingual-monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics*, 35, 1189-1211.
- Deuchar, M., & Quay, S. (2000). *Bilingual acquisition: Theoretical implications of a case study*. OUP.
- Dewaele, J. M. (2000). *Trilingual first language acquisition: exploration of a linguistic miracle*. La Chouette.
- Dijkstra, T. (2003). Lexical processing in bilinguals and multilinguals: the word selection problem. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *The Multilingual lexicon* (pp. 11-26). Kluwer Academic Publishers.
- Dixon, Q. (2011). The role of home and school factors in predicting English vocabulary among bilingual kindergarten children in Singapore. *Applied Psycholinguistics*, 32, 141-168.
- Doiz, A., & Lasagabaster, D. (2004). The effect of the early learning of the foreign language on written production. *The International Journal of Bilingualism*, 8, 527-542.
- Doyle, A., Champagne, M., & Segalowitz, N. (1978). Some issues in the assessment of linguistic consequences of early bilingualism. In M. Paradis (Eds.), *Aspects of bilingualism* (pp.13-20). Hornbeam Press.
- Dunn, L. M., & Dunn, L., M. (1997). *Peabody Picture Vocabulary Test (3rd ed.)*. Circle Pines.

References

- Dunn, L. M., Dunn, L. M., & Altarriba, D. (2010). *Test Peabody de vocabulario en imágenes*. TEA Ediciones.
- Fenson, L., Dale, P. S., Reznick, J. S., Donna, T., Bates, E., Hartung, J., Pethick, S. J. I., & Reilly, J. (1993/2007). *The MacArthur Bates Communicative Development Inventories: User's guide and technical manual*. Singular Publishing.
- Fiestas, C. E., & Pena, E. (2004). Narrative discourse in bilingual children. *Language, Speech and Hearing Services in Schools*, 35, 155-168.
- Foroodi, N. F., & Paradis, J. (2009). Crosslinguistic transfer in the acquisition of compound words in Persian-English bilinguals. *Bilingualism*, 12(4), 411-427.
- Frankenburg, W. K., Doods, J., & Archer, P. (1992). The Denver II: A major revision and restandardization of the Devnver Developmental Screening Test. *Pediatrics*, 89, 91-97.
- Friend, M., & Keplinger, M. (2003). An infant-based assessment of early lexicon acquisition. *Behaviour Research Methods, Instruments, and Computers*, 35(2), 302–309.
- Frota, S., Butler, J., Correia, S., Severino, C., Vicente, S., Vigario, M. (2016). Infant communicative development assessed with the European Portuguese MacArthur–Bates Communicative Development Inventories short forms. *First Language*, 36(5), pp 525-545.
- Gabrys- Barker, D. (2019). Applied linguistics and multilingualism. In D. Singleton & L. Aronin (Eds.), *Twelve lectures on multilingualism* (pp.165-17). Mm Textbooks.
- Galeote, M., Soto, P., Serrano, A., Pulido, L., Rey, R., & Martinez-Roa, P. (2006). Un nuevo instrumento para evaluar el desarrollo comunicativo y lingüístico de niños con síndrome de Down. *Revista Síndrome de Down*, 23, 20-26.
- Gallardo del Puerto, F. (2007). Is L3 phonological competence affected by learner's level of bilingualism? *International Journal of Multilingualism*, 4(1), 1-16.
- García Mayo, M. P. (2003). Age, length of exposure and grammaticality judgements in the acquisition of English as a foreign language. In M. P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 94-114). Multilingual Matters.
- García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Wiley/Blackwell
- Gatt, D., Baldacchino, R., & Dodd, B. (2020). Which measure of socioeconomic status best predicts bilingual lexical abilities and how? A focus on four-year-

References

- olds exposed to two majority languages. *Journal of Child Language*, 47, 737-765.
- Genesee, F. (1983). Bilingual education of majority- language children: The immersion experiments in review. *Applied Psycholinguistics*, 4, 1-46.
- Genesee, F., Boivin, I., & Nicoladis, E. (1996). Talking with strangers: A study of bilingual children's communicative competence. *Applied Psycholinguistics*, 17, 427-442.
- Gibson, T. A., Oller, D. K., Jarmulowicz, L., & Ethington, C. (2012). The receptive-expressive gap in the vocabulary of young second language learners: Robustness and possible mechanisms. *Bilingualism: Language and Cognition*, 15(1), 102-116.
- Greene, J. P. (1997). A meta-analysis of the Rossell and Baker review of bilingual education research. *Bilingual Research Journal*, 21(2), 103-122.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36, 3-15.
- Grosjean, F. (2010). *Bilingual: life and reality*. Harvard University Press.
- Harley, B. & Swain, M. (1984). An analysis of verb form and function in speech of French immersion pupils. *Working Papers in Bilingualism*, 14, 31-56.
- Harris, M., & Chasin, J. (1999). Developments in early lexical comprehension: a comparison of parental report and controlled test. *Journal of Child Language*, 22 (6), 453-460.
- Heaton, J. B. (1966). *Composition through pictures*. Longman
- Heilmann, J., Ellis W. S., Evans, J., & Hollar, C. (2005). Utility of the MacArthur Bates Communicative Development Inventory in identifying language abilities of late-talking and typically developing toddlers. *American Journal of Speech-Language Pathology*, 14(1), 40-51.
- Herdina, P., & Jessner, U. (2000). The dynamics of third language acquisition. In J. Cenoz & U. Jessner (Eds.), *English in Europe- The acquisition of a third language* (pp. 84-98). Multilingual Matters
- Herdina, P. & Jessner, U. (2002). *A dynamic model of multilingualism- Perspectives of change in psycholinguistics*. Multilingual Matters.
- Hernandez-Calafat, M. (2021). El cerebro del bebe multilingue. *Journal of Neuroeducation*, 1(2), 63-68.
- Hervig, A. (2001). Plurilingual lexical organization. In J. Cenoz, B. Hufeisen & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 115-137). Multilingual Matters.

- Hoff, E., & Elledge, C. (2005). Bilingualism as one of many environmental variables that affect language development. In J. Cohen, K T. McAlister, K. Rolstad & J. MacSwan (Eds.), *Proceedings of the 4th International Symposium on Bilingualism* (pp. 1034-1040). Cascadilla Press.
- Hoff, E., Core, C., Palce, S., Rumiche, R., Senor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39, 1-27.
- Hoffmann, C. (1985). Language acquisition in two trilingual children. *Journal of Multilingual and Multicultural Development*, 6(6), 479-495.
- Hoffmann, C., & Stavans, A. (2007). The evolution of trilingual code-switching from infancy to school age: the shaping of trilingual competence through dynamic language dominance. *International Journal of Bilingualism*, 11(1), 55-72.
- Holowka, S., Brosseau-Lapre, F., & Petitto, L. A. (2002). Semantic and conceptual knowledge underlying bilingual babies' first signs and words. *Language Learning*, 52, 205-262.
- Hufeisen, B. (2005). Models of multilingual competence. In A. Bonnet & P. Siemund (eds.), *Foreign language education in multilingual classrooms* (pp. 173-189). John Benjamins.
- Hufeisen B., & Jessner (2019), The psycholinguistics of multiple language learning and teaching. In D. Singleton & L. Aronin (Eds.), *Twelve lectures on multilingualism* (pp.). Mm textbooks
- Hyltenstam, K. (1988). Lexical characteristics of near-native second-language learners of Swedish. *Journal of Multilingual and Multicultural Development*, 9(1), 67-84.
- Hyltenstam, K. (1992). Non-native features of near-native speakers: On the ultimate attainment of childhood L2 learners. In R. J. Harris (Eds.), *Cognitive processing in bilinguals* (pp. 351-368). Elsevier.
- Johnson, C., & Lancaster, P. (1998). The Development of More than One Phonology: A Case Study of a Norwegian-English Bilingual Child. *The International Journal of Bilingualism*, 2(3), 265- 300.
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21(1), 60–99.
- Jessner, U. (2008). Teaching third languages: Findings, trends and challenges. *Language Teaching*, 41(1), 15-56.

- Junker, D., & Stockman, I. (2002). Expressive vocabulary in German-English bilingual toddlers. *American Journal of Speech-Language Pathology*, 11, 381-394.
- Kartushina, N. (2019). Third language learning affects bilinguals' production in both their native languages: A longitudinal study of dynamic changes in L1, L2 and L3 vowel production. *Journal of Phonetics*, 77, 12-25.
- Kastos, N. & Alexopoulou, D. (2011). Different patterns of L2 learning: language cognition and education. Bilingual Education Symposium University of Cambridge, 19-20 April 2011. Retrieved from <http://www.cie.org.uk/images/85383-different-patterns-of-l2-learning-language-cognition-and-education.pdf>
- Kauschke, C., & Hofmeister, C. (2002). Early lexical development in German: a study on vocabulary growth and vocabulary composition during the second and third year of life. *Journal of Child Language*, 29, 735-757.
- Kemp, C. (2009). Defining multilingualism. In L. Aronin & B. Hufeisen (Eds.), *The exploration of multilingualism* (pp. 11-26). John Benjamins.
- Kessler, C. (1984). Language acquisition processes in bilingual children. *Bilingual Education Paper Series*, 7 (6), 3-46.
- Kester, E. S., & Pena, E. D. (2002). Language ability assessment of Spanish-English bilinguals: future directions. *Practical Assessment, Research and Evaluation*, 8(4), 1-5.
- Kohnert, K (2002). Picture naming in early sequential bilinguals: A 1 year-follow-up. *Journal of Speech, Language and Hearing Research*, 45(1), 759-771.
- Kohnert, K., Kan, F. P., & Conboy, T. B. (2010). Lexical and Grammatical Associations in Sequential Bilingual Pre-schoolers. *Journal of Speech, Language and Hearing Research*, 53(1), 684-693.
- Kroll, J.F., & Dussias, P. E. (2006). The comprehension of words and sentences in two languages. In T. Bhatia & W.C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 169-200). Blackwell.
- Lambert, E. W., & Tucker, G.R. (1972). *Bilingual education of children: The St. Lambert experiment*. Newbury House Publishers.
- Lanza, E. (1997). Language mixing in infant bilingualism: a sociolinguistic perspective. Clarendon.
- Lanza E. (2001). Bilingual first language acquisition: A discourse perspective on language contact in parent-child interaction. In J. Cenoz & F. Genesee (Eds.), *Trends in bilingual acquisition* (pp. 201-320). John Benjamins.

- Lasagabaster, D. (2005). *Trilingualism at school: the place of English in the Basque education*. Retrieved from www.euskara.euskadi.eus/r593693/en/contenidos/informacion/artik20_1_trilinguismoa_07_06/en_trilingu/adjuntos/trilingualismatschool.pdf. Last accessed May 2022.
- Lasagabaster, D., & Doiz, A. (2003). Maturational constraints in foreign language writing. In M. L. García Lecumberri & M. P. García Mayo (Eds.), *Age and the acquisition of English as a foreign language* (pp. 136-160). Multilingual Matters.
- Lenneberg, E. H. (1967). *Biological foundations of language*. Wiley
- Leopold, W. (1949/1970). *Speech development of a bilingual child. A linguist's record*. AMS Press.
- Leopold, W. (1953). Patterning in children's language learning. *Language Learning*, 5, 1-14.
- Litcofsky, K. A., Tanner, D., & Van Hell, G. J. (2015). Effects of language experience, use, and cognitive functioning on bilingual word production and comprehension. *International Journal of Bilingualism*, 20(6), 666-683.
- Luyster, R., Qiu, S., Lopez, K., & Lord, C. (2007). Predicting outcomes of children referred for autism using the MacArthur-Bates communicative development inventory. *Journal of Speech, Language and Hearing Research*, 50(3), 667-681.
- Lyster, R. (1994). The effect of functional-analytic teaching in aspects of French immersion students' sociolinguistic competence. *Applied Linguistics*, 15, 263-287.
- Lyster, R. (2007). *Learning and teaching language through content: A counterbalanced approach*. John Benjamins.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analysing talk* (3rd ed.). Lawrence Erlbaum Associates.
- MacWhinney, B., & Snow, C. (1990). The Child Language Data Exchange System: An update. *Journal of Child Language*, 17(2), 457-472.
- Mahon, M., & Crutchley, A. (2006). Performance of typically-developing school-age children with English as an additional language (EAL) on the British Picture Vocabulary Scales II (BPVS II). *Child Language Teaching and Therapy*, 22(3), 1-19.
- Malvern, D., Richards, B., Chipere, N., & Duran, P. (2004). *Lexical Diversity and Language Development: Quantification and Assessment*. Macmillan.

