UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM INGLÊS:
ESTUDOS LINGUÍSTICOS E LITERÁRIOS

INTRALINGUAL SUBTITLES, INTERLINGUAL SUBTITLES, AND L2 VOCABULARY LEARNING: AN EXPLORATORY STUDY WITH BRAZILIAN EFL STUDENTS

RAFAEL MATIELO

Tese submetida à Universidade Federal de Santa Catarina em cumprimento parcial dos requisitos para obtenção do grau de

DOUTOR EM LETRAS

FLORIANÓPOLIS
Março, 2016

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Matielo, Rafael
Intralingual Subtitles, Interlingual Subtitles, and L2 Vocabulary Learning: An Exploratory Study with Brazilian EFL Students / Rafael Matielo ; orientadora, Roberta Pires de Oliveira ; coorientadora, Luciane Baretta. - Florianópolis, SC, 2016.
272 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Comunicação e Expressão. Programa de Pós Graduação em Inglês: Estudos Linguísticos e Literários.

Inclui referências

1. Inglês: Estudos Linguísticos e Literários. 2. legendas intralinguais. 3. legendas interlinguais. 4. inglês como língua estrangeira. 5. memória de trabalho. I. Pires de Oliveira, Roberta. II. Baretta, Luciane. III. Universidade Federal de Santa Catarina. Programa de Pós Graduação em Inglês: Estudos Linguísticos e Literários. IV. Título.

Esta Tese de Rafael Matielo, intitulada Intralingual Subtitles, Interlingual Subtitles, and L2 Vocabulary Learning: An Exploratory Study with Brazilian EFL Students, foi julgada adequada e aprovada em sua forma final, pelo Programa de Pós-Graduação em Inglês: Estudos Linguísticos e Literários, da Universidade Federal de Santa Catarina, para fins de obtenção do grau de

DOUTOR EM LETRAS

Área de concentração: Inglês e Literatura Correspondente
Opção: Língua Inglesa e Linguística Aplicada

_____
Profa. Dra. Anelise Reich Corseuil
Coordenadora

BANCA EXAMINADORA:

_____
Dra. Roberta Pires de Oliveira
Orientadora
Universidade Federal de Santa Catarina (UFSC)

_____
Dra. Luciane Baretta
Coorientadora
Universidade Estadual do Centro-Oeste (UNICENTRO) – Guarapuava

Dr. Celso Henrique Soufen Tumolo
Universidade Federal de Santa Catarina (UFSC)

Dra. Gicele Vergine Vieira Prebianca
Instituto Federal Catarinense (IFC)

Dra. Lêda Maria Braga Tomitch
Universidade Federal de Santa Catarina (UFSC)

Dra. Rosane Silveira
Universidade Federal de Santa Catarina (UFSC)

Dra. Viviane Maria Heberle
Universidade Federal de Santa Catarina (UFSC)

Florianópolis, 18 de março de 2016.

Ah, se pudéssemos contar…
As voltas que a vida dá!
Marcus Viana, 2001

ACKNOWLEDGEMENTS

I am fortunate to have met incredible people who were willing to make this road a little less bumpy. I hope these heartfelt words can help render my everlasting gratitude for each and every one of you listed here.

I am very appreciative of the guidance and the academic support that I have received from my advisor, Dr. Roberta Pires de Oliveira. Even though you boarded the boat when its course had already been set out, you have been extremely helpful and decisive in determining safe routes in this uncharted territory. Thank you for helping me restore my confidence as an academic and as a researcher.

I cannot find the words to describe my deepest gratefulness and honor to have been co-supervised by Dr. Luciane Baretta. I may never find anyone else in the academia who is as admirable, committed, fair, hardworking, human, inspiring, kind, loyal, responsible, and trustworthy as you are. You are undoubtedly one of the most important reasons why I have carried on until the very end, with sanity and humor [almost] intact.

I would like to thank the examining committee for having accepted to be part of the examination process: Dr. Celso Tumolo, Dr. Lêda Tomitch, Dr. Rosane Silveira, and Dr. Viviane Heberle from UFSC, and the external members, Dr. Claudia Finger-Kratochvil (UFFS) and Dr. Gicele Prebianca (IFC). I am positive that your contributions will resonate upon this and future works of mine.

I wish to express how thankful I am to Dr. Rosane Silveira, in particular, for welcoming me to her home in Summer 2014, during her short stop in Florianópolis in her sabbatical year, when she kindly went over some of the statistics. One day, I hope I can be as knowledgeable as you are. Your example of integrity and professionalism is one to which I look up and which everyone should recognize and follow.

I would like to thank the participants of this doctoral study enrolled in all of the English 5 levels in the Extracurricular (non-credit) Language Courses at UFSC, who accepted to participate in my research. My special thanks also go to the English teachers in charge of the groups and the raters of this study (whose identities are to be preserved for ethical reasons), who were so cooperating throughout the data collection and the rating process. Without all of you, this study would have never been possible.

I would also like to express my gratitude to academic peers and colleagues: Dani, Deise, Fernanda, Filipi, Kátia, Marisol, Maurílio, Milena, Nanda, Nayara, Pâmela, Paulo, Rodrigo, Stephanie, Thaís Collet, Thaís Schadech, and Vinícius. Thank you for not just talking the talk, but walking the walk. I would also like to thank Leandro, for coming to my rescue, saving me.

I would like to thank my students and co-workers at Complexo de Ensino Superior de Santa Catarina (CESUSC) for coping with my limitations and absences. You have made me feel so welcome in such a healthy higher education environment.

I also wish to thank CNPq for the scholarship I was granted throughout these three years, which has enabled me to focus on my research and travel around the country to attend conferences, sharing and building knowledge with fellow researchers from other institutions.

I am indebted to two special people: my first English teachers in my hometown, Cidinha and Fernanda, who inspired me to become an English teacher myself from the very early on and who told me repeatedly that I should study hard and seek solid professional development if I ever wanted to be serious about teaching English and working with the language. Thank you for the amazing lessons and for teaching me how to express myself using this language.

Finally, I want to express my profound appreciation to my family. I sense that words as ‘academia’, ‘dissertation’, and ‘conference’ have become almost forbidden in our conversations over the phone. Nonetheless, I am compelled to remind you that all of this has also been for you, in a strange, illogical way.

Suddenly, I was an undergraduate student at UFSC, in 2007, full of hopes and dreams. Three years later, I was majoring in Letras-Inglês. Then, suddenly, I was a graduate student, pursuing a Master’s Degree in English and Applied Linguistics at Programa de Pós-Graduação em Inglês (PPGI), in 2010. Suddenly again, I was a PhD student, in 2013. Suddenly… suddenly… and suddenly. This is just one way of looking at it. Another way of looking at how it has all unfurled – which I prefer – is by accepting the fact that I have learned what I was supposed to learn, met the people that I was supposed to meet, faced the challenges that I was supposed to face, and that I have collaborated with just the right people. Everything has unraveled exactly the way it was supposed to unravel. It has, I reckon, been written and planned all along. *Maktub*.

ABSTRACT

INTRALINGUAL SUBTITLES, INTERLINGUAL SUBTITLES, AND L2 VOCABULARY LEARNING: AN EXPLORATORY STUDY WITH BRAZILIAN EFL STUDENTS

This dissertation explores the effects of intralingual and interlingual subtitles on Brazilian EFL learners' development. More specifically, it seeks to explore (i) the effects of intralingual and interlingual subtitles on learners' general and specific video comprehension; (ii) the effects of intralingual and interlingual subtitles on learners' L2 vocabulary recognition; (iii) and a possible correlation amongst learners' Working Memory (WM) capacity, their level of general and specific video comprehension, and their level of L2 vocabulary recognition. A total of 36 intermediate-level EFL learners, enrolled at the Extracurricular Language Courses at Universidade Federal de Santa Catarina (UFSC), were divided into two experimental groups (intralingual subtitles and interlingual subtitles) and one control group (soundtrack only). In addition to profile data, participants' performance was also measured on three sets of tests: (i) a general and a specific video comprehension test; (ii) an L2 vocabulary recognition test, comprising a pre-test, a test, and a post-test; (iii) and two WM tests, a Reading Span Test (RST) and an Operation-Word Span Test (OSPAN). Data were analyzed both quantitatively and qualitatively. As regards the effects of subtitles on video comprehension, statistical tests and analyses performed indicated more beneficial effects when using intralingual subtitles over the other conditions, even though participants' performance in the experimental conditions was not statistically different from one another. As for the effects of subtitles on L2 vocabulary recognition, experimental conditions were not found to substantially foster it, and no statistically significant differences among the experimental groups and the control group were found. Across time, the results obtained on the L2 vocabulary recognition tests point out to more positive growth in performance obtained by the intralingual subtitles group, followed by the interlingual subtitles group, and then the control group. Finally, concerning the correlation tests performed, participants' WM capacity did not significantly interact with their performance on any of the tests, which may possibly be related to the study design at large.

Kewords: intralingual subtitles; interlingual subtitles; L2 comprehension; L2 vocabulary recognition; working memory.

Number of pages: 210 (excluding appendixes)
Number of words: 63.316 (excluding appendixes)
Number of pages: 272 (including appendixes)
Number of words: 68.520 (including appendixes)

RESUMO

LEGENDAS INTRALINGUAIS, LEGENDAS INTERLINGUAIS E APRENDIZAGEM DE VOCABULÁRIO EM L2: UM ESTUDO EXPLORATÓRIO COM ESTUDANTES BRASILEIROS DE ILE

Esta tese de doutorado explora os efeitos de legendas intralinguais e interlinguais no desenvolvimento de aprendizes brasileiros de ILE. Especificamente, este estudo busca explorar (i) os efeitos de legendas intralinguais e interlinguais na compreensão geral e específica de vídeo pelos aprendizes; (ii) os efeitos de legendas intralinguais e interlinguais no reconhecimento de vocabulário em L2 pelos aprendizes; (iii) e uma possível correlação entre a capacidade de Memória de Trabalho (MT) dos aprendizes, seu nível de compreensão geral e específica e seu nível de reconhecimento de vocabulário em L2. Um total de 36 aprendizes de nível intermediário de ILE, matriculados nos Cursos Extracurriculares de Línguas da Universidade Federal de Santa Catarina (UFSC), foram divididos em dois grupos experimentais (legendas intralinguais e legendas interlinguais) e um grupo controle (sem legendas). Além de informações sobre seus perfis, o desempenho dos participantes também foi medido por meio de três testes: (i) teste de compreensão geral e específica de vídeo; (ii) teste de reconhecimento de vocabulário em L2, contendo um pré-teste, um teste e um pós-teste; (iii) e dois testes de MT, sendo um Teste de Alcance de Leitura (TAL) e um Teste Soma-Palavra (TSP). Os dados foram analisados de forma quantitativa e qualitativa. Com relação aos efeitos das legendas na compreensão de vídeo, os testes estatísticos e análises feitas indicaram mais efeitos benéficos com legendas intralinguais, ainda que o desempenho dos participantes nas condições experimentais não tenha sido estatisticamente diferente entre si. Sobre os efeitos de legendas em reconhecimento de vocabulário em L2, as condições experimentais não pareceram promovê-lo substancialmente, não sendo encontradas diferenças estatisticamente significantes entre os grupos. Ao longo do tempo, os resultados obtidos nos testes de reconhecimento de vocabulário em L2 demonstraram maior crescimento no desempenho por parte do grupo de legendas intralinguais, seguido pelo grupo de legendas interlinguais e depois pelo grupo controle. Finalmente, no tocante aos testes estatísticos de correlação, a capacidade de MT não interagiu significativamente com o

desempenho dos participantes nos demais testes, estando possivelmente relacionado ao design do estudo.

Palavras-chave: legendas intralinguais; legendas interlinguais; compreensão em L2; reconhecimento de vocabulário em L2; memória de trabalho.

Número de páginas: 210 (excluindo anexos)
Número de palavras: 63.316 (excluindo anexos)
Número de páginas: 272 (incluindo anexos)
Número de palavras: 68.520 (incluindo anexos)

TABLE OF CONTENTS

CHAPTER I – INRODUCTION 22
1.1 Introductory Remarks 22
1.2 Statement of the Purpose and Significant of the Research 24
1.3 Organization of the Dissertation 26
CHAPTER II – INTRALINGUAL SUBTITLES, INTERLINGUAL SUBTITLES, COGNITIVE ASPECTS, AND L2 DEVELOPMENT 27
2.1 Intralingual Subtitles and L2 Development 27
2.2 Intralingual Subtitles, Interlingual Subtitles, and L2 Development 35
2.3 Summary of the Findings of the Literature on Subtitling and L2 Development 46
2.4 Other Studies on Subtitling and Cognitive Aspects 48
CHAPTER III – L2 VOCABULARY DEVELOPMENT AND WORKING MEMORY 57
3.1 L2 Vocabulary Development: Models and Perspectives 57
3.2 Empirical Studies on L2 Vocabulary Development 64
3.3 Working Memory 69
3.4 Empirical Studies on L2 Vocabulary Development and Working Memory 75
CHAPTER IV – METHOD 81
4.1 Objectives 81
4.2 Research Questions 81
4.3 Hypotheses 82
4.4 Participants 83
4.5 Materials: On the TV Series 85
4.6 Materials: On the Video Comprehension Test 86
4.7 Materials: On the L2 Vocabulary Recognition Tests 87
4.8 Materials: On the Working Memory Measures 88
4.8.1 The Reading Span Test (RST)..... 89
4.8.2 The Operation-Word Span Test (OSPAN)..... 91
4.9 Questionnaires 93

4.10 Data Collection Procedures and Research Design 95
4.11 Tests' Scoring and Statistical Procedures 96
4.11.1 Scoring of Video Comprehension Test (Parts A and B) 97
4.11.2 Scoring of the L2 Vocabulary Recognition Tests 98
4.11.3 Scoring of the Working Memory Measures 99
CHAPTER V – RESULTS 102
5.1 Chapter Outline 102
5.2 Video Comprehension Test 102
5.2.1 General Comprehension: Descriptive Statistics and Results 103
5.2.2 Specific Comprehension: Descriptive Statistics and Results 107
5.3 L2 Vocabulary Recognition Tests: Descriptive Statistics and Results..... 111
5.4 Working Memory Tests: Descriptive Statistics and Results 117
5.4.1 Scores on the Reading Span Test (RST) 118
5.4.2 Scores on the Operation-Word Span Test (OSPAN) .. 120
5.5 RST and OSPAN Correlations 122
5.6. Video Comprehension Test and Working Memory Correlations..... 128
5.6.1 General Comprehension and Working Memory Correlations 129
5.6.2 Specific Comprehension and Working Memory Correlations 137
5.7 L2 Vocabulary Recognition and Working Memory Correlations..... 145
5.7.1 L2 Vocabulary Recognition Pre-Test and Working Memory Correlations 145
5.7.2 L2 Vocabulary Recognition Test and Working Memory Correlations 148
5.7.3 L2 Vocabulary Recognition Test and Working Memory Correlations 151
5.8 Summary of Hypotheses and Results 154

CHAPTER VI – DISCUSSION 156
6.1 Intralingual Subtitles, Interlingual Subtitles, and L2 Comprehension 156
6.2 Intralingual Subtitles, Interlingual Subtitles, and L2 Vocabulary Recognition 164
6.3 Subtitling, L2 Comprehension, L2 Vocabulary Recognition, and Working Memory 174
6.3.1 Subtitling, L2 Comprehension, and Working Memory 175
6.3.2 Subtitling, L2 Vocabulary Recognition, and Working Memory 180
CHAPTER VII – CONCLUSION 185
7.1 Major Findings 185
7.2 Limitations and Suggestions for Further Research 187
7.3 Methodological Implications 189
7.4 Pedagogical Implications 191
7.5 Concluding Remarks 194
REFERENCES 196
APPENDIXES 212

LIST OF TABLES

Table 1 – Word Learnability Factors (Laufer, 1997) 64
Table 2 – Research Design 96
Table 3 – Tests of normality: General video comprehension..... 243
Table 4 – Tests of normality: Specific video comprehension 243
Table 5 – Inter-rater reliability for the L2 vocabulary recognition – pre-test: Intralingual subtitles group 248
Table 6 – Inter-rater reliability for the L2 vocabulary recognition – pre-test: Interlingual subtitles group 248
Table 7 – Inter-rater reliability for the L2 vocabulary recognition – pre-test: Control group 248
Table 8 – Inter-rater reliability for the L2 vocabulary recognition - test: Intralingual subtitles group 249
Table 9 – Inter-rater reliability for the L2 vocabulary recognition - test: Interlingual subtitles group 249
Table 10 – Inter-rater reliability for the L2 vocabulary recognition - test: Control group 249
Table 11 – Inter-rater reliability for the L2 vocabulary recognition – post-test: Intralingual subtitles group 250
Table 12 – Inter-rater reliability for the L2 vocabulary recognition – post-test: Interlingual subtitles group 250
Table 13 – Inter-rater reliability for the L2 vocabulary recognition – post-test: Control group 250
Table 14 – Tests of normality: L2 vocabulary recognition tests 251
Table 15 – Tests of normality: Working memory tests (RST and OSPAN) 258
Table 16 – General comprehension results 103
Table 17 – Kruskal-Wallis one-way ANOVA on general comprehension 105
Table 18 – Kruskal-Wallis one-way ANOVA on general comprehension: intralingual and interlingual subtitles groups 106
Table 19 – Kruskal-Wallis one-way ANOVA on general comprehension: interlingual subtitles and control groups 106

Table 20 – Kruskal-Wallis one-way ANOVA on general comprehension: intralingual subtitles and control groups 106
Table 21 – Specific comprehension results 107
Table 22 – Kruskal-Wallis one-way ANOVA on specific comprehension 109
Table 23 – Kruskal-Wallis one-way ANOVA on specific comprehension: intralingual and interlingual subtitles groups 109
Table 24 – Kruskal-Wallis one-way ANOVA on specific comprehension: interlingual subtitles and control groups 110
Table 25 – Kruskal-Wallis one-way ANOVA on specific comprehension: intralingual subtitles and control groups 110
Table 26 – L2 vocabulary tests results 111
Table 27 – Kruskal-Wallis one-way ANOVA on L2 vocabulary recognition tests 113
Table 28 – Friedman's Two-Way ANOVA Test on L2 vocabulary recognition: intralingual subtitles group 114
Table 29 – Friedman's Two-Way ANOVA Test on L2 vocabulary recognition: interlingual subtitles group 114
Table 30 – Friedman's Two-Way ANOVA Test on L2 vocabulary recognition: control group 114
Table 31 – Gains scores in L2 vocabulary recognition 116
Table 32 – RST scores 118
Table 33 – OSPAN scores 120
Table 34 – RST and OSPAN Correlations 123
Table 35 – General Comprehension and WM Correlations 129
Table 36 – Specific Comprehension and WM Correlations 137
Table 37 – L2 vocabulary recognition pre-test and WM correlations 145
Table 38 – L2 vocabulary recognition test and WM correlations ... 148
Table 39 – L2 vocabulary recognition post-test and WM correlations .151
Table 40 – Hypotheses and major results 155
Table 41 – General comprehension raw scores..... 267
Table 42 – Specific comprehension raw scores 269
Table 43 – L2 vocabulary recognition raw scores 271

LIST OF FIGURES

Figure 1 –Summary of the studies on subtitling and language learning – Part I 46
Figure 2 –Summary of the studies on subtitling and language learning – Part II 47
Figure 3 – Lexical representation (a) and processing (b) at the beginning of L2 lexical development 60
Figure 4 – Lexical representation (a) and processing (b) in L2 lexical development at the second stage 61
Figure 5 – Lexical representation (a) and processing (b) in L2 lexical development at the final stage 61
Figure 6 – You's (2011) model of vocabulary acquisition 63
Figure 7 – Baddeley & Hitch's (1974) multi-component model of WM 71
Figure 8 – Baddeley's (2000) WM model revision 73

LIST OF GRAPHS

Graph 1 – Scatterplot – RST strict and OSPAN strict correlations for the intralingual subtitles group 125
Graph 2 – Scatterplot – RST lenient and OSPAN lenient correlations for the intralingual subtitles group 125
Graph 3 – Scatterplot – RST strict and OSPAN strict correlations for the interlingual subtitles group 126
Graph 4 – Scatterplot – RST lenient and OSPAN lenient correlations for the interlingual subtitles group 126
Graph 5 – Scatterplot – RST strict and OSPAN strict correlations for the control group 127
Graph 6 – Scatterplot – RST lenient and OSPAN lenient correlations for the control group 127
Graph 7 – Scatterplot – General Comprehension Scores and RST strict correlations for the intralingual subtitles group 130
Graph 8 – Scatterplot – General Comprehension Scores and RST lenient correlations for the intralingual subtitles group 131
Graph 9 – Scatterplot – General Comprehension Scores and OSPAN strict correlations for the intralingual subtitles group 131
Graph 10 – Scatterplot – General Comprehension Scores and OSPAN lenient correlations for the intralingual subtitles group 132
Graph 11 – Scatterplot – General Comprehension Scores and RST strict correlations for the interlingual subtitles group 132
Graph 12 – Scatterplot – General Comprehension Scores and RST lenient correlations for the interlingual subtitles group 133
Graph 13 – Scatterplot – General Comprehension Scores and OSPAN strict correlations for the interlingual subtitles group 133
Graph 14 – Scatterplot – General Comprehension Scores and OSPAN lenient correlations for the interlingual subtitles group 134
Graph 15 – Scatterplot – General Comprehension Scores and RST strict correlations for the control group 134
Graph 16 – Scatterplot – General Comprehension Scores and RST lenient correlations for the control group 135

Graph 17 – Scatterplot – General Comprehension Scores and OSPAN strict correlations for the control group 135
Graph 18 – Scatterplot – General Comprehension Scores and OSPAN lenient correlations for the control group 136
Graph 19 – Scatterplot – Specific Comprehension Scores and RST strict correlations for the intralingual subtitles group 138
Graph 20 – Scatterplot – Specific Comprehension Scores and RST lenient correlations for the intralingual subtitles group 139
Graph 21 – Scatterplot – Specific Comprehension Scores and OSPAN strict correlations for the intralingual subtitles group 139
Graph 22 – Scatterplot – Specific Comprehension Scores and OSPAN lenient correlations for the intralingual subtitles group 140
Graph 23 – Scatterplot – Specific Comprehension Scores and RST strict correlations for the interlingual subtitles group 140
Graph 24 – Scatterplot – Specific Comprehension Scores and RST lenient correlations for the interlingual subtitles group 141
Graph 25 – Scatterplot – Specific Comprehension Scores and OSPAN strict correlations for the interlingual subtitles group 141
Graph 26 – Scatterplot – Specific Comprehension Scores and OSPAN lenient correlations for the interlingual subtitles group 142
Graph 27 – Scatterplot – Specific Comprehension Scores and RST strict correlations for the control group 142
Graph 28 – Scatterplot – Specific Comprehension Scores and RST lenient correlations for the control group 143
Graph 29 – Scatterplot – Specific Comprehension Scores and OSPAN strict correlations for the control group 143
Graph 30 – Scatterplot – Specific Comprehension Scores and OSPAN lenient correlations for the control group 144
Graph 31 – Scatterplot – L2 vocabulary recognition pre-test and WM correlations for the intralingual subtitles group 146
Graph 32 – Scatterplot – L2 vocabulary recognition pre-test and WM correlations for the interlingual subtitles group 147
Graph 33 – Scatterplot – L2 vocabulary recognition pre-test and WM correlations for the control group 147

Graph 34 – Scatterplot – L2 vocabulary recognition test and WM correlations for the intralingual subtitles group 149
Graph 35 – Scatterplot – L2 vocabulary recognition test and WM correlations for the interlingual subtitles group 150
Graph 36 – Scatterplot – L2 vocabulary recognition test and WM correlations for the control group 150
Graph 37 – Scatterplot – L2 vocabulary recognition post-test and WM correlations for the intralingual subtitles group 152
Graph 38 – Scatterplot – L2 vocabulary recognition post-test and WM correlations for the interlingual subtitles group 153
Graph 39 – Scatterplot – L2 vocabulary recognition post-test and WM correlations for the interlingual subtitles group 153

LIST OF APPENDIXES

Appendix A – Consent Form 212
Appendix B – Student Profile Questionnaire 215
Appendix C – L2 Vocabulary Recognition Pre-Test 219
Appendix D – L2 Vocabulary Recognition Test 221
Appendix E – L2 Vocabulary Recognition Post-Test 224
Appendix F – Reading Span Test 226
Appendix G – Reading Span Test Instructions 231
Appendix H – Operation-Word Span Test 233
Appendix I – Operation-Word Span Test Instructions 237
Appendix J – Teachers Profile Questionnaire 239
Appendix K – Video Retrospective Questionnaire 241
Appendix L – Tests of Normality for the Video Comprehension Test (General and Specific Parts) 243
Appendix M – Histograms and Boxplots for the Video Comprehension Test (General Comprehension) 244
Appendix N – Histograms and Boxplots for the Video Comprehension Test (Specific Comprehension) 246
Appendix O – Inter-rater Reliability for the L2 Vocabulary Recognition – Pre-Test 248
Appendix P – Inter-rater Reliability for the L2 Vocabulary Recognition – Test 249
Appendix Q – Inter-rater Reliability for the L2 Vocabulary Recognition – Post-Test 250
Appendix R – Tests of Normality for the L2 Vocabulary Recognition Tests 251
Appendix S – Histograms and Boxplots for the L2 Vocabulary Recognition Tests 252
Appendix T – Tests of Normality for the Working Memory Tests 258
Appendix U – Histograms and Boxplots for the Working Memory Tests 259
Appendix V – Participants' General Comprehension Raw Scores 267
Appendix W – Participants' Specific Comprehension Raw Scores 269
Appendix X – Participants' L2 Vocabulary Recognition Raw Scores 271

CHAPTER I
INTRODUCTION

1.1 Introductory Remarks

The field of Second Language Acquisition (henceforth SLA) has benefitted considerably from the worldwide scholar attention concerning the investigation of the use of subtitled and captioned videos in the classroom. Ever since Karen Price's groundbreaking work in 1983, a substantial body of knowledge of the effects of and the effects with subtitles and captions in language learning has been gathered. Throughout the last thirty years or so, researchers have steadily taken steps towards a deeper understanding of how leaners may benefit from subtitling in terms of their language development. Despite what may seem to be a long period of academic inquiry, much is still unknown (Vanderplank, 2015).

Such increasing interest in the potentials of subtitled and captioned video materials for language learning might stem from the fact that they "seem to provide a rich context for foreign language acquisition" (Koolstra & Beentjes, 1999, p. 51). Furthermore, some scholars have advocated for the use of the First Language (L1) as a valuable resource in the classroom (Atkinson, 1987; Auerbach, 1993; Lucindo, 2006; Romanelli, 2006, 2009; Schweers, 1999; Souza, 1999, to name but a few), which might help explain the use of interlingual subtitled materials for language teaching and learning purposes.

In order to make videos accessible to populations that do not fully know the language spoken in the dialogues of the videos or to facilitate students' overall comprehension, subtitles tend to be used. Regarding this translational aid, an important distinction merits attention: while interlingual subtitles[1] specifically refer to target-language texts, varying depending on the country, typically displayed at the bottom of the screen, intralingual subtitles refer to same-language subtitles (captions), which originally had the function to serve the hearing-impaired (Neuman & Koskinen, 1992).

[1] Interlingual subtitles are also commonly referred to as standard subtitles or simply as subtitles, that is, the type of translational aid involving one linguistic pair (when soundtrack and subtitles present two different languages). Intralingual subtitles, on the other hand, are also often referred to as captions in that they present the same language on both soundtrack and subtitles, though a few technical differences can be observed (Danan, 2004). However, for the purposes of this study, the terms interlingual subtitles and intralingual subtitles will be hereby mostly used, unless the original study reported here uses the term captions, which then will be kept whenever made reference to it.

The driving force behind the instructional use of subtitled video materials also seems to have been pushed forward by some other specific reasons. According to Neuman and Koskinen (1992), it is noteworthy to mention that: (i) videos' combination of sounds and pictures might enhance the relationship between words and meanings; (ii) videos have entertainment qualities that constitute a potential advantage over texts; and (iii) viewing could be perceived as a cognitively active experience – when suitable material is used (Anderson & Collins, 1988; Neuman, 1989). Moreover, from a learning styles perspective, subtitled videos might cater for both visual and auditory types of learners.

Studies[2] on the instructional and non-instructional use of interlingual and intralingual subtitled videos have addressed a number of topics that include, but are not limited to: the improvement of Foreign/Second Language (L2)[3] reading (Markham & Peter, 2003), L2 listening comprehension and L2 word recognition (Garza, 1991; Huang & Eskey, 1999; Koolstra & Beentjes, 1999; Peter, 2003; Markham, Peter & McCarthy, 2001; Stewart & Pertusa, 2004; Taylor, 2005; Winke, Gass & Sydorenko, 2010), L2 vocabulary development (Bianchi & Ciabattoni, 2008; Bird & Williams, 2002; d'Ydewalle; &Van De Poel, 1999; Garza, 1991; Koolstra & Beentjes, 1999; Neuman & Koskinen, 1992), the effects on implicit and explicit memory and cognitive processing (Bird & Williams, 2002), and the acquisition of L2 grammar (Bianchi & Ciabattoni, 2008; Van Lommel, Laenen & d'Ydewalle, 2006). It is worthwhile pointing out, however, that the literature lacks systematic studies on the use of interlingual and intralingual subtitles for language learning purposes involving the Brazilian population though one piece of research in the niche was in fact recently carried out by this researcher and colleagues (Matielo, Collet & D'Ely, 2013).

Though the significant number of studies previously mentioned signals a prolific niche of research in SLA, one might contend that the investigation of the use of subtitled video materials upon any domain of language learning does not seem to take a central position in the research agenda within the field. Furthermore, provided the lack of subtitling

[2] See Chapter III for a thorough review of the studies hereby listed.

[3] The importance, implications, and underlying assumptions that the terms Foreign Language and Second Language pose are not to be disregarded. Also, I am well aware of the differences between these two terms, as posed by Krashen (1982/2009), but I side with Ellis (2008), who argues that, for now, it is better not to treat these two situations as different since it still remains unknown how different the learning processes in each of them are. Hence, L2 will be used as an umbrella term in the present study. Moreover, the same rationale is adopted concerning Learning and Acquisition, which, in this study, will be treated as equal and used interchangeably for the sake of convenience.

studies involving the Brazilian population, the present research intends to explore the effects of intralingual and interlingual subtitles on language learning, more specifically on how these translational aids affect general comprehension and specific comprehension of the story narrated on screen. Finally, it is aimed at exploring the effects of intralingual and interlingual subtitles on L2 vocabulary recognition and whether they correlate with learners' Working Memory (WM) capacity[4].

1.2 Statement of the Purpose and Significance of the Research

The use of intralingual and interlingual subtitles to investigate language learning has been gaining prominence in the past years. From the 1980s to the present day, more than sixty papers have been published in different journals all around the world, which all report on experiments with different populations and target languages (Matielo, D'Ely & Baretta, 2015). Interestingly, only a handful of them[5] have encompassed Brazilian learners of English, which might mean that such a topic has not yet flourished among scholars in the field of SLA and Applied Linguistics.

To date, subtitles have generally been found to enhance language comprehension considerably, regardless of being interlingual or intralingual (d'Ydewalle & Van de Poel, 1999; Huang & Eskey, 1999; Koolstra & Beentjes, 1999; Markham, 1999; Markham & Peter, 2003; Danan, 2004; Stewart & Pertusa, 2004; Taylor, 2005; Caimi, 2006; Chang, 2006; Van Lommel, Laenen & d'Ydewalle, 2006; Sydorenko, 2010; Winke, Gass & Sydorenko, 2010; Zarei & Rashvand, 2011; Raine, 2013; Rokni & Ataee, 2014, to mention but a few).

Some studies, however, have not shown significant differences considering subtitling availability and its relationship with the specific language component being tested, such as general/listening comprehension (Bianchi & Ciabattoni, 2008; Latifi, Mobalegh & Mohammadi, 2011; Perez, Peters & Desmet, 2013; Sharif & Ebrahimian, 2013) and vocabulary development (Yuksel & Tanriverdi, 2009; Matielo, Collet & D'Ely, 2013; Zarei & Gilanian, 2013). As promising as the use of subtitles in language learning may seem, when looking at studies with interlingual and intralingual subtitles comparatively, the results yielded

[4] The concept of WM is properly addressed in section 3.3.

[5] A thorough search using CAPES database revealed no MA or PhD study on the use of subtitling for language learning purposes to date.

are therefore somewhat multifaceted, favoring the former in some cases and the latter in others.