- Mancilla-Martinez, J., Alexander-Pan, B., & Banu, V. S. (2010). Assessing the productive vocabulary of Spanish – English bilingual toddlers from low income families. *Applied Psycholinguistics*, 32(2), 333-357.
- Marianis, T., & Chondrogianni, V. (2010). Production of tense marking in successive bilingual children: when do they converge with their monolingual peers? *International Journal of Speech-Language Pathology*, 12(1), 19–28.
- Mariscal Altares, S., Nieva, S., & Lopez-Ornat, S. (2010). Observar y medir el desarrollo gramatical temprano en español. *Psicothema*, 22(1), 51-56.
- Martin, N. A. (2013). *Receptive One-Word Picture Vocabulary Test* (4th Edition): *Spanish bilingual Edition* (ROWPVT-4: SBE). ATPONLINE.
- Mayor, J., & Plunkett, K. (2010). A statistical estimate of infant total vocabulary size from CDI analysis. *Developmental Science*, 14(3), 1-17.
- McClain, J. B., Oh, M. H., Mancilla-Martinez, J. (2021). Questioning the monolingual norm with conceptually scored bilingual vocabulary assessments: Findings from a research-practice partnership. *Tesol Journal*, <https://doi.org/10.1002/tesj.585>.
- McLaughlin, B. (1984). Early bilingualism: Methodological and theoretical issues. In M. Paradis & Y. Lebrun (Eds.), *Early bilingualism and child development* (pp. 19-45). Swets and Zeitlinger.
- Meara, P.M., & Miralpeix, I. (2004). *D_Tools: The Manual*. Retrieved from <http://www.lognostics.co.uk/tools/>.
- Meara, P.M. & Miralpeix, I. (2017). *Tools for researching vocabulary*. Multilingual Matters.
- Meisel, J. (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). CUP.
- Meisel, J. (2001). The simultaneous acquisition of two first languages: Early differentiation and subsequent development of grammar. In J. Cenoz & F. Genessee (Eds.), *Trends in Bilingual Acquisition* (pp. 11-41). John Benjamins.
- Mikes, M. (1990). Some issues of lexical development in early bi- and trilinguals. In G. Conti-Ramsden & C. Snow (Eds.), *Children's Language vol. 7* (pp. 103-120). Gardner Press.
- Minami, M. (2005). Bilingual narrative development in English and Japanese- A form/function approach. In J. Cohen, K. T. McAlister, K. Rolstad & J. MacSwan (Eds.), *Proceedings of the 4th International Symposium on Bilingualism* (pp. 1618-1629). Cascadia Press.

- Miralpeix, I. (2006). The influence of age on vocabulary acquisition in English as a foreign language. In C. Muñoz (Eds.), *Age and rate of foreign language learning* (pp. 1-40). Multilingual Matters.
- Miralpeix, I. (2008). The influence of age on vocabulary acquisition in EFL: Estimating vocabulary size. AAL Annual Conference (American Association for Applied Linguistics). Washington, USA.
- Miralpeix, I (2020). L1 and L2 vocabulary size and growth. In Webb, S. (Eds.), *The Routledge handbook of vocabulary studies* (pp. 189-206). Routledge.
- Montanari, S. (2006). Language differentiation in early trilingual development: Evidence from a case study. Dissertation presented to the faculty of the Graduate School University in Southern California.
- Montanari, S. (2014). A case study of bi-literacy development among children enrolled in an Italian-English dual language program in Southern California. *International Journal of Bilingual Education and Bilingualism*, 17(5), 509-52.
- Muñoz, C. (2000). Bilingualism and trilingualism in school students in Catalonia, In J. Cenoz and U. Jessner (Eds.), *English in Europe- the acquisition of a third language* (pp. 157-178). Multilingual Matters.
- Muñoz, C. (2006). The effects of Age on foreign language learning: BAF project (2006). In C. Muñoz (Eds.), *Age and rate of foreign language learning* (pp. 1-40). Multilingual Matters.
- Muñoz, C. & Singleton, D. (2019). Age and Multilingualism. In D. Singleton & L. Aronin (Eds.) *Twelve lectures on multilingualism*. Multilingual Matters.
- Muñoz-Sandoval, A., Cummins, J., Alvarado, C. G., & Ruef, M. L. (2005). *Bilingual verbal ability test*. Riverside Publishing.
- Murrell, M. (1966). Language acquisition in a trilingual environment: Notes from a case study. *Studia Linguistica*, 20, 9-35.
- Murphy V. (2014). *Second Language Learning in the early school years*. OUP.
- Nicoladis, E., & Secco, G. (2000). The role of a child's productive vocabulary in the language choice of a bilingual family. *First Language*, 58, 3-28.
- Ogura, T. (1998). *The development and study of the MacArthur-Bates Communicative Development Inventories*. Kobe University Press.

References

- Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology Vol. I: Production* (pp. 93–112). Academic Press.
- Ortega Duran, M. (2017). Is the acquisition of multiple languages a challenge? *Verbeia, 1*(1), 143-159.
- Padilla, A., & Lindholm, K. (1984). Child bilingualism: The same old issues revisited. In J. L. Martinez & R. Mendoza (Eds.), *Chicano psychology* (pp. 369-408). Academic Press.
- Palma, P., Marin, M.F., Onishi, K.H., Titone, D. (2022). Learning, Inside and Out: linguistic knowledge and learning environment impact word learning in bilingual individuals. *Language Learning, 72*(4), 890-11016.
- Pavicic, V. (2008). *Vocabulary Learning Strategies and Foreign Language Acquisition*. Clevedon: Multilingual Matters.
- Pavlenko, A. (2008). Narrative Analysis. In W. Li & G. M. Moyer (Eds.), *The Blackwell guide to research methods in bilingualism and multilingualism* (pp. 311-325). Blackwell.
- Patterson, J. (1998). Expressive vocabulary development and word combinations of Spanish-English bilingual toddlers. *American Journal of Speech-Language Pathology, 7*, 46-56.
- Patterson, J. (1999). What bilingual toddlers hear and say: Language input and word combinations. *Communication Disorders Quarterly, 21*, 32-28.
- Patterson, J. (2000). Observed and reported expressive vocabulary and word combinations in bilingual toddlers. *Journal of Speech, Language and Hearing Research, 43*, 121-128.
- Patterson, J. (2002). Relationships of expressive vocabulary to frequency of reading and television experience among bilingual toddlers. *Applied Psycholinguistics, 23*(4), 493- 508.
- Peal, E., & Lambert, W.E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs, 76*, 1-23.
- Pearson, B. Z., & Fernandez, S. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language Learning, 44*(4), 617-653.
- Pearson, B. Z., Fernandez, S., & Oller, D.K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43*, 93-120.
- Pearson, B. Z., Fernandez, S., & Oller, D.K. (1995). Cross-language synonyms in the lexicons of bilingual infants: one language or two? *Journal of Child Language, 22*, 345-368.

- Pearson, B. Z., Fernandez, S., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, 18(1), 41-58.
- Perez-Pereira, M. (2008). Early Galician/ Spanish bilingualism: contrasts with monolingualism. In C. Vidal, M. Garau & A. Bel (Eds.), *A portrait of the young in the new multilingual Spain. Issues in the acquisition of two or more languages in multilingual environments* (pp. 39-62). Multilingual Matters.
- Perez-Vidal, C. (1995). La adquisición del inglés de un niño bilingüe catalán-inglés. Doctoral dissertation, University of Barcelona.
- Perez-Vidal, C., & Garau, M. J. (2002). A contrastive study of tense-aspect in the Catalan/English oral production of a young bilingual. *Atlantis*, 24(2), 169-182.
- Perez, D., Mendoza Lara, E., Carballo, G., Fresneda Lopez, M. D., Castro Munoz, J. (2012). Repetición de pseudopalabras en niños con síndrome de Down. *Revista de Lingüística, Filología y Traducción de la Pontificia Universidad Católica de Chile*, 26, 377-390.
- Pham, G. (2016). Pathways for learning two languages: lexical and grammatical associations within and across languages in sequential bilingual children. *Bilingualism: Language and Cognition*, 19(5), 928-238.
- Pinto Artigao, J. (2013). Cross-linguistic influence at lexical level. A study with Moroccan learners of Portuguese as an L3/LN. *Revista Nebrija de Lingüística Aplicada*, 12, 3-50.
- Pfenniger, S. E. (2014). The misunderstood variable: Age effects as a function of type of instruction. *Studies in Second Language Learning and Teaching*, 3, 529-556.
- Platt, A. (2010). Examining the appropriateness of the PPVT-III as a measure of vocabulary development in linguistically diverse kindergarten children. Doctoral Thesis, University of Toronto.
- Poulin-Dubois, D., & Goodz, N. (2001). Language differentiation in bilingual infants: Evidence from babbling. In J. Cenoz & F. Genesee (Eds.), *Trends in bilingual acquisition* (pp. 95-106). John Benjamins.
- Poulin-Dubois, D., Bialystok, E., Blaye, A., Polonia, A., & Yott, J. (2013). Lexical access and vocabulary development in very young bilinguals. *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behaviour*, 17(1), 57-70.
- Puric, D., Vuksanovic, J., & Chondrogianni, V. (2017). Cognitive advantages of immersion education after 1 year: effects of amount of exposure. *Journal of Experimental Child Psychology*, 159, 256-309.

References

- Quay, S. (1995). The bilingual lexicon: implications for studies of language choice. *Journal of Child Language*, 22(2), 369-387.
- Quay, S. (2001). Managing linguistic boundaries in early trilingual development. In J. Cenoz & F. Genesee (Eds.), *Trends in bilingual acquisition* (pp. 149-200). John Benjamins.
- Restrepo, M. A., Schwanenflugel, P. J., Blake, J., Neuhart-Pritchett, S., Cramer, E.S., & Ruston, P. H. (2006). Performance of the PPVT III and EVT applicability of the measures with African American children and European American children. *Language Speech and Hearing Services in Schools*, 37, 17-21.
- Reznick, J. S., & Schwartz, B. B. (2001). When is an assessment an intervention? Parent perception of infant intentionality and language. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40, 11-17.
- Saborit, C., & Julian, J. (2005). *L'avaluació del llenguatge infantil- ELI*. Universitat Jaume I de Castelló: Col.lecció Educació.
- Sapardo, K. (2013). Maturational constraints on lexical acquisition in a second language. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude and ultimate L2 attainment* (pp. 43-66). John Benjamins.
- Schoonen, R., & Verhallen, M. (2008). The assessment of deep word knowledge in young first and second language learners. *Language Testing*, 25(2), 211-236.
- Serrat, E., Sanz-Torrent, M., Badia, I., Aguilar, E., Olmo, R., Lara, F., Llorenc, A., & Serra, M. (2010). La relación entre el aprendizaje léxico y el desarrollo gramatical. *Infancia y Aprendizaje*, 33(4), 435-448.
- Serrat, E., Aguiar-Mediavilla E., Sanz-Torrent, E.M., Andreu, L., Amado, A., Badia, I., Serra, M. (2022). *Inventaris del desenvolupament d'habilitats comunicatives MacArthur-Bates en català: Guia d'ús i manual tècnic*. UOC.
- Singh, L., Hui, T. J., Chan, C., & Golinkoff, R. M. (2014). Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation. *Developmental Science*, 17(1), 94-109.
- Singleton, D. (2000). *The language and the lexicon*. OUP.
- Singleton, D. (2003). Perspectives of the multilingual lexicon: a critical synthesis. In J. Cenoz, B. Hufeisen & U. Jessner (Eds.), *The multilingual lexicon* (pp. 167-176). Kluwer Academic Publishers.
- Singleton, D., & Aronin, L. (2019). *Twelve lectures on multilingualism*. Multilingual Matters.
- Sinka, I., & Schletter, C. (1998). Morphosyntactic development in bilingual children. *International Journal of Bilingualism*, 2(3), 301-326.

- Slavin, R. & Cheung, A. (2005). A synthesis of research of reading instruction for English language learners. *Review of Educational Research*, 75(2), 247-284.
- Snow, C. E., & Hoefnagel-Höhle, M. (1978). Age differences in second language acquisition. In E. M. Hatch (Eds.), *Second language acquisition. A book of readings* (pp. 333-344). Newbury House.
- Stavans, A., & Muchnik, M (2007). Language production in trilingual children: Insights into code switching and code mixing. *Sociolinguistic Studies*, 1(3), 483-511.
- Sun, H., Chin, N.G. S., O'Brien, B. A., & Fritzche, T. (2020). Child, family and school factors in bilingual pre-schoolers' vocabulary development in heritage languages. *Journal of Child Language*, 47, 817-843.
- Taeschner, T. (1983). *The sun is feminine: A study on language acquisition in bilingual children*. Springer.
- Tong, F., Lara, L. R., Kirby, B., Mathes, P., & Kwok, O. (2008). Accelerating early academic oral English and structures English immersion programs. *American Educational Research Journal*, 45(4), 1011-1044.
- Torregrossa, J., Eisenbeib, S., & Bongartz, C. (2023). Boosting bilingual metalinguistic awareness under dual language activation: Some implication for bilingual education. *Article accepted in Language Learning*, online first.
- Turnbull, M., Cormier, M., & Bourque, J., (2012), The First Language in Science Class: A quasi- experimental study in language immersion. *The Modern Language Journal*, 95(1), 182-198.
- Tulviste, T., & Schults, A. (2020). Parental reports of communicative development at the age of 36 months. *First Language*, 40(1), 64-83.
- Turell, M. T. (1997). Our trilingual experience: A Catalan/ English/ Spanish trilingual family in Catalonia. *The Bilingual Family Newsletter*, 14(4), 1-7.
- Vihman, M. (1985). Language differentiation by the bilingual infant. *Journal of Child Language*, 12(1), 297-234.
- Vila, I. (1984). Yo siempre hablo catalán y castellano: datos de una investigación en curso sobre la adquisición del lenguaje en niños bilingüe. In M. Signán (Eds.), *Adquisición precoz de una segunda lengua* (pp. 31-51). Publicacions i edicions de la Universitat de Barcelona.
- Verhallen, M., & Schoonen, R. (1998). Lexical knowledge in L1 and L2 of third and fifth graders. *Applied Linguistics*, 19(4), 452-470.

References

- Volterra, V., & Taeschner, T. (1978). The acquisition of language by bilingual children. *Journal of Child Language*, 5(2), 311-326.
- Wang, X. L. (2008). *Growing up with three languages: Birth to eleven*. Multilingual Matters.
- Washington, J.A., & Craig, H.K. (1992). Performance of at-risk, African American preschoolers on the Peabody Picture Vocabulary Test- III. *Language, Speech & Hearing Services in Schools*, 30(1), 75- 81.
- Weisleder, A., Friend, M., Sin Mei Tsui, A., & Marchman, V.A. (2023). Using parent report to measure vocabulary in young bilingual children: a scoping review. *Language Learning*, online first.
- Williams, K. T. (1997). *Expressive Vocabulary Test*. Circle Pines.
- Woodcock, R. (1991). *Woodcock Language Proficiency Battery*. Allen TX.
- Wreker, J. F., & Byers, H.K. (2008). Bilingualism in infancy: first steps in perception and comprehension. *Trends in Cognitive Science*, 12(4), 144.151.
- Yamamoto, M. (2001). *Language use in interlingual families: A Japanese-English sociolinguistic study*. Multilingual Matters.
- Zwanziger, E., Allen, S. E. M., & Genesee, F. (2005). Cross linguistic influence in bilingual acquisition: subject omission in learners of Inuktitut and English. *Journal of Child Language*, 32, 893-909.

APPENDICES

APPENDIX 1: EUROPEAN INTERNATIONAL SCHOOL REGISTRATION FORM (ATTESTING LANGUAGE DOMINANCE OF THE IG PARTICIPANTS)

This form is to be accessed online at: <https://eisbarcelona.openapply.com/roi>

Personal Details (Applicant) | *Datos personales del solicitante* | *Dades personals dels sol·licitants*

Student's Surname

Apellidos | *Cognoms*

Student's First Name

Nombre Del Candidato/a | *Nom*

Date of Birth

Fecha de nacimiento | *Data naixement*

Place of Birth

Lugar de nacimiento | *Lloc de naixement*

First Language (Mother Tongue)

Primer Idioma (Lengua Materna) | *Primer Idioma (Llengua Materna)*

Otro Idioma 1 | *Altres Idiomes 1*

Nacionalidad | *Nacionalitat*

Colegio actual | *Escola actual*

Academic Level You Are Asking For

Curso para el que Solicita Plaza | *Nivell Acadèmic que sol·licita*

Academic Level You Are Asking For

Select an Option

P1 (1 year old)/ P2 (2 years old)/ P3 (3 years old)/ P4 (4 years old)/ P5 (5 years old) /Prim 1 (6 years old) /Prim 2 (7 years old)/ Prim 3 (8 years old)/ Prim 4 (9 years old) /Prim 5 (10 years old) /Prim 6 (11 years old)/ ESO 1 (12 years old) /ESO 2 (13 years old) /ESO 3 (14 years old)/ ESO 4 (15 years old)/ Baccalaureate 1 (16 years old)/ Baccalaureate 2 (17 years old)/ American High School Diploma Year 9 /American High School Diploma Year 10

Enrolment Year

Año de inscripción | Any de matrícula

Select an Option 2023-2024 /Academic Year 2024-2025 /

How did you hear about us?

¿Cómo nos has conocido?/ Com ens has conegut?

How did you hear about us?

Select an Option

Google Social Media Family and friends Press

Information of the Parent/Guardian |

Datos de Contacto (Padre/ Madre/ Tutor Legal) | Dades de Contacte (Mare/Pare/Tutor legal)

1. Parents' or Legal Guardians' Information (Padres o tutores legales | Pares o tutors legals)

Relationship to Applicant

Parentesco con solicitante | Parentiu amb sol.licitant

Select an Option

Father/ Mother/ Stepfather/ Stepmother/ Grandmother /Grandfather /Sister/ Brother/ Uncle/ Aunt Other Guardian/ Legal Guardian /Consultant/Recruiter

Surname

Apellidos | Cognoms

First Name

Nombre | Nom

Phone Number

Teléfono de contacto | Telèfon de contacte

E-mail Address

Correo electrónico | Correu electrònic

Set Password

Password Confirmation

2. Parents' or Legal Guardians' Information (Padres o tutores legales | Pares o tutors legals)

Relationship to Applicant

Parentesco con solicitante | Parentiu amb sol.licitant

Select an Option

Father Mother Stepfather Stepmother Grandmother Grandfather Sister Brother Uncle Aunt Other Guardian Legal Guardian Consultant/Recruiter

Appendices

Surname

Apellidos | Cognoms

First Name

Nombre | Nom

Phone Number

Teléfono de contacto | Telèfon de contacte

E-mail Address

Correo electrónico | Correu electrònic

Preferred Language of Communication

English/Spanish/Catalan

APPENDIX 2: MBCDI

MBCDI Words and Gestures

**MacArthur-Bates
Communicative Development Inventories**

Child's name: _____ Gender: _____
Birthdate: _____ Today's date: _____



**MacArthur-Bates CDI
Words and Gestures**

Copyright © 2007 The CDI Advisory Board.
All rights reserved.
Distributed by Paul H. Brookes Publishing Co.
1800-438-3775 • www.brookespublishing.com

Larry Fenson, Ph.D.
Virginia A. Marchman, Ph.D.
Donna J. Thal, Ph.D.
Philip S. Dale, Ph.D.
J. Steven Reznick, Ph.D.
Elizabeth Bates, Ph.D.

This form may be scored by hand or with a desktop image scanner and special software.

2. Animal Names (Real or Toy) (36)

	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
animal	☐	☐	☐	☐	☐	☐
cow	☐	☐	☐	☐	☐	☐
goose	☐	☐	☐	☐	☐	☐
pig	☐	☐	☐	☐	☐	☐
bear	☐	☐	☐	☐	☐	☐
deer	☐	☐	☐	☐	☐	☐
horse	☐	☐	☐	☐	☐	☐
pony	☐	☐	☐	☐	☐	☐
bee	☐	☐	☐	☐	☐	☐
dog	☐	☐	☐	☐	☐	☐
kitty	☐	☐	☐	☐	☐	☐
puppy	☐	☐	☐	☐	☐	☐
bird	☐	☐	☐	☐	☐	☐
donkey	☐	☐	☐	☐	☐	☐
sheep	☐	☐	☐	☐	☐	☐
frog	☐	☐	☐	☐	☐	☐
duck	☐	☐	☐	☐	☐	☐
fish	☐	☐	☐	☐	☐	☐
elephant	☐	☐	☐	☐	☐	☐
monkey	☐	☐	☐	☐	☐	☐
teddy bear	☐	☐	☐	☐	☐	☐
butterfly	☐	☐	☐	☐	☐	☐
mouse	☐	☐	☐	☐	☐	☐
tiger	☐	☐	☐	☐	☐	☐
cat	☐	☐	☐	☐	☐	☐
owl	☐	☐	☐	☐	☐	☐
turkey	☐	☐	☐	☐	☐	☐
chicken	☐	☐	☐	☐	☐	☐
giraffe	☐	☐	☐	☐	☐	☐
penguin	☐	☐	☐	☐	☐	☐
turtle	☐	☐	☐	☐	☐	☐

3. Vehicles (Real or Toy) (9)

	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
airplane	☐	☐	☐	☐	☐	☐
car	☐	☐	☐	☐	☐	☐
stroller	☐	☐	☐	☐	☐	☐
bicycle	☐	☐	☐	☐	☐	☐
firetruck	☐	☐	☐	☐	☐	☐
train	☐	☐	☐	☐	☐	☐
bus	☐	☐	☐	☐	☐	☐
motorcycle	☐	☐	☐	☐	☐	☐
truck	☐	☐	☐	☐	☐	☐

4. Toys (8)

	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
ball	☐	☐	☐	☐	☐	☐
block	☐	☐	☐	☐	☐	☐
bubbles	☐	☐	☐	☐	☐	☐
pen	☐	☐	☐	☐	☐	☐
balloon	☐	☐	☐	☐	☐	☐
book	☐	☐	☐	☐	☐	☐
doll	☐	☐	☐	☐	☐	☐
toy	☐	☐	☐	☐	☐	☐