The picture becomes even blurrier when adding the simultaneous processing of audio and subtitles to the equation. Since reading subtitles has been found to be an automatic process, regardless of one's familiarity with them, knowledge of the foreign language in the soundtrack or its availability – as confirmed in studies with eye-movement recordings (d'Ydewalle & Gielen, 1992; Bisson, Van Heuven, Conklin & Tunney, 2012; Winke, Gass & Sydorenko, 2013) – it then seems reasonable to assume that part of one's attentional resources when watching a subtitled film is allocated in reading and processing the subtitles, whereas the rest is ideally distributed between the processes of watching and attending to the story, processing the motion picture, and arguably attending to the auditory input. Hence, the ability to successfully attend to the whole set of input – the simultaneous use of spoken (audio/soundtrack) and written (interlingual and intralingual subtitles) input modes – merits further research.

One possible variable that might account for the extent to which learners are able to attend to the input in the video while engaged in video watching activities is WM, due to the fact that such a construct refers to an integrated system that temporarily stores and manipulates information during the performance of a given cognitive task (Baddeley, 1992; Baddeley, 2011). Since attending to audio, video, and subtitles simultaneously can be taken to mean a highly cognitive demanding task, the systematic exploration of the correlation between learners' WM capacity and their ability to both understand the content of the video and recognize vocabulary previously presented in the video seems to be relevant and necessary.

In this vein, exploring whether intralingual or interlingual subtitles seem to be more beneficial to Brazilian L2 leaners of English to aid their understanding of the content of the video watched and whether their level of understanding and word recognition is associated with their WM capacity is extremely relevant, especially considering the pedagogical insights derived from such an investigation. Given that both learners and teachers often times use subtitled videos for language learning purposes, and the fact that very little is known about the effects of such combinations on this population – Brazilian L2 learners of English – the present doctoral dissertation hopes to contribute to the field of SLA by exploring these matters.

1.3 Organization of the Dissertation

This dissertation comprises six chapters along with this introductory Chapter. Chapters II and III provide the theoretical background for the present study and have been divided for theme readability purposes. More specifically, Chapter II focuses on the relevant literature on interlingual and intralingual subtitles and L2 development. It addresses studies that have reported on the instructional and experimental use of subtitling, both intralingual and interlingual, for language learning purposes. More specifically, this section revisits important findings derived from the research on subtitling and both comprehension and vocabulary development.

Chapter III, in turn, addresses key aspects in vocabulary development. It also centers on characterizing WM – an important variable investigated in this study –, and reviews some studies reported in the literature that have looked at the correlation between WM and vocabulary development.

Chapter IV presents and describes the method and procedures employed for data collection and data analysis. It provides detailed description of the participants, the selection and development of instruments for data collection, the research questions guiding this study, the data collection procedures and research design, tests' scoring and statistical procedures adopted to examine the data gathered.

Chapter V presents the results of the statistical analyses carried out on the raw data collected. In Chapter VI, the discussion of the results is provided, beginning with the general comprehension and specific comprehension results aided by the different subtitling conditions. After that, the discussion centers on the results of the vocabulary recognition tests. Finally, in Chapter VI, a possible a relationship between the comprehension test, the L2 vocabulary recognition test, and the measures obtained in the WM tests is explored. In the end of Chapter VI, a summary of the main findings is also provided as the research questions and hypotheses that guide the present research are revisited.

Finally, Chapter VII recaps the main results of this investigation in light of the objectives set out in the beginning of this PhD dissertation. Moreover, the section addresses the theoretical and pedagogical implications, limitations of this study, suggestions for further research, and a few concluding remarks.

CHAPTER II
INTRALINGUAL SUBTITLES, INTERLINGUAL SUBTITLES, COGNITIVE ASPECTS, AND L2 DEVELOPMENT

This Chapter addresses key studies on the experimental and instructional use of intralingual and interlingual subtitles on L2 development. It starts by reviewing studies that have focused on the use of intralingual subtitles only, followed by studies that have comparatively examined the effects of intralingual and interlingual subtitles. Then, for readability purposes, a summary of the main findings is provided to help visualize the variety of puzzling results reported in the literature. Finally, relevant literature on subtitling and cognitive related aspects is reviewed with a few to providing extra insightful information on the effects of subtitling processing and language development.

2.1 Intralingual Subtitles and L2 Development

Evaluating the use of captioned video materials in advanced language classes was the crux of the discussion put forward by Garza (1991) in the beginning of the 1990s. He reports on a study addressing captioned videos used as a pedagogical tool in advanced Russian and English as Second Language (ESL) classes. Garza established a set of criteria for the selection and production of the video test materials, including whether the videos contained unmarked and appropriate language, as well as their grammatical and lexical complexity, interest value to university-level students, and a variety of salient speech functions. Whether the videos presented a high audio/video correlation was also taken into account, with a view to providing "a kind of 'visual glossary' for difficult or obscure lexicon" (p. 241).

Verbatim captions (word for word) and a multiple-choice based test were adopted to check for content comprehension, general and specific. Questions required the participants to provide informational paraphrases, basic deductions, or synonym identification of low-frequency words, after having watched the video segments twice. The study involved 40 adult Russian as a Foreign Language learners – native speakers of English – who were divided into two groups (the test group, with captions, and the control group, without captions) and 70 adult ESL learners – who spoke nine different native languages – equally divided into test and control groups. Five participants were randomly chosen from each test group to engage in a five-minute oral interview to retell any of the video segments.

Results are suggestive of the positive effects of captions on comprehension since the striking number of correct answers suggests that the presence of captions significantly increased the amount of comprehensible input to the foreign language learners. Captions facilitated learners' comprehension of the video they watched. However, when comparing the performance of the groups, it is clear that the presence of captions had a more significantly positive effect over the Russian group when they watched for the second time with captions. Yet, the number of correct responses of the ESL non-captioned group was still higher than the number of correct responses by the Russian group when watching for the second time.

Similarly, Neuman and Koskinen (1992) looked at incidental word learning by language minority bilingual children. Based on the insight that captioned television may provide comprehensible input – a key ingredient for second language acquisition (Krashen, 1982/2009) – and can be seen as a "multisensory, largely entertaining medium" (p. 96), the scholars investigated the learning of incidental science vocabulary and concepts. The participants – 129 bilingual seventh and eighth graders – were divided into four groups: (a) captioned television; (b) non-captioned television; (c) reading along and listening to text; (d) textbook only. In order to explore the effects of learning words in context, television segments from *3-2-1 Contact*, a Children's Television Workshop science production was selected given its motivational display of scientific concepts, its special appeal to the audience, and its appropriateness to seventh and eighth graders. *3-2-1 Contact*, a Children's Television Workshop

Participants were given pre-tests, a weekly 10-word recognition test to measure their ability to recognize words, a weekly written retelling of the week's lesson to measure the frequency of the target words in the their writing, a sentence anomaly test to measure their ability to understand target-words in context, and a final 90-item multiple choice test to measure knowledge of all target word meanings.

The captioned TV group outscored the reading text group for all three units on word recognition, though they were not statistically significant for Units 1 or 3. Regarding the results of the sentence anomaly tests, differences among all four groups indicated a similar trend, favoring captioning. As for unit tests, results revealed that the captioned TV group outperformed the reading text group, but when comparing the captioned TV group and the traditional TV group, results were only statistically significant for Unit 2. These results indicate that specific features inherent to the videos in each unit must have played a role in the students'

comprehension of the science topics. As they speculate, "the visual representation of words in video form is an important contributor to students' increased word knowledge" (Neuman & Koskinen, 1992, p. 102).

Interestingly, scores on the word meaning posttest revealed that the captioned TV group outperformed the other groups because bilingual students made significant gains in vocabulary knowledge without any formal instruction. Furthermore, analysis of participants' weekly written recall of science concepts favored the captioned TV group for the participants used target-words more often than other groups.

The participants in the captioned TV group consistently achieved higher mean scores than the other groups on all word knowledge tests, though these differences were not always statistically significant. As well as confirming the beneficial role of captioning for incidental vocabulary acquisition, this study raised word and picture-related aspects, which need to be taken into account not only by researchers when designing such experiments, but also by practitioners when selecting video materials for the classroom. Comprehensible input, facilitated in a bimodal input form (audio + image), and the helpfulness of the context seem to have played a key role in the participant's vocabulary acquisition.

Huang and Eskey (1999) analyzed the effects of captioned TV on the listening comprehension of 30 intermediate ESL learners with a view to contrasting with learners watching traditional TV. The study also aimed to check whether there would be vocabulary and phrase acquisition by the participants. The authors investigated whether certain factors, such as starting age of ESL instruction, length of time in the United States, length of ESL instruction, time in private language schools, time with tutors, and time traveling in English speaking countries, would correlate with listening comprehension.

The participants watched a seven-minute episode of Family Album U.S.A. (FAU) twice, a television series designed in 1992 for ESL classroom teaching (non-authentic). A 10-minute TOEFL-based format listening test, consisting of spoken statements and short conversations, was used to test participants' listening comprehension of sixteen multiple-choice test items, each of them with three possible answers.

The captioning group outscored the control group. Additionally, participants reported that captioning helped them understand the story better, aiding them regarding their vocabulary/phrase acquisition, improving their language listening skill, and that it was an enjoyable way to learn English. Nevertheless, factors regarding their personal learning experiences showed no correlation with the listening comprehension test.

Also, this study demonstrates that captioned materials can have a positive impact on students' listening comprehension, though it still is unclear at this point whether captioned material would have a similar impact on other populations, such as beginning learners.

One of the few studies to take into consideration the non-instructional setting in language learning and watching subtitled/captioned videos equation is that of Koolstra and Beentjes (1999), which examined foreign language vocabulary acquisition by 246 Dutch fourth and sixth graders[6] watching subtitled TV at home. Participants watched a 15-minute episode on grizzly bears of the series called The New Wilderness. Three different experimental conditions were employed: (a) watching an English television program with Dutch subtitles; (b) watching the same English program without subtitles; and (c) watching a Dutch television program. At the end of the experiment, children were also asked about their habits in terms of watching subtitled videos at home.

The measures comprised an English vocabulary matching test with two parts. In the first part, children listened to an audiotape on which a native (American) speaker spoke 45 English words and they had to select a picture depicting the word pronounced for 33 items. The items could include an action, an animal, an object or a body part. Participants had to choose the English word spoken from four pictures. The second part consisted of an exact same test, with 12 items only and four pictures that included an object, an animal or action, with no body part.

An English target vocabulary test was also used as a measure, which consisted of a multiple-choice test asking for the Dutch translation of 35 words depicted in the video. Moreover, the participants in the two conditions with English TV program also received a word recognition test, based on a 30-item auditory word recognition test, in which 20 of them were actually presented in the soundtrack and 10 that were not, but which could have been, given their content.

Vocabulary recognition scores were higher in the subtitled condition and scores in the no subtitles condition were higher than in the control group. The analysis also yielded a main effect of grade, with sixth graders outperforming fourth graders. Concerning word recognition, more English words were recognized after participants watched the subtitled television program in comparison to the no subtitles condition,

[6] In the Netherlands, in primary schools, about one hour per week is spent teaching English from Grade 5 on. This is particularly important to note since fourth graders were then unlikely to have had formal English instruction by the time they participated in the study.

and sixth graders outperformed fourth graders in this test as well. There was indeed a correlation between high frequency of watching subtitled programs at home often and higher English vocabulary scores.

As Koolstra and Beentjes (1999) claim, "the findings indicate that young children can acquire elements of a foreign language through watching subtitled television programs" (p.58). Thus, one of the greatest contributions is that word recognition in the no subtitles condition was not superior to the subtitled condition, suggesting that the belief that reading subtitles might distract viewers from hearing English words may be just a belief. Another contribution regards implicit learning[7] since Dutch fourth graders did show some knowledge of the English language though they had never been taught English before. These findings lend support to the benefits of watching subtitled television programs at home for language learning purposes, especially considering that more English vocabulary was acquired by those who watched subtitled TV at home more often.

In tune with previous research (Garza, 1991; Neuman & Koskinen, 1992), Markham (1999) confirmed to some extent what other studies had suggested: "the availability of captions significantly improved the ESL students' ability to recognize words" (p. 321). In order to explore captions as an aid to comprehension and vocabulary development, he examined their effects on 118 advanced university-level ESL students'[8] listening word recognition. The materials comprised a 13-minute episode of marine biology information on whales and a 12-minute episode of an interview-based discussion between two people on the history of the civil rights movement in the United States. While the first video presented a high audio/video correlation, the second video presented a low audio/video correlation. A 50-item listening multiple-choice test was administered after watching each video once. The participants listened to a sentence and had to choose the correct word from alternatives given. Demographic data on the participants were taken into account as well, such as the

---

[7] According to Ellis (2008), implicit learning is a result of "the acquisition of knowledge about the underlying structure of a complex stimulus environment by a process which takes place naturally, simply, and without conscious operations" (p. 2). On the other hand, explicit learning involves "a more conscious operation where the individual attends to particular aspects of the stimulus array and volunteers and tests hypotheses in a search for structure" (p. 3).

[8] According to Markham (1999), the students were from fifteen different countries, but 71% of them were Chinese, Japanese or Korean, therefore Asians. It is important to be mindful of such information since "generalizing the results of this research to other ESL populations should be done very cautiously" (p. 322).

number of years in the U. S., academic major, L1, gender, prior knowledge of the topic, and educational background.

Caption availability was found to improve the participants' listening ability to recognize words. In relation to the effects of captioning and passage content, the results were unexpected: The civil rights video, with a low audio/video correlation, was found to be less challenging regardless of caption availability. Demographic variables were not found to be significant.

Markham (1999) did not focus on students' global comprehension, but word recognition skills only, and the findings confirm the potential benefits that captioning can have to second language learning, and this study "presents specific positive evidence concerning the influence of second-language captions directly on second-language listening" (p. 326).

In order to shed light upon key issues related to the benefits and limitations of using intralingual subtitled videos for language learning purposes, Taylor (2005) carried out a study involving beginning language learners, a population that had been frequently disregarded in the literature. The author also attempted to unveil what processing strategies the students seemed to use while watching a verbatim captioned material.

After a pilot study, 85 university-level, beginning[9] second-semester Spanish learners participated in the actual experiment, in which they watched a non-authentic 10-minute video segment narrating the history and the consumption of various foods of Spain and Latin America. Test materials consisted of: (a) a free written story recall (in English); (b) a 15-item multiple-choice recall in English; and (c) a sheet of paper in which participants explained their strategies when using captions, video, and audio for comprehension purposes.

Overall, when analyzing the results on the free recall and multiple-choice recall, no significant differences were found comparing the groups. However, students with more time of Spanish study outperformed those with less time in the intralingual subtitles group. Taylor asserts that the difference "was not surprising, assuming that the students with more years of study had more reading and listening comprehension practice" (p.425). As for the control group, no difference was found between more 'real' beginners and 'false' beginners. Nonetheless, a surprising finding is that when comparing real beginners in the intralingual subtitles group and the control group, scores were found to be higher for the latter. "It

[9] Taylor (2005) clarifies that despite the fact that students were beginners, their length of previous Spanish study ranged from 8 months to 5 years.

seems that captions did not aid first-year students in comprehension, and in fact, seem to have been detrimental to their understanding of the video" (p. 425), which does not seem to be the case with false beginners because captioning had a neutral effect on them.

As far as processing strategies are concerned, while 35% of the first-year captioning group participants reported that captions were distracting or confusing and that they struggled when trying to devote attention to the three channels (audio, video, and subtitles) simultaneously, only 11% of third-year students in the same group reported similar difficulties. Also, 23% of 'real' beginners reported being able to devote attention to the three channels as opposed to 50% of 'false' beginners of the same group. As Taylor points out, despite their difficulties, students tended to express a positive attitude toward the use of captions, although some of them reported having ignored the audio and focused on the captions mostly.

Yuksel and Tanriverdi (2009) looked at vocabulary gains by comparing the performance of two groups – intralingual subtitles and control – in a pre-test/post-test research design. A vocabulary knowledge scale adapted from Wesche and Paribakht (1996) was used before and after the treatment. The participants were 120 Turkish intermediate university-level students enrolled at Kocaeli University taking intensive English classes to fulfill their program language requirements. They were majoring in different areas, such as Finance, Business Administration, Engineering, Journalism, and Education. The participants had taken an in-house English proficiency exam before enrolling in the classes and thus no other proficiency exams were administered.

The participants were shown the first 9min14sec of an episode from the TV Series Seinfeld twice, a series that has been popular in Turkey, according to the authors. The target-words from the video to assess vocabulary gains, which all belonged to the same word category (noun), were chosen on the basis of the students' proficiency level. After informal piloting, ten of them were kept for the actual study. Wesche and Paribakht's version of the vocabulary knowledge scale contained the following questions to be administered right after the treatment:

- I don't remember having seen this word before.
- I have seen this word before, but I don't know what it means.
- I have seen this word before, and I think it means (synonym or translation).
- I know this word. It means … (synonym or translation)

▪ I can use this word in a sentence: (Write a sentence)
(If you do this section, please also do Section IV)

The results that emerged pointed out to gains in vocabulary in both intralingual subtitles group and control group, when comparing their performance in the pre-test and post-test. However, the groups did not differ statistically, which means that intralingual subtitles availability did not cause participants to outperform the control group. Nevertheless, the participants reported having seen most of the target-words before but were not able to recall them or attribute some kind of meaning to them. After having encountered them in context, that is, after having seen and listened to them twice in the series, they were then able to remember and in some cases even demonstrate they had learned some of target-words.

Another important aspect to observe is that self-report scales to measure word learning, such as the one adopted in Yuksel and Tanriverdi's (2009) study, have been criticized by some scholars due to the fact that they might not be reliable enough to address novel word learning or that learning learners' developing L2 vocabulary knowledge cannot be meaningfully represented by a single linear scale (Read, 2000).

Winke, Gass and Sydorenko (2010) investigated the effects of intralingual subtitles a special focus on different languages – learners of Arabic, Chinese, Russian, and Spanish. In order to dive deeper into underinvestigated aspects, the authors looked at: (i) whether intralingual subtitles would elicit more comprehension by Spanish learners; (ii) effects of captioning order, that is, whether intralingual subtitles would be more effective the first or the second time when a video was watched twice (for all languages); (iii) whether proficiency differences affect the benefits of intralingual subtitles derived from captioning order.

A total of 150 foreign language university learners participated in the study, who were second or fourth year learners of Spanish and Russian, and second year learners of Arabic and Chinese. Participants watched documentaries about three animals – salmon, dolphins, and bears – and then were given vocabulary tests which had the same target-words despite the different target languages after being pre-tested. Half of the vocabulary words were presented orally first, while the other half were presented in written form and vice-versa. Participants were asked to translate the target words into English, which had the same target words despite the different target languages. Also, participants answered multiple-choice questions in English about the main points of the story to check for their overall comprehension.

Participants who saw the videos with intralingual subtitles both times outperformed those who saw the videos without captions both times on the vocabulary test with written input. Similar results were found on the vocabulary test with aural input. Intralingual subtitles then resulted in significantly higher comprehension test scores than the control group, in tune with previous research (Garza, 1991; Markham & Peter, 2003). As far as the results of the effect of order are concerned, the participants who saw captions in the first viewing performed significantly higher on the aural vocabulary test than those who saw captions on the second viewing, but the effect of order was not significant for the written vocabulary or the comprehension test. Yet, as for the effect of order, there seems to be a trend differentiating Russian and Spanish learners: captions first seemed to be more beneficial, as opposed to Arabic and Chinese where captions second appeared to be more helpful.

One aspect to be highlighted concerns the emerging themes from the interviews with the participants, in which they claim that they need multiple input modalities and that captions reinforced what they heard, affecting their attention to the input. However, some of them found captions to be "attention depleting" (Winke, Gass & Sydorenko, 2010, p. 78). This is relevant information for it enhances the question as to the extent to which learners are capable of devoting their attentional resources to the bimodal input to which they are exposed to while engaged in video watching.

2.2 Intralingual Subtitles, Interlingual Subtitles, and L2 Development

Comparative studies as regards the effects of the use of intralingual subtitles and interlingual subtitles upon L2 learning also emerge in the late 1990's and beginning of 2000s. Until then, studies carried out to investigate the benefits and drawbacks of subtitling on language learning had typically addressed intralingual subtitles (captions, as to what they were often times referred), only. The scenario starts to change as researchers began to devote academic attention to comparative research on the effects of intralingual and interlingual subtitles related to various L2 learning domains. In this subsection, I review some of these insightful investigations.

Thought-provoking results derived from the study carried out by d'Ydewalle and van de Poel (1999) on incidental language acquisition by children with subtitled material. The authors claim that for acquisition to happen, the channel in which the foreign language is presented –

soundtrack or subtitles– must be processed. They also claim that reading the subtitles is an automatic process, regardless of one's familiarity with them, knowledge of the foreign language in the soundtrack or its availability, which had been previously indicated in studies involving eye-movement recordings (d'Ydewalle & Gielen, 1992).

In order to investigate the acquisition of French and Danish as foreign languages – closer languages to the participants' first language, which was Dutch – by 327 third, fourth, fifth, and six graders of a primary school in Belgium, this study included tests on vocabulary, morphology, and syntax. A 10-min long movie named Young Deer was presented in four different experimental conditions: Dutch subtitles and French soundtrack, French subtitles and Dutch soundtrack, Dutch subtitles and Danish soundtrack, and Danish subtitles and Dutch soundtrack. The control condition received a fifth version with Dutch in both soundtrack and captions.

Participants took three different tests: One to assess vocabulary acquisition, containing 20 content words whose correct translations were to be chosen out of three alternatives. All target words were somewhat salient in the film (they all appeared four times at least), an important aspect to observe. As for the syntax test, participants were asked to choose the correct construction of 10 elementary sentences from three alternatives for each that differed in word order. In relation to the tests, because not all language components were equally salient in the auditory and the visual mode, the tests contained different visually versus auditory salient test items.

Regarding vocabulary, presenting the foreign language in the soundtrack improved acquisition in the visual and auditory tests; when the foreign language was presented in the subtitles, performance was only improved in the visual test. As for the syntax test, there was no evidence for foreign language acquisition. As regards the morphology test, performance in the auditory test was found to be better when the foreign language was available in the soundtrack.

Based on the findings, some issues emerge. As the authors point out, "similarity with the native language may cause interference in acquiring a foreign language; however, the similarity could also facilitate the acquisition process" (d'Ydewalle & van de Poel, 1999, p. 240). With few exceptions, this was evident when the results for the Danish and French tests are scrutinized and considering that Danish is closer to Dutch. Results also suggest that when there are significant differences in the experimental conditions, it was always to the advantage of the condition with the foreign language in the soundtrack, possibly linked to

"efficiency of information processing" (p. 242). This lends support to the need for the inclusion of specific features of individual differences in studies that focus on the effects of subtitled and captioned materials on language learning.

Markham, Peter and McCarthy (2001) sought to examine the effects of intralingual and interlingual subtitles on 169 intermediate university-level Spanish as a Foreign Language students, with a view to verifying their listening and reading comprehension. The high audio/video correlation[10] listening passage was a DVD episode of about seven minutes with information about preparation for the Apollo 13 NASA mission. The participants were distributed in the three treatment conditions: interlingual subtitles, intralingual subtitles, and control group. Participants were given 10 minutes to write a summary of the passage in English to check their understanding and content recall. The written accounts were assessed on the basis of unit ideas, elaborations, and distortions. Also, participants were given a 10-item, multiple-choice test in English, which reflected the content and the level of difficulty of the video they watched.

As regards the written accounts, results revealed that the absence of subtitles prevented the participants from recalling much of the information and suggested a lower level of comprehension in relation to other conditions; regarding the multiple-choice test, results demonstrated that participants in the interlingual subtitles groups outperformed those in the intralingual ones, who in turn outperformed the ones in the no subtitles group.

The scholars posit that "university-level second language students typically have better reading comprehension than listening comprehension skills, and that they would most likely benefit from viewing very difficult target language video material with native-language captions first" (Markham, Peter & McCarthy, 2001, pp. 443-444). After that, intralingual subtitles could be used and finally no subtitles whatsoever, which would then probably allow for a greater level of comprehension.

In a very similar study, Markham and Peter (2003) investigated the effects of using intralingual subtitles and interlingual subtitles on 213 intermediate university-level Spanish as a Foreign Language students, with a view to examining their listening and reading comprehension. The

[10] Audio/video correlation, in studies dealing with subtitling and language learning, refers to whether the material presented on screen is supported by the content of the soundtrack presented to its viewers.

same video passage used in Markham, Peter and McCarthy (2001) was selected for this study. When asked about their knowledge of the Apollo 13 mission, 48 participants reported having little or no knowledge of the topic. These were the participants who were then distributed in the three treatment conditions: interlingual subtitles, captions, and no subtitles/captions. Participants answered a 20-item Spanish multiple-choice comprehension test on the passage to verify their content comprehension.

The results revealed that interlingual subtitles were more beneficial for the students' comprehension than intralingual subtitles or control group (an exact same finding had been obtained by Markham, Peter and McCarthy, 2001). The authors argue that this particular finding might indicate that "the English language reading input is obviously a powerful contributor to general comprehension" (Markham & Peter, 2003, p. 339). The authors speculate that the participants' Spanish language reading comprehension level may have also contributed to the general comprehension these students presented. More importantly, this specific element could be taken to mean "a factor in enhancing the participants' listening comprehension" (p. 339).

One of the main contributions from both studies concerns the speculation that literate adult or secondary-level foreign language students would probably demonstrate enhanced listening comprehension if exposed to interlingual (native language) subtitles first, and then, as their foreign language literacy skills develop, they would be able to comprehend more difficult materials with intralingual subtitles or captions, and finally they could view and understand more challenging video materials with no subtitles or captions. This argument seems reasonable considering one's language learning path.

Stewart and Pertusa (2004) attempted to explore the beneficial aspects to interlingual and intralingual subtitles for listening comprehension. The experiment involved 95 university-level Spanish as a Foreign Language students who watched two full-length movies with either Spanish captions or English subtitles. Participants were given an intermediate-level multiple-choice vocabulary quiz, containing items predicted to be unfamiliar for their proficiency level, to check their recognition level. A total of seven intermediate-level conversation classes participated in the research during two semesters.

Results of the first semester showed more positive effects for the use of Spanish captions than English subtitles regarding vocabulary recognition. Results of the second semester demonstrated similar trends, except for one instance in which English subtitles group outscored the

Spanish captions group. While positive effects of using interlingual subtitles were found in Markham, Peter and McCarthy (2001) and Markham and Peter (2003), more positive effects for the use of intralingual subtitles were found in Stewart and Pertusa (2004).

Regardless of the inconclusive results in terms of the vocabulary recognition test, after seeing both films with either intralingual subtitles or interlingual subtitles, participants reported in an anonymous questionnaire that they felt they had learned more because they were able to listen to the word and see it written on screen in the L2. They also reported that they would prefer to watch films with captions in the future, and relying on the Spanish captions did not seem to make them nervous, according to the questionnaires. In this sense, students' responses to the use of intralingual subtitles may signal beliefs that students hold, which may potentially lead to habit formation. Interestingly, most of the students who watched the films with English subtitles reported that they were able to understand the word when listening to them, but they felt they needed to see them written (in the L2) to establish a better connection. Most of these students reported to prefer to watch movies with Spanish captions in the future.

In short, Stewart and Pertusa (2004) advocate for the use of intralingual subtitles, grounded in the students' beliefs and the partially positive effects observed in the results obtained from the study. Their main argument is that "the use of English subtitles does not encourage [Spanish] learners to use their previously acquired listening skills" (p. 438), taking into account this specific population and their level of proficiency. This is what may have contributed to trigger the debate in these studies, which raises the need for reflection on the choice of the type of subtitles to be used in the instructional setting and the impact and shortcomings that both options may bring about in the students' development.

A comprehensive study conducted by Bianchi and Ciabattoni (2008) addressed short and long-term effects of intralingual and interlingual subtitles in terms of vocabulary acquisition and pragmatic use of language in which a computerized setting was adopted. The authors justify the use of a computerized setting for this kind of investigation in order "to simulate a real home-video scenario where a student watches a film on DVD and takes advantage of the text aids provided (captions or subtitles)" (p.75).

The participants, 85 Italian ESL learners in the 18-45 age range, were classified into groups based on their proficiency level and the treatment condition as follows: beginners – with intralingual subtitles,

with interlingual subtitles, and control group; intermediate – with intralingual subtitles, with interlingual subtitles, and control group; and finally advanced – with intralingual subtitles, with interlingual subtitles, and control group. After being pre-tested on aspects of grammatical, vocabulary, and pragmatic use of lexicogrammatical phrases a week before, the participants watched a series of clips from two films in English (*Fantasia* and *Harry Potter and the Philosopher's Stone*) in their correspondent treatment condition and, at the end of each clip, answered multiple-choice content questions. The participants could watch the clip twice to review their answers. Finally, a post-test was administered a week later to examine the long-term effects of intralingual and interlingual subtitles.

At beginner's level, results revealed that participants in the interlingual subtitles group outperformed those in the other treatment conditions in both films' content comprehension, despite a more marked difference in score in the case of *Fantasia*. Regarding intermediate level participants, the interlingual subtitles group performed better than the other two groups, although the difference between the intralingual subtitles group and control showed different patterns depending on the film. As far as advanced participants' results are concerned, the interlingual subtitles group performed slightly better than the intralingual subtitles group and significantly better than the control group.

In short, the interlingual subtitles group outperformed the other groups in content comprehension, regardless of the proficiency level and type of film. Bianchi and Ciabattoni (2008) argue that "this result is expected given that subtitling is processed automatically and content comprehension can logically be facilitated by text in the mother tongue" (p. 78).

In terms of vocabulary, best results were obtained by the control group at beginners' level, followed by the interlingual subtitles group and finally the intralingual subtitles group. In relation to intermediate participants, all groups obtained good results for *Harry Potter*, in which the intralingual subtitles group outperformed the others, but not for *Fantasia* because the best results were achieved by the interlingual subtitles group. Finally, with reference to advanced participants, no patterns were found, since all groups' results were almost identical for both films, although the scores were higher for the vocabulary related to *Harry Potter*.

The three proficiency levels did not exhibit a pattern in terms of the three treatment conditions. Nonetheless, it is interesting to note that the results for *Harry Potter* were higher for all groups in all treatment

conditions, in which "semantic match among the different communication channels was higher" (Bianchi & Ciabattoni, 2008, p. 79). Moreover, captions were found to be less effective for lower proficiency levels than subtitles.

Considering language in use, beginners' scores were similar to beginners' in the vocabulary test, that is, the control group outperformed the experimental groups. Also, results were better for experimental conditions in the case of *Harry Potter*. As for intermediate participants, scores were better for the control group as well, followed by the captions group and then the subtitles group (regarding *Harry Potter*); however, in relation to *Fantasia*, the control group outscored the experimental groups, but subtitles were more effective than captions in this case. Finally, in this type of task, advanced participants obtained higher scores with intralingual subtitles. Thus, when the participants' results are put together and film type is disregarded, an interesting picture emerges: "a gradual passage from text aids in general and captions in particular limiting comprehension in lower proficiency groups to the complete opposite with advanced students" (Bianchi & Ciabattoni, 2008, p. 82).

In terms of long-term acquisition, vocabulary and language in use aspects were taken into account by comparing the participants' pre-test and post-test scores. First, in terms of vocabulary gains, subtitles were found to be more effective than captions or no subtitles/captions when proficiency and film type are disregarded. When these two components are considered, different patterns emerge: beginners seem to have profited more from captions regarding *Harry Potter* and no subtitles/captions regarding *Fantasia*; and intermediate and advanced participants seem to have profited more from subtitles for both films. Moreover, as for long-term acquisition of pragmatic use of language, beginners did not tend to acquire when subtitles or captions were made available. As for intermediate students, subtitles had a better effect on this type of acquisition, followed by captions and the control group. The same pattern was found for advanced learners.

Bianchi and Ciabattoni (2008) shed light on several interesting issues, especially because different populations were compared across different tasks, taking into account short and long-term effects of interlingual and intralingual subtitles. Though puzzling results were obtained, they "may be connected to the intrinsic differences between such activities in terms of nature and cognitive effort" (p. 87), which deserves further scrutiny.

Another study to corroborate the positive effects of intralingual subtitles over interlingual subtitles is that of Hayati and Mohmedi's

(2011). In their study, 90 intermediate Iranian EFL learners, with an average age of 22, majoring in an English Teaching program were divided into three groups: intralingual subtitles (English), interlingual subtitles (Persian) and control group (no subtitles). The participants' mother tongue was Persian.

The participants took a proficiency test, watched part of a documentary film on natural disasters developed by BBC called 'Wild Weather' and finally took a comprehension test. The test contained six sets of multiple-choice tests, with 10 items each, derived from the video to test their level of comprehension. Each question contained language that was present in the episode and a final comprehension test was also administered. There were statistically significant differences between all of the groups and the results indicated that the intralingual subtitles group outperformed the interlingual subtitles group, who in turn outperformed the control group.