5. Food and Drink (30)

	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
apple	☐	☐	☐	☐	☐	☐
cheerios	☐	☐	☐	☐	☐	☐
fish	☐	☐	☐	☐	☐	☐
peas	☐	☐	☐	☐	☐	☐
banana	☐	☐	☐	☐	☐	☐
cheese	☐	☐	☐	☐	☐	☐
food	☐	☐	☐	☐	☐	☐
pizza	☐	☐	☐	☐	☐	☐
bread	☐	☐	☐	☐	☐	☐
chicken	☐	☐	☐	☐	☐	☐
ice cream	☐	☐	☐	☐	☐	☐
raisin	☐	☐	☐	☐	☐	☐
butter	☐	☐	☐	☐	☐	☐
coffee	☐	☐	☐	☐	☐	☐
juice	☐	☐	☐	☐	☐	☐
spaghetti	☐	☐	☐	☐	☐	☐
cake	☐	☐	☐	☐	☐	☐
cookie	☐	☐	☐	☐	☐	☐
meat	☐	☐	☐	☐	☐	☐
bacon	☐	☐	☐	☐	☐	☐
candy	☐	☐	☐	☐	☐	☐
cracker	☐	☐	☐	☐	☐	☐
egg	☐	☐	☐	☐	☐	☐
drink	☐	☐	☐	☐	☐	☐
milk	☐	☐	☐	☐	☐	☐
water	☐	☐	☐	☐	☐	☐
cornflakes	☐	☐	☐	☐	☐	☐
egg	☐	☐	☐	☐	☐	☐
orange	☐	☐	☐	☐	☐	☐

MacArthur-Bates Communicative Development Inventories: Words and Gestures
Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-438-3775 • 410-337-9992 • www.brookespublishing.com

Part I: Early Words

A. First Signs of Understanding
Before children begin to speak, they show signs of understanding language by responding to familiar words and phrases. Below are some common examples. Does your child do any of these?

	yes	no
1. Respond when name is called (e.g., by turning and looking at source).	☐	☐
2. Respond to "no" by stopping what he/she is doing, at least for a moment.	☐	☐
3. React to "Here's mommy/daddy" by looking around for them.	☐	☐

B. Phrases (28)
In the list below, please mark the phrases that your child seems to understand.

	understands	understands	understands	
Are you hungry?	☐	Daddy/mommy's home. ☐	Give me a kiss. ☐	Sit down. ☐
Are you tired/sleepy?	☐	Do you want more? ☐	Go get _____ ☐	Get it out. ☐
Be careful.	☐	Don't do that. ☐	Good job/better. ☐	Here it is. ☐
Be quiet.	☐	Don't touch. ☐	Hold still. ☐	Time to go right night. ☐
Clap your hands.	☐	Get up. ☐	Let's go bye-bye. ☐	Throw the ball. ☐
Change/diaper.	☐	Give it to mommy. ☐	Look/look here. ☐	This little piggy. ☐
Come here/come on.	☐	Give me a hug. ☐	Open your mouth. ☐	Want to go for a ride? ☐

C. Starting to Talk

	Never	Sometimes	Often
1. Some children like to "jargon" or imitate things that they've just heard (including new words that they are just learning, and/or parts of sentences, for example, repeating "work now" after mother says "Mommy's going to work now"). How often does your child imitate words?	☐	☐	☐
2. Some children like to go around naming or labeling things, as though proud of knowing the names and wanting to show this. How often does your child do this?	☐	☐	☐

D. Vocabulary Checklist
The following is a list of typical words in young children's vocabularies. For words your child understands but does not yet say, place a mark in the first column (understands). For words that your child not only understands but also uses, place a mark in the second column (understands and says). If your child uses a different pronunciation of a word (for example, "tuffler" for "spatula" or "baet") for "spaghetti", mark the word anyway. Remember, this is a "catalogue" of words that are used by many different children. Don't worry if your child knows only a few right now.

1. Sound Effects and Animal Sounds (12)

	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
baa baa	☐	☐	grrr	☐	oach	☐
clap clap	☐	☐	meow	☐	quick quick	☐
cockadoodoo	☐	☐	moo	☐	uh uh	☐
	☐	☐		☐	yum yum	☐

MacArthur-Bates Communicative Development Inventories: Words and Gestures
Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-438-3775 • 410-337-9992 • www.brookespublishing.com

10. Outside Things and Places to Go (27)

	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
backyard	☐	☐	moon	☐	☐	school	☐
beach	☐	☐	outside	☐	☐	showel	☐
church*	☐	☐	park	☐	☐	sky	☐
flower	☐	☐	party	☐	☐	slide	☐
garden	☐	☐	pool	☐	☐	snow	☐
home	☐	☐	rain	☐	☐	star	☐
house	☐	☐	rack	☐	☐	store	☐

*for word used in your family

11. People (20)

	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
aunt	☐	☐	brother	☐	☐	grandma*	☐
baby	☐	☐	child	☐	☐	lady	☐
babysitter	☐	☐	daddy*	☐	☐	man	☐
babysitter's name	☐	☐	girl	☐	☐	mommy*	☐
boy	☐	☐	grandma*	☐	☐	child's own name	☐

*for word used in your family

12. Games and Routines (19)

	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
bath	☐	☐	hello	☐	☐	no	☐
breakfast	☐	☐	hi	☐	☐	patty cake	☐
bye or bye-bye	☐	☐	lunch	☐	☐	peekaboo	☐
dinner	☐	☐	nap	☐	☐	please	☐
don't	☐	☐	night night	☐	☐	shh/shush/hush	☐

13. Action Words (55)

	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
bite	☐	☐	bump	☐	☐	dance	☐
blow	☐	☐	clean	☐	☐	draw	☐
break	☐	☐	close	☐	☐	feed	☐
bring	☐	☐	cry	☐	☐	drive	☐

MacArthur-Bates Communicative Development Inventories: Words and Gestures

Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-438-0775 • 410-337-9590 • www.brookespublishing.com

	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
get	☐	☐	look	☐	☐	say	☐
give	☐	☐	love	☐	☐	see	☐
go	☐	☐	open	☐	☐	show	☐
help	☐	☐	play	☐	☐	sing	☐
hit	☐	☐	pull	☐	☐	sleep	☐
hug	☐	☐	push	☐	☐	smile	☐
hurry	☐	☐	put	☐	☐	splash	☐
jump	☐	☐	read	☐	☐	stop	☐
kick	☐	☐	ride	☐	☐	swim	☐
kiss	☐	☐	run	☐	☐	swing	☐

14. Words About Time (8)

	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
day	☐	☐	morning	☐	☐	none	☐
later	☐	☐	night	☐	☐	today	☐
	☐	☐		☐	☐	tomorrow	☐
	☐	☐		☐	☐	tonight	☐

15. Descriptive Words (37)

	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
all gone	☐	☐	dark	☐	☐	hot	☐
asleep	☐	☐	dirty	☐	☐	hungry	☐
bad	☐	☐	dry	☐	☐	hurt	☐
big	☐	☐	empty	☐	☐	little	☐
blue	☐	☐	fast	☐	☐	naughty	☐
broken	☐	☐	fine	☐	☐	old	☐
careful	☐	☐	gentle	☐	☐	pretty	☐
clean	☐	☐	good	☐	☐	red	☐
cold	☐	☐	happy	☐	☐	scared	☐
cute	☐	☐	hard	☐	☐	soft	☐

MacArthur-Bates Communicative Development Inventories: Words and Gestures

Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-438-0775 • 410-337-9590 • www.brookespublishing.com

16. Pronouns (11)

	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
his	☐	☐	it	☐	☐	my	☐
her	☐	☐	me	☐	☐	that	☐
I	☐	☐	either	☐	☐	this	☐

17. Question Words (6)

	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
how	☐	☐	when	☐	☐	who	☐
what	☐	☐	where	☐	☐	why	☐

18. Prepositions and Locations (11)

	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
away	☐	☐	in	☐	☐	on	☐
back	☐	☐	inside	☐	☐	out	☐
down	☐	☐	off	☐	☐	there	☐

19. Quantifiers (8)

	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says	under-stands	under-stands and says
all	☐	☐	more	☐	☐	not	☐
another	☐	☐	none	☐	☐	other	☐
	☐	☐		☐	☐	same	☐
	☐	☐		☐	☐	some	☐

MacArthur-Bates Communicative Development Inventories: Words and Gestures

Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-438-0775 • 410-337-9590 • www.brookespublishing.com

Part II: Actions and Gestures

A. First Communicative Gestures

When infants are first learning to communicate, they often use gestures to make their wishes known. For each item below, mark the line that describes your child's actions right now.

	Not Yet	Sometimes	Often
1. Extends arm to show you something he/she is holding.	☐	☐	☐
2. Reaches out and gives you a toy or some object that he/she is holding.	☐	☐	☐
3. Points (with arm and index finger extended) at some interesting object or event.	☐	☐	☐
4. Waves bye-bye on his/her own when someone leaves.	☐	☐	☐
5. Extends his/her arm upward to signal a wish to be picked up.	☐	☐	☐
6. Shakes head "no".	☐	☐	☐
7. Nods head "yes".	☐	☐	☐
8. Gestures "hush" by placing finger to lips.	☐	☐	☐
9. Requests something by extending arm and opening and closing hand.	☐	☐	☐
10. Blows kisses from a distance.	☐	☐	☐
11. Smacks lips in a "yum yum" gesture to indicate that something tastes good.	☐	☐	☐
12. Struggles to indicate "all gone" or "where'd it go".	☐	☐	☐

B. Games and Routines

Does your child do any of the following?

	Yes	No
1. Play peekaboo.	☐	☐
2. Play patty cake.	☐	☐
3. Play "no big".	☐	☐
4. Play chasing games.	☐	☐
5. Sing.	☐	☐
6. Dance.	☐	☐

MacArthur-Bates Communicative Development Inventories: Words and Gestures

Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-438-0775 • 410-337-9590 • www.brookespublishing.com

C. Actions with Objects

Does your child do or try to do any of the following?

	Yes	No
1. Eat with a spoon or fork.	☐	☐
2. Drink from a cup containing liquid.	☐	☐
3. Comb or brush own hair.	☐	☐
4. Brush teeth.	☐	☐
5. Wipe face or hands with a towel or cloth.	☐	☐
6. Put on hat.	☐	☐
7. Put on a shoe or sock.	☐	☐
8. Put on a necklace, bracelet, or watch.	☐	☐
9. Lay head on hands and squeeze eyes shut as if sleeping.	☐	☐
10. Blow to indicate something is hot.	☐	☐
11. Hold plane and make it "fly".	☐	☐
12. Put telephone to ear.	☐	☐
13. Sniff flowers.	☐	☐
14. Push toy car or truck.	☐	☐
15. Throw a ball.	☐	☐
16. Pour pretend liquid from one container to another.	☐	☐
17. Stir pretend liquid in a cup or pan with a spoon.	☐	☐

D. Pretending to be a Parent

Here are some things that young children sometimes do with stuffed animals or dolls. Please mark the actions that you have seen your child do.

	Yes	No
1. Put to bed.	☐	☐
2. Cover with blanket.	☐	☐
3. Feed with bottle.	☐	☐
4. Feed with spoon.	☐	☐
5. Brush/tomb its hair.	☐	☐
6. Pat or burp it.	☐	☐
7. Push in stroller/buggy.	☐	☐
8. Rock it.	☐	☐
9. Kiss or hug it.	☐	☐
10. Try to put shoe or sock or hat on it.	☐	☐
11. Wipe its face or hands.	☐	☐
12. Talk to it.	☐	☐
13. Try to put diaper on it.	☐	☐

MacArthur-Bates Communicative Development Inventories: Words and Gestures
Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-638-3775 • 410-337-9580 • www.brookespublishing.com

9

E. Imitating Other Adult Actions

(Using real or toy implements)

Does your child do or try to do any of the following?

	Yes	No
1. Sweep with broom or mop.	☐	☐
2. Put key in door or lock.	☐	☐
3. Pound with hammer or mallet.	☐	☐
4. Attempt to use saw.	☐	☐
5. "Type" at a typewriter or computer keyboard.	☐	☐
6. "Read" (opens book, turns page).	☐	☐
7. Vacuum.	☐	☐
8. Water plants.	☐	☐
9. Play musical instrument (e.g., piano, trumpet).	☐	☐
10. "Drive" car by turning steering wheel.	☐	☐
11. Wash dishes.	☐	☐
12. Clean with cloth or duster.	☐	☐
13. Write with a pen, pencil, or marker.	☐	☐
14. Dig with a shovel.	☐	☐
15. Put on glasses.	☐	☐

10

MacArthur-Bates Communicative Development Inventories: Words and Gestures
Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-638-3775 • 410-337-9580 • www.brookespublishing.com

Child Report Form CDI: Words and Gestures



ID number _____ Date of report _____ Date CDI completed _____

Child's name _____ Child's birth date _____
FIRST MIDDLE LAST

Parent/guardian _____ Gender _____ Child's age in months _____
FIRST LAST

PART I: EARLY WORDS

First Signs of Understanding

Responds when name is called: ☐ yes ☐ no _____
Responds to "no no": ☐ yes ☐ no _____
Responds to "there's mommy/daddy": ☐ yes ☐ no _____

Percentage of yes
answers at this child's age

Phrases Understood: Number: _____ (of 28) Percentile: _____

Starting to Talk

Imitation: ☐ yes ☐ no _____
Labeling: ☐ yes ☐ no _____

Percentage of affirmative
answers at this child's age

Vocabulary Checklist

Words Understood: Number: _____ (of 396) Percentile: _____
Words Produced: Number: _____ (of 396) Percentile: _____

PART II: ACTIONS AND GESTURES

Early Gestures (Sections A and B) Number: _____ (of 18) Percentile: _____

Later Gestures (Sections C through E) Number: _____ (of 45) Percentile: _____

Total Gestures (Sections A through E) Number: _____ (of 63) Percentile: _____

MacArthur-Bates Communicative Development Inventories: Words and Gestures
Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-638-3775 • 410-337-9580 • www.brookespublishing.com

11

MBCDI Words and Sentences

**MacArthur-Bates
Communicative Development Inventories**

Child's name: _____ Gender: _____
Birthdate: _____ Today's date: _____





**MacArthur-Bates CDI
Words and Sentences**

Copyright © 2007 The CDI Advisory Board.
All rights reserved.
Distributed by Paul H. Brookes Publishing Co.
1-800-638-3775 • 410-337-9580
www.brookespublishing.com

**Larry Fenson, Ph.D.
Virginia A. Marchman, Ph.D.
Donna J. Thal, Ph.D.
Philip S. Dale, Ph.D.
J. Steven Reznick, Ph.D.
Elizabeth Bates, Ph.D.**

This form may be scored by hand or with a desktop image scanner and special software.

1

Part I: Words Children Use

A. Vocabulary Checklist
Children understand many more words than they say. We are particularly interested in the words your child SAYS. Please go through the list and mark the words you have heard your child use. If your child uses a different pronunciation of a word (for example, "yaller" instead of "giraffe" or "baan" for "banana"), mark the word anyway. Remember that this is a "catalogue" of all the words that are used by many different children. Don't worry if your child knows only a few of these right now.

1. Sound Effects and Animal Sounds (12)

bee bee	☐	grrr	☐	ouch	☐	vrooom	☐
choo choo	☐	meow	☐	quack quack	☐	woof woof	☐
cockadoodadoo	☐	moo	☐	uh ah	☐	yum yum	☐

2. Animals (Real or Toy) (43)

alligator	☐	cow	☐	horse	☐	puppy	☐
animal	☐	deer	☐	kitty	☐	rooster	☐
ant	☐	dog	☐	lamb	☐	sheep	☐
bear	☐	donkey	☐	lion	☐	squirrel	☐
bee	☐	duck	☐	monkey	☐	teddybear	☐
bird	☐	elephant	☐	mouse	☐	tiger	☐
bug	☐	fish	☐	mouse	☐	turtle	☐
bunny	☐	frog	☐	owl	☐	weir	☐
butterfly	☐	giraffe	☐	penguin	☐	wolf	☐
cat	☐	goose	☐	pig	☐	zebra	☐
chicken	☐	hen	☐	pony	☐		

3. Vehicles (Real or Toy) (14)

airplane	☐	car	☐	clod	☐	tricycle	☐
bicycle	☐	firetruck	☐	skoller	☐	truck	☐
boat	☐	helicopter	☐	tractor	☐		
bus	☐	motorcycle	☐	train	☐		

4. Toys (18)

ball	☐	bubbles	☐	glue	☐	puzzle	☐
balloon	☐	chalk	☐	pen	☐	story	☐
bat	☐	crayon	☐	pencil	☐	toy	☐
block	☐	doll	☐	play dough	☐		
book	☐	game	☐	present	☐		

5. Food and Drink (58)

apple	☐	corn	☐	lollipop	☐	pretzel	☐
applesauce	☐	cracker	☐	meat	☐	pudding	☐
banana	☐	donut	☐	melon	☐	pumpkin	☐
beans	☐	drink	☐	milk	☐	raisin	☐
bread	☐	egg	☐	muffin	☐	salt	☐
butter	☐	fish	☐	noodles	☐	sandwich	☐
cake	☐	food	☐	nuts	☐	sauce	☐
candy	☐	french fries	☐	orange	☐	soda/pop	☐
carrots	☐	grapes	☐	pancake	☐	soup	☐
cereal	☐	green beans	☐	peanut butter	☐	spaghetti	☐
cheerios	☐	gum	☐	peas	☐	strawberry	☐
cheese	☐	hamburger	☐	pickle	☐	toast	☐
chicken	☐	ice	☐	pizza	☐	tuna	☐
chocolate	☐	ice cream	☐	popcorn	☐	vanilla	☐
coffee	☐	jello	☐	popcycle	☐	vitamins	☐
cookie	☐	jelly	☐	potato	☐	water	☐
cookie	☐	jelly	☐	potato chip	☐	yogurt	☐

6. Clothing (28)

beads	☐	dress	☐	pagamas	☐	snaker	☐
belt	☐	gloves	☐	pants	☐	snowsuit	☐
bib	☐	hat	☐	scarf	☐	sock	☐
boots	☐	jacket	☐	shirt	☐	sweater	☐
button	☐	jeans	☐	shoe	☐	tights	☐
coat	☐	mitten	☐	shorts	☐	underpants	☐
diaper	☐	necklace	☐	slipper	☐	zipper	☐

7. Body Parts (27)

ankle	☐	eye	☐	knee	☐	shoulder	☐
arm	☐	face	☐	leg	☐	tooth	☐
belly button	☐	feet	☐	lips	☐	toe	☐
buttocks/bottom*	☐	finger	☐	mouth	☐	tongue	☐
cheek	☐	hair	☐	nose	☐	tummy	☐
chin	☐	hand	☐	owle/boo boo	☐	vagina*	☐
ear	☐	head	☐	penis*	☐		

*or word used in your family

8. Small Household Items (50)

basket	☐	dish	☐	mop	☐	spoon	☐
blanket	☐	fork	☐	nail	☐	tape	☐
bottle	☐	garbage	☐	napkin	☐	telephone	☐
box	☐	glass	☐	paper	☐	tissue/kleenex	☐
bowl	☐	glasses	☐	penny	☐	toothbrush	☐
broom	☐	hammer	☐	picture	☐	towel	☐
brush	☐	jar	☐	pillow	☐	trash	☐
bucket	☐	keys	☐	plant	☐	tray	☐
camera	☐	knife	☐	plate	☐	vacuum	☐
can	☐	lamp	☐	purse	☐	walker	☐
clock	☐	light	☐	radio	☐	watch	☐
comb	☐	medicine	☐	scissors	☐		
cup	☐	money	☐	soap	☐		

9. Furniture and Rooms (33)

basement	☐	orb	☐	play pen	☐	stairs	☐
bedroom	☐	door	☐	porch	☐	stove	☐
bathroom	☐	drawer	☐	potfy	☐	table	☐
bed	☐	dryer	☐	refrigerator	☐	TV	☐
bedroom	☐	garage	☐	rocking chair	☐	washing machine	☐
bench	☐	high chair	☐	room	☐	window	☐
chair	☐	kitchen	☐	shower	☐		
closet	☐	living room	☐	sink	☐		
couch	☐	oven	☐	sofa	☐		

10. Outside Things (31)

backyard	☐	lawn mower	☐	sidewalk	☐	stone	☐
cloud	☐	moon	☐	sky	☐	street	☐
flag	☐	pool	☐	slide	☐	sun	☐
flower	☐	rain	☐	snow	☐	swing	☐
garden	☐	rock	☐	snowman	☐	tree	☐
grass	☐	roof	☐	sprinkler	☐	water	☐
hose	☐	sandbox	☐	star	☐	wind	☐
ladder	☐	shovel	☐	stick	☐		

MacArthur-Bates Communicative Development Inventories: Words and Sentences
Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-638-3775 • 410-337-9580 • www.brookespublishing.com

3

MacArthur-Bates Communicative Development Inventories: Words and Sentences
Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-638-3775 • 410-337-9580 • www.brookespublishing.com

4

237

11. Places to Go (22)

beach	☐	farm	☐	park	☐	woods	☐
camping	☐	gas station	☐	party	☐	work	☐
church*	☐	home	☐	picnic	☐	yard	☐
circus	☐	house	☐	playground	☐	zoo	☐
country	☐	movie	☐	school	☐		
downtown	☐	outside	☐	store	☐		

*or word used in your family

12. People (29)

aunt	☐	cowboy	☐	lady	☐	pet's name	☐
baby	☐	daddy*	☐	mailman	☐	police	☐
baby-sitter	☐	doctor	☐	man	☐	sisiter	☐
baby-sitter's name	☐	fireman	☐	mommy*	☐	teacher	☐
boy	☐	friend	☐	nurse	☐	uncle	☐
brother	☐	girl	☐	child's own name	☐		
child	☐	grandma*	☐	people	☐		
clown	☐	grandpa*	☐	person	☐		

*or word used in your family

13. Games and Routines (25)

bath	☐	go potty	☐	patty cake	☐	thank you	☐
breakfast	☐	hi	☐	peekaboo	☐	this little piggy	☐
bye	☐	hello	☐	please	☐	turn around	☐
call (on phone)	☐	lunch	☐	shh/hush/hush	☐	yes	☐
dinner	☐	nap	☐	sleeping	☐		
give me fixed	☐	night night	☐	snack	☐		
gonna get you	☐	no	☐	so big!	☐		

14. Action Words (103)

bite	☐	catch	☐	clay	☐	dump	☐
blow	☐	chase	☐	cat	☐	eat	☐
break	☐	clap	☐	dance	☐	fall	☐
bring	☐	clean	☐	draw	☐	feed	☐
build	☐	climb	☐	drink	☐	find	☐
bump	☐	close	☐	drive	☐	finish	☐
buy	☐	cook	☐	drop	☐	fit	☐
carry	☐	cover	☐	dry	☐	fix	☐

(continued)

MacArthur-Bates Communicative Development Inventories: Words and Sentences
Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-638-3775 • 410-337-9580 • www.brookespublishing.com