Conflicting results in relation to what has been mostly reported in the literature were obtained in Latifi, Mobalegh and Mohammadi (2011), whose goal was to investigate the effects of the use of intralingual and interlingual subtitles on 36 Iranian EFL learners' listening comprehension. All of the participants spoke Persian as an L1, and were divided into three groups: intralingual subtitles, interlingual subtitles, and control.

All participants took two different listening sections of IELTS as pre-test and post-test to assess the effects of subtitling on general listening comprehension. Moreover, 15 multiple choice comprehension tests were administered, and each of the them was related to one of the 2-minute parts of the movie 'Alvin and Chipmunks.' A total of 15 comprehension tests were designed and administered at the end of each session to test for immediate effects of subtitling. The tests had 10 questions on the key points discussed in the dialogues of that section of the film.

Interestingly, the results for the multiple choice comprehension tests indicated that interlingual subtitles group outperformed the intralingual subtitles group, who in turn outperformed the control group. The differences were statistically significant between the intralingual subtitles and control groups, and between the interlingual subtitles and control groups. There were no statistically significant differences between the two experimental groups.

As far as general improvement in listening comprehension is concerned, a comparison of the groups' performance in IELTS Test revealed that the control group outperformed the experimental groups. Statistically significance was found between the control group and the

interlingual subtitles group; no statistically significance differences were found between the experimental groups.

The results in Latifi, Mobalegh and Mohammadi raise an interesting question as to the long-term effects of subtitling exposure. The authors claim that studies exploring the use of subtitled materials do not typically assess participants' listening comprehension on materials that are not subtitled per se. In other words, their contention it might be related to the fact that studies involving interlingual and intralingual subtitles do not evaluate listening comprehension alone because reading while listening is a condition that is almost imposed on the participants in these studies. The only time listening comprehension is being 'purely' measures is when no text aid is provided, that is, is most control conditions. This reasoning is line with what the authors posit in that the results obtained in studies that favor the use of subtitles over control groups usually assess vocabulary development mostly rather than the development of listening comprehension per se.

Finally, Latifi, Mobalegh and Mohammadi argue that the better performance in terms of listening comprehension achieved by the control group "can also be attributed to the problem that the opponents of subtitles mention – the distracting effects of subtitles" (p. 27). On this issue, Zanon (2006) points out that there may be two handicaps related to the constant use of subtitles. One refers to too much concentration on reading, causing dialogues to be ignored or forgotten; the other problem relates to the difficulty to break the habit of reading once learners are somehow used to doing so when watching subtitled materials.

Zarei and Rashvand (2011) designed an experiment to research the impact of verbatim – including everything that is spoken, such as fillers, hesitations, pauses – and non-verbatim subtitles – including only relevant information for video comprehension. A total of 120 Iranian intermediate university-level EFL learners participated in the study, who were divided into four groups: Interlingual verbatim subtitles, interlingual non-verbatim subtitles, intralingual verbatim subtitles, and intralingual non-verbatim subtitles. The research carried by Zarei and Rashvand was aimed at verifying possible differences between the translational aid concerning L2 vocabulary comprehension and production.

Participants watched a summarized version of the film 'She's the man'. Prior to watching, they were given a vocabulary test containing 100 words to assess their prior knowledge, which required them to write a translation (in Persian) or synonym for the words. Additionally, two 3-item vocabulary post-tests were also administered to the participants after

they had watched the movie to measure L2 vocabulary comprehension (in multiple-choice format) and L2 vocabulary production (in cloze format).

Results obtained suggest more positive effects for intralingual subtitles than interlingual subtitles in terms of L2 vocabulary comprehension. Additionally, in both the interlingual and intralingual subtitles groups, the participants of the group receiving non-verbatim subtitles outperformed those with verbatim subtitles. Tests run revealed that there were no statistically significant differences between the intralingual and interlingual subtitles groups in terms of vocabulary comprehension, but non-verbatim subtitles did interact with more vocabulary comprehension.

Results obtained also suggest the participants in the intralingual subtitles groups outperformed those in the interlingual subtitles groups regarding L2 vocabulary production. Moreover, in either of the interlingual and intralingual subtitles groups, the participants of the group with non-verbatim subtitles outperformed those with verbatim subtitles. Post-hoc tests also revealed that intralingual subtitles were more conducive to L2 vocabulary production regardless of whether they were verbatim or non-verbatim.

Matielo, Collet, and D'Ely (2013) report on a study with 27 Brazilian EFL learner enrolled in a non-credit extracurricular course at *Universidade Federal de Santa Catarina* (UFSC). The study was aimed at investigating the effects of interlingual and intralingual subtitled video on vocabulary recognition. The intermediate-level participants were divided into three groups: interlingual subtitles, intralingual subtitles, and control.

Participants answered a questionnaire in order to provide the researchers with more information on their learning profile and education background. Moreover, participants watched a 20-minute episode of the TV series *The Big Bang Theory,* selected on the basis of appropriateness. Participants took a vocabulary pre-test, one session prior to video watching, took a pre-test, containing 20 target-words and 25 distractors. After the video watching session, participants took an immediate vocabulary recognition test[11], with the target-words only, which required them to either translate the target-words into Portuguese or explain them in Portuguese or English. In the next session of the data collection,

[11] In addition to a vocabulary recognition test, participants also answered general and specific comprehension questions about the video they watched. The scores on this portion of the test were not included in the article, though.

participants were administered a delayed post-test to check whether possible gains in vocabulary recognition had remained over time.

Results showed more immediate gains in the intralingual subtitles group than interlingual subtitles group or control group. Nonetheless, the results obtained were not statistically significant, possibly due to the small sample investigated in the study. When analyzing the scores on the delayed post-test, the analyses reveal gains across all of the three groups. According to the authors, this possibly serves as an indication that the participants may have looked up the words after the experimental sessions.

When looking at gain scores more closely, the results that emerge point to more gains in the intralingual subtitles groups than the other groups, in spite of the lack of statistically significant results. The authors stress that a possible trend for better results in the intralingual subtitles groups might be linked to the participants' language instructors' frequent use of subtitled videos in the classes for which they (the language teachers) affirm using intralingual subtitles most if not all the time. Results obtained with the participants' profile questionnaire also confirm that most of them are habitual intralingual subtitles users.

Raine (2013) investigated the effectiveness of a different combination that had not been addressed before. In his study, in addition to examining the effects of intralingual subtitles and interlingual subtitles, he also included dual subtitles as an experimental condition. A total of 39 female Japanese university-level students who were pre-intermediate English speakers participated in this study. The students were enrolled in different non-English majors and were taking English lessons focused on the four main language skills.

The participants were divided into four groups: intralingual subtitles, interlingual subtitles, dual subtitles, and control. A DVD of Jessi Arrington's *Wearing Nothing New* TED talk (Arrington, 2011) was used. Participants were also administered a modified version of the vocabulary knowledge scale (Wesche & Paribakht, 1996). Raine (2013) explains that the six target-words were selected by the availability of a direct L1 translation in the Japanese version of the transcript and the unlikely familiarity by the participants. The target-words belonged to different word groups (two nouns, three adjectives, and one verb).

In general, results obtained revealed that the dual subtitles group, with both English and Japanese subtitles simultaneously show on screen, outperformed the other groups. The interlingual subtitles group outperformed the intralingual subtitles group, who in turn outperformed the control group (no subtitles). However, whether the differences

between the groups are statistically significant is impossible to determine since no post-hoc tests were run.

2.3 Summary of the Findings of the Literature on Subtitling and L2 Development

The last two sections of this Chapter sought to provide an overview of some of the relevant literature on the use of subtitled materials on domains of language development. As these studies point out, the picture that emerges from the results does not seem to be a clear one, since sometimes more optimal results are achieved with interlingual subtitles, whereas other times with intralingual subtitles. Nonetheless, these results suggest that both interlingual and intralingual subtitles can be beneficial for L2 learners.

In light of the 18 empirical studies reviewed in these two sections so far, an interesting, though puzzling picture conjures up[12], which is represented in Figures 1 and 2:

Author(s) and Year	Main Focus of the Study	Participants	Type(s) of Subtitles	Main Results
Garza (1991)	Content Comprehension	Advanced Russian and ESL Learners (University-level)	Intralingual Subtitles and Control	Intralingual subtitles over control
Neuman & Koskinen (1992)	Vocabulary recognition, knowledge and use	Language Minority Bilingual Learners	Intralingual Subtitles and Control	Intralingual subtitles, mostly
d'Ydewalle & van de Poel (1999)	Vocabulary, morphology, and syntax	3rd, 4th, 5th, and 6th Dutch graders	Interlingual Subtitles, Reversed Subtitles, and Control	Interlingual subtitles over control for vocabulary; Not for syntax
Huang & Eskey (1999)	Listening Comprehension	Intermediate ESL Learners (University-level)	Intralingual Subtitles and Control	Intralingual subtitles over control
Koolstra & Beentjes (1999)	Vocabulary-picture recognition	Beginning Dutch fourth and sixth graders learners of English	Intralingual Subtitles and Control	Intralingual subtitles over control
Markham (1999)	Vocabulary recognition	Advanced University-Level ESL Learners (University-level)	Intralingual Subtitles and Control	Intralingual subtitles over control
Markham, Peter & McCarthy (2001)	Listening and reading comprehension	Intermediate Spanish as a Foreign Language learners (University-level)	Intralingual Subtitles, Interlingual Subtitles, and Control Group	Intralingual and interlingual subtitles in written accounts over Control; Interlingual over intralingual and control in multiple-choice test
uMarkham & Peter (2003)	Listening and reading comprehension	Intermediate Spanish as a Foreign Language learners (University-level)	Intralingual Subtitles, Interlingual Subtitles, and Control Group	Interlingual over intralingual subtitles and control in content comprehension
Stewart & Pertusa (2004)	Listening and reading comprehension	Intermediate Spanish as a Foreign Language learners (University-level)	Intralingual Subtitles, Interlingual Subtitles, and Control Group	Intralingual over interlingual subtitles and control in content comprehension
Taylor (2005)	Listening and reading comprehension	Beginning Spanish as a Foreign Language Learners (University-level)	Intralingual Subtitles and Control Group	No differences were found; real beginners: control group outperformed intralingual subtitles'

Figure 1: Summary of the studies on subtitling and language learning – reviewed in this Chapter – part I, adapted from "Intralingual Subtitles and Interlingual Subtitles: Exploring their Effects on Brazilian EFL Learners"

[12] I acknowledge that a substantial part of this Chapter has already been published in a state-of-the-art paper (see Matielo, D'Ely & Baretta, 2015).

by R. Matielo, 2016, Unpublished Doctoral Dissertation, Universidade Federal de Santa Catarina.

Bianchi & Ciabattoni (2008)	Listening and reading comprehension; Grammatical knowledge; Vocabulary; Lexicogrammatical phrases	Beginning, Intermediate, and Advanced Italian ESL learners (University-level)	Intralingual Subtitles, Interlingual Subtitles, and Control Group	Content comprehension: Beginners – Interlingual in both films Intermediate – Interlingual in one of the films and intralingual in one of the films Advanced – interlingual over intralingual, and intralingual over control in both films Vocabulary: Beginners – control over other groups Intermediate – intralingual over other groups in one film; interlingual over other groups in the other film Advanced – no differences Lexicogrammatical phrases: Beginners - control over other groups Intermediate: control over other groups in one film, but interlingual subtitles over intralingual subtitles in the other film Advanced – intralingual over other groups.
Yuksel & Tanriverdi (2009)	Vocabulary recognition, knowledge and use	Intermediate Turkish EFL Learners (University-level)	Intralingual Subtitles and Control Group	No differences were found.
Winke, Gass & Sydorenko (2010)	Vocabulary recognition	Beginning and Intermediate Learners of Arabic, Chinese, Russian, and Spanish (University-level)	Intralingual Subtitles and Control Group	Intralingual over control for all languages, but Arabic and Chinese learners benefitted from intralingual subtitles more in the second viewing.
Hayati & Mohmedi (2011)	Content Comprehension	Intermediate Iranian EFL Learners (University-level)	Intralingual Subtitles, Interlingual Subtitles, and Control Group	Intralingual over interlingual subtitles and interlingual over control.
Latifi, Mobalegh and Mohammadi (2011)	Content comprehension and listening comprehension	Intermediate Iranian EFL Learners (University-level)	Intralingual Subtitles, Interlingual Subtitles, and Control Group	Content comprehension: interlingual over intralingual, and intralingual over control; Listening comprehension: control over intralingual and interlingual.
Zarei & Rashvand (2011)	Vocabulary comprehension and production	Intermediate Iranian EFL Learners (University-level)	Intralingual Subtitles (verbatim and non-verbatim) and Interlingual Subtitles (verbatim and non-verbatim).	Vocabulary comprehension: intralingual over interlingual, non-verbatim; Vocabulary production: intralingual over interlingual, non-verbatim.
Matielo, Collet & D'Ely (2013)	Vocabulary recognition	Intermediate Brazilian EFL Learners	Intralingual Subtitles, Interlingual Subtitles, and Control Group	No differences were found.
Raine (2013)	Vocabulary recognition	Pre-Intermediate Japanese EFL Learners	Intralingual Subtitles, Interlingual Subtitles, Dual Subtitles, and Control	Dual Subtitles over other groups; Interlingual Subtitles over Intralingual Subtitles and Control.

Figure 2: Summary of the studies on subtitling and language learning – reviewed in this Chapter – part II, adapted from "Intralingual Subtitles and Interlingual Subtitles: Exploring their Effects on Brazilian EFL Learners" by R. Matielo, 2016, Unpublished Doctoral Dissertation, Universidade Federal de Santa Catarina.

In short, it is possible to gather that subtitles have proven effective, regardless of their type of translational aid. With early studies, intralingual subtitles were mostly found to be more useful in aiding comprehension or language development in all of the studies that were reviewed in this Chapter, except for two recent ones (Taylor, 2005; Yuksel & Tanriverdi, 2009), which yielded no statistically significant differences between intralingual subtitles and control groups.

As for comparative studies that emerged from the 2000s on, the situation is not quite clear and presents greyer areas. First of all, in two studies (Stewart & Pertusa, 2004; Hayati & Mohmedi, 2011), better results were achieved with intralingual subtitles, mostly, whereas in one of the studies (Markham & Peter, 2003), better results were achieved with interlingual subtitles. Some studies presented better results with one or

the other depending on the language component being tested or proficiency group (Markham, Peter & McCarthy, 2001; Bianchi & Ciabattoni, 2008; Latifi, Mobalegh and Mohammadi, 2011; Zarei & Rashvand, 2011). As far as video comprehension and vocabulary development are concerned, there does not seem to be an agreement as to which type of translational aid can be more beneficial.

When singling out studies involving university-level participants from studies involving other populations, the number adds up to a total of 13 studies. When analyzing the results obtained by these studies with a focus on those that looked at a direct comparison of intralingual subtitles and their absence, out of the six studies, the scenario we get is that four of them favored the presence of intralingual subtitles, whereas two of them found no differences between experimental and control groups.

When analyzing the results obtained by the studies involving university-level participants, with a focus on those that examined the effects of intralingual subtitles, interlingual subtitles and no subtitles comparatively, we picture we get is that out of the seven studies in this category indicates that intralingual subtitles were found to be more effective in language development than interlingual subtitles and/or control. A few discrepancies were found, but they have already been properly discussed in this Chapter. The discrepancies mostly relate to the type of test being used and the language component being tested. It is, nevertheless, important to note that most studies reviewed in this Chapter involve intermediate-level participant, with varying L1 and L2 backgrounds.

In light of the inconclusive results reported in this Chapter, there is still a need for further scrutiny regarding the effects of intralingual and interlingual subtitles for language development, especially when one considers the fact that certain populations of EFL learners have been underinvestigated, Brazilian EFL learners being the case in point.

2.4 Other Studies on Subtitling and Cognitive Aspects

In this section of the Chapter, the selected pieces of research hereby reviewed offer important insights derived from investigations on the use of subtitles for language learning development and cognitive related aspects. Some of them are very much influence by recent technological advancements, which have enabled researchers to examine underexplored features related to the effects of subtitles, with the use of eye-tracking different methodologies, for instance.

Bird and Williams (2002) investigated an issue that for long had prevailed among the speculations of the effects of subtitled materials on language learning, that is, whether soundtrack gets processed when subtitled and captioned videos are watched. The study addressed the effects of single modality – either sound or text – and bimodal input – sound and text – presentation on word learning, with an explicit focus on word learning. Measures involved enhancements in spoken word recognition efficiency (implicit memory) and recognition memory related to word retention (explicit memory).

It is important to highlight that this study was in part motivated by complex and delicate issues. One of them, which is likely to be one of the most important issues, concerns the results of several studies involving the use of subtitling to aid comprehension and/or learning of some language-related aspect. Bird and Williams (2002) argue that the results favoring the use of (intralingual) subtitles, especially in terms of global and specific comprehension, are far from being surprising because the text (subtitles) presents "the easiest path to comprehension, and the auditory input might be ignored without loss of information required to successfully completing a written test" (p. 2). This implies, they argue, that it is still unclear whether subtitles improve listening comprehension or in fact hinder it.

Two experiments were carried out by the authors. Experiment 1 involved native speakers of English who ranged in age between 21 and 36 years old and advanced nonnative speakers of English, who ranged in age between 18 and 24. The first experiment comprised three phases: Phase 1 – reaction times to familiar and unfamiliar words, a measure used to decode speed; Phase 2 – reaction times for 'old' previously presented and 'new' items, which represents a measure of implicit memory; and Phase 3 – recognition memory for items presented in Phase 1, thus representing a measure of explicit memory. The study also examined repetition priming for unfamiliar words in order to verify the effects of bimodal presentation on implicit learning of novel word forms.

Results showed that auditory lexical decisions on familiar words were equally primed by prior bimodal and sound-only presentation. No priming effects for nonwords were found. Also, the results suggested that the participants were able to attend to and process both text and sound. Nonetheless, bimodal input failed to show any significant advantage in relation to sound only. This is an extremely important finding in that it lends support to the notion that the addition of text was not conducive to better gains in learning in relation to single-modality input.

Experiment 2 involved 24 advanced learners of English, whose ages ranged from 19 to 26 years. In this experiment, the scholars employed a rhyme judgment task on 56 nonwords in which participants have to quickly decide whether the second word presented in a pair rhymed with the first one. Results revealed that participants showed more implicit memory facilitation for nonwords in the bimodal condition, that is, when nonwords were present in both audio and text, than in sound-only condition.

As far as implicit memory is concerned, both experiments yielded different results. While repetition priming was not enhanced when text was added to auditory input in the first experiment, bimodal input enhanced implicit learning of nonwords in the second experiment, when participants had to judge whether nonwords rhymed. As Bird and Williams (2002) contend, "it appears that bimodal presentation beneficially affected implicit memory only when new phonological forms needed to be encoded" (p. 17). The authors argue that bimodal input might be beneficial for implicit memory, but it might be limited to cases where the input and its phonological forms cannot be established based on sound only.

In relation to explicit memory, the results demonstrated that bimodal input played a beneficial role in the recognition memory task. One of the most important contributions of the present study is that "the bimodal condition created no apparent interference with auditory processing and learning" and "bimodal inputs can be attended to and used to bolster both the implicit and explicit aspects of vocabulary learning" (p. 18). This is a particularly important finding, given the speculations on whether bimodal input would have an aiding impact on implicit and explicit learning of (novel) spoken word forms.

Finally, another poignant aspect to be highlighted from the study regards the fact that "it is possible that an advantage of bimodal presentation over text or sound alone would be more pronounced when both the sound and text input are ambiguous when taken in isolation" (Bird & Williams, 2002, p. 18). For instance, when the sound is not so clear and words with ambiguous or difficult spelling patterns are used in the subtitles. The authors claim that each modality may then be expected to compensate for deficiencies in the other. All in all, results obtained are suggestive that providing both input modalities – sound and text – may facilitate recognition memory.

Perego, Del Missier, Porta, and Mosconi (2010) carried out a study to analyze the cognitive processing of a subtitled movie excerpt by triangulating a variety of measures: eye-movement data, word

recognition, and visual scene recognition. The authors intended to test the hypothesis that the processing of subtitled film is cognitively effective, that is, one should be able to understand the film content without a significant tradeoff between the processing of images (visual input) and text (written input). The authors contend that the underlying hypothesis that motivated this study stems from the fact that when watching subtitled material, "attention needs to be flexibly allocated on parallel information sources during this task" (Perego, Del Missier, Porta, & Mosconi, 2010, p. 250). Hence, they hypothesized that when attention is more focused on the subtitles per se, image processing would be less effective; the opposite would also be true. They explain that this hypothesis is generally consistent with attentional theories that postulate early selection of information channels.

Conversely, a few scholars seem to hold a different view, arguing that reading subtitles and processing both the visual and written input modalities are usually highly efficient and automatized cognitive activities (d'Ydewalle & Gielen, 1992; Lang, 2000; Zhou, 2004), though the populations investigated in those studies had been acquainted with watching subtitled video materials. Thus, the overall conclusion from these opposite views is that participants who are more proficient in (or acquainted with) subtitle processing might also be more proficient in image processing, due to individual differences in attentional resources or executive control (Perego et al, 2010).

In their study, Perego, Del Missier, Porta, and Mosconi (2010) also set out to investigate whether subtitle segmentation would impact on the processing (ill-segmented vs. well-segmented subtitles), therefore slowing down reading processes and causing a decrease in both text and scene recognition performances.

A total of 42 participants took part in the experiment. They were Italian native speakers, all of whom with normal or corrected-to-normal vision. Interestingly, participants reported not being habitual viewers of subtitled videos, they had no knowledge of the language in the audio (Hungarian), and were assigned to two groups: ill-segmented subtitles (low-quality) and well-segmented subtitles (high-quality). They watched a 15-minute video excerpt taken from a Hungarian drama, with Italian subtitles. They were also administered a multiple-choice, general comprehension questionnaire with 12 questions about the video. Moreover, they were given a word recognition completion test, containing 28 sentences to which they had to complete with one given word. A scene recognition test was also administered, which consisted of

a judgment of 60 freeze-frames that participants had to judge whether they were part of the film or not.

Participants' eye movements were recorded using a Tobii 1750 eye-tracking system, integrating all of its components into a 17-inch monitor. According to the authors, with an accuracy of 0.5 degrees and a relatively high freedom of movement, the system is a satisfying eye-movement recording method for natural-use settings, thus ensuring head-motion compensation and very low drift effects.

The results of the study revealed that participants had a better performance on the gist comprehension test than the word recognition test. Interestingly, subtitle quality did not interact with word recognition. Participants also had a very good performance on scene recognition, which did not interact with subtitle quality either. Additionally, results obtained revealed no tradeoff between subtitle processing and image processing, thus suggesting that the participants watching the subtitled film did process its content and subtitles effectively.

Eye-movement recordings revealed very insightful results. The number of fixations in the subtitle area was three times greater than the number of fixations in the upper area, regardless of subtitle segmentation type. However, fixations on the subtitle area tended to be shorter than the fixations on the upper area. The results obtained with eye-movement recordings therefore suggest that participants spent more time reading the subtitles in order to understand the story and only then did they move on to the visual input (scenes). Perego and colleagues (2010) contend that this might lend support to the idea that viewers – at least the ones participating in the study – showed a pattern of attentional allocation, thus willingly deciding to pay more attention to the written input, rather than the visual one.

Another study to make use of eye-tracking methodologies in order to look into subtitling processing is that Kruger and Steyn's (2013), which offers novel insights from an experiment carried to investigate subtitle reading behavior and performance. According to the authors, because there has been no reliable indexes of reading behavior for dynamic texts, they decided to first formulate and validate an index that is capable to measure the reading of a dynamic text, such as subtitles on a film. In a nutshell, the index is a product of the number of unique fixations per standard word in any subtitle by each individual viewer participating in the study and the average forward saccade length of the viewer on this subtitle per length of the standard word in the text as a whole.

A total of 31 first year Psychology students, who spoke English as an L2 with varying L1 (they spoke one of the indigenous languages of

South Africa) participated in the experiment. The participants were randomly divided into two groups: intralingual subtitles and control group. They watched six videos (Psychology lectures) and then completed a multiple choice comprehension test with 40 questions. There was no statistically significant difference between the groups' performance on the test.

The results obtained by correlating all the data demonstrated that although no significant difference was found between the performance of those participants in the intralingual subtitles condition and those in the control condition, it became evident that participants in the test group who actually did read the subtitles performed better on the comprehension test than those who saw the videos with the subtitles but did not read the subtitles.

Kruger and Steyn (2013) claim that language learning courses "stand to benefit greatly from the use of reading over dynamic texts" (p. 118), but "to utilize such texts, instructional design has to interrogate exactly how children and adults process dynamic texts" (p. 119). Moreover, what the study indicated is that word count does not seem to impact significantly on subtitle reading and that the number of lines in a subtitle does not seem to cause problems related to reading and comprehension. It is, however, paramount to bear in mind that this study did not inform whether participants were habitual viewers of subtitled videos, nor does it provide information on the participants' proficiency level in English. One might be led to believe that, provided the level of instruction (higher education) and the type of material adopted (Psychology lectures in English), participants must have been somewhat proficient in the language, with sound educational background.

Winke, Gass, and Sydorenko (2013) investigated caption-reading behavior also using eye-tracking methodologies. A total of 33 second-year English-speaking learners of Arabic, Chinese, Russian, and Spanish, enrolled in the 4th semester in university-level courses at Michigan State University, participated in the study. All of the participants had been placed in the 4th semester in terms of the language classes they would take by an in-house placement test.

Participants watched two 3 to 5-minute videos, one about salmon migration – judged to be familiar by the participants' language instructors – and one about bears – judged to be unfamiliar by the participants' language instructors. Videos were noted to be of equal difficulty by the participants. The videos contained audio in the corresponding target language and intralingual subtitles in the corresponding target language as well (Arabic, Chinese, Russian, and Spanish). A multiple-choice

general comprehension test was administered after viewing, but they were excluded from the analyses. Participants' eye movements were tracked and recorded with the EyeLink 1000 (2009) system.

Results revealed that, in general, participants fixated on the intralingual subtitles 68% of the time that the captions were on screen. Results also revealed no statistically significant difference between the fixations on screen considering the two videos used and that the language being learned did not interact with the time spent on fixating on the captions. However, there was indeed an effect of video (video familiarity) on caption viewing depending on the L2 being learned. More specifically, learners of Arabic, Russian, and Spanish spent similar amount of time reading the captions on both videos. The Chinese learners, in contrast, spent less time reading the captions when watching the video with familiar content.

Another interesting result is the language being learned impacted on the learners' use of the intralingual subtitles in the study. Learners of Spanish spent less time on captions than learners of Russian, who, in turn, spent less time on captions than learners of Arabic. On this aspect, Winke, Gass, and Sydorenko (2013) posit that Chinese learners may have probably spent more time fixating on captions because of the characters that are difficult to process. A possible explanation for the great time spent on captions by Arabic learners, according to the authors, might be due to the complex morphology in Arabic. They also speculate that "because the pronunciation of Arabic words is highly dependent on the word's root, the frequent modification of the root may render word (meaning) identification difficult, especially in the aural mode" (p. 265). In writing, that is, in the intralingual subtitles, one can see the roots and then gain some sense of the word's meaning. The authors do not discuss why Russian learners spent more time reading the captions than Spanish learners, but one may infer that the aforementioned reasons might also hold some truth, given our common knowledge of the structures of the two languages.

The results gathered in Winke, Gass, and Sydorenko (2013) are aligned with those reported in Winke et al (2010), which were previously addressed in this Chapter, pointing out to a distance between the L1 and L2 that can affect the way learners use captions to help their understanding of the video. Furthermore, the results obtained in Winke, Gass, and Sydorenko (2013) indicate that more attention in allocated in the subtitled when the processing of the aural input seems to be more difficult. Simultaneous distribution of attention across audio, video, and

subtitles might enhance one's understanding of the video, though it might also cause a cognitive overload.

The last study selected to be included in this section of the Review of Literature was designed by Perez, Peters, and Desmet (2015), which looked at the effects of type of captioning (full and keyword) on leaners' learning and processing of novel French words, combing and relating to data gathered by eye-tracking methodologies. A total of 51 Dutch-speaking undergraduate students at a Flemish university took part in the experiment, who were taking a mandatory economics class. However, eye-tracking data of 34 participants were properly recorded, who all had normal, uncorrected eyesight.

Two authentic short French clips from a Swiss and Belgian current affairs program for native speakers of French were selected. The first clip, with 6 minutes, was about the LEGO factory and its marketing strategy and history. The second clip, with 2 minutes and 30 seconds, was about a brewery in the north of France, its production process, and export strategy. The clips contained either full captions (in French) or keywords captions (in French too). The vocabulary test, consisting of 51 items – target-words and distractors – required participants to recognize the word form and its meaning and state in which clip it was found, if not a distractor. Moreover, participants were administered a general comprehension test, containing open-ended and multiple-choice questions about both clips. A Tobii X120 standalone stationary eye-tracker was used to register participants' eye movements in a 19-inch monitor.

The authors also decided to include another variable in the study: test announcement, that is, informing participants in a given experimental group that there would be a test after they had been provided with some stimulus. This meant that the study would then comprise four different groups: full captioning incidental, key captioning incidental, full captioning intentional, and key captioning intentional. This allowed the authors to compare the participants' performance on the tests by examining whether the participants awareness of the fact that there would be a follow-up vocabulary test would influence their performance. Additionally, this would enable the authors to check it there would be differences regarding the test announcement and a possible interaction with the type of captioning condition under which participants were.

Analyses revealed that the group in the keyword captioning intentional condition outperformed the other groups in terms of word form recognition, clip association, and word meaning parts of the test. Because participants actually performed better in relation to word

meaning than word form recognition, the authors suspect that there was guessing on the participants' part, especially when taking into account the fact that no pseudowords were included in the test. The analyses also revealed that participants in the two keyword captioning groups outperformed the participants in the other two groups (full captioning intentional and incidental).

As far as eye-movement data are concerned, gaze duration or initial processing analyses revealed that participants in the keyword captioning groups spent more time on the subtitles than the full captioning groups. Perez, Peters, and Desmet (2015) posit that this is not a surprising finding since there was less amount of text to be processed, that is, fewer words in the captions, participants were then able to go back and analyze it more, therefore spending more time on it. Eye-movement data analyses, in terms of second pass reading time, also interacted positively with the participants belonging to intentional groups (both full captioning and keyword captioning). This might lend support to the fact that test announcement yielded the need for participants to reanalyze the content of the captions more closely, thus encouraging them to spend additional time on that area.

Perhaps the greatest contention that Perez, Peters, and Desmet (2015) make is that eye-movement can predict word learning. Also, they argue that students who received a test announcement must have made a conscious extra effort to reanalyze captions' area and this may be considered "indicative of increased intention to commit the word to memory, which resulted in greater learning gains" (p. 324). Finally, from a pedagogical standpoint, this study lends support to the importance of test announcement when designing or preparing video-based activities in the context of instructed SLA. Another aspect that is equally important is that it underpins the value and the potential of keyword captioning as a tool to enhance attention and vocabulary development through video-based activities.

This Chapter focused exclusively on reviewing the relevant literature that explores the use of intralingual and interlingual subtitles on L2 development. The next Chapter focuses on the selected literature on vocabulary development, including important theoretical perspectives on vocabulary recognition in SLA and Working Memory (WM), a variable investigated in the present study.

CHAPTER III
L2 VOCABULARY DEVELOPMENT AND WORKING MEMORY

This Chapter lays theoretical foundations on L2 vocabulary development, starting with theoretical perspectives and models of L2 vocabulary acquisition, followed by a review of key empirical studies. Additionally, it presents the main theoretical underpinnings related to an important variable investigated here: Working Memory (WM). This Chapter also provides an overview of some of the research carried out at the interface of L2 vocabulary development and WM.

3.1 L2 Vocabulary Development: Models and Perspectives

Since Lado's (1956) early paper on patterns of difficulty in vocabulary, it is possible to argue that we have come a long way towards a broader, more encompassing understanding of the importance of vocabulary in L2 development. In keeping with the theoretical, technological, and methodological advancements in the field, vocabulary development has awakened the interest of scholars in SLA, questioning long-held assumptions, especially as to the role vocabulary plays in the picture of learning a second language (Laufer, 1997), therefore constituting itself a telltale matter.

Yet, other questions seem to be as urgent and they are worthwhile focusing on. For one, Ellis (1997) briefly comments on what he believes learning a new word is. According to him, "minimally we must recognize it as a word and enter it into our mental lexicon" (p. 123), but he himself admits that this does not entail simple processes. Ellis is actually referring to the learning of novel word, in the sense that it requires the acquisition of its form, its input and output lexical specifications, its collocations and sentence use, the recognition and understanding of its semantic and conceptual properties, and the mapping of word form labels onto meaning representations. These may all result from conscious (or explicit) and unconscious (or implicit) learning processes.