5

get	☐	listen	☐	see	☐	take	☐
give	☐	look	☐	shake	☐	talk	☐
go	☐	love	☐	share	☐	taste	☐
hate	☐	make	☐	show	☐	tear	☐
have	☐	open	☐	sing	☐	think	☐
hear	☐	paint	☐	sit	☐	throw	☐
help	☐	pick	☐	skate	☐	tickle	☐
hide	☐	play	☐	sleep	☐	touch	☐
hit	☐	pour	☐	slide	☐	walk	☐
hold	☐	pretend	☐	roll	☐	water	☐
hop	☐	put	☐	spill	☐	walk	☐
horry	☐	push	☐	splash	☐	wash	☐
jump	☐	put	☐	stand	☐	watch	☐
kick	☐	read	☐	stay	☐	wipe	☐
kiss	☐	ride	☐	stop	☐	wish	☐
knock	☐	rip	☐	sweep	☐	work	☐
lick	☐	run	☐	swim	☐	write	☐
like	☐	say	☐	swing	☐		

15. Descriptive Words (63)

allgone	☐	dry	☐	last	☐	scared	☐
asleep	☐	empty	☐	little	☐	sick	☐
awake	☐	fast	☐	long	☐	sleepy	☐
bad	☐	fine	☐	loud	☐	slow	☐
better	☐	first	☐	mad	☐	soft	☐
big	☐	full	☐	naughty	☐	sticky	☐
black	☐	gentle	☐	new	☐	stuck	☐
blue	☐	good	☐	nice	☐	thirsty	☐
broken	☐	green	☐	natty	☐	tiny	☐
brown	☐	happy	☐	old	☐	tired	☐
careful	☐	hard	☐	orange	☐	wet	☐
clean	☐	heavy	☐	poor	☐	white	☐
cold	☐	high	☐	pretty	☐	windy	☐
cute	☐	hot	☐	quiet	☐	yellow	☐
dark	☐	hungry	☐	red	☐	yucky	☐
dirty	☐	hurt	☐	sad	☐		

MacArthur-Bates Communicative Development Inventories: Words and Sentences
Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-638-3775 • 410-337-9580 • www.brookespublishing.com

6

16. Words About Time (12)

after	☐	later	☐	now	☐	tomorrow	☐
before	☐	morning	☐	time	☐	tonight	☐
day	☐	night	☐	today	☐	yesterday	☐

17. Pronouns (25)

he	☐	me	☐	their	☐	we	☐
her	☐	name	☐	them	☐	you	☐
hers	☐	my	☐	these	☐	your	☐
him	☐	myself	☐	they	☐	yourself	☐
his	☐	our	☐	this	☐		
i	☐	she	☐	those	☐		
it	☐	that	☐	us	☐		

18. Question Words (7)

how	☐	when	☐	which	☐	why	☐
what	☐	where	☐	who	☐		

19. Prepositions and Locations (26)

about	☐	beside	☐	next to	☐	there	☐
above	☐	by	☐	of	☐	to	☐
around	☐	down	☐	off	☐	under	☐
at	☐	for	☐	on	☐	up	☐
away	☐	here	☐	on top of	☐	with	☐
back	☐	inside/in	☐	out	☐		
behind	☐	into	☐	over	☐		

20. Quantifiers and Articles (17)

a	☐	any	☐	not	☐	the	☐
all	☐	each	☐	none	☐	too	☐
a lot	☐	every	☐	other	☐		
an	☐	more	☐	same	☐		
another	☐	much	☐	some	☐		

MacArthur-Bates Communicative Development Inventories: Words and Sentences
Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-638-3775 • 410-337-9580 • www.brookespublishing.com

7

21. Helping Verbs (21)

am	☐	do	☐	is	☐	were	☐
are	☐	does	☐	let me	☐	will	☐
be	☐	don't	☐	need/had to	☐	would	☐
can	☐	gonna/going to	☐	try/try to	☐		
could	☐	gonna/got to	☐	wanna/want to	☐		
couldn't ya	☐	halls/have to	☐	was	☐		

22. Connecting Words (6)

and	☐	but	☐	so	☐
because	☐	if	☐	then	☐

B. How Children Use Words

	Not Yet	Sometimes	Often
1. Does your child ever talk about past events or people who are not present? For example, a child who saw a parade last week might later say parade, clown, or band.	☐	☐	☐
2. Does your child ever talk about something that's going to happen in the future, for example, saying "tchoo choo" or "bye-bye" before you leave the house for a trip, or saying "baww" when you are going to the park?	☐	☐	☐
3. Does your child talk about objects that are not present such as asking about a missing or absent toy, referring to a pet out of view, or asking about someone not present?	☐	☐	☐
4. Does your child understand if you ask for something that is not in the room, for example, by going to the bedroom to get a teddy bear when you say "where's the bear?"	☐	☐	☐
5. Does your child ever pick up or point to an object and name an object person to whom the object belongs? For example, a child might point to mommy's shoe and say "mommy's".	☐	☐	☐

MacArthur-Bates Communicative Development Inventories: Words and Sentences
Copyright © 2007 The CDI Advisory Board. All rights reserved.
Do not reproduce without permission of Paul H. Brookes Publishing Co. • 1-800-638-3775 • 410-337-9580 • www.brookespublishing.com

8

Part II: Sentences and Grammar

A. Word Endings/Part 1

	Not Yet	Sometimes	Often
1. To talk about more than one thing, we add an "s" to many words. Examples include cars (for more than one car), shoes, dogs, and keys. Has your child begun to do this?	☐	☐	☐
2. To talk about ownership, we add an "'s", for example, Daddy's key, kitty's dish, and baby's bottle. Has your child begun to do this?	☐	☐	☐
3. To talk about activities, we sometimes add "ing" to verbs. Examples include looking, running, and crying. Has your child begun to do this?	☐	☐	☐
4. To talk about things that happened in the past, we often add "ed" to the verb. Examples include looked, opened, and pushed. Has your child begun to do this?	☐	☐	☐

B. Word Forms

Following are some other words children learn. Please mark any of these words that your child uses.

Nouns	
children	☐ men ☐ teeth ☐
feet	☐ mice ☐

Verbs	
sle	☐ fell ☐ made ☐
saw	☐ flew ☐ ran ☐
bought	☐ got ☐ sat ☐
broke	☐ had ☐ saw ☐
came	☐ heard ☐ took ☐
drank	☐ held ☐ went ☐
drove	☐ lost ☐

C. Word Endings/Part 2

Young children often place the wrong endings on words. For example, a child might say "buntle good home." Mistakes like this are often a sign of progress in language. In the following lists, please mark all the mistakes of this kind you have heard your child say recently.

Nouns	
blockies	☐ foots ☐ mousees ☐ lottees ☐
childrens	☐ mans ☐ shoeses ☐ tooths ☐
childs	☐ mans ☐ sockses ☐
feets	☐ mices ☐ beeths ☐

Verbs	
abed	☐ bringed ☐ brokeed ☐ doed ☐
blewed	☐ boyed ☐ comed ☐ expined ☐
blowed	☐ brokeed ☐ comed ☐ drinkeed ☐

(continued)

eated	☐	gotted	☐	listed	☐	satted	☐
faled	☐	haved	☐	maled	☐	stted	☐
flad	☐	heaved	☐	ramed	☐	taked	☐
gotted	☐	hoided	☐	runned	☐	wented	☐
goed	☐	looad	☐	seed	☐		☐

Has your child begun to combine words yet, such as "muther cracker", or "beegie blee?"

Not Yet ☐ Sometimes ☐ Often ☐

If you answered not yet, please stop here. If you answered sometimes or often, please continue.

D. Examples

Please list three of the longest sentences you have heard your child say recently.

- _____
- _____
- _____

E. Complexity

In each of the following pairs, please mark the one that sounds MOST like the way your child talks right now. If your child is saying sentences even more complicated than the two provided, just pick the second one.

1. Two shoes.	☐	11. (Talking about something that already happened) Daddy pick me up.	☐	24. I no do it.	☐
2. Two shoes.	☐	12. (Talking about something that already happened) Kitty go away.	☐	25. I like read stories.	☐
3. Daddy's car.	☐	13. (Talking about something happening right now) Kitty sleep.	☐	26. Don't read book.	☐
4. (Talking about something happening right now) Kitty sleep.	☐	14. That my truck.	☐	27. Turn on light.	☐
5. (Talking about something happening right now) I make tower.	☐	15. Baby crying.	☐	28. I want that.	☐
6. (Talking about something that already happened) I fall down.	☐	16. That's my truck.	☐	29. I want that one you get.	☐
7. More cookie!	☐	17. Baby crying.	☐	30. Want cookies and milk.	☐
8. These my teeth.	☐	18. You to it?	☐	31. Cookie mummy.	☐
9. Baby blanket.	☐	19. Can you fix it?	☐	32. Baby want eat.	☐
10. (Talking about something that already happened) Doggie kiss me.	☐	20. Read me story, Mommy.	☐	33. Lookit me!	☐
	☐	21. No wash doll.	☐	34. Lookit me dancing!	☐
	☐	22. Don't wash doll.	☐	35. Lookit!	☐
	☐	23. Want more juice.	☐	36. Where's my doll name Sam?	☐
	☐	24. Want juice in there.	☐	37. We made this.	☐
	☐	25. There's a kitty.	☐	38. Me and Paul made this.	☐
	☐	26. Go bye-bye.	☐	39. I sing song.	☐
	☐	27. Wanna go bye-bye.	☐	40. Baby crying cus shw's sad.	☐
	☐	28. Where mommy go?	☐		☐
	☐	29. Where did mommy go?	☐		☐
	☐	30. Coffee hot.	☐		☐
	☐	31. That coffee hot.	☐		☐

Child Report Form
CDI: Words and Sentences



ID number _____ Date of report _____ Date CDI completed _____

Child's name _____ Child's birth date _____

Parent/guardian _____ Gender _____ Child's age in months _____

PART I: WORDS CHILDREN USE

Vocabulary Checklist

Words Produced: _____ Number: _____ (of 680) Percentile: _____

How Children Use Words _____ Percentage of affirmative answers at this child's age _____

Past: ☐ yes ☐ no _____

Future: ☐ yes ☐ no _____

Absent Object (Production): ☐ yes ☐ no _____

Absent Object (Comprehension): ☐ yes ☐ no _____

Absent Owner: ☐ yes ☐ no _____

PART II: SENTENCES AND GRAMMAR

Word Endings/Part 1

Percent of affirmative answers at this child's age _____

Plural (-s): ☐ yes ☐ no _____

Possessive ('s): ☐ yes ☐ no _____

Progressive (-ing): ☐ yes ☐ no _____

Past tense (-ed): ☐ yes ☐ no _____

Word Forms

Number: _____ (of 25) Percentile: _____

Word Endings/Part 2

Number: _____ (of 45) Percentile: _____

Combining ☐ yes ☐ no _____

Examples

Length in morphemes of child's three longest sentences (M3L): 1. _____ 2. _____ 3. _____

M3L (mean): _____ Percentile: _____

Complexity

Number of times the more complex sentence is selected: _____ (of 37) Percentile: _____

APPENDIX 3: PPVT III (SCORING SHEET AND EXAMPLE)

Nombre y apellidos: _____

Sexo: ☐ M ☐ V Ciudad: _____

Provincia: _____ Curso: _____

Idioma habitual: ☐ Castellano ☐ Otro: _____ (especificar: extranjero o lengua regional)

Profesor: _____ Examinador: _____

Hoja de anotación

Razón de la evaluación: _____

Otras informaciones del examinando: _____

CÁLCULO DE LA EDAD CRONOLÓGICA

Año: ____ Mes: ____ Día: ____

Fecha de la evaluación: ____/____/____

Fecha de nacimiento: ____/____/____

Edad cronológica: ____ años ____ meses ____ días

*Igualar los días entre:

Puntuación directa

Conjunto 1: ____ Conjunto 5: ____ Conjunto 9: ____ Conjunto 13: ____

Conjunto 2: ____ Conjunto 6: ____ Conjunto 10: ____ Conjunto 14: ____

Conjunto 3: ____ Conjunto 7: ____ Conjunto 11: ____ Conjunto 15: ____

Conjunto 4: ____ Conjunto 8: ____ Conjunto 12: ____ Conjunto 16: ____

TOTAL ERRORES ____

Puntuaciones transformadas

CI: ____

Percentil: ____

Eneatipo: ____

Desarrollo

Edad equivalente: ____

Anote el número del elemento techo, es decir, el número del último elemento del conjunto que ha acertado. Rote a ese número el número total de errores cometidos por el examinando desde el conjunto base al conjunto techo. El resultado es la puntuación directa.