For Ellis (1997), the learning of novel vocabulary relates to chunking, consisting of the development of a set of associative connections that happens in long-term memory (LTM), through which words are grouped together to facilitate their understanding, processing and consequent learning. In this sense, novel vocabulary learning relates to chunking in that it underlies the achievement of automaticity (as a result of repetition, learning, and practice). Melton (1963, as cited in Ellis,

1997), showed that for letter or digit sequences, the more stimuli were repeated in the short-term memory (STM), the greater the LTM [13] for these items was. Hence, it has been fairly agreed among these scholars that the influences of STM and LTM underlie the development of vocabulary learning and language automaticity.

Ellis (1997) contends that STM is a reliable predictor of long-term acquisition of L1 vocabulary and syntax, though the results are not necessarily generalizable to L2. Phonological short-term memory (PSTM), a specialized memory system that sets up long-term representations for novel word learning, however, is considered a reliable predictor of later vocabulary acquisition in both L1 and L2. The scholar also explains that, unlike children, older young leaners or adults have already developed rich conceptual and semantic systems in their L1, which makes it possible for them to map out L2 word form onto pre-existing meanings and equivalents.

Regarding the links between meaning and form in vocabulary learning, Ellis argues that imagery and semantic mnemonic (memory-based) strategies are perhaps the most effective strategies for both L1 and L2 vocabulary learning, rather than repetition, sentence reading or sentence generation with the target words. Conversely, Nation (2006) argues for the positive effects of repetition on vocabulary learning, since the results showed that greater gains in knowledge are found when learners encounter unknown words at least ten times in context, although they might need more than ten repetitions to develop full knowledge of a word.

Laufer (1997) argues for the important role vocabulary plays in any language learning process, being the bulk of communication and language learning, without which communication is impossible. Siding with Gass and Selinker (1994, as cited in Laufer, 1997), she believes the lexicon to be the most important component for learners. Laufer (1997) also highlights what is involved in the learning of a new word, which can be roughly summarized as follows: (i) form in terms of pronunciation and spelling; (ii) word structure: morpheme; (iii) syntactic pattern of a word in a sentence; (iv) referential (belonging to a specific group, context, or universe), affective (polarity of feelings and emotions), and pragmatic (contextually dependent) meaning; (v) lexical relations of a word; (vi) common collocations.

---

[13] The terms *short-term memory* and *long-term memory* will be properly addressed in subsection 3.3, along with the concept of WM.

The literature of L2 vocabulary development does not provide so many models that attempt to explain how one develops vocabulary in a second language. In fact, de Bot, Paribakht, and Wesche (1997) claim that "at present, there is no widely accepted model of this kind" (p. 309). For the purposes of the theoretical discussions on this matter, two L2 vocabulary acquisition models have been included in this section of the review of literature. They illustrate different, but complementary perspectives on the development of this highly important component of L2 development.

Firstly, understanding how L2 vocabulary development occurs can be exemplified by looking at Jiang's (2000) psycholinguistic model of L2 vocabulary acquisition. His model sees L2 vocabulary acquisition as consisting of three stages, all of which are influenced by the idea that there are constraining conditions under which L2 vocabulary seems to be acquired, such as limited contextualized exposure to the necessary input and the presence of an already existing body of semantic and lexical systems (in the L1), which may interweave with L2 acquisition processes.

Jiang also claims that in L1 vocabulary development, one needs to understand the meaning and the properties of the word, whereas in L2 vocabulary development, the major task is to remember the word. As a consequence, "little semantic, syntactic, and morphological information is created and established within the lexical entry in the process" (p. 50), thus characterizing the first, initial stage of L2 vocabulary acquisition, which refers to a formal stage of lexical development with representational differences concerning L1 and L2 processes.

Nonetheless, Jiang (2000) clarifies that little semantic, syntactic, and morphological information in the lexical entry does not necessarily equate with unavailability of such information to the learners. The author states that the meaning and grammatical features of L2 vocabulary may be activated by L2-L1 links, by means of translation or lexical associations, for instance. More importantly, Jiang contends that "the recognition of an L2 word activates its L1 translation equivalent" (p. 51), thus making such information about the word available for the learner in order to assist comprehension. As one develops in the L2 acquisition processes, stronger associations are developed between L2 and L1 vocabulary. Figure 3 represents the first stage of L2 vocabulary development according to the model.

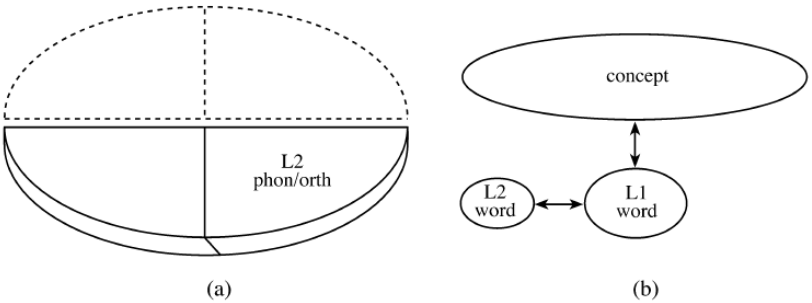


Figure 3: Lexical representation (a) and processing (b) at the beginning of L2 lexical development, taken from "Lexical Representation and Development in a Second Language", by N. Jiang, 2000, *Applied Linguistics, 21,* p. 48.

The second stage in L2 vocabulary acquisition is achieved when the lemma (abstract conceptual form of a word[14]) space of an L2 word is taken up by the lemma information from its L1 translation. In this stage in L2 vocabulary acquisition, L1 lemma information is responsible for mediating the processing that goes into an L2 word.

Moreover, the second stage in L2 vocabulary acquisition comprises a weak established connection between L2 words and their conceptual representations. Jiang (2000) argues that the rationale behind such a weak connection either stems from the fact that the lemma information is copied from L1, rather than created during the learning of L2 words or by the fact that the representation of the information that is copied off the L1 lemma is weak itself because part of the information is lost in translation. Figure 4 represents the second stage of L2 vocabulary development according to the model.

[14] Lemma, according to Kempen and colleagues (Kempen & Huijbers, 1983; Kempen & Hoenkamp, 1987) can also be understood as an entry containing the semantic-syntactic properties of a lexical item, whereas lexeme specifies its morphophonological properties.

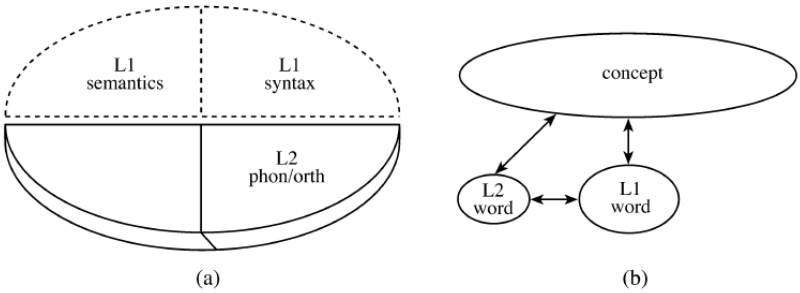


Figure 4: Lexical representation (a) and processing (b) in L2 lexical development at the second stage, taken from "Lexical Representation and Development in a Second Language", by N. Jiang, 2000, *Applied Linguistics, 21,* p. 53.

The third and final stage of the model involves semantic, syntactic, and morphological specifications of an L2 word, which is "extracted from exposure and use and integrated into the lexical entry" (Jiang, 2000, p. 53). The author considers this final stage to be an L2 integration stage in that there will be a high degree of similarity in terms of an L1 and an L2 lexical entry regarding both representation and processing. Figure 5 represents the final stage of L2 vocabulary development the model.

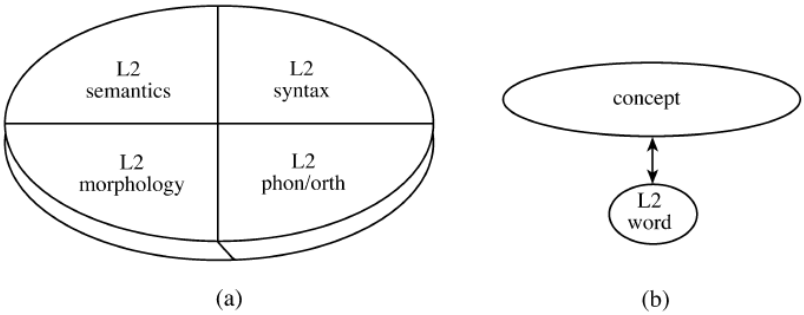


Figure 5: Lexical representation (a) and processing (b) in L2 lexical development at the final stage, taken from "Lexical Representation and Development in a Second Language", by N. Jiang, 2000, *Applied Linguistics, 21,* p.53.

It is important to be mindful that Jiang's L2 vocabulary model intends to describe the evolvement of an L2 word in the learning process rather than characterize an individual's lexical competence as a whole.

Additionally, according to this model, L1 and L2 vocabulary acquisition essentially differ due to limitations that are directly connected with the lack of sufficient contextualized input and the confluence of the established semantic and lexical systems in the L1 with the L2.

The second model hereby reviewed that has attempted to characterize L2 vocabulary development is that of You's (2011). In her study, the author devised a model for vocabulary acquisition based on a series of empirical evidence in the literature[15]. You's model takes into account 12 important factors that have stood out in her thorough analyses of the findings on L2 vocabulary acquisition. These factors are considered as being positively influential in novel word learning and they are summarized as follows:

- Word repetition or frequency of exposure;
- Marginal glosses and dictionary use;
- Explicit instruction of target-words;
- Presence of contextual clues;
- Level of text comprehension;
- Learner proficiency[16];
- Pictorial input along with text;
- Reading-while-listening mode is more effective than reading only condition;
- L1 translation;
- Types of tasks (form-oriented rather than message-oriented only);
- Word class: nouns are easier to acquire than verbs and adjectives;
- Absence of semantic elaboration: synonym generation negatively affects L2 word acquisition.

[15] It is important to bear in mind that the studies reviewed by You (2011) and the L2 vocabulary acquisition model she puts forth comprise evidence on L2 vocabulary acquisition exclusively through reading.

[16] Interestingly, a body of evidence is presented by You (2011), suggesting that less proficient learner are less affected by word frequency than higher proficient learners.

The model, integrating the above-mentioned factors as well as the reading material, target vocabulary, and the learner, can be visualized in Figure 6:

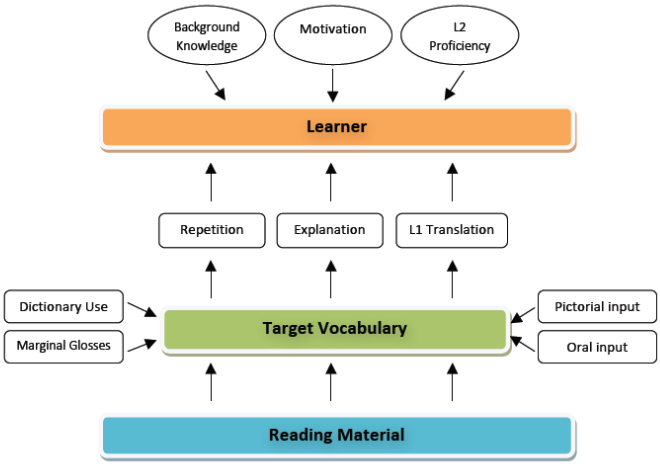


Figure 6: You's (2011) model of vocabulary acquisition, taken from "Factors in Vocabulary Acquisition through Reading", by Y. You, 2011, *ITJ, 8,* p. 53.

You's (2011) L2 vocabulary acquisition model places the reading material, target vocabulary, and learner as its main elements. The model is influenced by the author's view in which "reading, as revealed through the previous studies, is one of the most effective sources to acquire new words" (p. 55), but emphasizes that the development of learner's strategies has to accompany all of the stages in her/his L2 vocabulary development. Finally, the author stresses the importance of keeping these influential factors in mind in order to look at L2 vocabulary development more holistically and pedagogically.

The two models of L2 vocabulary development reviewed in this subsection of the present Chapter are indicative of the complex processes that underlie L2 vocabulary development. The models – presenting different perspectives that may deepen or even sharpen our understanding of such intricate processes imbued with L2 vocabulary development – can have theoretical and pedagogical ramifications. They may influence the design of instructional materials and pedagogical practices that can inform how one approaches L2 vocabulary development instruction or research.

3.2 Empirical Studies on L2 Vocabulary Development

Empirically driven, Laufer (1997) formulated a list of factors influencing word learnability, which can be classified into facilitating, difficulty-inducing or unclear effect factors. Table 1 summarizes essential information related to the three categories as follows:

Table 1

Word Learnability Factors (Laufer, 1997)

Facilitating Factors	Difficulty-inducing Factors	Unclear Factors
familiar morphemes, phonemic regularity, fixed stress, consistency of sound-script relationship, inflectional and derivational regularity, morphological transparency, register neutrality, and homonyms.	foreign morphemes, phonemic irregularity, variable stress and vowel change, incongruence in sound-script relationship, inflectional and derivational complexity, deceptive morphological transparency, register restrictions, idiomaticity, and polysemy.	word length, concreteness, and abstractness

In addition to the factors that influence word learnability, Laufer (1997) argues that little research has focused vocabulary learning strategies employed by language learners while consciously trying to acquire novel words. Such coarse research on vocabulary learning strategies (Ahmed, 1989, as cited in Schmitt, 1997; Sanaoui, 1995, as cited in Schmitt, 1997) has revealed that learners are either typically aware of their vocabulary learning strategies or not. In case they are, they tend to structure their learning, engage in conscious learning moments, and review and practice their target words. Furthermore, the literature suggests that teaching vocabulary learning strategies is highly important when the target words are low-frequency words. This was observed in Nation (1994), who argues that for these words, both teachers and students might benefit if the latter are taught at least three strategies: guessing from context, mnemonic techniques, and using word parts.

In relation to strategies used by students to learn novel vocabulary, Schmitt (1997) carried out a survey with Japanese L1 speakers who had taken or were taking EFL classes in Japan. Surveys were administered to four groups: Junior high school students, high school students, college

students, and adult learners. In each group, there were a total of 150 surveys, bringing the total number to 600.

The taxonomy Schmitt came up with consisted of the following strategies: (i) Discover meaning (bilingual dictionary, guessing meaning from context, asking a classmate for meaning); (ii) Consolidate meaning (verbal repetition, written repetition, study the spelling, say new word aloud, take notes in class, study the sound of a word, word lists); (iii) Check for L1 cognate; (iv) Use physical action; (v) Use cognates when studying; (vi) Use semantic maps; (vii) Teachers check flash cards for accuracy.

The survey revealed that in Japan there is a strong affinity for the bilingual dictionary, being the most used strategy of all. Additionally, 74% of respondents reported that they guessed meaning from context. The only other frequently used strategy was asking classmates, at 73%. Repetition of a word's verbal or written form (presumably thinking of and focusing on its meaning) rank top at the list, which can be partially attributed to the study style encouraged by the Japanese school system, with constant repetitions. It is not surprising that cognates are relatively unused as a strategy, given the dissimilarities between English and Japanese.

Erten and Tekin (2008) report on a study that investigated the effects of vocabulary recall in semantically related sets and semantically unrelated sets. Early studies on vocabulary presentation showed that when similar words are presented together, there seems to be an interfering effect on learning (McGeoch & McDonald, 1931, as cited in Erten & Tekin, 2008).

In their experiment, 55 fourth graders in Turkey, with limited formal instruction in English, were presented with 80 new English words in two different sets: two semantically related sets of 20 words and two semantically unrelated sets of 20 words. Students received instruction on the lexical items. The items were selected on the basis of length, semantic relations, cognateness, idiomaticity, and concreteness.

The results showed that participants learned more words in semantically unrelated sets. Also, the results showed that participants learned words faster in semantically unrelated sets too. The authors argue that although lexical items appear to be arranged in the mental lexicon around semantic bonds, "the learning of new lexical items may involve a different route of mental processing" (Erten & Tekin, 2008, p. 416). They also explain that, when presented with words in semantically related sets, participants may have had to refine their existing lexical bonds through a restructuring process, which may have caused confusion motivated by a

cross-association between similar items. This would entail what is called interference in memory (Baddeley, 1997, as cited in Erten & Tekin, 2008).

The authors stress the importance of rethinking language instruction textbooks in terms of how vocabulary is presented. Since the results in their study corroborate previous findings in favor of presenting novel vocabulary in semantically unrelated sets, they posit that it might be more effective to present them as such. Also, they emphasize the need to present new vocabulary in meaningful contexts to foster vocabulary instruction and acquisition. As for sets of semantically related vocabulary, Erten and Tekin suggest that they might be introduced in different units of study, in thematic units, and should be recycled often.

It is common for L2 language teachers to encourage learners to use the language their students are learning outside the classroom and to incentivize them to try and find situations to expose themselves to authentic input from which they can profit in terms of language gains. Milton (2008) reviews a few studies[17] concerning vocabulary acquisition outside the classroom environment in terms of informal tasks carried out by learners to foster vocabulary uptake.

What the author found in the studies was that vocabulary learning success is not only related to the amount of new words acquired, but also to the time spent on the informal tasks, outside the classroom. However, learners may end up acquiring odd, infrequent words that might not prove to be very effective in their language development. Moreover, Milton brings up a highly important issue from the studies he reviewed, which relates to conscious or deliberate effort to learn new words, regardless of the fact that learners are engaged in language-related activities outside a formal instruction environment.

The author concludes by stating that informal tasks carried out outside the classroom "may aid long-term retention of words" (p. 235), since they seem to be, most of the times, practical, effective, motivating, and enjoyable for learners. Milton also reminds us of the importance of classroom input and direction on the part of the language instructor, which, combined with learners' motivation to gain vocabulary outside the classroom, may prove to be effective in the long run.

[17] See Horst and Meara (1999) for an example of one of the studies revisited by Milton (2008). The other studies the author addresses in his review are unpublished.

Determining the effects of active translation[18] on short-term incidental (no instruction or test announcement) recall of unknown L2 words was the crux of the investigation undertaken by Hummel (2010). A total of 191 university-level French native speakers who were intermediate-level speakers of English were divided into three groups: French to English active translation; English to French active translation; and exposure and copy exercise.

During instrumentation, all participants in all conditions received a list containing 15 unfamiliar English nouns to provide an L1 equivalent. Then, they received a testing booklet with words and their corresponding translation in French. Then, participants were given the testing booklet corresponding to their experimental group, and they were not informed about subsequent tests on the target items.

In each testing booklet, participants were provided with the proper translation equivalent for each of the 15 lexical items since each L2 word was juxtaposed with its L1 translation. Furthermore, participants in the active translation conditions were provided with a sentence context for each target lexical item, a French sentence for the French to English translation group, and an English sentence for the English to French translation group. After reading each sentence, they were required to translate it.

As for the exposure and rote-copy condition, participants were exposed to the L2 target lexical items and their French equivalents. Half of the rote-copy condition participants were also supplied with L2 (English) sentences containing the target word and each sentence's L1 (French) translation equivalent, and their task was to copy each of the sentences. The other half of the rote-copy condition participants was provided with an L2 (English) sentence containing the target word and their task was to copy the sentence.

Results indicated that all the groups showed short-term increase in vocabulary recall in relation to the pre-test scores. Participants in the rote-copy condition outperformed both experimental groups. Hummel argues that the better performance by participants in the copy condition suggests that "exposure to L1 and L2 sentence equivalents – especially when attention is focused on sentences as when asked to copy them – appears

[18] Although Hummel (2010) does not provide an explanation to what she means by *active* translation, one can be led to assume that she refers to this task as translation as we know it. Moreover, the rationale behind the use of *active* translation may be considered as "allowing deeper and more elaborated processing and therefore may facilitate retention" (p. 63).

to contribute to more effective L2 vocabulary retention compared to a translation exercise" (p. 68). It appears that by copying the words, participants' attention is drawn to the form and the act of copying them may result in a separate motor trace in memory, which seems to be assistive in word retrieval. Hummel also claims that active translation might be "even more resource consuming" (p. 69) and might have required much concentration, attention, and cognitive processing in the activity of translating the equivalents that attention was actually taken away from the target-words. This would constitute an information overload. No statistically significant differences were found between the translation groups.

Finally, Hummel (2010) discusses the issue of incidental learning, positing that ultimately it is the learner's choice to direct intentional effort and attentional resources to the learning material presented, be it either in the instructional or experimental contexts. The author argues that all the conditions may be applicable to the classroom to facilitate short-term vocabulary recall, but "students' L2 vocabulary acquisition may be particularly advantaged by exposure to sentence translation equivalents and participation in a copy exercise" (p. 70), thus combining both translation and rote-copying exercises.

In line with Milton's (2008) work, Willis and Ohashi (2012) sought to research whether a combination of variables – word cognateness, frequency, and length – predict difficulty in L2 word learning and retention. A total of 69 first- and second-year students in the Departments of Communication, Linguistics, Mathematics and Science, and Psychology at a women's university in Tokyo, Japan who had been studying English formally for at least 7 years at school and university participated in the experiment, with varying proficiency levels.

A vocabulary size test (Nation & Beglar, 2007) was used, requiring participants to answer multiple-choice questions such as the one below, with a total of 70 words, belonging to different frequency levels:

(1) soldier: He is a soldier.
(a) person in a business
(b) student
(c) person who uses metal
(d) person in the army

As far as cognateness is concerned, the experiment comprised the use of a binary measure, that is, cognate/non-cognate words, whereas word length was measured in terms of the number of letters, phonemes,

and syllables, given the dissimilarities between the linguistic pair at stake (English and Japanese).

Results obtained show high correlations between word difficulty and the predictor variables. In other words, cognateness correlated the most with word learning, followed by frequency, and finally word length, thus indicating that it was easier for the participants to learn and retain cognates and more frequent lexical items. However, the results also pointed out to a negative correlation in some of the variables, indicating that it is harder to learn and or retain longer words than shorter ones.

Willis and Ohashi argue that for non-cognate words, learners need more encounters for long-term retention. Moreover, the scholars admit that certain features might not be as easy to be manipulated in the classroom to ensure learning, such as cognateness and word length, but emphasize that word frequency is something that can be more easily manipulated during lesson design and the planning of activities. Finally, the authors claim that there should be room for form-focused activities that present target-words with repeated meaningful encounters.

What the empirical studies reviewed in this section demonstrate is that L2 vocabulary development seems to be affected by a plethora of factors that include, but are not limited to: familiar or foreign morphemes, phonemic regularity, fixed stress, consistency of sound-script relationship, inflectional regularity, derivational regularity, morphological transparency, register neutrality, and homonyms. Also, word learnability seems to be affected by word length, cognateness, and frequency. Though these studies do not point out to a clear convergence of factors, the results in them can be taken as indications of relevant aspects to be kept in mind for both instructional and experimental approaches to L2 vocabulary development by language practitioners and researchers in the field of SLA.

The variety of studies on vocabulary development reviewed in this Chapter so far all point out to a recurring aspect of human cognition that seems to permeate and influence the pace and the success related to cognitive processes of recognizing, recalling, attending to, and learning novel vocabulary, which is Working Memory (WM). In this sense, the next subsection of this Chapter addresses this variable investigated in the present study, followed by another subsection that presents empirical studies that have shown (or not) a correlation between such a variable and L2 vocabulary development.

3.3 Working Memory

The term Working Memory (WM) has been defined differently by a number of scholars from different fields of knowledge and sometimes in the past used to be equated with STM depending on the theoretical perspective adopted (Baddeley & Hitch, 1994). In this study, a few theoretical definitions are in order, starting with STM. The concept of STM memory regards primary memory (Broadbent; 1958; Atkinson & Shiffrin, 1968 as cited in Chen & Cowan, 2009). It reflects faculties of the human mind that can store a limited amount of information for a certain period in a very accessible state, not necessarily in conscious awareness. LTM, on the other hand, differs from STM in terms of duration and capacity (Chen & Cowan, 2009). Regarding duration, the difference indicates that items in short-term storage decay from this type of storage as a function of time, whereas a capacity difference indicates that there is a limit as to how many items short-term storage can hold.

WM directly relates to the control of specific cognitive mechanisms, such as attention, processing, and other regulatory functions, entailing the access to LTM information (Baddeley, 2000). Additionally, Baddeley characterizes WM as a limited capacity system, which is capable of temporarily maintaining and storing information. Because of that, WM supports human thought processes by offering an interface between perception, LTM and action. Since WM has been suggested in the literature to vary on an individual basis, a number of studies have investigated WM to verify the extent to which it might function as a predictor of language learning outcomes, for instance (Daneman & Merikle, 1996; Engle, Kane & Tuholski, 1999; Engle, Tuholski, Laughlin & Conway, 1999; Unsworth & Spillers, 2010; to mention but a few).

The origins of WM date back to Atkinson's and Shiffrin's (1968) model[19], a multi-store or modal model that was extremely successful in terms of the amount of research it went on to generate. Building up on that model, Baddeley and Hitch (1974) began to develop an alternative model to depict WM in more detail, arguing that STM, as represented by Atkinson's and Shiffrin's (1968) model, was not very complete. Atkinson's and Shiffrin's model postulated that STM can hold limited

[19] Today, over 10 different WM models exist, and they essentially tend to differ in terms of how information is thought to be retained, processed, manipulated, and stored (Myiake & Shah, 1999). Though I am well aware of the existing myriad of WM models, this section will only review the multi-component WM model put forward by Baddeley and colleagues, chosen on the basis of theoretical affiliation.

amounts of information for short periods of time with relatively little processing, thus being considered a unitary system without any subsystems. Baddeley and Hitch (1974) did not concur on that matter, since they claimed that WM does not consist of a unitary store system, but rather a multiple-component storage system. Baddeley and Hitch's WM model is presented in Figure 7.

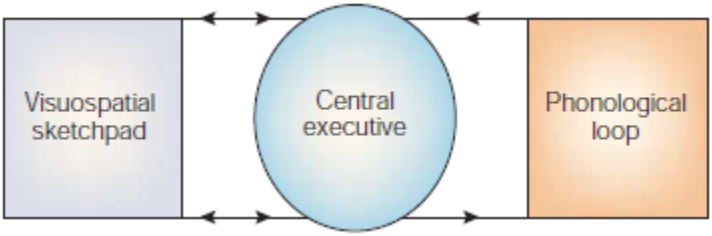


Figure 7. Baddeley & Hitch's (1974) multi-component model of WM, taken from "Working Memory: Looking Back and Looking Forward", by A. Baddeley, 2003, *Nature Reviews Neuroscience, 4,* p. 830.

With time, Baddeley and Hitch's model also underwent a few theoretical refinements, resulting in a maintained tripartite structure model with more specified functions in terms of the central executive (Baddeley & Logie, 1999). This WM model comprises multiple specialized components, with a system that deals with temporary storage and online information processing, in other words, a system that addresses things (information) that we are doing and handling as of now. Furthermore, the system is under attentional control, suggesting that we are responsible for directing and allocating our attentional resources. This system underpins our capacity for complex thoughts that rely heavily on temporary storage and manipulation of information.

Baddeley and Logie's (1999) view on WM holds that this construct allows people to comprehend and mentally instantiate their environment, retain information about their (immediate) past experiences. Moreover, the scholars contend that it supports the acquisition of new knowledge. Baddeley and Logie's WM model consists of a control system of constrained attentional capacity, the central executive, which is aided by two subsidiary storage or 'slave' systems, namely: the phonological loop, in charge of storing and rehearsing speech-based information, and the visuospatial sketchpad, which, in a nutshell, consists of a workplace for holding and manipulating spatial and visual information.

As regards the central executive – the most complex component of WM – it is postulated to be a supervisory system in charge of the

selection, initiation, and termination of processing routines, such as encoding, storing, and retrieving information. Baddeley and Hitch (1994) equate the central executive with the supervisory attentional system (SAS), proposed and described by Norman and Shallice (1980) and by Shallice (1982). According to Shallice (1982), SAS is a limited capacity system used for a variety of purposes, including tasks that involve planning or decision-making processes, trouble shooting in situations in which the automatic processes seem to be running into difficulty, novel situations, dangerous or technically complex situations, and situations entailing strong habitual responses or temptations.

Baddeley and Logie (1999) consider the central executive to be responsible for controlling the two slave systems, focusing and switching attention, although they mentioned that it is becoming clear that this is not unitarily controlled, activating representations within LTM, but not being involved with temporary storage.

One of the components of the so-called slave systems is the phonological loop. Baddeley (1992) states that the phonological loop seems to be the simplest and most well-understood and investigated component of working memory and comprises a phonological store that can hold acoustic or speech-based information for up to 2 seconds, coupled with an articulatory control process. At least four specific laboratory-based findings related to the phonological loop are consistent to date: (i) the acoustic similarity effect: immediate ordered recall of items is poorer with similar than dissimilar sounds (similarity of meaning does not have this effect); (ii) the irrelevant speech effect: immediate serial recall of sequences of visually presented verbal items is impaired if certain task irrelevant background sounds are presented simultaneously; (iii) the word length effect: memory span for words is inversely related to their spoke duration, which provides evidence to the subvocal rehearsal process; (iv) articulatory suppression: the process of inhibiting memory performance by speaking while being presented with an item to recall. When suppression occurs, the acoustic similarity effect is abolished.

The other component is the visuospatial sketchpad, which is assumed to hold visual information. It is used in the temporary storage and manipulation of spatial and visual information, such as remembering shapes and colors (the visual cache), or the location or speed of objects in space. The visuospatial sketchpad is also considered to be crucial in tasks involving planning of spatial movements (through the inner scribe), like planning one's way through a complex building. The visuospatial sketchpad can be further categorized into separate visual, spatial, and possibly kinesthetic (movement) components.

After several theoretical refinements, Baddeley (2000) introduced a new component to the already existing model of WM, the episodic buffer. It consists of a multimodal temporary store that can deal with different modalities of information, binding together information from different sources within the WM system. Information about a scene may entail visual information, speech sounds, and movement, and it is hypothesized that the episode buffer may be in charge of to join this information into a coherent memory episode. The episodic buffer might work with the idea of chunking in that the more the information can be bound together, the greater the capacity of the buffer. Figure 8 shows Baddeley's (2000) WM model revision.

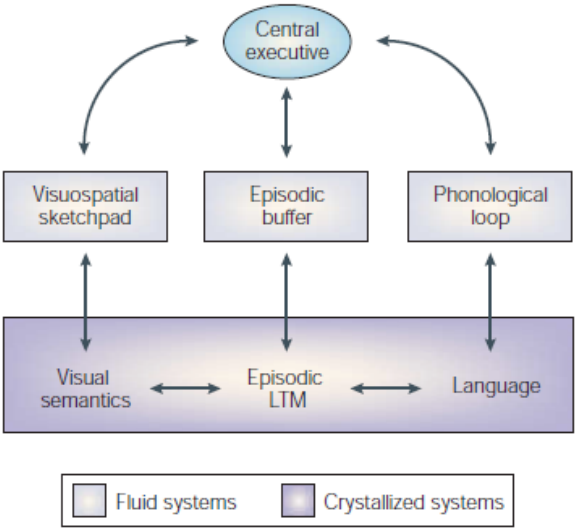


Figure 8. Baddeley's (2000) WM model revision, taken from "Working Memory: Looking Back and Looking Forward", by A. Baddeley, 2003, *Nature Reviews Neuroscience, 4,* p. 835.

Baddeley (2000) explains that the episodic buffer is "a temporary storage system that is able to combine information from the loop, the sketchpad, long-term memory, or indeed from perceptual input, into a coherent episode" (p. 148). Moreover, the episodic buffer is potentially what gives us our experience of consciousness, since in order to experience consciousness one must be able to keep track of one's current

experience, but also be able to reflect upon it in real-time. Nevertheless, not much is known about the episodic buffer as of now.

The episodic buffer was proposed as an attempt to account for a variety of empirical data that did not seem to be resolved with the previous tripartite model. In their article, Repovš and Baddeley (2006) discuss a few issues that the episodic buffer was proposed to address. One of the issues was that the existing model did not provide a solid explanation as to how individuals were able to recall only five unrelated words and up to 16 related words (using sentences), which means that the tripartite model did not offer substantial explanation that could elucidate the advantage of recall when meaningful relationship between words existed. Secondly, the model could not account for how subsystems of WM relate to and interface with LTM. Thirdly, the original model failed to explain the binding of information from the two slave systems as to how and where the combination occurs. Lastly, the WM model could not suitably account for the wealth of research exploring individual differences in WM, such as those verified in WM span tasks designed by Daneman & Carpenter (1980; 1983).

Repovš and Baddeley (2006) explain that the episodic buffer needs exploring and that can be done so by means of measures of capacity and interference tasks. Such exploration, as the authors contend, may prove even more challenging than the examination of the phonological loop and visuospatial sketchpad, since tasks that can reliably test the underlying mechanisms of the episodic buffer will necessarily have to require complex integration of information.

According to Baddeley (1992), WM research has mostly developed under two different, but complementary approaches. The first approach is named the dual-task and neuropsychological approach, which focus on the analysis of the structure of WM itself, emphasizing its slave subsystems. In practical terms, this approach usually includes the study of evidence of neuropsychology and the application of dual tasks, requiring participants to memorize and store digits in a digit-span task while performing other cognitive tasks, for instance.

The second approach, called the psychometric correlational, refers to the correlation between individual differences in WM and the performance of cognitive abilities (Daneman & Carpenter, 1980, 1983). In this paradigm of research on WM, it is thought that WM is a good predictor of individual differences and that individuals with larger WM capacity present better performance in cognitive tasks than those individuals with smaller WM capacity. In this perspective, the two functions of working memory – storage and processing of information

(Baddeley, 1992; Daneman, 1991) – compete while high cognitive skills are being performed (Daneman & Carpenter, 1980; 1983). The methodology underlying this approach involves tasks and correlate performance in these tasks with performance in other high cognitive tasks. The present study is carried out within the psychometric approach to the study of WM and its possible relationship with other cognitive tasks, namely language learning tasks, for instance.

Today, over 10 different WM models exist, which signal how prolific investigations have been in the past years. Simply put, the existing WM models essentially tend to differ in terms of how information is believed to be retained, processed, manipulated, and stored (Myiake & Shah, 1999). As stated before, this section of the review of the literature has focused on the multi-component WM perspective put forward by Baddeley and colleagues, chosen on the basis of theoretical affiliation. In other words, considering the characteristics of the model, how each of them seems to be associated with specific types of WM function, the nature of the present investigation, and the types of tests used to gather data, which require the processing of auditory and visual channels (see Method section for more explanation on the tests used in this study), this study then interprets all the information related to one of the variables hereby investigated – WM – through the theoretical underpinnings put forth by Baddeley and colleagues.

3.4 Empirical Studies on L2 Vocabulary Development and Working Memory

The plausible relationship between WM and vocabulary development has been somewhat extensively investigated, and positive results with both L1 and L2 studies have been reported in the literature (Gathercole & Baddeley, 1990; Baddeley, Gathercole & Papagno, 1998; Service, 1992; Masoura & Gathercole, 1999, to mention but a few).

One example is Gathercole and Baddeley (1990), who investigated the role of phonological memory, in charge of the reception, analysis, and processing of sound-based elements in language. In their study, skills of a group of six children who presented some type of disordered language development and histories of low progress at acquiring first language were examined. These children were compared against two control groups. A series of standardized psychometric tests were administered to the experimental group. Also, receptive vocabulary was assessed by the Short Form of the British Picture Vocabulary Scale (Dunn, Dunn, Whetton & Pintilie, 1982). Reading abilities were assessed by the reading

test of the British Abilities Scales (Elliott, Murray & Pearson, 1983). Oral comprehension was checked using Bishop's (1982) Test for Reception of Grammar. Nonverbal intelligence was tested using Raven's Coloured Progressive Matrices (Raven, 1984). Results comparing the experimental group and the control groups suggest that a deficit in phonological short-term memory (PSTM), in terms of storage in WM, may explain the language-disordered children's poor memory performance on the tests.

Service (1992) established a link between non-word repetition test-performance on the learning of English vocabulary by Finnish schoolchildren. In the experiment, three tasks were employed to predict English learning by Finnish children over three years. In the non-word repetition task, participants were required to repeat out aloud tape-recorded non-words that sounded like Finnish or English. In the non-word copying task, participants saw strings of letters that looked like Finnish or English words and copied them when they had disappeared from their eyesight. When comparing syntactic-semantic structures, participants had to find the syntactically matching pairs from two sets of Finnish sentences. Results showed that repetition and copying accuracy and the ability to compare syntactic-semantic structures predicted English learning. Service argues that the participants' ability to represent unfamiliar phonological material in their WM underpins their acquisition of new vocabulary items.

Baddeley, Gathercole, and Papagno (1998) present compelling evidence that the phonological loop is crucial in learning the novel phonological forms of new words. The scholars review evidence in typical adults and children, as well as neuropsychological patients. Digit span and nonword repetition tasks are usually the most used measure of verbal STM ability in children.

Masoura and Gathercole's (1999) study with forty-five Greek foreign language learners of English (children) also provided interesting results. Children's native vocabulary knowledge was assessed using measures obtained by receptive and productive tests. As regards foreign vocabulary measures, data were gathered by requiring children to translate words (from English to Greek and from Greek to English). Nonword repetition tests were also administered. The results demonstrated that their PSTM measures in both Greek and English shared high associations with their knowledge of vocabulary in both languages.

Speciale, Ellis, and Bywater (2004) present interesting results from two experiments that investigated individual cognitive differences affecting L2 vocabulary acquisition. In the first experiment, 38 undergraduate Psychology majors, native speakers of English, with no

knowledge of German, participated in the study. This experiment was designed to measure participants' nonword repetition ability to determine their phonological store capacity and their phonological sequence learning ability. The results of these tasks were associated with participants' performance on the German vocabulary test, which consisted of a learning phase, a receptive phase, and finally a productive phase. Information on participants' L1 vocabulary knowledge was also gathered.

Scores on the test demonstrated a correlation between participants' phonological memory measures and their performance on both the German receptive and German productive vocabulary tasks. More espefically, phonological sequence learning was found to be a significant predictor of receptive vocabulary learning, whereas phonological sequence learning and phonological store capacity were found to contribute to productive vocabulary learning.

In the second experiment carried out by Speciale and colleagues (2004), in the context of a longitudinal study of classroom Spanish foreign language vocabulary acquisition, 44 beginning undergraduate students of Spanish taking part in a 10-week Spanish course participated in the study. Participants' ability to learn phonological regularities were measured, as well as their store capacity using the nonword repetition task. In addition to the nonword repetition task and phonological sequence learning task, participants took a Spanish receptive vocabulary test, a Spanish nonword repetition task, and had their performance assessed in a Spanish written exam at the end of their course.

In sum, the results obtained from the second experiment confirm the relationship between phonological short-term memory ability and L2 vocabulary acquisition. Results from both experiments indicated that the combined effect of the phonological store capacity and the ability to learn phonological regularities was more directly related to both productive and receptive L2 lexical competence than the capacity of the store alone. The results obtained in Speciale et al, in that phonological memory was associated with L2 vocabulary acquisition, were also obtained in a study carried out by French & O'Brien (2008), who found that L2 vocabulary acquisition by French-speaking children in an intensive English program

correlated with their phonological memory capacity, though their L2 grammatical acquisition did not necessarily depend on their lexical knowledge.

In a very recent study, Verhagen and Leseman (2016) reported on a study whose results support earlier studies comparing the effects of phonological memory on the acquisition of L2 grammar and vocabulary.

In their study, the authors differentiate verbal short-term memory – the capacity to store verbal information – from verbal WM – the ability to process verbal information while storing it. The study involved 45 5-year-old children, learning Dutch as an L1, and 63 Turkish L2 learners of Dutch. In the study, participants' verbal short-term memory and verbal WM were measured in order to correlate with L2 grammar and vocabulary tests.

In a nutshell, the results obtained in this study provide support to studies comparing the effects of verbal short-term memory and verbal WM in the same population investigated. Verhagen and Leseman explain that both components of verbal memory are "significantly related to individual differences in vocabulary and grammar learning" (p. 80). However, verbal short-term memory was found to be associated with L2 vocabulary and grammar acquisition, whereas verbal WM was found to correlate with L2 grammar acquisition only.

Interestingly, some studies have reported no correlation between WM and L2 vocabulary development. For one, a study carried out by Cheung (1996) with 84 Hong Kong seventh-graders in high school, who spoke Cantonese Chinese as their L1, with around 9 years of English formal instruction, demonstrated conflicting results. The participants took an English vocabulary test, an English reading comprehension test, a nonverbal intelligence task, a nonword span task, and a simple word span task.

Unlike findings on the links between WM and L2 vocabulary acquisition (Gathercole & Baddeley, 1990; Baddeley, Gathercole & Papagno, 1998; Service, 1992; Masoura & Gathercole, 1999), the results obtained in the study did not show a correlation between participants' phonological memory and L2 word learning. In fact, WM was only associated with a subgroup of the participants with low vocabulary knowledge. Cheung speculates that one possible reason for the absence of such a connection refers to the WM model arrangement itself. In other words, it might be the case that novel verbal material can only be transferred into long-term memory depending on the efficiency in terms of rehearsal process. "Working memory in general and the rehearsal process in particular, therefore, become an important 'front gate' through which verbal information must pass to gain registration in long-term memory" (p. 871), Cheung explains.

According to the author, another plausible explanation for the lack of correlation found between WM and L2 vocabulary development in the study carried out relates to a possible interaction between phonological memory and long-term knowledge. In other words, Cheung contends,

based on empirical evidence, that "vocabulary size indicates long-term experience with the language" (p. 871), which would help explain why there was no correlation between WM and participants with higher vocabulary size. The puzzling results found by Cheung pose an unforeseen hypothesis, which relates to the possibility that "there may be a point in second-language development beyond which phonological memory will just fail to predict vocabulary size" (p. 872).

Another study that found no correlation between WM and L2 vocabulary development is that of Akamatsu's (2008), who investigated the effects of training on word-recognition processing of EFL learners. The author also investigated whether word-recognition processing correlated with word frequency and WM capacity. The stance taken by the author is that of word recognition as a vital component in automatizing reading processes given its interactive nature of lower level processing – for instance, retrieving the orthographical, phonological, syntactic, and semantic information of a printed word – and higher level processing – processing a stream of words for comprehension while preserving the semantic and syntactic information of previous sentences in WM.

In terms of the development of L2 word recognition, Segalowitz and Segalowitz (1993, as cited in Akamatsu, 2008) explain that it comprises two qualitatively different changes: *simple speedup* and *automatization*. An initial stage in the development of L2 word-recognition skills consist of simple speedup of performance. Simply put, beginning L2 readers' increase in speed and accuracy of word-recognition processing is "the result of simply speeding up controlled processing" (p. 178), whereas automatization "is associated with restructuring a word-recognition mechanism or with increasing cognitive efficiency in word-recognition processing" (p. 178).

Akamatsu's study involved 49 first-year education majors who had all received around six years of formal English education by the time of the study. Participants took a reading comprehension and a vocabulary test to determine their proficiency and vocabulary knowledge. As for the training, the word-recognition training was given as part of regular class period across seven training sessions, being one per week, which required them, for instance, to take word chains task (drawing a line between a string of subsequent letters to form words) as fast as they could.

Results revealed that WM did not correlate with participants' reaction time in word-recognition processing. With regards to frequency, the results suggest that improvement in word-recognition speed with low-frequency words was associated with automatization whereas that with high-frequency words was associated with simple speedup. Akamatsu

explains that the universal direct access hypothesis would help provide an insight into the results since familiar words are recognized via visual mediation or sight-word reading, whereas less familiar words tend to be processed on a phonological basis where knowledge of letter−sound correspondences and spelling patterns plays a very important crucial role. Finally, Akamatsu (2008) demonstrated that training had a positive effect on word-recognition processing in terms of speed and accuracy.

All in all, studies reviewed in this subsection of the Chapter, apart from the two exceptions cited in the literature – Cheung (1996) and Akamatsu (2008) –, serve as strong indications of the role that WM – especially the phonological loop – plays on (L2) vocabulary development, possibly because it may as well enable the establishment of "stable, long-term mental representations of novel phonological material" (Martin & Ellis, 2012).

The next Chapter will focus explicitly on the methodological rationale that guides the present research. It explains the foundation and justifications behind the methodological decisions made as regards instrumentation and data collection, as well as provides detailed explanation on the steps undertaken for data collection, data analyses, and the statistical procedures adopted.

CHAPTER IV
METHOD

This chapter provides a detailed account on the methodological rationale that has guided this study, by firstly outlining the objectives in section 4.1 and research questions in section 4.2. Section 4.3 then presents the hypotheses drawn from the research questions and insights gathered from the literature. Information about the participants is provided in section 4.4. Sections 4.5, 4.6, 4.7, and 4.8 address the materials as well as tests designed and adapted for data collection. Section 4.9 focuses on the questionnaires used to collect additional data on the participants and their performances on the tests. Section 4.10 provides detailed explanation on the procedures employed during data collection, as well as the research design. Finally, section 4.11 concentrates on the scoring and statistical procedures regarding the tests administered in the present research.

4.1 Objectives

The present study, which is of a mixed design nature – qualitative and quantitative –, is aimed at investigating the effects of the use of intralingual and interlingual subtitles on Brazilian intermediate English as a Foreign Language (EFL) learners' L2 development. More specifically, this study aims to examine whether intralingual and interlingual subtitles aid content comprehension – general and specific – as well as they help promote L2 vocabulary recognition in a pre-test, test, and post-test research design. Furthermore, this study seeks to verify the extent to which learners' level of comprehension and L2 vocabulary recognition correlates with their WM capacity in any subtitling condition.

4.2 Research Questions

The objectives of the present study can be translated into the following research questions (RQ):

- (RQ1): How do intralingual and interlingual subtitles affect learners' video comprehension, as measured by immediate post-viewing test?

- (RQ2): How do intralingual and interlingual subtitles affect learners' L2 vocabulary recognition, as measured by pre-test, test, and post-test?

- (RQ3): How does learners' L2 vocabulary recognition, in case there is any, differ across time for any experimental condition?

- (RQ4): Does leaners' WM capacity correlate with their performance on the video comprehension test and/or their performance on the L2 vocabulary recognition test in any subtitling condition?

4.3 Hypotheses

The objectives of the present study, the research questions presented above, and insights from the literature have all given rise to four hypotheses, each of them referring to each of the RQs previously shown, respectively:

- Hypothesis 1: Participants in any of the subtitling conditions – intralingual and interlingual subtitles – will outperform participants in the control condition in terms of their level of video comprehension.

- Hypothesis 1a: Participants in the intralingual subtitles condition will outperform those in the interlingual subtitles condition in terms of their level of video comprehension.

- Hypothesis 2: Participants in any of the subtitling conditions – intralingual and interlingual subtitles – will outperform participants in the control condition in terms of their performance of L2 vocabulary recognition.

- Hypothesis 2a: Participants in the intralingual subtitles condition will outperform those in the interlingual subtitles condition in terms of their performance of L2 vocabulary recognition.

- Hypothesis 3: Scores on the participants' L2 vocabulary recognition test will vary from test to pre-test and from post-test to test, considering any of the subtitling conditions and possible gain scores in the three test moments.

- Hypothesis 4: Participants' WM capacity will significantly interact with participants' performance in the video

comprehension test and L2 vocabulary recognition test in that participants with higher WM capacity will have better scores on the video comprehension test and the L2 vocabulary recognition test.

4.4 Participants

Originally, 84 students, Portuguese native speakers enrolled in level 5 in the Extracurricular[20] (non-credit) Language Courses at Universidade Federal de Santa Catarina (UFSC), Florianópolis-SC, Brazil, were invited to participate in the research, which corresponded to the entire student population with this level of proficiency enrolled in the Extracurricular (non-credit) Language Courses at the university at the time. Though 64 students accepted to participate in the study, throughout the data collection sessions (see section 4.8 for more information on the research design), 28 participants either quit the research or missed at least one of the sessions, whose data were then discarded.

In the end, the number of participants who completed all the stages in the data collection period and whose data were submitted to analysis was 36, in that 20 are female and 16 are male, all of which are Brazilian intermediate[21] EFL learners in the 18-60 age range (mean age of 22 years old), chosen on the basis of their proficiency level. It is worthwhile noting that studies on the effects of subtitled video materials on any language learning aspect have typically mostly comprised intermediate learners of the language, and an assumption behind that is that these participants are usually at a threshold proficiency level that enables them to read the subtitles on screen in the foreign language, which may not be possible with beginning learners since subtitles are displayed on screen for 2-4 seconds only and they might have not automatized reading and processing skills yet.

Participants belonged to six different Level 5 English classes (English 5B, 5C, 5E, 5F, 5G, 5H), each two classes taught by one teacher.

---

[20] The Extracurricular (non-credit) Language Courses at UFSC are open to undergraduate and graduate students enrolled at UFSC or any other higher education institution in the area, as well as faculty members, and members of the community. Students can enroll twice a year and can take a placement test to determine their proficiency level in any of the language courses offered: English, French, German, Italian, Japanese, Portuguese as a Second Language, and Spanish. In relation to English courses, language instructors consist of undergraduate and graduate students taking the Letras program, who are supervised by two coordinators.

[21] It is noteworthy to point out that proficiency level was not systematically assessed by the researcher, since it is assumed that all students enrolled in the same group were intermediate. Nonetheless, all of the students enrolled in the language courses take an in-house placement test.

Each two classes were randomly assigned to one of the treatment groups or the control group, namely: Intralingual Subtitles Group (English 5B and English 5H), with 12 participants; Interlingual Subtitles Group (English 5C and English 5E), with 12 participants; and the Control Group – no subtitles (English 5F and English 5G), with 12 participants. Participants were informed about all the stages in the data collection in the first meeting, when they were invited to participate in the research, signed the Consent Form[22] (see Appendix A), and had the opportunity to clarify doubts related to the goals and design of the study. However, participants were not told about specific details that could bias their answers during any of the stages in the data collection, though they were informed that they could ask more specific question once the data collection was over.

Information gathered on the participants' profile through the administration of a student profile questionnaire designed specifically for the present study (see section 4.7 for a complete description of the questionnaires used in this study and Appendix B to see the student profile questionnaire) revealed that, by the time of data collection, most of them (75%) had been studying English for at least three or four years, had been having contact with the English language mostly by attending EFL classes, as well as by listening to music and watching films and TV series. The questionnaire also revealed that most participants (55.6%) had never been to an English speaking country before and most of them informed that they had been studying English because they seek professional and personal development (97.22%).

As for their TV series viewing habits, the profile questionnaire revealed that 83.36% of the participants watch films or TV series at least twice or three times a week and, among the mostly watched TV series are *Friends*, *Games of Thrones* and *How I Met Your Mother* (69.25%). The majority of all participants (86.11%) reported watching TV series with interlingual subtitles, whereas a few of them (13.89%) reported watching TV series with intralingual subtitles. No participant reported using dubbing or no translational aid. The participants reported that TV series are sometimes used (50%) and frequently used (50%) in their English classes, especially to address cultural issues, introduce the topic of a unit or develop their listening comprehension skills (75%).

[22] The research project was submitted to the university's Ethics Committee and an approval was obtained. It is registered under the code 36597314.9.0000.0118. Participants' teachers also signed a consent form.

All of the participants who answered the question on whether they considered that TV series could be used to enhance their L2 knowledge reported to believe that TV series can be used, among other things, to learn new vocabulary, which represented 97.22% of the population's answers in the study. This particular piece of information help unveil participants' attitudes towards the use of subtitled video materials in and outside the classroom with a view to improving the development of their language skills.

4.5 Materials: On the TV Series

The participants watched a 20-minute episode of the American TV series *The Big Bang Theory*, which premiered in 2007. In Brazil, the show is broadcast with Portuguese subtitles on Warner channel and is a critically acclaimed show. The sitcom depicts Leonard Hofstadter and Sheldon Cooper, two brilliant physicists who are best friends and roommates. They are also friends with their co-workers Howard Wolowitz, a mechanical engineer, and Rajesh Koothrappali, an astrophysicist. The gang spends their time working on their individual work projects, playing video games, watching science-fiction movies, or reading comic books. As they are self-professed nerds, they have little or no luck with women. When Penny, a pretty woman and an aspiring actress that works as a waitress, moves into the apartment next to Sheldon and Leonard's, the latter has another aspiration in life, that is, to try and get Penny to be his girlfriend[23].

The Big Bang Theory was selected to be used in the present investigation based on a series of criteria. First of all, the series and the specific episode adopted in this research were used in Matielo, Collet & D'Ely (2013). Additionally, they were also chosen on the basis of its genre, since comedy was thought to be appealing and appropriate to the target audience. From the profile questionnaire administered in the very first session in the data collection, 31 out of the 36 participants reported watching sitcoms, thus suggesting their familiarity with the genre.

The episode used in the study was "The Grasshopper Experiment", the eighth episode on the first season. In this episode, Raj's parents back in New Delhi have set him up with a childhood acquaintance of his named Lalita Gupta, who has just moved to California, hoping to get them married. Raj, who cannot speak to or address any woman, does not want to meet with Lalita. Meanwhile, Penny, wanting to practice her

[23]Information retrieved from http://www.imdb.com/title/tt0898266.

bartending skills, uses the guys as guinea pigs, which opens up a whole new world for Raj: he now realizes he can speak to Lalita (or any other woman for that matter) when under the influence of alcohol. Lalita, however, ends up having a connection on the date, but with Sheldon, instead[24]. The episode was chosen since it contained a complete story line and did not require students to be familiar with the series or previous episodes.

4.6 Materials: On the Video Comprehension Test

In order to assess participants' general and specific comprehension of the video in terms of its content, three 'why' questions were posed and participants were instructed orally to answer them in Portuguese or in English and were assured that grammatical errors would be disregarded. Essentially, in order to successfully answer the general comprehension questions, participants were required to have some level of understanding of the story and how it unfolded throughout the episode. All of the events in the episode were highly intertwined, which meant that participants had to have some level of general comprehension as regards the connections among the goings-on in the episode to be able to respond to the questions. Concerning the specific details of the episode, the questions concentrated on requiring the participants to judge five statements about the story narrated on screen on whether they were true or false. The statements provided details about the characters, their relationships, likes and dislikes, and attitudes.

The video comprehension test was administered along with the L2 vocabulary recognition test (which is explained in the next subsection). The video comprehension test consisted of parts A – general comprehension, and B – specific comprehension. The video comprehension test and L2 vocabulary test were administered simultaneously in the interest of time so as to disrupt teachers' classroom routines as minimum as possible (see appendix D for a copy of the video comprehension test).

In fact, in addition to the interest in testing how intralingual subtitles and interlingual subtitles could assist learners' general and specific comprehension of the video material, the researcher decided to include a portion of video comprehension test because it could unveil how much of the story was understood, the participants' involvement with the task, and their seeming attitude towards it. It is noteworthy to stress that

[24]Information retrieved from http://www.imdb.com/title/tt1127389.

during all of the data collection sessions, the researcher attempted to observe whether participants would laugh at the story when jokes were made – which were constant, given its genre (comedy) – or whether they would remain in silence or show some facial expressions that could indicate confusion or difficulty. What the researcher observed was that regardless of the experimental group or control group, participants seemed to enjoy watching the TV series episode, laughed at the jokes indicating that they had understood them, and seemed to enjoy the task while engaged in it.

4.7 Materials: On The L2 Vocabulary Recognition Tests

The L2 vocabulary recognition tests contained three parts: a pre-test, a test, and a post-test. The pre-test was aimed at assessing the participants' previous knowledge of the target vocabulary. They received a 20-word list in English in which 10 of them were distractors. They were asked to write their meaning, a synonym or an explanation in Portuguese or English using their own words. The target words were chosen taking into account factors influencing word learnability (Laufer, 1997), addressed in the review of literature in section 3.1. For instance, some words were chosen because of their facilitated word learning aspect, such as familiar morphemes (e.g., *pointless* and *membership*), whereas other words were chosen due to their difficulty-inducing factors, such as the presence of foreign morpheme (e.g., *obnoxious*), and some were selected as neutral, such as those related to concreteness or abstractness of a lexical item (e.g. *wrath*)[25]. Moreover, words are related to the themes portrayed in the episode, but are not semantically related (Erten & Tekin, 2008).

Regarding the number of times the target-words appear in the selected episode, half of them was uttered and was shown in the subtitles – both intralingual and interlingual – twice (*slot, membership, guinea pigs, showdown,* and *wrath*), whereas the other half of the target-words was uttered or shown in the subtitles once (*embodiment, pointless, pushy, obnoxious,* and *resemblance*). The short exposure to the input is hereby acknowledged, although one has to have in mind that the video length is also short (20 minutes).

[25] All of the examples have been taken from the actual test, devised and used in this study.

Moreover, other important criteria[26] considered in the selection of the target-words relates to whether the words actually appeared in the interlingual and intralingual subtitles and whether they were somehow relevant to the story being narrated on screen, which could facilitate participants' processing and future recognition. Distractors, on the other hand, contained words that were likely to be familiar on the basis of the semantic familiarity (e.g., *affection* and *mint*) and unfamiliarity (e.g., *award* and *moisturizer*), taking into account their proficiency level. Another aspect that was not controlled for was word frequency. Additionally, the pre-test administration was not controlled for time in order to avoid imposing a burden on the participants, though most of them finished it under 10 minutes (see appendix C for a copy of the pre-test).

The L2 vocabulary recognition test required the participants to analyze 10 target-vocabulary word list from the pre-test in English (excluding the distractors) and write their meaning, a synonym or an explanation in Portuguese or English using their own words. The test administration was not controlled for time to avoid stress or making them nervous, though most participants finished it under 20 minutes, excluding the viewing time. See Appendix D for a copy of the test, which also contains the video comprehension test.

In the post-test, the participants were provided with a test identical to the test in order to check whether they were able to recognize the words they encountered when watching the video by the time the test was administered. The participants were asked to write their meaning, a synonym or an explanation in Portuguese or English using their own words. The post-test administration was not controlled for time not to cause stress on the participants, though most of the participants finished it under 10 minutes (see Appendix E for a copy of the post-test). As in the L2 vocabulary test, distractors were not included because the focus was on checking whether participants would still recognize the target-words they encountered in the video a week after viewing had occurred.

4.8 Materials: On the Working Memory Measures

This subsection addresses the two WM measures (Reading Span Test – RST and Operation-Word Span Test – OSPAN) used in this

[26] This researcher has worked at the Extracurricular Language Courses at Universidade Federal de Santa Catarina for seven years, and is therefore very familiar with the student population at beginning, intermediate, and advanced levels, including students enrolled in proficiency preparation courses. This knowledge of and familiarity with the population and their learning context have also indirectly influenced the selection of words.

investigation. It focuses on describing the rationale behind the tests and the methodological modifications made to their originals. It also describes in detail test administration procedures.

4.8.1 The Reading Span Test (RST)

The RST was originally designed by Daneman and Carpenter (1980) to explore the relationship between individual differences in WM capacity of native speakers of English and their comprehension of sentences in addition to the recall of the last words of a group of sentences. At the time, the underlying assumption with the RST related to the fact that WM capacity can correlate with individuals' ability to process and store information simultaneously (and only for a brief period of time) while performing another cognitively demanding task, which may or may not involve reading. According to the rationale behind the RST, the larger the participant's WM capacity, the better performance s/he can have on the RST and, in turn, the results can be attributed to predict outcomes in terms of reading comprehension/reading performance (Daneman & Carpenter, 1980).

As previously discussed in Chapter III, the psychometric correlational view or approach to WM tackles the correlation between individual differences in WM and the performance of cognitive abilities (Daneman & Carpenter, 1980, 1983). The present study follows this paradigm of research on WM in the sense that it hypothesizes that WM may be a good predictor of individual differences and their performance in the tests devised and used in this investigation.

Different versions of the RST are used all over the world, usually incorporating modifications to suit a given experimental population. In the present study, the Brazilian version of the RST adopted was a modification of the one used in Torres (2003), which was administered in Portuguese (participants' L1) to avoid confounds with participants' L2 proficiency level and to avoid floor effects (scoring too low) due to task difficulty. This study's version of the test comprised 42 unrelated sentences (see Appendix F for the RST version adopted in the present study), ranging from 13 to 17 words in length, presented in sets of 2, 3, 4, and 5 sentences.

Moreover, there were nine sentences in the practice session, presented in a single set of 2, 3 and 4 sentences so participants could get familiarized with the test and feel more comfortable before taking it. Each sentence in the testing session was displayed on a black computer screen, with a 20pts white Times New Roman font, arranged in 3 sets of 2

sentences, 3 sets of 3 sentences, 3 sets of 4 sentences, and 3 sets of 5 sentences. The end of each set was indicated by a black screen with white question marks on it, whose number indicated the number of words the participant was supposed to recall.

In order to illustrate the RST, an example of one set of three sentences is provided as follows (words in bold refer to words that should be recalled):

Set 1

1- Tornou-se cada vez mais comum ver adultos usando aparelhos nos **dentes**.
2- Em quase cinquenta anos de política, ACM acumulou fortuna e **poder**.
3- A crise de energia de falta só é não brasileira **chuvas**.
4- Cientistas analisam imagens das galáxias para traçar o mapa evolutivo do **cosmo**.
5- O maior símbolo da resistência uma foi colonização a contra africana **mulher**.
6- Ao chegar ao campo de batalha, os soldados da Grécia antiga bebiam **vinho**.

Each participant in the RST was given written instructions in Portuguese (see Appendix G for the instructions for the RST), though they were asked before practice and testing sessions if they had any questions concerning the test, therefore providing them with the chance to clarify any doubts that might have remained. The participant read each sentence out loud, in order to minimally ensure its processing, and then judged whether the sentence was grammatical or acceptable in terms of its syntax, an important methodological modification of the original test in tune with previous studies (Turner & Engle, 1989; Harrington & Sawyer, 1992; Budd, Whitney & Turley, 1995; Torres, 2003; Bailer, Tomitch & D'Ely, 2013). Fifty per cent (50%) of the sentences made sense syntactically and semantically (e.g. *Tornou-se cada vez mais comum ver adultos usando aparelhos nos dentes*), while the remaining fifty per cent (50%) were created by reversing the last 4 to 6 words immediately prior to the final word (e.g., *A crise de energia de falta só é não brasileira chuvas*).

The participant was supposed to try to memorize the last word of each sentence in the set and, when question marks were shown on screen, the participant had to try to recall the last words of all sentences in that set, respecting the exact order in which they appeared (Friedman &

Miyake, 2004), verbalizing them. Both practice and testing sessions of the RST were audiotaped using audio editor Gold Wave© 5.70 on the same computer used to administer the test, with a backup recording using an iPhone 4S Voice Memo©. The recordings were saved for later transcription and scoring since simultaneous note taking whilst the participant was taking the test could disrupt their concentration, therefore compromising data quality. The test was not controlled for time (differently from Torres, 2003, who adopted a six second time frame) because informal piloting revealed that participants felt too much pressure with the time frame. The researcher was in charge of pressing the key to go to the next sentence or slide only when he realized the participant was finished with that specific slide or part of the test.

Two important modifications in the test were made as a result of informal piloting prior to test administration for data collection purposes. This study's version of the test did not use sets containing 6 sentences given the fact that the five participants in the informal piloting could not remember the words when sets of six sentences were shown on screen, neither in the correct order nor out of order of presentation. Additionally, at the end of the test, these participants from the informal piloting reported finding the test too long and too tiresome. Another important modification was the color of the computer screen: instead of using a white color, this study's version of the test used a black color on the screen, since four out of the five participants in the informal piloting reported being extremely uncomfortable with a white screen during the entire practice and testing sessions and suggested a black screen instead. The font color was adapted to the screen color as well, therefore using white instead of black to suit the background of the slides.

4.8.2 The Operation-Word Span Test (OSPAN)

The OSPAN was originally designed by Turner and Engle (1989) to investigate the hypothesis that WM capacity is not language specific and can thus be generalized to any cognitively complex task, since it has been suggested to be a reliable measure of WM capacity (Conway et al., 2005). The test essentially consists of asking the participant to solve simple mathematical operations while trying to recall a set of unrelated words.

The Brazilian Portuguese version of the test adopted in the present research was designed by Prebianca (2009) based on Turner and Engle's (1989) test. The words in the test were disyllabic, unlikely to be unknown by native speakers of Portuguese in the age range under study (e.g. *papel*

and *tinta*). The OSPAN was also administered in Portuguese to avoid confounds with participants' L2 proficiency level. As in the RST, the OSPAN consisted of 42 operation strings along with Portuguese words (see Appendix H for the OSPAN version used in the present study), written in white and placed right in the middle of a computer black screen. From the 42 trials, 19 strings presented correct operations whereas 23 strings displayed incorrect operations.

The test was organized in 3 test blocks of 4 sets each, but instead of displaying sets of two to five words/operations, their display was randomized in order to prevent the participant from trying to predict the number of words s/he would have to recall. Block 1 had 4 sets of 3, 5, 3 and 2 trials, respectively; block 2 had 4 sets of 5, 4, 2 and 4 trials; and block 3 had 4 sets of 3, 5, 2 and 4 trials. Furthermore, there were 14 trials in the practice session, presented in a set of 2, 3, 4 and 5 operation strings and words, so participants could get familiarized with the test and feel more comfortable before taking it.