Elemento techo: ____ Total errores: ____ Puntuación directa: ____

68% ± 5 90% ± 7 95% ± 9

CI de ____ a ____

Percentil de ____ a ____

Eneatipo de ____ a ____

CURVA NORMAL

Realice una marca sobre las líneas correspondiente al CI obtenido. Dibuje una línea recta en vertical que corte con las otras escalas (véase el manual para mayor información).

Opcionalmente pueden representarse también los intervalos de confianza operativos. Para ello, elija dos líneas verticales sobre todos los ejes, uno a cada lado del CI obtenido. Para el intervalo de confianza del 95% utilice las puntuaciones CI (véase el manual si se desean otros ejes).

Autores: L.M. Dunn, L.M. Dunn y B. Arribas
Copyright © 2004 by TEA Ediciones, S.A. - Edito: TEA Ediciones, S.A. - Freix Benicardes Salgán, 24 28034 MADRID - Este ejemplar está impreso en papel reciclado.
Si se presentan estos en este registro en una reproducción legal. En beneficio de la propiedad y en el uso propio, NO LA UTILICE - Prohibida la reproducción total o parcial. Todos los derechos reservados - Printed in Spain. Impreso en España por Imprenta Castalia, S.L., Agostín Calvo, 40, 28040 MADRID.

1 2 años y medio-3 años

Escoba 2 E

Arón 1 E

Beber 3 E

Pala 4 E

Columpiarse 4 E

Lámpara 4 E

Dinero 3 E

Helicóptero 2 E

Valla 3 E

Llave 4 E

Tambor 3 E

Subir 1 E

Total errores ____

4 años

Vaca 1 E

Nadar 1 E

Vacio 1 E

Excavar 2 E

Granjero 3 E

Accidente 2 E

Nido 3 E

Lanzar 4 E

Sobre 2 E

Castillo 2 E

Medir 4 E

Canguro 2 E

Total errores ____

5 años

Fruta 1 E

Cadena 2 E

Cactus 3 E

Puercoespín 1 E

Bostezar 2 E

Colera 4 E

Decorado 4 E

Zorro 3 E

Garras 1 E

Discretiv 1 E

Astronauta 3 E

Serrar 4 E

Total errores ____

6-7 años

Tronco 2 E

Enorme 3 E

Paracaidas 3 E

Entregar 1 E

Globo 2 E

Calculadora 2 E

Gotear 4 E

Colmena 1 E

Lijar 2 E

Estatua 4 E

Aterrizada 1 E

Rectángulo 1 E

Total errores ____

8-9 años

Molestar 1 E

Lima 2 E

Isla 2 E

Seleccionar 1 E

Par 3 E

Ángulo 1 E

Reptil 2 E

Manchulla 4 E

Acantilado 1 E

Terror 3 E

Dirigir 2 E

Morsa 3 E

Total errores ____

10-11 años

Palmera 1 E

Degredador 3 E

Embudo 2 E

Repostar 2 E

Ajustable 2 E

Reedor 3 E

Colisionar 1 E

Terno 1 E

Ártico 2 E

Calcular 1 E

Trillizos 4 E

Contaminar 3 E

Total errores ____

12-16 años

Arquero 2 E

Manifiesto 2 E

Compositor 2 E

Oasis 1 E

Cítrico 2 E

Lubricar 1 E

Velocímetro 3 E

Brebaje 1 E

Izar 1 E

Reprimenda 1 E

Porcelana 2 E

Cuántico 2 E

Total errores ____

10

Barandilla 2 E

Brújula 3 E

Instruir 3 E

Carente 4 E

Infinito 4 E

Coreográfico 1 E

Confidencia 1 E

Cuía 3 E

Équido 4 E

Válvula 3 E

Cosechar 4 E

Gemir 1 E

Total errores ____

11

Bobina 4 E

Aislamiento 1 E

Caballero 4 E

Reflexión 2 E

Tapicería 4 E

Artefacto 1 E

Acicalarse 2 E

Erudito 4 E

Berlina 4 E

Fachada 1 E

Estabones 4 E

Picticio 1 E

Total errores ____

12 17 años o más

Estéril 2 E

Primate 4 E

Sesgado 3 E

Reponer 3 E

Península 4 E

Perpendicular 3 E

Diario 2 E

Obelisco 4 E

Cavilar 2 E

Incandescente 4 E

Incisivo 2 E

Culinario 3 E

Total errores ____

13

Hurtar 2 E

Dromedario 2 E

Encarcelar 1 E

Bovino 2 E

Estambre 3 E

Vestigio 2 E

Preceptor 1 E

Friccionar 2 E

Mercantil 3 E

Nu 1 E

Zarpa 1 E

Amazona 3 E

Total errores ____

14

Filtrar 1 E

Pentágono 1 E

Avzorar 3 E

Dársena 4 E

Converger 1 E

Receptáculo 1 E

Perforación 4 E

Vitreo 3 E

Remontar 3 E

Caducifolio 4 E

Anegar 3 E

Abrasive 1 E

Total errores ____

15

Palmipedo 4 E

Cizalla 3 E

Marsupial 4 E

Conifera 4 E

Temeraria 2 E

Entomólogo 2 E

Balaustre 4 E

Pecuniarío 3 E

Inocular 1 E

Repujado 4 E

Yantar 3 E

Paquidermo 2 E

Total errores ____

16

Friso 1 E

Calibrador 4 E

Selénico 3 E

Amarilla 2 E

Mielgo 1 E

Roturar 1 E

Conflagración 3 E

Gravar 3 E

Nopal 2 E

Motillar 4 E

Bancal 3 E

Ósculo 2 E

Total errores ____

13 4 años

37 6-7 años

73 10-11 años

133 17 o más años

PPVT-III

Nombre y apellidos: _____

Sexo: ☐ M ☐ F Ciudad: _____

Provincia: _____

Centro: _____

Idioma habitual: ☐ Castellano ☒ Otro

Profesor: _____

Curso: **P4C**

Examinador: _____

Hoja de anotación

Razon de la evaluación: _____

Otras informaciones del examinando: _____

CÁLCULO DE LA EDAD CRONOLÓGICA

Año: **13** Mes: **05** Día: **23**

Fecha de la evaluación: **13** **05** **23**

Fecha de nacimiento: **08** **08** **19**

Edad cronológica: **4** **9** **04**

Puntuación directa **32**

Puntuaciones transformadas

CI: **87**

Percentil: **19**

Eneatipo: **3**

Desarrollo

Edad equivalente: **3,10**

Conjunto 1	1	Conjunto 5		Conjunto 9		Conjunto 13
Conjunto 2	4	Conjunto 6		Conjunto 10		Conjunto 14
Conjunto 3	3	Conjunto 7		Conjunto 11		Conjunto 15
Conjunto 4	8	Conjunto 8		Conjunto 12		Conjunto 16
TOTAL ERRORES: 16						

Añote el número del elemento techo, es decir, el número del último elemento del conjunto leído. Reste a ese número el número total de errores cometidos por el examinando desde el conjunto base al conjunto leído. El resultado es la puntuación directa.

Elemento techo: **48** Total errores: **16** Puntuación directa: **32**

$48 - 16 = 32$

Realice una marca sobre la línea correspondiente al CI obtenido. Marque una línea recta en vertical que cruce con las otras escalas (véase el manual para mayor información).

Opcionalmente pueden representarse también los intervalos de confianza optativos. Para ello, dibuje dos líneas verticales sobre todas las escalas, una a cada lado del CI obtenido. Para el intervalo de confianza del 68% utilice 13 puntuaciones CI (véase el manual si se desean otras opciones).

0,13% 2,14% 13,59% 34,13% 34,13% 13,59% 2,14% 0,13%

-4σ -3σ -2σ -1σ X +1σ +2σ +3σ +4σ

40 50 60 70 80 90 100 110 120 130 140 150 160

1 2 3 4 5 6 7 8 9

Autores: U.M. Dunn, L.M. Dunn y B. Arribas.
Copyright © 2006 by TEA Ediciones, S.A. • Edición: TEA Ediciones, S.A.; Frey Bernardino Sahagún, 14; 28038 MADRID • Este material está impreso en DOS TINTAS.
Si se presentan otro en tinta negra es una reproducción ilegal. En beneficio de la cultura y en el uso propio, NO LA UTILICE. Prohibida la reproducción.

1 2 años y medio - 3 años

Escoba	2	E
Avión	1	E
Beber	3	E
Pala	4	E
Columpiarse	4	E
Lámpara	4	E
Dinero	3	E
Helicóptero	2	E
Valla	3	E
Llave	4	E
Tambor	3	E
Subir	1	E

Total errores

2 4 años

Vaca	1	E
Nadar	1	E
Vacio	1	E
Excavar	2	E
Granjero	3	E
Accidente	2	E
Huido	3	E
Lanzar	4	E
Sobre	2	E
Castillo	3	E
Medir	4	E
Canguro	2	E

Total errores

3 5 años

Fruta	1	E
Cadena	2	E
Cactus	3	E
Puercoespín	1	E
Bostezar	2	E
Cabra	4	E
Decorado	4	E
Zorro	3	E
Garras	1	E
Discutir	1	E
Astronauta	1	E
Serrar	1	E

Total errores

4 6 - 7 años

Tronco	2	E
Enorme	3	E
Paracaidas	3	E
Entregar	1	E
Globo	2	E
Calculadora	2	E
Goteo	1	E
Calmeña	1	E
Lijar	2	E
Estatua	4	E
Aterrorizada	1	E
Rectángulo	1	E

Total errores

5 8 años

Marco	1	E
Equinaje	2	E
Escritura	1	E
Animar	1	E
Vehículo	4	E
Abrillantar	1	E
Apio	1	E
Ovalo	1	E
Hortaliza	3	E
Peludo	4	E
Premiar	1	E
Cerebro	2	E

Total errores

6 8 - 9 años

Molestar	1	E
Lima	2	E
Isla	2	E
Seleccionar	1	E
Par	3	E
Ángulo	1	E
Reptil	2	E
Mandíbula	4	E
Acanthado	1	E
Terror	3	E
Dirigir	2	E
Morsa	1	E

Total errores

7 10 - 11 años

Palmera	1	E
Depredador	3	E
Embudo	2	E
Repostar	2	E
Ajustable	2	E
Ruedar	3	E
Colisionar	1	E
Termo	1	E
Artico	2	E
Calcular	1	E
Trillizos	4	E
Contaminar	3	E

Total errores

8 12 años

Ramo	4	E
Olraje	2	E
Sahr	4	E
Vaina	3	E
Clasificar	1	E
Parra	1	E
Disecionar	2	E
Planeador	1	E
Suculento	1	E
Pelicano	1	E
Yate	4	E
Acoger	3	E