Each participant was given written instructions in Portuguese (see Appendix I for the instructions on the OSPAN), though they were asked before practice and testing sessions if they had any questions concerning the test, therefore providing them with the chance to clarify any doubts that might have remained. The participant read each of the mathematical operation out loud, in order to minimally ensure its processing, and then judged the result as correct or incorrect, and then read the word next to the operation out loud as well, trying to keep this word in mind for later recall. The participant was supposed to try to memorize the last word of each operation/word string in the set and, when question marks were shown on screen, the participant had to try to recall the words, respecting the exact order in which they appeared, verbalizing them. An example of a set of three operations and words to be recalled is provided as follows:

Mathematical Operations	**Words to be recalled**
Block 1	
(10 ÷ 2) – 3 = 2 ?	carta
(10 ÷ 10) – 1 = 2 ?	lençol
(7 ÷ 1) + 2 = 7 ?	terra

As in the RST, both practice and testing sessions of the RST were audiotaped using audio editor Gold Wave© 5.70 on the same computer used to administer the test, with a backup recording using an iPhone 4S Voice Memo©. The recordings were saved for later transcription and scoring because simultaneous note taking whilst the participant was

taking the test could disrupt their concentration, therefore compromising data quality. Additionally, the OSPAN was not controlled for time and the researcher was in charge of pressing the key to go to the next operation/word string or slide only when he realized the participant was finished with that specific slide or part of the test.

It is important to highlight that each participant was tested on both RST and OSPAN individually, outside their regular language classroom, in order to ensure maximum concentration and silence in the room where data collection was taking place. However, due to an issue of schedule availability on the part of the participants, the researcher decided to collect both tests in the same session, but they were controlled for order effect (Robinson & Gilabert, 2007), which means that half of the participants took RST first and then OSPAN and the other half took OSPAN first and then RST.

4.9 Questionnaires

Profile questionnaires[27] were administered in order to gather additional information on the participants[28], including general questions on their English learning history, their goals in studying the language, and whether they had already lived in an English speaking country before. Participants were also asked specific questions about their use of subtitled TV series, including questions on the frequency with which they watch them, which ones they usually watch, the kind of translational aid they often use (dubbing, intralingual subtitles, interlingual subtitles or none), and their perceptions on the use of TV series in their English language classes. Participants were also asked an open-ended question in the end with a view to eliciting their views on whether subtitled TV series could have any kind of impact upon their English language learning within or outside the classroom.

The teachers were also given a questionnaire in Portuguese in order to provide more information about their education and experience with English language teaching. They were also inquired about their pedagogical use of videos in the classroom. Data gathered from the teachers' profile questionnaire revealed that one of them holds a PhD in English and Applied Linguistics by *Programa de Pós-Graduação em Inglês* at UFSC and had been teaching English for eight years at the

[27] A Participants' profile questionnaire can be seen in Appendix B and teachers' profile questionnaire can be viewed in Appendix J.
[28] The data from the profile questionnaire are presented in section 4.4.

Extracurricular courses, whereas the other two are 8th semester undergraduate students enrolled in the Letras-Inglês Program at UFSC and had been teaching English for a shorter period of time, around three years now. Teachers' responses to the questionnaire and informal conversations with them also revealed that the three teachers are very concerned about their students' linguistic development and well-being, and perceive teaching as a challenging, though rewarding professional activity.

Concerning the teachers' pedagogical views and practices related to the use of subtitled video materials in their classes, the responses provided did not differ substantially, since all of them mentioned that they usually use videos to explore culture issues, alter classroom routines and dynamics, introduce or close topic or transition into a new theme or content. Teachers also mentioned that the decision to use intralingual subtitles or interlingual subtitles (or none) depends on the type of activity and their goals. Nevertheless, their responses converged to the more frequent use of interlingual subtitles with beginners, intralingual subtitles with intermediate and upper-intermediate students, and no subtitles with advanced leaners of English. The justification was mostly based on the teachers' own assessment of their students' proficiency level, the types of activities used, and the students' response/reaction to them.

In relation to the video comprehension and L2 vocabulary recognition tests, a retrospective questionnaire with four questions was administered right after the post-test in order to gather more information on the students' perceptions of the test as a whole. Students were asked whether they considered the tests difficult, whether they felt they kept any of the words in mind after any of the experiment sessions, whether they checked (looked up or asked someone about) any of the words they encountered in the experiment sessions and finally whether they believed TV series can be a valuable resource to learn L2 vocabulary. In the last question, participants were also asked to report again on their subtitling preferences, bearing in mind that they had already had the opportunity to answer this question in their profile questionnaire administered in the very first session of the experiment.

Finally, a WM retrospective questionnaire was administered in order to obtain more data on the participants' perceptions of both RST and OSPAN tests (see Appendix K for a copy of the WM retrospective questionnaire). Participants were asked whether they considered the WM tests difficult, whether they consider to have a good memory, whether they feel they can attend to more than one thing simultaneously on their

everyday life, and whether they felt tired or bothered by anything during the WM testing session[29].

4.10 Data Collection Procedures and Research Design

Data collection took place during the regularly scheduled class periods in the regular classrooms with a view to minimizing disruption of the cooperating teachers' classroom schedules. In class, the researcher collected the data over three sessions: in the first session, a term of consent, the student profile questionnaire, and the pre-test were administered. In this session, the participants also provided their e-mail and telephone number so the researcher could contact each of them in order to schedule their WM tests.

In the second session, prior to video watching, the participants were provided with some background information regarding the TV series, such as the plot summary and a quick introduction to the characters for story contextualization purposes using a PowerPoint© presentation that took a total of 2 minutes, in Portuguese to avoid misunderstandings. Then, each group of participants were shown the video, which lasted for 20 minutes, under the different experimental conditions (intralingual subtitles, interlingual subtitles, and control group – no subtitles) in separate classrooms during their regularly scheduled classes. The researcher was present in all of the classes, one at a time. In this very same session, the video comprehension test and L2 vocabulary recognition test were administered. The sheet of paper containing the test itself was only handed out after participants had finished watching the video to avoid influencing their responses.

In the third and last session, the post-test was administered, as well as a retrospective questionnaire. There was an exact a one-week gap between pre-test, test, and post-test administration. WM tests were carried out throughout three weeks, but they only started on the second week of the data collection because the first week was necessary to organize the schedules for WM data collection. After WM tests, participants were specifically instructed not to tell classmates about their own WM tests. During the three-meeting data collection, the researcher allowed other

[29] Responses to the retrospective questionnaires of the video comprehension test, L2 vocabulary recognition test, and WM tests will be addressed when results to these tests are properly tackled, in the next Chapter.

students in the class who decided not to participate in the research to take part in the activities, but their data were discarded from the study[30].

The video was shown with the use of a standard DVD player with a digital image display projector and external speakers in a relatively small classroom with a large screen, with regular viewing and listening conditions. No time frame was established in any treatment conditions or data collection sessions in the classroom in order to ensure that the participants felt at ease to answer the questions. It is important to highlight that the participants were informed in advance that their participation in the research was voluntary and that their performance on the tests was unrelated to their course grades.

One final aspect concerning the data collection procedures merits mentioning. Due to the researcher's and students' conflicting schedules, ten WM tests were collected by two undergraduate research assistants. It is very important to stress that both research assistants were properly instructed on how to collect both RST and OSPAN tests and the tests were counterbalanced in terms of administration (50% of the participants received the RST first whereas the other 50% received the OSPAN first).

Table 2 outlines the data collection schedule and research design:

Table 2

Research Design

1st Classroom Meeting (Week 1)	2nd Classroom Meeting (Week 2)	3rd Classroom Meeting (Week 3)	Individual Meeting (Weeks 2-4)
Consent Form + Pre-Test + WM Tests' Scheduling + Profile Questionnaire	TV Series Contextualization + Video + Test	Post-Test + Retrospective Questionnaire	RST and OSPAN tests + Retrospective Questionnaire

4.11 Tests' Scoring and Statistical Procedures

The video comprehension test, the L2 vocabulary recognition tests, and the WM capacity tests received different scoring treatments. In the

[30] These participants who took the tests during their regular class periods, but did not wish to participate in the actual research were instructed to contact the researcher later if they desired to obtain feedback on their performance.

next subsections of this Chapter, I focus on presenting the criteria, procedures, methodological decisions made to score the tests adopted in the present investigation, starting off with the score of the video comprehension test, followed by the L2 vocabulary recognition tests, and finally the WM tests. Explanation on the correlations run are also provided in the end.

4.11.1 Scoring of the Video Comprehension Test (Parts A and B)

The video comprehension test, the L2 vocabulary recognition tests, and the WM capacity tests received different scoring treatments. Regarding the video comprehension test and the L2 vocabulary recognition tests, three different raters with similar educational and professional backgrounds, chosen on the basis of relevance to this study, were invited to assist in this process. Rater A is a 28-year-old experienced English teacher who holds an MA in English and Applied Linguistics and who is currently pursuing a PhD in the field of SLA. Rater B is also a 28-year-old experienced English teacher who holds an MA in English and Applied Linguistics and who is currently pursuing a PhD in the field of SLA. Rater C is a 33-year-old PhD student in the field of Translation Studies who has also worked as a language instructor. Raters accepted to rate tests in their entirety based on a copy of the pre-test, test, and post-test without the participants' identification. Raters watched the TV episode used in the present study before scoring any of the tests and individual conversations between the researcher and the raters about the TV series and the episode were held whenever necessary.

In relation to the video comprehension test, the raters were instructed to disregard any grammatical errors in the participants' responses (some of them decided to answer the general comprehension questions in English and had a few grammatical mistakes in spelling or conjugation). The raters were instructed to assign one point to each of the three questions in Part A in case they felt that the answer provided by the participant was acceptable. In case the answer was not acceptable in the raters' perception, no points were assigned to that given question. In relation to Part B, the specific comprehension part, the same procedure was followed. In both Parts A and B of the video comprehension test, there was total agreement among the raters as to the responses provided by the participants, that is, there was no discrepancy in the ratings. Hence, no Cronbach's Alpha correlation tests were run in order to check for inter-rater reliability.

Regarding the video comprehension test – general and specific parts, a Shapiro-Wilk's test ($p<.05$) and a visual inspection of the histograms, Q-Q plots, and box plots showed that the data were not approximately normally distributed, with varying skewness and kurtosis (see Appendix L for the tests of normality and Appendixes M and N for the histograms and box plots).

4.11.2 Scoring of the L2 Vocabulary Recognition Tests

The L2 vocabulary recognition tests were scored strictly. In the pre-test, each rater credited the participant with 1 point for each acceptable synonym, explanation or translation of the target-words. Distractors were excluded from the raters' duty of scoring for they are not included in the analyses. In the test, raters assigned 1 point for each acceptable synonym, explanation or translation of the target-words, as in the pre-test. Finally, in the post-test, raters also assigned 1 point for each acceptable synonym, explanation or translation of the target-words, as they did in the pre-test and in test.

Cronbach's Alpha correlation tests were run in order to check for inter-rater reliability in terms of the rating of each of the L2 vocabulary tests (pre-test, test, and post-test) for each of the groups (intralingual subtitles, interlingual subtitles, and control), separately because raters did not agree with the responses provided. As Larson-Hall (2010) explains, in general, the acceptable level of Cronbach's alpha varies from 0.70 to 0.80. In this study, regarding the L2 vocabulary recognition pre-test, the results obtained in inter-rater reliability tests revealed that the rating was very or highly reliable (intralingual subtitles group, $\alpha= 0.85$; interlingual subtitles group, $\alpha= 0.75$; control, $\alpha= 0.98$) (see Tables 5, 6, and 7 in Appendix O). In relation to the L2 vocabulary recognition test, the results obtained in the inter-rater reliability tests revealed that the rating was highly reliable (intralingual subtitles group, $\alpha= 0.92$; interlingual subtitles group, $\alpha= 0.95$; control, $\alpha= 0.99$) (see Tables 8, 9, and 10 in Appendix P). Finally, in relation to the L2 vocabulary recognition post-test, the results obtained in the inter-rater reliability tests also revealed that the rating was very or highly reliable (intralingual subtitles group, $\alpha= 0.88$; interlingual subtitles group, $\alpha= 0.97$; control, $\alpha= 0.98$) (see Tables 11, 12, and 13 in Appendix Q).

Thus, as far as L2 vocabulary recognition tests are concerned, since the results given by the outcome raters were considered very or highly reliable, the outcome rates of the three raters generated the mean scores

for each participants in the pre-test, test, and post- test, for each treatment and control condition.

Regarding the L2 vocabulary recognition tests – pre-test, test, and post-test, a Shapiro-Wilk's test ($p<.05$) and a visual inspection of the histograms, Q-Q plots, and box plots showed that the data were not approximately normally distributed, with varying skewness and kurtosis (see Appendix R for the normality tests and Appendix S for the histograms and box plots).

4.11.3 Scoring of the Working Memory Measures

In relation to WM measures, both RST and OSPAN were scored strictly and leniently. For the strict scoring of RST test, participants' reading span was calculated at the level at which s/he was accurate on at least two trials of a given set of sentences (Daneman & Carpenter, 1980). Also, in accordance with Turner and Engle (1989), an approximately 85% accuracy rate was required in terms of participants' judgment of sentence grammaticality or syntax acceptability to ensure the processing component of the task, which represented 36 out of 42 sentences[31]. A half point was given when the participant passed one trial at a certain level. For instance, a participant recalled correctly all the words in the right order in the three sets of 2 sentences and in the three sets of 3 sentences recalled correctly just one trial, this participant received half point, being her/his span 2,5, which is where scoring would then terminate.

Participants' answers on RST test were also scored leniently. In this method of scoring, participants were given points for any set for which s/he recalled all sentence final words, irrespective of the order of recall, as long as s/he recalled all of the words belonging to two of the trials in the given set, and a half point was given when the participant passed one trial at a certain level, which is where scoring would then terminate. The approximately 85% accuracy rate was also kept in the lenient scoring of RST test as well to maintain the processing component of the task.

Similarly, participants' answers on OSPAN test were scored strictly and leniently, following distinct procedures. In relation to the strict score of the test, 1 point was credited to each word recalled in the exact order of presentation, that is, in a test set of three trials, for instance, a participant who was able to solve at least two math operations correctly

[31] All of the 36 participants in the present study judged at least or more than 36 sentences correctly in terms of grammaticality and syntax acceptability.

and then recalled their corresponding words obeying the order of presentation was given 2 points. In this scoring procedure, both operation solving and word recall were taken into account.

The participants' responses to OSPAN test were also scored leniently in that they are not as strict in terms of the participants' processing efficiency. In line with previous studies (Turner & Engle, 1989; Kane, Beckley, Conway & Engle, 2001; Kane, Conway, Hambrick & Engle, 2007; Prebianca, 2009, among others), a criterion of 85% accuracy on correctly solving all of the mathematical operations of the entire test was required since it is useful in ensuring that participants do not trade-off between processing the mathematical operations of the test and storing the words that come along with them (Unsworth, Heitz, Schrock & Engle, 2005). In this sense, all words recalled in the exact order of presentation and which obeyed the criterion of 85% accuracy were credited 1 point, which means that each participant could have gotten up to 6 wrong mathematical operations (out of a total of 42 trials). Thus, if the participant responded to a specific operation of a particular set incorrectly but was able to accurately recall the word following that operation, s/he was credited 1 point as long as s/he had not reached six errors yet.

Concerning the statistical procedures employed for WM measures, a Shapiro-Wilk's test ($p<.05$) did reveal that some of the variables were approximately normally distributed, namely: RST strict for the interlingual group, RST strict for the control group, RST lenient for the interlingual group, OSPAN strict for the intralingual, Interlingual and control groups, and OSPAN lenient for the intralingual, interlingual, and control groups. The other WM variables – RST for the intralingual group and RST lenient for the intralingual and control groups were not approximately normally distributed (see Appendix T for the normality tests and Appendix U for the histograms and box plots).

Hence, in light of normality tests run on the data, the goals, the RQs, and the hypotheses of this research, the following non-parametric statistical procedures were employed in the study:

- In order to answer (RQ1): How do intralingual and interlingual subtitles affect learners' video comprehension, as measured by immediate post-viewing test? and (RQ2): How do intralingual and interlingual subtitles affect learners' L2 vocabulary recognition, as measured by pre-test, test, and post-test?, Kruskal-Wallis One-Way ANOVA tests were run;

- In order to address (RQ3): How does learners' L2 vocabulary recognition, in case there is any, differ across time for any experimental condition?, Friedman's Two-Way Analysis of Variance was run;

- In order to focus on (RQ4): Does leaners' WM capacity correlate with their performance on the video comprehension test and/or their performance on the L2 vocabulary recognition test in any subtitling condition?, a series of individual Spearman's Rank Order Correlation Coefficient tests were run.

This Chapter entailed the methodological reasoning upon which this study relies. It is hoped that the methodological choices made can contribute to the organic nature, soundness, and academic rigor that was intended with this research. Next, Chapter V will present the main results and discuss the main findings of this study.

CHAPTER V
RESULTS

5.1 Chapter Outline

This Chapter reports on the results of the statistical tests run and the analyses performed with a view to addressing each of the RQs and hypotheses presented in the Method in light of the objectives set out in the beginning of this dissertation. To recap, the present study has sought to examine the effects of intralingual and interlingual subtitles on Brazilian intermediate EFL learners' L2 development by assessing their performance on a video comprehension test, containing general and specific comprehension parts, and L2 vocabulary recognition tests. Furthermore, this study also sought to verify whether learners' level of video comprehension and/or L2 vocabulary recognition correlates with their WM capacity, as measured by the RST and OSPAN tests.

To begin with, subsection 5.2 focuses on the results of the video comprehension test – both general and specific comprehension parts. Subsection 5.3 turns to the results obtained on the L2 vocabulary recognition tests – pre-test, test, and post-test. After that, subsection 5.4 lays out the results of the WM tests. Following these subsections, subsection 5.5 presents the results obtained with the correlation tests run between the RST and OSPAN tests, measures of WM capacity. Subsection 5.6 provides the results of correlation tests run between the video comprehension (general and specific parts) and WM (RST and OSPAN). Similarly, subsection 5.7 also explores the results of correlation tests, though this time between the L2 vocabulary recognition and WM tests. Finally, at the end of this Chapter, in subsection 5.8, a summary of the hypotheses and major results is provided.

5.2 Video Comprehension Test

This subsection of the Chapter focuses on the results of the statistical tests run and analyses performed in order to address RQ1 and hypotheses 1 and 1a (see Method for more information on the RQs and hypotheses). Subsection 5.2.1 will specifically entail the descriptive statistics and the results obtained on the general comprehension portion of the video comprehension test, whereas subsection 5.2.2 will present the results obtained on the specific comprehension portion of the video comprehension test.

5.2.1 General Comprehension: Descriptive Statistics and Results

As regards the video comprehension test, participants were assessed in terms of their general and specific comprehension of the video. Table 16 displays the results obtained on the general comprehension part of the test, which consisted of three open-ended questions to assess participants' broader understanding of the video content and storyline (see Appendix V for the participants' general comprehension raw scores).

Table 16

General comprehension results

Group		Statistic	Std. Error
Intralingual Subtitles (n=12)	Mean	2.67	.142
	SD	.492	
	Minimum	2	
	Maximum	3	
	Skewness	-.812	.637
	Kurtosis	-1.650	1.232
Interlingual Subtitles (n=12)	Mean	2.25	.250
	SD	.866	
	Minimum	1	
	Maximum	3	
	Skewness	-.567	.637
	Kurtosis	-1.446	1.232
Control (n=12)	Mean	1.75	.250
	SD	.866	
	Minimum	1	
	Maximum	3	
	Skewness	-.567	.637
	Kurtosis	-1.446	1.232

Note. n = sample size; Std. Error = standard error; SD = standard deviation; Minimum number of points to be obtained in the test: 0; Maximum number of points to be obtained in the test: 3.

As indicated in Table 16, participants' mean scores in the general comprehension part of the test as regards the video content were found to be different considering the two experimental conditions and the control condition. Participants in the intralingual subtitles condition (M = 2.67) outperformed those in the interlingual subtitles condition (M = 2.25), who

in turn outperformed those in the control condition (M = 1.75). It is interesting to notice that 2 points was the minimum score obtained by the intralingual subtitles group, whereas 1 point was the minimum score obtained by the interlingual subtitles and control groups. In all groups there were participants scoring the maximum number of points in this portion of the test, though no participant scored 0 points.

By inspecting the results more closely, both raw scores and box plots (see Appendix M for the box plots and appendix V for the raw scores), one notices the emergence of a few interesting figures that deserve further scrutiny. In relation to the intralingual subtitles groups, 25% of the participants scored 2 points, which corresponds to the lowest score obtained by this group on the test, whereas 75% of the same group had the highest and maximum score on the test. As for the interlingual subtitles group, 50% of the group had the highest and maximum score on the test, whereas 25% of the participants scored 2 points, and 25% of the participants in this group had the lowest score on the test, that is, 1 point. Finally, regarding the control group, the figures change considerably, since only 25% of the participants in this group had the highest and maximum score on the test, while 25% of them scored 2 points, and 50% of the participants had the lowest score on the test, that is, 1 point.

The numbers provided in Table 16 signal differences in terms of the participants' ability to attend to and successfully understand the gist of the video they had watched. Furthermore, differences in the mean scores obtained by the groups were found, in spite of the fact that they are more visually salient when comparing experimental to control groups, but not the experimental groups themselves (intralingual and interlingual subtitles).

As previously explained in the Method, in order to explore possible differences in terms of the performance of the three groups on the video comprehension test (general and specific parts), Kruskal-Wallis One-Way ANOVA tests were run. To start off, the results of the Kruskal-Wallis One-Way ANOVA tests run on the results of the general comprehension portion of the test are presented in Table 17:

Table 17

Kruskal-Wallis one-way ANOVA on general comprehension

	GCScores
Chi-Square	6.806
df	2
Asymp. Sig.	.033

Note. GCScores = General comprehension scores; df = degrees of freedom; Asymp. Sig. = asymptotic significance.

The results obtained with the Kruskal-Wallis One-Way ANOVA test run on the general comprehension part of the test indicated that the groups are statistically different from one another and that there was a significant effect of availability of subtitles on the participants' performance on the general comprehension of the video ($H(2) = 6.806$, $p = .033$). An effect size[32] of 19% was found in the data, which indicates the percentage of the variability in the general comprehension scores that was accounted for by the availability of subtitles.

Even though the Kruskal-Wallis One-Way ANOVA test does inform that the three groups are statistically different, it does not inform where the difference is. Further separate Kruskal-Wallis One-Way ANOVA tests[33] were therefore run as post-hoc tests between two groups at a time. This allows the researcher to spot where the statistically significant differences among them lie. The results obtained by the post-hoc tests are shown in Tables 18, 19, and 20:

[32] In this study, an effect size was calculated by taking the chi-square value divided by n-1 (Larsen-Hall, 2010).

[33] The same results can be obtained by running Mann-Whitney U Tests, given that the data are not approximately normally distributed. The researcher ran further separate Kruskal-Wallis One-Way ANOVA tests for the sake of convenience given that SPSS does offer its own post-hoc under the same name.

Table 18

Kruskal-Wallis one-way ANOVA on general comprehension: intralingual and interlingual subtitles groups

n	12
Chi-Square	1.392
df	1
Asymp. Sig.	.238

Note. n = sample size; df = degrees of freedom; Asymp. Sig. = asymptotic significance.

Table 19

Kruskal-Wallis one-way ANOVA on general comprehension: interlingual subtitles and control groups

n	12
Chi-Square	1.917
df	1
Asymp. Sig.	.166

Note. n = sample size; df = degrees of freedom; Asymp. Sig. = asymptotic significance.

Table 20

Kruskal-Wallis one-way ANOVA on general comprehension: intralingual subtitles and control groups

n	12
Chi-Square	6.799
df	1
Asymp. Sig.	.009

Note. n = sample size; df = degrees of freedom; Asymp. Sig. = asymptotic significance.

By running three separate Kruskal-Wallis One-Way ANOVA tests on the general comprehension part of the test, it was made possible to accurately determine where the statistical differences among the groups are. As shown in Table 18, the differences between the experimental groups – intralingual and interlingual subtitles – are not statistically

significant (H(2) = 1.392, $p > .05$), though the former obtained higher mean scores than the latter. Similarly, as shown in Table 19, the differences between the interlingual subtitles group and the control group are not statistically significant either (H(2) = 1.917, $p > .05$). However, when observing the results of the statistical test run between the intralingual subtitles and the control groups displayed in Table 20, it is possible to state that this is where the statistically significant difference is (H(2) = 6.799, $p = .009$). Thus, the Kruskal-Wallis One-Way ANOVA test confirms that these are the two groups that are statistically different – intralingual subtitles and control groups – and reveals that the availability of intralingual subtitles impacted positively on the participants' level of general comprehension of the video as compared to the control group.

5.2.2 Specific Comprehension: Descriptive Statistics and Results

Participants were also assessed on their specific comprehension level of the video in terms of their ability to attend to and understand specific details about the video content and the storyline. This portion of the video comprehension test contained five statements of a true or false nature, as explained in the Method. Participants' performance on the specific comprehension test is displayed in Table 21 (see Appendix V for the participants' specific comprehension raw scores):

Table 21

Specific comprehension results

Group		Statistic	Std. Error
Intralingual Subtitles (n=12)	Mean	4.67	.188
	SD	.651	
	Minimum	3	
	Maximum	5	
	Skewness	-1.930	.637
	Kurtosis	3.165	1.232
Interlingual Subtitles (n=12)	Mean	4.33	.256
	SD	.888	
	Minimum	3	
	Maximum	5	
	Skewness	-.797	.637
	Kurtosis	-1.269	1.232
	Mean	3.83	.241

Control (n=12)	SD	.835	
	Minimum	3	
	Maximum	5	
	Skewness	.354	.637
	Kurtosis	-1.447	1.232

Note. n = sample size; Std. Error = standard error; SD = standard deviation; Minimum number of points to be obtained in the test: 0; Maximum number of points to be obtained in the test: 5.

As shown in Table 21, participants' mean scores on the specific comprehension part of the test about the video are dissimilar considering the two experimental conditions and the control condition of the experiment. Participants in the intralingual subtitles condition (M = 4.67) outperformed those in the interlingual subtitles condition (M = 4.33), who in turn outperformed those in the control condition (M = 3.83). In all three groups, 3 and 5 were the minimum and maximum scores obtained on the test, respectively.

In analyzing the results, raw scores, and box plots (see Appendix N for the box plots and appendix W for the raw scores), a few interesting numbers surface. In relation to the intralingual subtitles groups, 75% of the participants had the highest and maximum score on the test, whereas 16.67% of the participants scored 4 points, and 8.33%[34] of them had the lowest score, that is, 3 points. As far as the interlingual subtitles group are concerned, 58.33% of the participants had the highest and maximum score on the test, whilst 16.67% scored 4 points, and finally 25% had the lowest score on the test in this group, that is, 3 points. With regards to the control group, the numbers also differ when comparing to the experimental groups', since only 25% of the group had the highest and maximum score on the test, that is, 5 points, 33.34% of them scored 4 points, and 41.65% of them had the lowest score on the test, that is, 3 points. All groups presented participants scoring the maximum number of points in this portion of the test, though no participant scored 0 points.

Now, because there is observable variance concerning the performance of the three groups as regards the specific comprehension on the video comprehension test, a Kruskal-Wallis One-Way ANOVA test

[34] Though SPSS considers this to be an outlier (8.33% = 1 participant, Participant 12), the researcher decided not to discard this participant from the sample because the tests results reported in Tables 21, 22, and 23 were not significantly impacted and this outlier was not masking the existence of any other outlier on the data set. Also, the researcher decided to keep this participant in the sample pool to avoid a higher rate of participant mortality (see subsection 4.4 in the Method).

with the three groups together was performed. This statistical procedure allows the researcher to inspect whether the groups are indeed statistically different. The results obtained by the statistical test are displayed in Table 22:

Table 22

Kruskal-Wallis one-way ANOVA on specific comprehension

	SCScores
Chi-Square	6.113
df	2
Asymp. Sig.	.047

Note. SCScores = Specific comprehension scores; df = degrees of freedom; Asymp. Sig. = asymptotic significance.

The results from the Kruskal-Wallis One-Way ANOVA test run on the specific comprehension part of the test indicated that the three groups are statistically different and the test revealed that there was a significant effect of availability of subtitles on the participants' performance on the specific comprehension of the video ($H(2) = 6.113$, $p = .047$). An effect size of 19% was also found in the data, which indicates the percentage of the variability in the specific comprehension scores that was accounted for by the availability of subtitles, as it did on the general comprehension portion of the test.

Further separate Kruskal-Wallis One-Way ANOVA tests were then run between two groups at a time to precisely determine which groups are statistically different from one another. The results obtained by the tests are shown in Tables 23, 24, and 25:

Table 23

Kruskal-Wallis one-way ANOVA on specific comprehension: intralingual and interlingual subtitles groups

	SCScores
Chi-Square	.939
df	1
Asymp. Sig.	.332

Note. SCScores = Specific comprehension scores; df = degrees of freedom; Asymp. Sig. = asymptotic significance.

Table 24

Kruskal-Wallis one-way ANOVA on specific comprehension: interlingual subtitles and control groups

	SCScores
Chi-Square	2.012
df	1
Asymp. Sig.	.156

Note. SCScores = Specific comprehension scores; df = degrees of freedom; Asymp. Sig. = asymptotic significance.

Table 25

Kruskal-Wallis one-way ANOVA on specific comprehension: intralingual subtitles and control groups

	SCScores
Chi-Square	5.998
df	1
Asymp. Sig.	.014

Note. SCScores = Specific comprehension scores; df = degrees of freedom; Asymp. Sig. = asymptotic significance.

Taken together, the results obtained with the three separate Kruskal-Wallis One-Way ANOVA tests on the specific comprehension part of the test allowed to determine where the statistically significant differences among the groups are. As specified in Table 23, the differences between the experimental groups – intralingual and interlingual subtitles – are not statistically significant ($H(2) = .939$, $p > .05$), though the former obtained higher mean scores than the latter. Likewise, as shown in Table 24, the differences between the interlingual subtitles group and the control group are not statistically significant either ($H(2) = 2.012$, $p > .05$). However, when observing the results of the statistical test between the intralingual subtitles and the control groups in Table 25, statistically significant differences between these two groups are confirmed ($H(2) = 5.998$, $p = .014$).

Thus, because the Kruskal-Wallis One-Way ANOVA test does inform that these two groups are statistically different – intralingual subtitles and control – such a difference can be arguably explained by the presence of intralingual subtitles, which positively impacted on the

intralingual subtitles group' level of specific comprehension of the video in comparison to the control group.

In light of the statistical results reported in subsection 5.2, hypothesis 1 – on whether *participants in any of the subtitling conditions would outperform participants in the control condition in terms of their level of video comprehension* – is thus partially supported. The performance of only one of the experimental groups (intralingual subtitles) was statistically different in relation to the other experimental group (interlingual subtitles) and the control group.

Hypothesis 1a – on whether *participants in the intralingual subtitles condition would outperform those in the interlingual subtitles condition in terms of their level of video comprehension* – is not supported. Even though the results demonstrated that participants' performance in the intralingual subtitles condition in terms of general and specific comprehension of the video material was better than those in the interlingual subtitles condition, the two experimental groups were not statistically different from one another.

5.3 L2 Vocabulary Recognition Tests: Descriptive Statistics and Results

This subsection reports on the results of the statistical tests run and analyses performed to refer to RQs 2 and 3, as well as hypotheses 2, 2a, and 3 (see Method for more information on the RQs and hypotheses). The results obtained by the two experimental groups and the control group on the L2 vocabulary recognition pre-test, test, and post-test are displayed in Table 26 (see Appendix X for the L2 vocabulary recognition raw scores and Appendix S for the histograms and box plots):

Table 26

L2 vocabulary tests results

Groups	Statistics	Pre-Test Score	Test Score	Post-Test Score
Intralingual Subtitles (n=12)	Mean	.58	.94	1.02
	SD	.621	.826	.846
	Min.	.00	.00	.00
	Max	1.67	2.67	2.67

Interlingual Subtitles (n=12)	Mean	.47	.66	.63
	SD	.593	.898	.926
	Min.	2.00	3.00	3.00
	Max	1.11	1.16	1.41
Control (n=12)	Mean	1.11	1.16	1.41
	SD	1.25	1.34	1.29
	Min.	.00	.00	.00
	Max	3.67	3.67	4.00

Note. n = sample size; SD = standard deviation; minimum and maximum obtainable scores: 0-10 points.

According to Table 26, the performance of the three groups on the L2 vocabulary recognition tests points out to different mean scores on the pre-test, test, and post-test. The pre-test, test, and post-test scores allowed for a minimum of 0 and a maximum of 10 points each (see subsection 4.11.2 of the Method for the scoring procedures of the L2 vocabulary recognition tests), and a quick look at the participants' scores reveals that the scores obtained by the groups are different.

Looking at the pre-test scores alone, it is possible to state that the three groups are slightly different from start: while the performance of the experimental groups did not differ much apparently (M = .58 and M = .47, for the intralingual and interlingual subtitles groups, respectively), the performance of the control group was found to be better (M = 1.11). As for the test scores, it is possible to perceive that the control group (M = 1.16) outperformed both experimental groups, though a better performance by the intralingual subtitles group (M = .94) over the interlingual subtitles groups (M = .66) was also found. Finally, in relation to the post-test scores, a similar trend is observed in that the control group (M = 1.41) outperformed the intralingual subtitles group (M = 1.02), who in turn outperformed the interlingual subtitles group (M = .63).

Examining the minimum and maximum scores obtained on the L2 vocabulary pre-test, test, and post-test by the two experimental groups and the control group provides an idea of the test difficulty. The minimum score obtained on the tests was 0 points by the intralingual subtitles group and the control group (on the pre-test, test, and post-test), whereas the highest score obtained on the tests was 4 points, which was obtained by the control group on the post-test.

A careful inspection of the histograms and box plots was informative since it revealed the presence of outliers (See Appendix S for the histograms and box plots). On the pre-test, Participant 22 (interlingual subtitles group) was considered an outlier, obtaining a score of 2 points

when the mean score of the participant's group is .47. Moreover, the same participant can be considered an outlier on the post-test, when s/he obtained a score of 3 points in a group whose mean score is .63. Lastly, on the post-test as well, Participant 30 (control group) was considered an outlier, scoring 4 points in a group whose mean score is 1.41.

Nevertheless, when the time came to compare the groups to check whether they differed statistically, the results obtained by the statistical tests revealed that the groups were not statistically significant among themselves, with or without the aforementioned outliers. The removal of outliers confirmed that they were not masking the existence of other outliers in the data. Thus, the researcher decided to keep Participants 22 and 30 in the pool, also bearing in mind the high participant mortality rate previously reported in the Method.

The next step was to verify whether the apparent differences in performance by the three groups on the three testing moments – pre-test, test, and post-test – were statistically significant. To this end, a Kruskal-Wallis One-Way ANOVA test was run, and the results obtained with the statistical test are shown in Table 27:

Table 27

Kruskal-Wallis one-way ANOVA on L2 vocabulary recognition tests

	Pre-Test	Test	Post-Test
Chi-Square	1.310	.916	3.830
df	2	2	2
Asymp. Sig.	.520	.633	.147

Note. df = degrees of freedom; Asymp. Sig. = asymptotic significance.

Despite the differences in mean scores shown in Table 26 obtained by the groups throughout the three testing moments – pre-test, test, and post-test – a Kruskal-Wallis One-Way ANOVA test showed that the differences between the groups did not reach statistical significance and therefore did not reveal a significant effect of subtitle availability on L2 vocabulary recognition ($H(2) = .1.310$, $p > .05$ for the pre-test; $H(2) = .916$, $p > .05$ for the test; $H(2) = 3.830$, $p > .05$ for the post-test). Therefore, no post-hoc tests were run. Moreover, a small effect size (ranging from 2% on the test to 10% on the post-test) was found in the data, which refers to the percentage of the variability in the L2 vocabulary recognition tests that suggests that availability of subtitles did not seem

to play a determining role in the participants' performance on the L2 vocabulary recognition tests.

In order to gather insights into the variation of the participants' performance on the L2 vocabulary recognition across time, three separate Friedman's Two-Way ANOVA tests were run for each of the two treatment groups and the control group. The tests were aimed at verifying whether the groups' L2 vocabulary recognition performance statistically differed across time considering their pre-test, test, and post-test moments. The results obtained are reported in Tables 28, 29, and 30:

Table 28

Friedman's Two-Way ANOVA Test on L2 vocabulary recognition: intralingual subtitles group

n	12
Chi-Square	3.588
df	2
Asymp. Sig.	.166

Note. n = sample size; df = degrees of freedom; Asymp. Sig. = asymptotic significance.

Table 29

Friedman's Two-Way ANOVA Test on L2 vocabulary recognition: interlingual subtitles group

n	12
Chi-Square	.963
df	2
Asymp. Sig.	.618

Note. n = sample size; df = degrees of freedom; Asymp. Sig. = asymptotic significance.

Table 30

Friedman's Two-Way ANOVA Test on L2 vocabulary recognition: control group

n	12
Chi-Square	4.455
df	2
Asymp. Sig.	.108

Note. n = sample size; df = degrees of freedom; Asymp. Sig. = asymptotic significance.

Altogether, the results herein obtained with the separate Friedman's Two-Way ANOVA tests on the L2 vocabulary recognition tests (pre-test, test, and post-test) revealed no statistically significant difference across time for the intralingual subtitles group ($\chi^2(3) = 3.588$, $p > .05$), interlingual subtitles group ($\chi^2(3) = .963$, $p > .05$), and control group ($\chi^2(3) = 4.455$, $p > .05$). This means that the treatment – the availability of subtitles – did not have a statistically significant effect on the sample investigated in this experiment, that is, a statistically significant change on the L2 vocabulary recognition test scores across time.

Based on the statistical results reported in subsection 5.3, hypothesis 2 – on whether *participants in any of the subtitling conditions – intralingual and interlingual subtitles – would outperform participants in the control condition in terms of their performance of L2 vocabulary recognition* – is not supported, given that participants in the control group obtained higher mean scores in the L2 vocabulary recognition pre-test, test, and post-test, in spite of the lack of statistical significance. As for the hypothesis 2a – on whether *participants in the intralingual subtitles condition would outperform those in the interlingual subtitles condition in terms of their performance of L2 vocabulary recognition* – the results obtained by the statistical tests do not provide enough evidence to support it either.

Even though Friedman's Two-Way ANOVA tests revealed no statistically significant differences concerning pre-test, test, and post-test scores on the L2 vocabulary recognition tests, the researcher decided to further inspect gain scores in order to gather more information about the participants' performance. In computing the variance in gain scores, it is possible to generate valid data on which group benefitted the most from the treatment, even if statistical significance was not achieved. The results of gain scores comparisons from test to pre-test and post-test to test are presented in Table 31:

Table 31

Gains scores in L2 vocabulary recognition

	Participant	Test to Pre-Test Gain Scores	Post-Test to Test Gain Scores
Intralingual Subtitles (n=12)	P1	1.00	.00
	P2	.00	.00
	P3	1.00	.00
	P4	.00	-.33
	P5	.00	.67
	P6	.67	.00
	P7	-.33	.00
	P8	-.33	.33
	P9	.67	-.67
	P10	1.00	.00
	P11	.33	1.33
	P12	.33	-.33
Interlingual Subtitles (n=12)	P13	.00	-.33
	P14	.33	.33
	P15	.00	.33
	P16	.00	.00
	P17	.00	.00
	P18	-.33	.00
	P19	.00	.00
	P20	-.33	-.33
	P21	1.00	.00
	P22	1.00	.00
	P23	.67	-.33
	P24	.00	.00
Control (n=12)	P25	.00	.00
	P26	.00	.00
	P27	.00	.00
	P28	-.67	.00
	P29	.00	.67
	P30	.00	.33
	P31	.33	-.33
	P32	.00	.00
	P33	.00	1.00
	P34	1.00	.33
	P35	.00	.00
	P36	.00	1.00

Note. n = sample size.

Examining the data displayed in Table 31 allows for the detection of positive, negative, and neutral gain scores considering the L2 vocabulary recognition pre-test, test, and post-test for the experimental conditions – intralingual and interlingual subtitles – and the control condition. The numbers indicate that, as regards the test/pre-test comparison, 7 participants exhibited positive gain scores in the intralingual subtitles group, whereas 4 participants had positive gain scores in the interlingual subtitles group, and only 2 participants showed positive gain scores in the control group. Despite the positive gain scores, a few participants obtained negative gain scores, as in the case of 2 participants in the intralingual and interlingual subtitles condition and 1 in the control. Furthermore, 3 participants in the intralingual subtitles condition, 6 participants in the interlingual subtitles condition, and 9 participants in the control condition obtained no gain scores in the test/pre-test gain score comparison.

Turning to the numbers in the post-test/test comparison, it is possible to observe that 3 participants demonstrated positive gain scores in the intralingual subtitles group, while 2 participants had positive gain scores in the interlingual subtitles group, and 5 participants showed positive gain scores in the control group. Likewise, regardless of the positive gain scores obtained, some participants did obtain negative gain scores, as in the case of 3 participants in the intralingual and interlingual subtitles condition and 1 in the control. Additionally, 6 participants in the intralingual subtitles condition, 7 participants in the interlingual subtitles condition, and 6 participants in the control condition obtained no gain scores in the post-test/test gain score comparison.

All in all, the variance in gain scores in terms of test/pre-test and post-test/test comparisons do not provide enough statistical evidence to support hypothesis 3 – on whether *scores on the participants' L2 vocabulary recognition test would vary from test to pre-test and from post-test to test, considering any of the subtitling conditions and possible gain scores in the three test moments*. The gain score variance across time and the implications it bears will be discussed in the next Chapter.

5.4 Working Memory Tests: Descriptive Statistics and Results

This subsection reports on the descriptive statistics and results obtained on the WM tests used in the present investigation, namely the Reading Span Test (RST) and the Operation-Word Span Test (OSPAN), according to the scoring criteria previously described in the Method.

Firstly, the results on the RST will be presented in subsection 5.4.1, followed by the results on the OSPAN in subsection 5.4.2.

5.4.1 Scores on the Reading Span Test (RST)

As previously clarified in the Method, participants' performance on the WM tests were scored strictly and leniently. Table 32 shows the raw scores on the RST, following both strict and lenient scoring procedures, as well the mean scores, standard deviation, minimum score obtained on the test, and maximum score obtained on the test (Appendix U for the histograms and box plots):

Table 32

RST scores

	Participant	RST Strict Score	RST Lenient Score
Intralingual Subtitles (n=12)	P1	2	2
	P2	2	2
	P3	2	2
	P4	2	2
	P5	2.5	2.5
	P6	3	3
	P7	2	2.5
	P8	2	2
	P9	0.5	0.5
	P10	3	3
	P11	2.5	2.5
	P12	2.5	2.5
	Mean	2.16	2.20
	SD	.651	.655
	Minimum	0.5	0.5
	Maximum	3	3
Interlingual Subtitles (n=12)	**P13**	**3.5**	**3.5**
	P14	**4.5**	**4.5**
	P15	2.5	2.5
	P16	**5**	**5**
	P17	2.5	3
	P18	**3.5**	**3.5**
	P19	3	3
	P20	2	2
	P21	2.5	2.5

	P22	2.5	2.5
	P23	0.5	0.5
	P24	2.5	2.5
	Mean	2.87	2.91
	SD	1.17	1.16
	Minimum	0.5	0.5
	Maximum	5	5
Control (n=12)	P25	0	2
	P26	0.5	0.5
	P27	2	2
	P28	2	2
	P29	0	0
	P30	0.5	0.5
	P31	2.5	2.5
	P32	2.5	2.5
	P33	2	2
	P34	2	2.5
	P35	0.5	0.5
	P36	3	3
	Mean	1.45	1.66
	SD	1.07	1.00
	Minimum	0	0
	Maximum	3	3

Note: n = sample size; SD = standard deviation; higher spans are highlighted in bold.

Overall, the scores on the RST indicate that approximately 53% of the participants either scored 2.5 or 3.0 (considering both scoring procedures), whereas around 17% of the participants scored 0 or 0.5. Interestingly, only one participant scored 4.5 and only one participant scored 5.0 as well, considering the strict scoring procedure, for the sake of illustration. Because there does not seem to be an agreement in the field as to what in fact constitutes a higher or lower span participant, this researcher sides with Tomitch (2003a) and thus takes any score equal to or above 3.5 to mean higher spans. Hence, the results in the RST, when looked at from the strict scoring perspective, reveal 4 higher spans (around 11% of the sample) and 32 lower spans (89% of the sample).

All of the higher spans in this study (Participants 13, 14, 16, and 18) happen to belong to the interlingual subtitles group, a fact that was not controlled for since the researcher could not have anticipated the number of higher and lower spans to be found in each group to then (re)distribute participants across experimental and control groups by using the research design adopted, since data collection for the video tasks

were collected in the participants' regularly scheduled classes. In order to balance the existence of higher spans across groups, data collection on WM tests would have had to be collected first, before the administration of the classroom tests. Groups would have had to be organized only after taking the WM tests, but that would mean that the video comprehension test and the L2 vocabulary recognition tests would have had to be collected in several different sessions, thus not in the participants' regularly scheduled classes.

Following the strict and lenient scoring procedures, the interlingual subtitles group (M = 2.87 and M = 2.91, respectively) has higher mean scores on the test than the participants in the intralingual subtitles group (M = 2.16 and M = 2.20, respectively), who, in turn, has higher mean scores on the test than the participants in the control group (M = 1.45 and M = 1.66, respectively).

It is noteworthy to mention that an inspection of the histograms and box plots for the RST results revealed the presence of two outliers (Participants 9 and 23). SPSS considered them as outliers for deviating significantly in relation to the other participants within their group, which reflects the variability in the measurement. These outliers were not masking the existence of other outliers in the data. Their data have not been discarded due to a high participant mortality rate and the lack of statistical interference with the other tests performed.

5.4.2 Scores on the Operation-Word Span Test (OSPAN)

Participants' performance on the OSPAN test was scored strictly and leniently as well. Table 33 displays the raw scores on the OSPAN, following both strict and lenient scoring procedures, as well as the mean scores, standard deviation, minimum score obtained on the test, and maximum score obtained on the test (Appendix U for the histograms and box plots):

Table 33

OSPAN scores

	Participant	OSPAN Strict Score	OSPAN Lenient Score
Intralingual Subtitles (n=12)	P1	10	22
	P2	25	28
	P3	19	20
	P4	08	28

	P5	14	27
	P6	32	36
	P7	28	33
	P8	21	33
	P9	11	29
	P10	26	35
	P11	14	25
	P12	21	33
	Mean	19.08	29.08
	SD	7.739	5.089
	Minimum	8	20
	Maximum	32	36
Interlingual Subtitles (n=12)	**P13**	**42**	**42**
	P14	**42**	**42**
	P15	28	28
	P16	27	37
	P17	10	20
	P18	29	32
	P19	**41**	**41**
	P20	27	29
	P21	34	34
	P22	32	35
	P23	36	36
	P24	26	27
	Mean	31.16	33.58
	SD	8.98	6.734
	Minimum	10	20
	Maximum	42	42
Control (n=12)	P25	17	30
	P26	13	23
	P27	21	30
	P28	05	24
	P29	12	14
	P30	19	23
	P31	36	37
	P32	21	32
	P33	22	27
	P34	19	32
	P35	26	31
	P36	26	29
	Mean	19.75	27.66
	SD	7.875	5.974
	Minimum	5	14
	Maximum	36	37

Note: n = sample size; SD = standard deviation; higher spans are highlighted in bold.

Scores on the OSPAN test also amount to a similar picture to the one portrayed by the RST results. While more than 50% of the participants scored more than half of the possible points, that is, more than 21 points, 9% of the participants scored 10 points or less, a score that represents almost 25% of the possible points on the test, and only 5.5% of the participants scored 42 points (maximum score possible on the test), all of them following the strict scoring procedure (see section 4.11.2 in the Method for the scoring procedures employed on the OSPAN test).

Due to the fact that there does not seem to be an agreement in the field as to what in fact constitutes a higher or lower span participant on the OSPAN test, this researcher sides with Prebianca (2009) and thus takes any score equal to or above 38 to mean higher spans and any score equal to or below 37 to mean lower spans. Hence, the results on the OSPAN test, when looked at from the strict scoring perspective, reveal only 3 higher spans (around 8%) and 33 lower spans (around 92%).

When comparing the RST and OSPAN scores, two of the higher span participants on the RST are the same ones who were classified as higher span on the OSPAN test (Participants 13 and 14), following the strict scoring procedure. The other two higher span participants on the RST (Participants 16 and 18) were not classified as such on the OSPAN, but it should be noted that Participants 16 got 37 points on the OSPAN test when scored through the lenient scoring procedure, a score that would be extremely close to the cut-off point of classification as a higher span. As for Participant 18, their performance on the OSPAN test was not as good as on the RST, though this participant got 32 points on the OSPAN when considering the lenient scoring procedure, a score that is not far from the cut-off point of classification as a higher span.

Again, all of the three higher spans on the OSPAN happen to belong to the interlingual subtitles group. As previously explained, the possible emergence of higher spans within the same group was not controlled for since the researcher did not collect the WM tests prior to the administration of the other tests.

5.5 RST and OSPAN Correlations

Traditionally, WM tests, such as the RST and OSPAN tests, have been adopted in experimental research because they have been considered to deliver reliable measures of WM capacity (Conway, Kane, Bunting, Hambrick, Wilhelm & Engle, 2005). Since they are thought to provide

solid measures of the cognitive capacity to store information for a brief period of time while manipulating it, their relationship with the performance on other cognitive tasks – such as reading, watching a subtitled film or learning another language, for instance – seems to be a fruitful one to be investigated.

In this sense, in order to have some perspective into the correlative nature of these two tests' results, the researcher decided to investigate the relationship of the results obtained with the RST and the OSPAN tests in an attempt to check the extent to which the scores obtained on these tests by the population hereby investigated correlate. This would also provide, to some degree, some indicative evidence that both tests could measure the same construct.

Table 34 shows the results obtained with the correlation tests performed between the scores on both RST and OSPAN tests. As explained in the Method, in subsection 4.11.3, some of the data related to the WM measures were approximately normally distributed and some of the data were not. Consequently, Pearson's correlations were run among the variables that are approximately normally distributed, whereas Spearman's correlations were performed when at least one variable is not approximately normally distributed:

Table 34

RST and OSPAN Correlations

	Correlation Test	RST Strict x OSPAN Strict	RST Lenient x OSPAN Lenient
Intralingual Subtitles (n=12)	Spearman's Coefficient r_s	.467	.516
	p value	.126	.086
	Correlation Test	RST Strict x OSPAN Strict	RST Lenient x OSPAN Lenient
Interlingual Subtitles (n=12)	Pearson's Coefficient r	.149	.308
	p value	.644	.330
	Correlation Test	RST Strict x OSPAN Strict	RST Lenient x OSPAN Lenient
Control (n=12)	Pearson's Coefficient r	.406	.705*
	p value	.190	.010

Note. n = sample size; RST = Reading Span test; OSPAN = Operation-Word Span Test; p = significance level. *Correlation is significant at the .05 level.

Spearman's Rank Order Correlation Coefficient tests and Pearson Product-Moment Correlation Coefficient tests revealed a positive, though not always statistically significant correlation between the answers to both RST and OSPAN tests at $\alpha = .05$. The lowest correlation found was between RST and OSPAN in the strict scoring procedure for the interlingual subtitles group ($r = .149$, $n = 12$, $p > .05$), whereas the highest correlation found was between RST and OSPAN in the lenient scoring procedure for the control group ($r = .705$, $n = 12$, $p = .01$), which is the only significant correlation in the tests reported in Table 34.

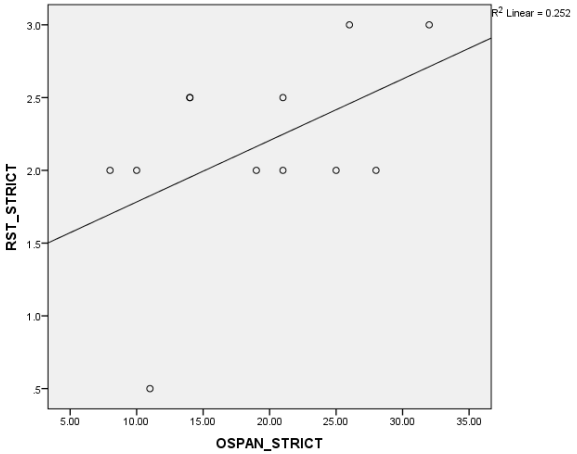
An extremely important observation about these numbers is in order. In this study, the researcher ran correlation tests among the variables for each group separately and the results revealed that, except for the RST Lenient and OSPAN Lenient correlation for the control group, all of the other correlations are not statistically significant. However, Bailer (2011) reported significant correlations between the same variables in her study, which dealt with a different population. The only aspect that needs to be carefully observed is that the correlation tests run in that study comprised all of the participants from experimental and control groups altogether, thus representing a different methodological decision from the one taken here[35].

In order to graphically visualize the correlations reported in Table 33, scatterplots representative of those tests are now presented. Graphs 1 and 2 present scatterplots of the correlation tests run with the intralingual subtitles group; Graphs 3 and 4 present scatterplots of the correlation tests performed with the interlingual subtitles group; finally, Graphs 5 and 6 display scatterplots of the correlations tests run with the control group:

[35] The researcher ran correlation tests among the WM variables with all participants together and indeed obtained significant correlations for all tests (ranging from r = .525, n = 36, p = .001 to r = .952, n = 36, p = .000).

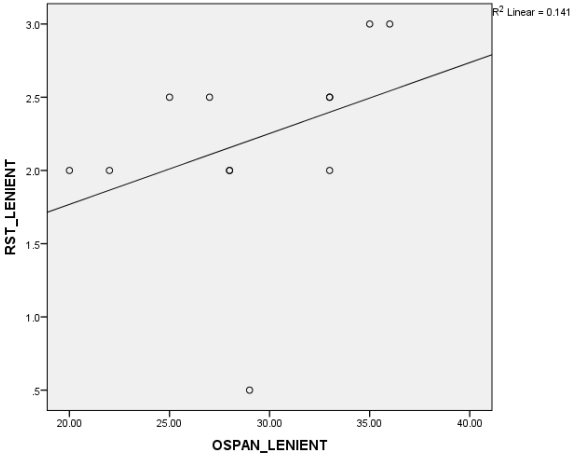
Graph 1

Scatterplot – RST strict and OSPAN strict correlations for the intralingual subtitles group



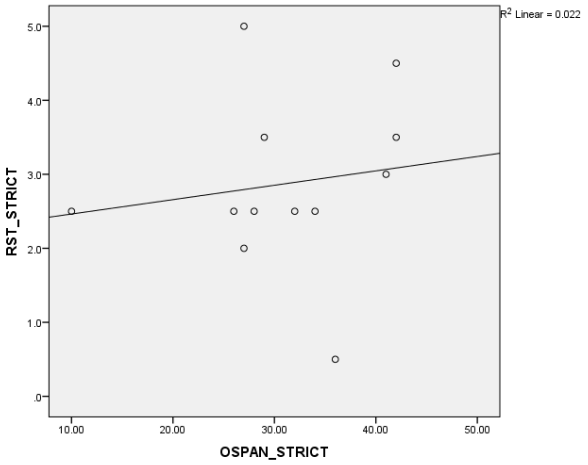
Graph 2

Scatterplot – RST lenient and OSPAN lenient correlations for the intralingual subtitles group



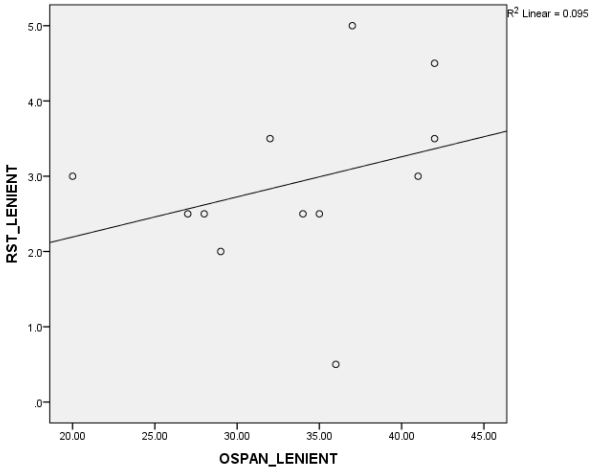
Graph 3

Scatterplot – RST strict and OSPAN strict correlations for the interlingual subtitles group



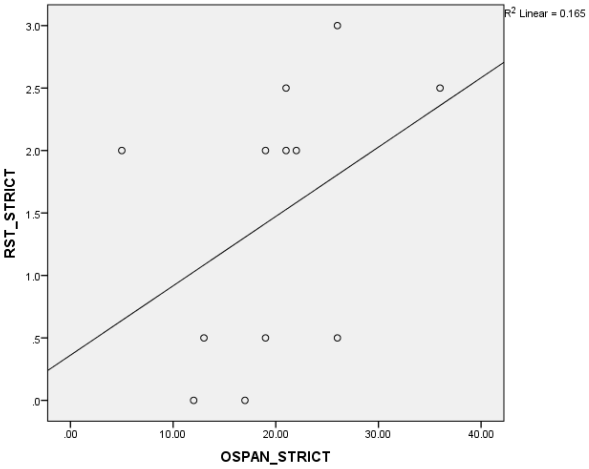
Graph 4

Scatterplot – RST lenient and OSPAN lenient correlations for the interlingual subtitles group



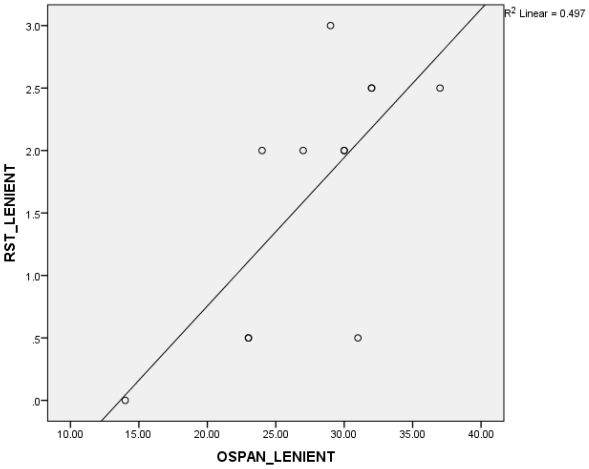
Graph 5

Scatterplot – RST strict and OSPAN strict correlations for the control group



Graph 6

Scatterplot – RST lenient and OSPAN lenient correlations for the control group



The scatterplots contained in Graphs 1-6 all indicate that there definitely is a trend in the distribution of correlative scores from lower/mid left towards upper right, which is suggestive of a positive

correlation between the tests, even though the statistical tests did not reveal statistical significance in any but one of the correlations. The linear trend lines in the scatterplots also reveal a positive linear relationship between the WM tests, even though it signals a weak relationship between them, especially when observing the dispersion of scores from the line of best fit in the scatterplots, except for the correlation between RST lenient and OSPAN lenient results for the control group, which, in this case, was found to be indicative of correlating positively and strongly.

Finally, it is noteworthy to mention that the R^2 linear values obtained[36] explain little of the covariance in the results between the tests. In other words, these values stand for a relatively small percentage of the influence in change of one variable over the other. The smallest R^2 value obtained was between the RST strict and OSPAN strict correlation for the interlingual subtitles group ($R^2 = 0.02$), whereas the greatest R^2 value obtained was between the RST lenient and OSPAN lenient statistically significant correlation for the control group ($R^2 = 0.49$). A discussion of the WM tests results as well as the correlation tests will be presented in the next Chapter. The next subsection focuses on the correlation tests performed between the video comprehension test (general and specific parts) and the WM measures.

5.6 Video Comprehension Test and Working Memory Correlations

This subsection of the Chapter provides details about the results of the statistical tests run and analyses performed in order to address RQ4 and hypothesis 4 (see Method for more information on the RQs and hypotheses). Subsection 5.6.1 will specifically entail the descriptive statistics and the results obtained on the correlation tests between the general comprehension portion of the video comprehension test and the WM tests, whereas subsection 5.6.2 will present the results obtained on the correlation tests between the specific comprehension part of the video comprehension test and the WM tests.

[36] It should be noted that these values are automatically calculated by SPSS v. 20 when designing a correlation scatterplot, regardless if the variables follow normal distribution or not.

5.6.1 General Comprehension and Working Memory Correlations

In order to check whether and the extent to which the performance of the participants on the general comprehension part of the video comprehension test held a significant relationship with their performance on the WM tests, correlation tests were run. As previously mentioned in the Method, on the general comprehension part of the test, the results obtained presented data that are not approximately normally distributed (see Appendix L for the tests of normality and Appendixes M and N for the histograms and box plots). Thus, Spearman's Rank Order Correlation Coefficient tests were performed, and the results are presented in Table 35:

Table 35

General Comprehension and WM Correlations

	Correlation Test	GCScore x RST Strict	GCScore x RST Lenient	GCScore x OSPAN Strict	GCScore x OSPAN Lenient
Intralingual Subtitles (n=12)	Spearman's Coefficient r_s	.000	.054	.231	-.336
	p value	1.000	.867	.470	.286
	Correlation Test	GCScore x RST Strict	GCScore x RST Lenient	GCScore x OSPAN Strict	GCScore x OSPAN Lenient
Interlingual Subtitles (n=12)	Spearman's Coefficient r_s	.556	.517	.148	.302
	p value	.061	.085	.645	.340
	Correlation Test	GCScore x RST Strict	GCScore x RST Lenient	GCScore x OSPAN Strict	GCScore x OSPAN Lenient
Control (n=12)	Spearman's Coefficient r_s	.006	.247	-.177	.297
	p value	.986	.439	.581	.348

Note. n = sample size; RST = Reading Span test; OSPAN = Operation-Word Span Test; p = significance level; GCScore = general comprehension score.

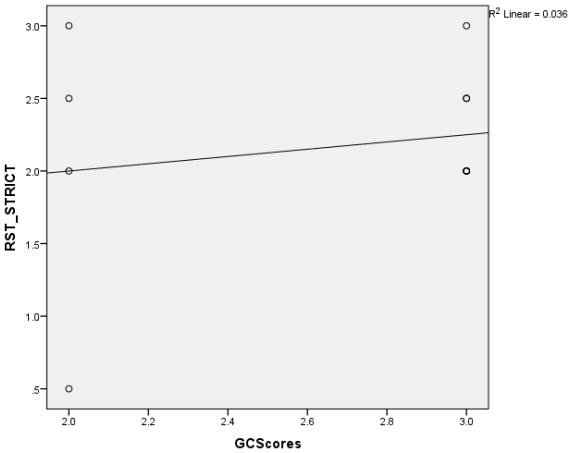
Overall, Spearman's Rank Order Correlation Coefficient tests revealed mostly a positive, though not statistically significant correlation between the participants' performance on the general comprehension portion of the video comprehension test and the WM tests at $\alpha = .05$. The

correlation tests also revealed two instances of negative correlations, such as in the OSPAN and the general comprehension scores test with the intralingual subtitles group (lenient scoring) and the control group (strict scoring). Nevertheless, these correlations are not statistically significant. The results also indicated a total absence of correlation between the general comprehension scores and the OSPAN Strict variables with the intralingual subtitles groups ($r_s = .000$, $n = 12$, $p > .05$).