Total errores

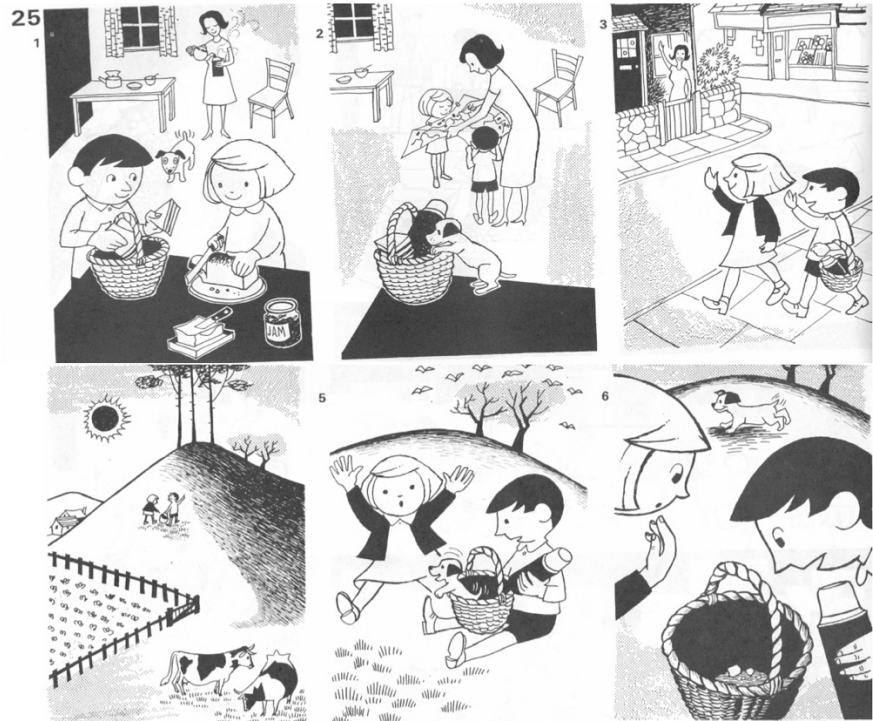
9 12 - 16 años

Arquera	2	E
Mamífero	2	E
Compositor	2	E
Oasis	1	E
Citrón	2	E
Lubricar	1	E
Velocímetro	3	E
Ilebeje	1	E
Izar	1	E
Reprimenda	1	E
Porcelana	2	E
Cuspidoso	2	E

Total errores

APPENDIX 4: THE DOG STORY

STORY 2



APPENDIX 5: DATA-EDITING CONVENTIONS FROM THE ORAL PRODUCTION TASKS AT THE END OF T3

Original data	Conversion used
mom	mum
mother	
mommy	
mummy	
bye-bye	goodbye
bye	
goodbye	
she's	she is
they're	they are

APPENDIX 6: EXAMPLE OF ORAL PRODUCTION TASK TRANSCRIPTIONS

Transcriptions for the IG

Example 1

Clan file

@Begin

@Language en

@Participants MAX XXX Subject, CRI Cristina Investigator

@ID en/belc/MA/5;03/ male/imm/ Subject

@Coder Cristina

@Date 12-JUN-2014

@Recording Track 9

location

the boy the girl and the mum are in the #house the girl prepares a ## sandwich the dog is ## near the basket the boy and the girl say #bye they go on a picnic the girl is## happy the dog is in the basket the boy and the girl look inside the basket

@End

1. Raw data

The boy the girl and the mom are in the house. The girl prepares a sandwich. The dog is near the basket. The boy and the girl say bye. They go on a picnic. The girl is happy. The dog is in the basket. The boy and the girl look inside the basket.

2. Plain text

the boy the girl and the mum are in the house the girl prepares a sandwich the dog is near the basket the boy and the girl say goodbye they go on a picnic the girl is happy the dog is in the basket the boy and the girl look inside the basket

Example 2

Clan file

@Begin

@Language en

@Participants SIL Silvia_XXX Subject, CRI Cristina Investigator

@ID en/belc/SIL/5;08/ female/imm/ Subject

@Coder Cristina

@Date 12-JUN-2014

@Recording track1

location

here is a boy a girl and a dog.## the boy and the girl are making a ####picnic. the dog is in the basket. the boy and the girl say goodbye to their mum and then sit down on the grass. ## the boy and girl look in the basket. the dog has eaten the food.

@End

1. Raw data

Here is a boy a girl and a dog. The boy and the girl are making a picnic. The dog is in the basket. The boy and the girl say goodbye to their mum and then sit down on the grass. The boy and girl look in the basket. The dog has eaten the food

2. Plain text

here is a boy a girl and a dog the boy and the girl are making a picnic the dog is in the basket the boy and the girl say goodbye to their mum and then sit down on the grass the boy and girl look in the basket the dog has eaten the food.

APPENDIX 7: RESULTS OF THE NORMALITY TESTS

Table 45: Normality tests for the MBCDI total scores for P1 (n= 24), P2 (n=19), P3 (n=8).

Measures	Data collection time	Group	n	Sig	
				Kolmogorov-Smirnov	Shapiro-Wilk
PhsU	T1	P1	24	.02	.00
		P2	19	.20	.31
		P3	8	.20	.19
	T2	P1	24	.04	.00
		P2	19	.10	.89
		P3	8	.20	.08
	T3	P1	24	.20	.85
		P2	19	.20	.51
		P3	8	.20	.27
WdsU	T1	P1	24	.15	.11
		P2	19	.06	.00
		P3	8	.17	.22
	T2	P1	24	.13	.06
		P2	19	.17	.200
		P3	8	.18	.27
	T3	P1	24	.03	.08
		P2	19	.20	.82
		P3	8	.20	.36
WdsP	T1	P1	24	.20	.75
		P2	19	.04	.14
		P3	8	.20	.23
	T2	P1	24	.20	.61
		P2	19	.20	.11
		P3	8	.20	.67
	T3	P1	24	.20	.50
		P2	19	.19	.50
		P3	8	.20	.32
M3L	T1	P1	24	.00	.00
		P2	19	.00	.00
		P3	8	.04	.00
	T2	P1	24	.00	.00
		P2	19	.00	.00
		P3	8	.00	.00

Measures	Data collection time	Group	n	Sig	
				Kolmogorov-Smirnov	Shapiro-Wilk
SC	T3	P1	24	.00	.00
		P2	19	.00	.00
		P3	8	.06	.01
	T1	P1	24		
		P2	19		
		P3	8		
	T2	P1	24	.07	.37
		P2	19	.20	.10
		P3	8	.03	.17
		P1	24	.00	.02
		P2	19	.20	.35
		P3	8	.20	.39

Shaded cells indicate absence of testing due to age.

Table 45: Cont.

Table 46: Normality tests for the Spanish and Catalan PPVT test scores for P1 (n= 24), P2 (n=19), P3 (n=8).

Measures	Data collection time	Group	n	Sig	
				Kolmogorov-Smirnov	Shapiro-Wilk
Spanish PPVT	T1	P1	24	.20	.37
		P2	19	.20	.85
		P3	8	.20	.60
	T2	P1	24	.20	.67
		P2	19	.20	.77
		P3	8	.20	.20
	T3	P1	24	.20	.33
		P2	19	.20	.87
		P3	8	.20	.11
Catalan PPVT	T1	P1	24	.20	.10
		P2	19	.13	.08
		P3	8	.20	.43
	T2	P1	24	.20	.62
		P2	19	.18	.25
		P3	8	.20	.33

Measures	Data collection time	Group	n	Sig	
				Kolmogorov- Smirnov	Shapiro- Wilk
		P1	24	.20	.74
	T3	P2	19	.20	.37
		P3	8	.20	.57

Table 46: Cont.

Table 47: Normality tests for the MBCDI total scores for the IG (n= 51) and RG1 (n=15).

Measures	Data collection time	Group	n	Sig.	
				Kolmogorov- Smirnov	Shapiro- Wilk
PhsU	T1	IG	51	.06	.05
		RG1	15	.20	.80
	T2	IG	51	.09	.08
		RG1	15	.08	.07
	T3	IG	51	.20	.55
		RG1	15	.05	.06
WdsU	T1	IG	51	.06	.07
		RG1	15	.14	.11
	T2	IG	51	.07	.08
		RG1	15	.11	.09
	T3	IG	51	.20	.22
		RG1	15	.20	.07
WdsP	T1	IG	51	.06	.31
		RG1	15	.06	.01
	T2	IG	51	.20	.23
		RG1	15	.62	.05
	T3	IG	51	.08	.07
		RG1	15	.20	.33
M3L	T1	IG	51	.00	.00
		RG1	15	.00	.00
	T2	IG	51	.00	.00
		RG1	15	.00	.00
	T3	IG	51	.00	.00
		RG1	15	.00	.00
SC	T1	IG	51	.07	.18
		RG1	15	.09	.08
	T2	IG	51	.07	.18
		RG1	15	.06	.08
	T3	IG	51	.08	.07
		RG1	15	.06	.09

Table 48: Normality tests for the Spanish/Catalan PPVT for the IG (n= 51) and RG1 (n=15).

Measures	Data collection time	Group	n	Sig.	
				Kolmogorov-Smirnov	Shapiro-Wilk
Spanish PPVT	T1	IG	51	.20	.38
		RG1	15	.06	.07
	T2	IG	51	.20	.45
		RG1	15	.20	.47
	T3	IG	51	.20	.44
		RG1	15	.09	.18
Catalan PPVT	T1	IG	51	.29	.06
		RG1	15	.09	.10
	T2	IG	51	.20	.53
		RG1	15	.20	.47
	T3	IG	51	.20	.45
		RG1	15	.20	.15

Table 49: Normality tests for the measures in the oral production task at the end of T3 for the IG (n= 51) and RG2 (n=44)

Measures	Group	n	Kolmogorov-Smirnov	Shapiro-Wilk
Tokens	IG	51	.09	.59
	RG2	44	.06	.08
Types	IG	51	.05	.14
	RG2	44	.49	.57
Families	IG	51	.02	.01
	RG2	44	.00	.00
Cw	IG	51	.20	.16
	RG2	44	.10	.32
Fw	IG	51	.20	.67
	RG2	44	.06	.20
N	IG	51	.20	.17
	RG2	44	.20	.82
V	IG	51	.51	.29
	RG2	44	.07	.06
Adj	IG	51	.00	.00
	RG2	44	.00	.00
Adv	IG	51	.00	.00
	RG2	44	.00	.00
Pron	IG	51	.00	.00
	RG2	44	.00	.00
Num	IG	51	.00	.00
	RG2	44	.00	.00
Prep	IG	51	.00	.00
	RG2	44	.00	.00
Art	IG	51	.63	.24
	RG2	44	.61	.10
Conj	IG	51	.30	.21
	RG2	44	.60	.40
GI	IG	51	.54	.32

Measures	Group	<i>n</i>	Kolmogorov-Smirnov	Shapiro-Wilk
M3L	RG2	44	.64	.59
	IG	51	.08	.22
	RG2	44	.14	.05
MLU	IG	51	.10	.08
	RG2	44	.200	.10

Table 49: Cont.

APPENDIX 8: MAPLE BEAR YEARLY SYLLABUSES

Nursery 1st year	Junior kindergarten 2nd year	Senior kindergarten 3rd year
What can I learn about myself?	All about me	All about me
What are shapes?	Shapes, patterns and sequences	Exploring with my five senses
What is special about my five senses?	Dogs	Our neighbourhood
What floats and what sinks?	Beans	Trees
What is special about air?	Elephants	Water water oceans are alive
The four seasons?	Egg layers	Farm and nutrition
Animals	Transportation	Fairy tales and dragons
The food I eat	Bees	The solar system