In order to graphically visualize the correlations reported in Table 35, scatterplots of those tests are presented as follows. Graphs 7, 8, 9, and 10 show scatterplots of the correlation tests run with the intralingual subtitles group; Graphs 11, 12, 13, and 14 present scatterplots of the correlation tests performed with the interlingual subtitles group; lastly, Graphs 15, 16, 17, and 18 display scatterplots of the correlations tests run with the control group:

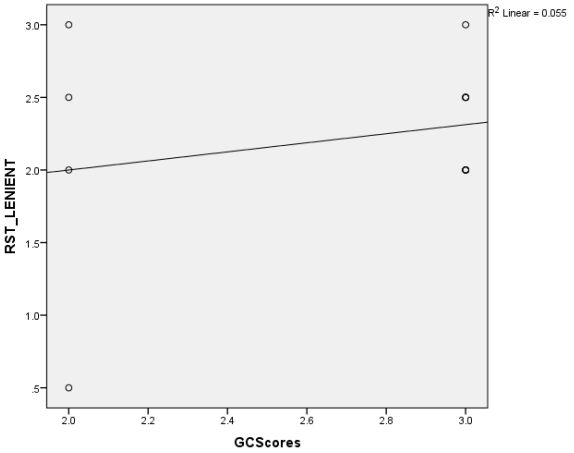
Graph 7

Scatterplot – General Comprehension Scores and RST strict correlations for the intralingual subtitles group



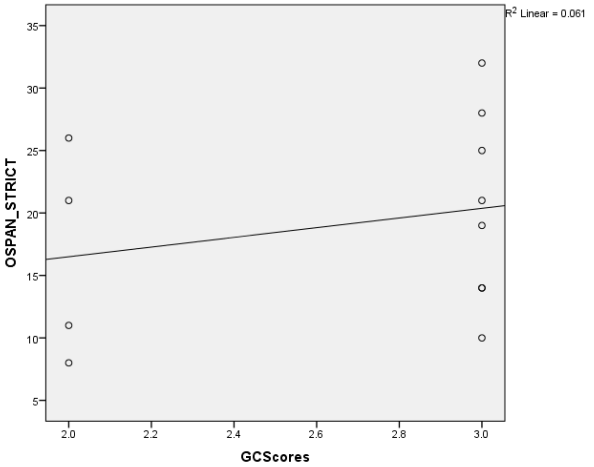
Graph 8

Scatterplot – General Comprehension Scores and RST lenient correlations for the intralingual subtitles group



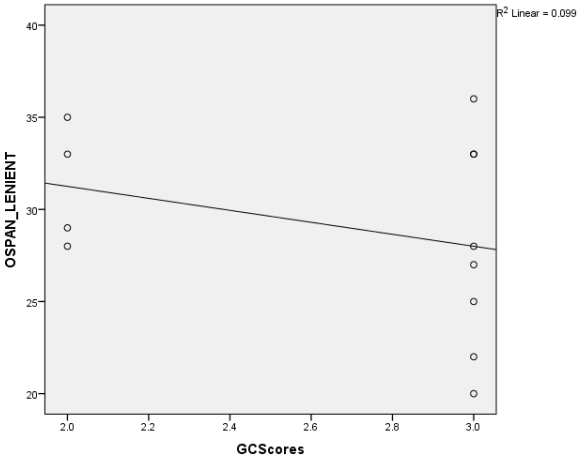
Graph 9

Scatterplot – General Comprehension Scores and OSPAN strict correlations for the intralingual subtitles group



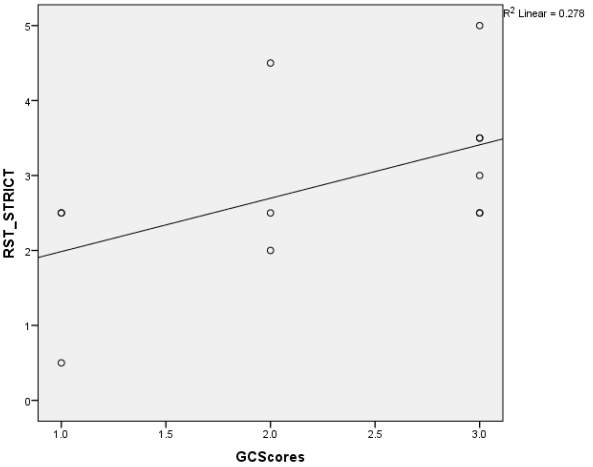
Graph 10

Scatterplot – General Comprehension Scores and OSPAN lenient correlations for the intralingual subtitles group



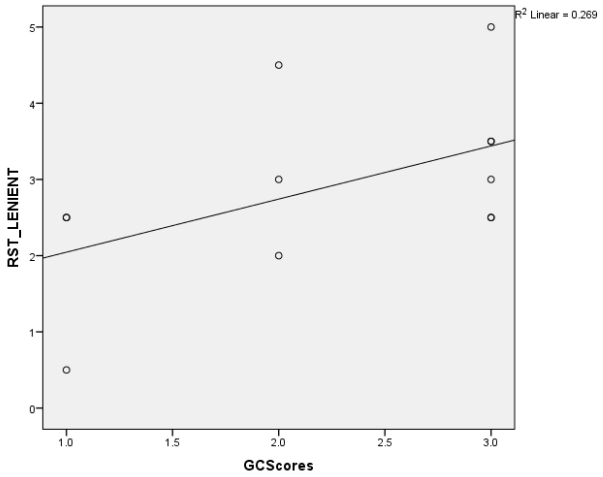
Graph 11

Scatterplot – General Comprehension Scores and RST strict correlations for the interlingual subtitles group



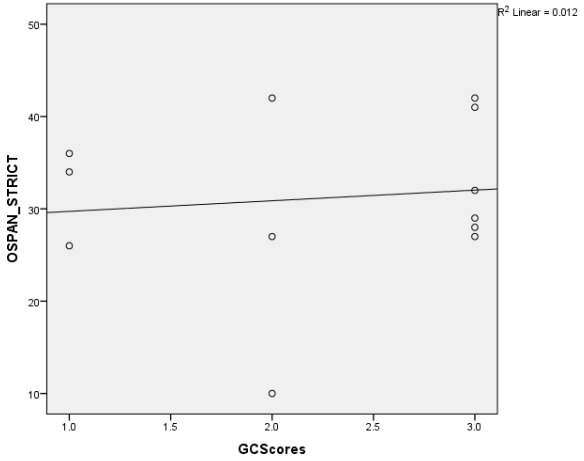
Graph 12

Scatterplot – General Comprehension Scores and RST lenient correlations for the interlingual subtitles group



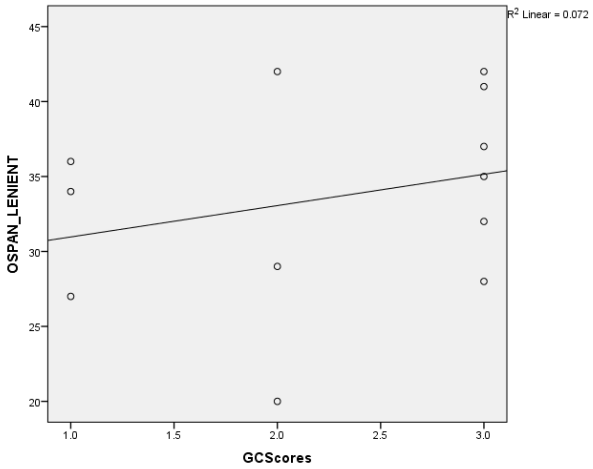
Graph 13

Scatterplot – General Comprehension Scores and OSPAN strict correlations for the interlingual subtitles group



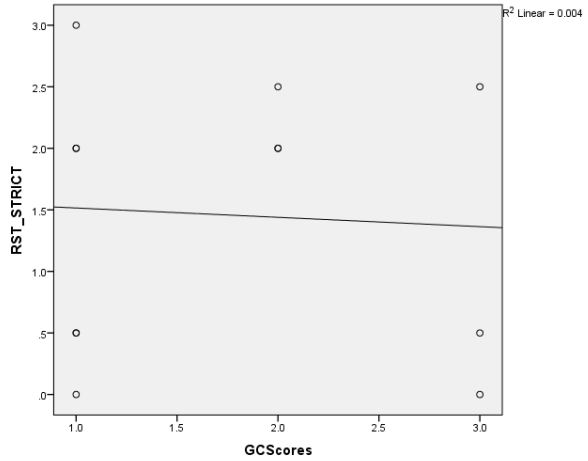
Graph 14

Scatterplot – General Comprehension Scores and OSPAN lenient correlations for the interlingual subtitles group



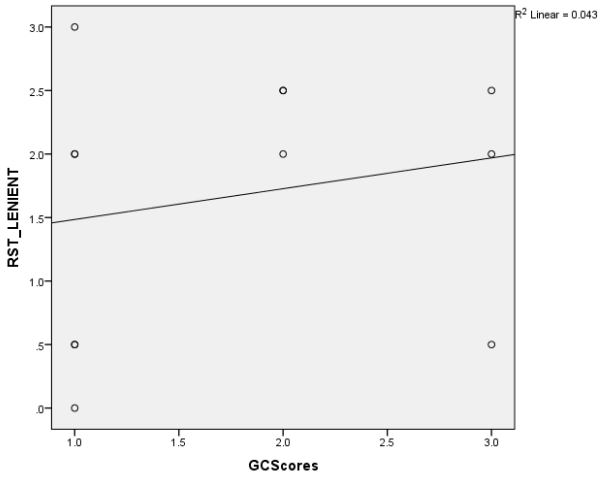
Graph 15

Scatterplot – General Comprehension Scores and RST strict correlations for the control group



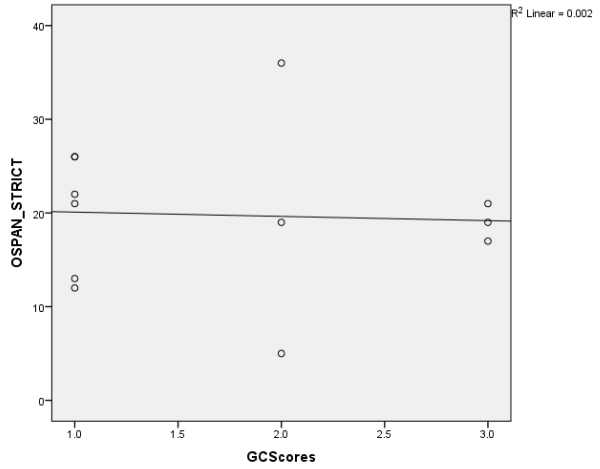
Graph 16

Scatterplot – General Comprehension Scores and RST lenient correlations for the control group



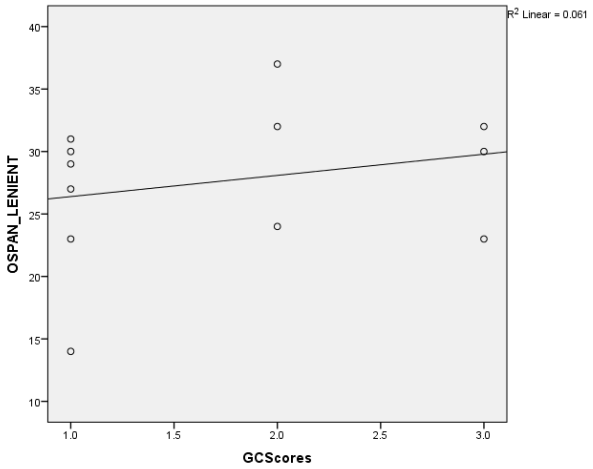
Graph 17

Scatterplot – General Comprehension Scores and OSPAN strict correlations for the control group



Graph 18

Scatterplot – General Comprehension Scores and OSPAN lenient correlations for the control group



The scatterplots just presented in Graphs 7-18 suggest that there is a very weak correlation between the scores obtained on the general comprehension portion of the video comprehension test and the WM tests, one that is not statistically significant in any of the treatment groups – intralingual and interlingual subtitles – and the control group.

As for the R^2 linear values, it is quite interesting to notice that the numbers obtained mostly reveal an extremely small percentage of the influence in change of one variable over the other. The smallest R^2 value obtained was between the general comprehension scores and OSPAN strict correlations for the control group ($R^2 = 0.002$), whereas the greatest R2 value obtained was between the general comprehension scores and RST strict correlations for the interlingual subtitles group ($R^2 = 0.27$). In light of the correlations or lack thereof, it is possible to posit that in this data set, there is a trending weak relationship between the participants' general comprehension and their WM capacity. These results will be addressed in the next Chapter.

5.6.2 Specific Comprehension and Working Memory Correlations

In order to check whether and the degree to which the performance of the participants on the specific comprehension part of the video comprehension test holds any relationship with their WM capacity, correlation tests were run. As previously explained in the Method, on the specific comprehension test, the results presented data that are not approximately normally distributed (see Appendix L for the tests of normality and Appendixes M and N for the histograms and box plots). Hence, Spearman's Rank Order Correlation Coefficient tests were run, and the results are presented in Table 36:

Table 36

Specific Comprehension and WM Correlations

	Correlation Test	SCScore x RST Strict	SCScore x RST Lenient	SCScore x OSPAN Strict	SCScore x OSPAN Lenient
Intralingual Subtitles (n=12)	Spearman's Coefficient r_s	-.338	-.271	-.293	-.520
	p value	.283	.295	.355	.083
	Correlation Test	SCScore x RST Strict	SCScore x RST Lenient	SCScore x OSPAN Strict	SCScore x OSPAN Lenient
Interlingual Subtitles (n=12)	Spearman's Coefficient r_s	.147	.050	.063	.099
	p value	.647	.877	.845	.760
	Correlation Test	SCScore x RST Strict	SCScore x RST Lenient	SCScore x OSPAN Strict	SCScore x OSPAN Lenient
Control (n=12)	Spearman's Coefficient r_s	-.146	.000	.271	.167
	p value	.651	1.000	.393	.605

Note. n = sample size; RST = Reading Span test; OSPAN = Operation-Word Span Test; p = significance level; SCScore = specific comprehension score.

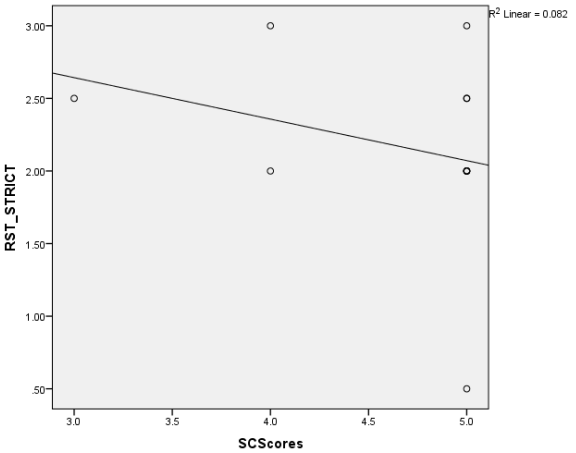
In a nutshell, Spearman's Rank Order Correlation Coefficient tests performed revealed varied results across the groups. In relation to the intralingual subtitles group, the correlations found were negative; as for the interlingual subtitles group, very weak positive correlations were

found; finally, regarding the control group, the correlations obtained were either negative or weak positive ones. Nonetheless, the correlations across all groups are not statistically significant considering the participants' performance on the specific comprehension portion of the video comprehension test and the WM tests ($p > .05$).

The correlations reported in Table 35 can be graphically visualized on the scatterplots that follow, which also bring the R^2 values found for each of the correlation tests. These values yield the corresponding percentage of the probable influence of one variable over the other. Graphs 19, 20, 21, and 22 present scatterplots of the correlation tests run with the intralingual subtitles group; Graphs 23, 24, 25, and 26 present scatterplots of the correlation tests performed with the interlingual subtitles group; lastly, Graphs 27, 28, 29, and 30 display scatterplots of the correlations tests run with the control group:

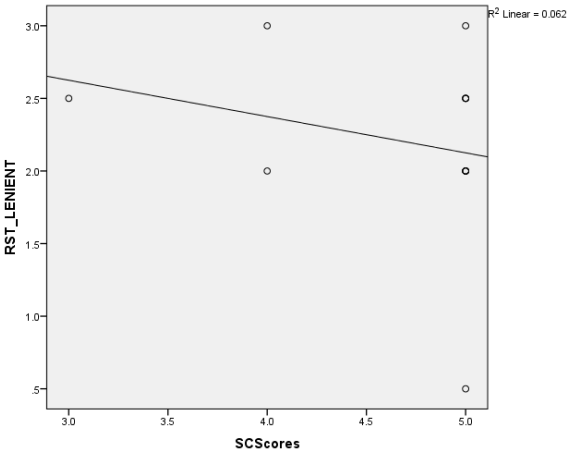
Graph 19

Scatterplot – Specific Comprehension Scores and RST strict correlations for the intralingual subtitles group



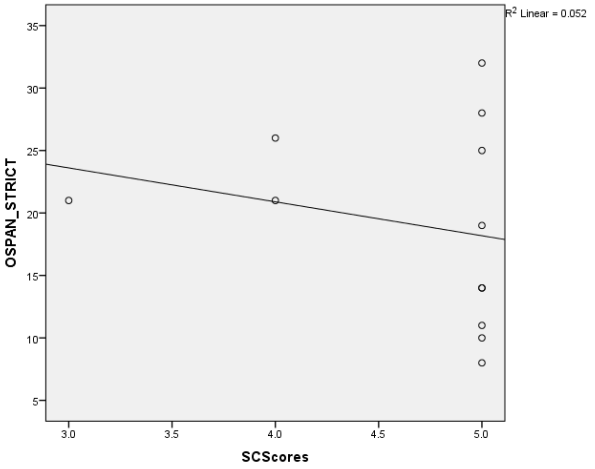
Graph 20

Scatterplot – Specific Comprehension Scores and RST lenient correlations for the intralingual subtitles group



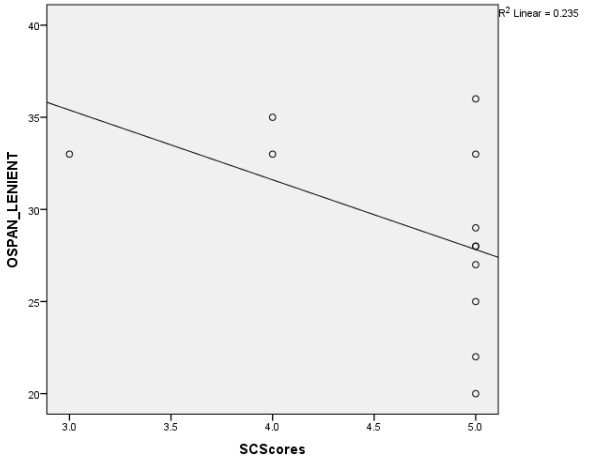
Graph 21

Scatterplot – Specific Comprehension Scores and OSPAN strict correlations for the intralingual subtitles group



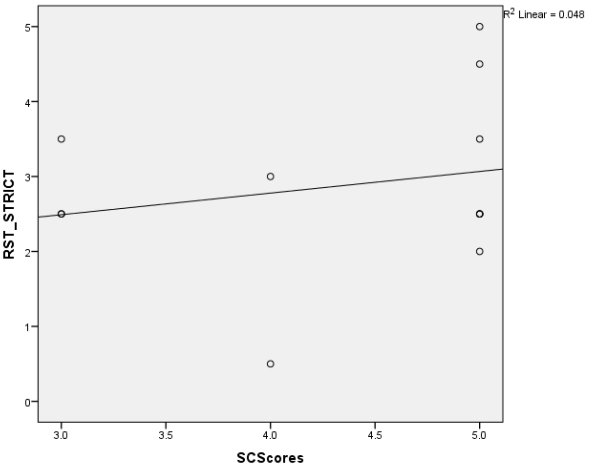
Graph 22

Scatterplot – Specific Comprehension Scores and OSPAN lenient correlations for the intralingual subtitles group



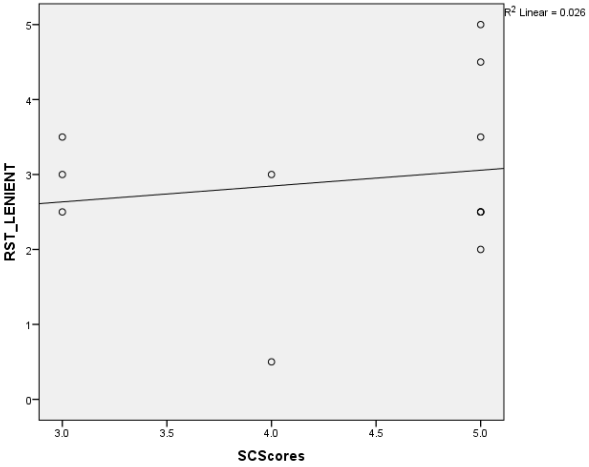
Graph 23

Scatterplot – Specific Comprehension Scores and RST strict correlations for the interlingual subtitles group



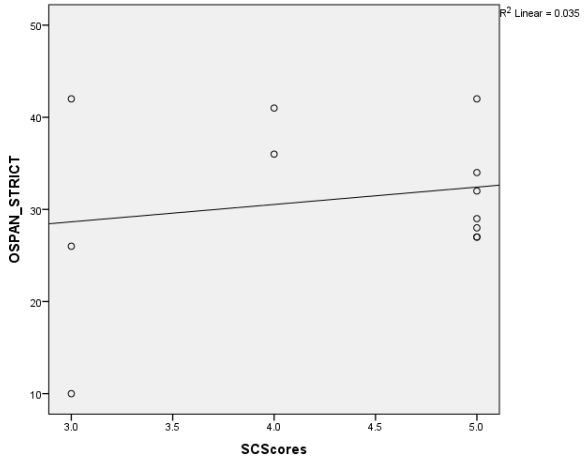
Graph 24

Scatterplot – Specific Comprehension Scores and RST lenient correlations for the interlingual subtitles group



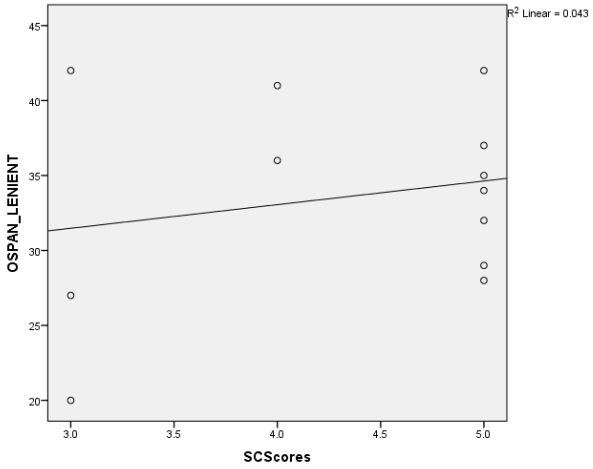
Graph 25

Scatterplot – Specific Comprehension Scores and OSPAN strict correlations for the interlingual subtitles group



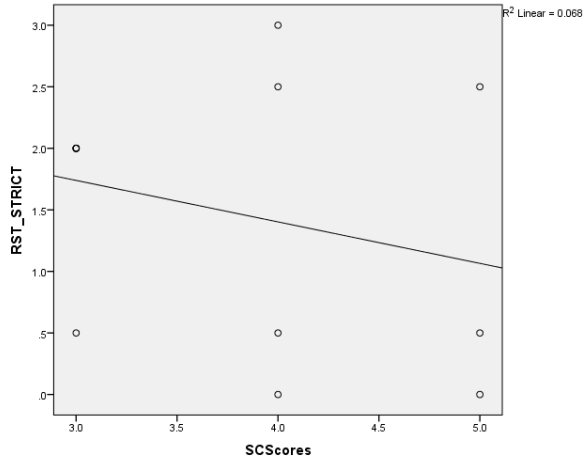
Graph 26

Scatterplot – Specific Comprehension Scores and OSPAN lenient correlations for the interlingual subtitles group



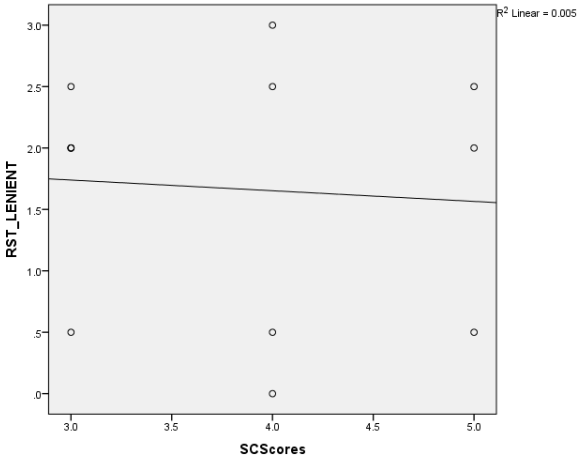
Graph 27

Scatterplot – Specific Comprehension Scores and RST strict correlations for the control group



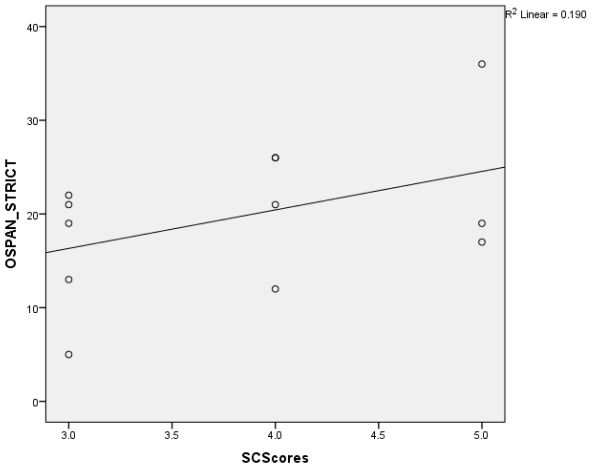
Graph 28

Scatterplot – Specific Comprehension Scores and RST lenient correlations for the control group



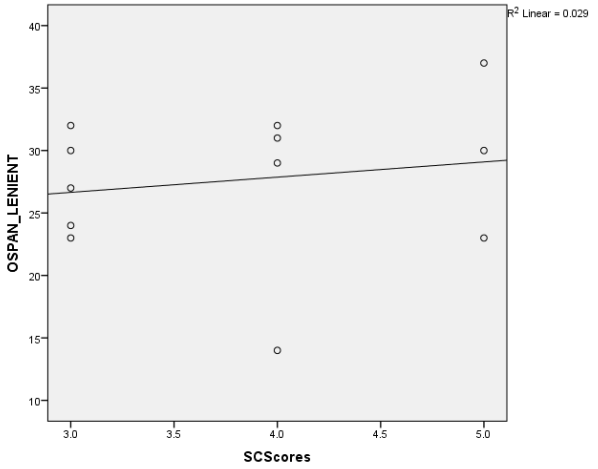
Graph 29

Scatterplot – Specific Comprehension Scores and OSPAN strict correlations for the control group



Graph 30

Scatterplot – Specific Comprehension Scores and OSPAN lenient correlations for the control group



All of the scatterplots displayed in Graphs 19-30 are representative of the general weak correlations found between the scores on the specific comprehension portion of the video comprehension tests and the WM tests. Moreover, the R^2 linear values reveal an extremely small percentage of the influence in change of one variable over the other. For instance, in this study, these values have been found to vary from less than 1%, in the case of the specific comprehension scores and RST lenient correlations for the control group, to 23%, in the case of the specific comprehension scores and OSPAN lenient correlations for the intralingual subtitles group.

In this vein, the results yielded by the correlation tests performed between the specific comprehension portion of the video comprehension tests and the WM tests suggest an extremely weak relationship between them, one that is not statistically significant ($p > .05$). These results will be revisited and discussed in the next Chapter.

5.7 L2 Vocabulary Recognition and Working Memory Correlations

This subsection of the Chapter centers on the results of the statistical tests and analyses performed in order to continue addressing RQ4 and hypothesis 4 (see Method for more information on the RQs and hypotheses). The results reported in the upcoming subsections are an attempt to investigate a possible relationship between the participants' performance on the L2 vocabulary recognition pre-test, test, and post-test and their WM tests.

Subsection 5.7.1 will present the descriptive statistics and the results of the correlation tests of the L2 vocabulary recognition pre-test and the WM tests; subsection 5.7.2 will focus on the results of the correlation tests of the L2 vocabulary recognition test and the WM tests; and subsection 5.7.3 provides the results of the correlation tests of the L2 vocabulary recognition post-test and the WM tests.

5.7.1 L2 Vocabulary Recognition Pre-Test and Working Memory Correlations

Firstly, the results of the correlation tests performed between participants' L2 vocabulary recognition pre-test and their WM tests will be reported. The results are shown in Table 37:

Table 37

L2 vocabulary recognition pre-test and WM correlations

	Correlation Test	Pre-Test x RST Strict	Pre-Test x RST Lenient	Pre-Test x OSPAN Strict	Pre-Test x OSPAN Lenient
Intralingual Subtitles (n=12)	Spearman's Coefficient r_s	-.419	-.440	-.262	-.102
	p value	.175	.152	.411	.753
	Correlation Test	Pre-Test x RST Strict	Pre-Test x RST Lenient	Pre-Test x OSPAN Strict	Pre-Test x OSPAN Lenient
Interlingual Subtitles (n=12)	Spearman's Coefficient r_s	-.194	-.112	.178	.130
	p value	.546	.729	.580	.687

	Correlation Test	Pre-Test x RST Strict	Pre-Test x RST Lenient	Pre-Test x OSPAN Strict	Pre-Test x OSPAN Lenient
Control (n=12)	Spearman's Coefficient r_s	.169	.264	-.184	.286
	p value	.599	.407	.568	.367

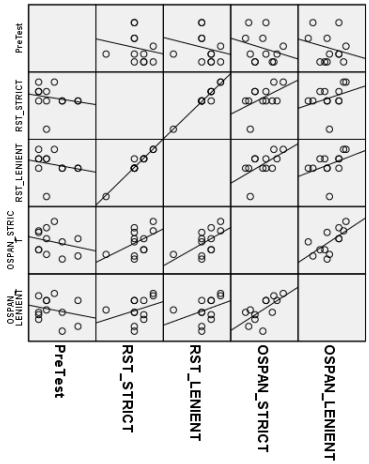
Note. n = sample size; RST = Reading Span test; OSPAN = Operation-Word Span Test; *p* = significance level.

As shown in Table 37, the results obtained with Spearman's Rank Order Correlation Coefficient tests revealed varied results for the three groups, converging into mostly moderate, weak negative correlations and a few weak positive correlations, all of which are not statistically significant ($p > .05$). In other words, the results of the correlation tests are suggestive of an absence of significant relationship between the participants' performance on the L2 vocabulary recognition pre-test and the WM tests.

The correlations reported in Table 37 can be visualized in the following Graphs containing matrix scatterplots (with all variables), which are arranged by group:

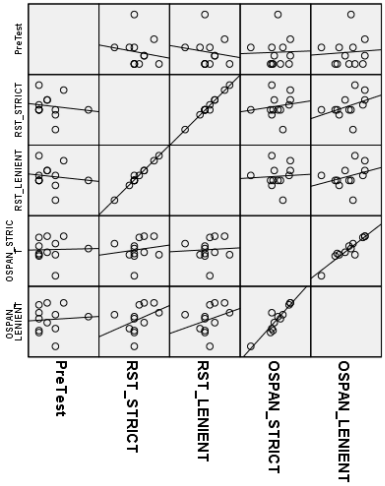
Graph 31

Scatterplot – L2 vocabulary recognition pre-test and WM correlations for the intralingual subtitles group



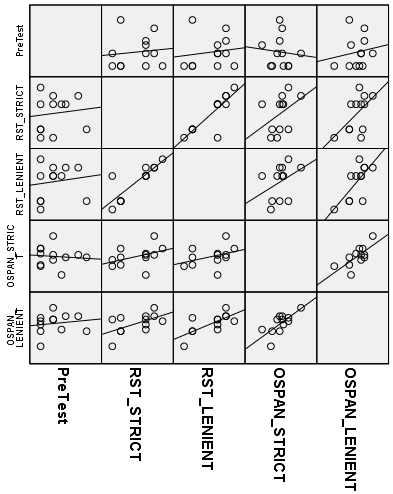
Graph 32

Scatterplot – L2 vocabulary recognition pre-test and WM correlations for the interlingual subtitles group



Graph 33

Scatterplot – L2 vocabulary recognition pre-test and WM correlations for the control group



The matrix scatterplots displayed in Graphs 31, 32, and 33 are illustrative of the lack of significant association between the participants' performance on the L2 vocabulary recognition pre-test and their performance on the WM tests. This can also be clearly perceived by looking at how the results are dispersed from the line of best fit on the each of the WM tests' axis related to the pre-test's axis. Furthermore, the R^2 linear values obtained reveal an extremely small percentage of the influence of the variables on one another. For instance, in this study, these values have been found to vary from less than 3 to 6%, in the intralingual subtitles group, and from nearly 0 to 2% in the interlingual subtitles group and control group.

5.7.2 L2 Vocabulary Recognition Test and Working Memory Correlations

Now, I turn to the results of the correlation tests performed between participants' L2 vocabulary recognition test and their WM tests, which are shown in Table 38:

Table 38

L2 vocabulary recognition test and WM correlations

Intralingual Subtitles (n=12)	Correlation Test	Test x RST Strict	Test x RST Lenient	Test x OSPAN Strict	Test x OSPAN Lenient
	Spearman's Coefficient r_s	-.119	-.254	-.209	-.100
	p value	.713	.426	.515	.756
Interlingual Subtitles (n=12)	Correlation Test	Test x RST Strict	Test x RST Lenient	Test x OSPAN Strict	Test x OSPAN Lenient
	Spearman's Coefficient r_s	-.323	-.281	.360	.230
	p value	.305	.377	.251	.473
Control (n=12)	Correlation Test	Test x RST Strict	Test x RST Lenient	Test x OSPAN Strict	Test x OSPAN Lenient
	Spearman's	.203	.318	-.029	.405

Coefficient r_s				
p value	.526	.314	.928	.192

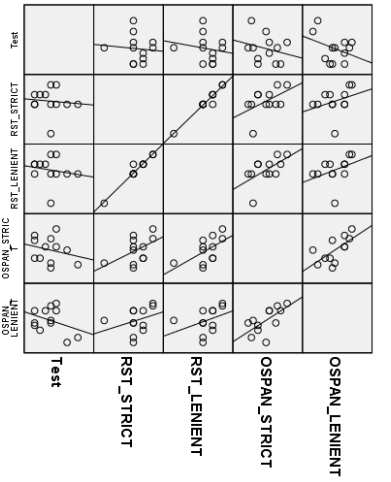
Note. n = sample size; RST = Reading Span test; OSPAN = Operation-Word Span Test; *p* = significance level.

As indicated in Table 38, the results obtained with Spearman's Rank Order Correlation Coefficient tests revealed varied results across the groups, converging into mostly weak negative correlations, though there are a few positive weak and moderate correlations. Nonetheless, all of correlations signal a lack of statistical significance ($p > .05$). This can be interpreted as an absence of noteworthy relationship between the participants' performance on the L2 vocabulary recognition test and on the WM tests.

The correlations reported in Table 38 can be visualized in the following Graphs containing matrix scatterplots (with all variables), which are arranged by group:

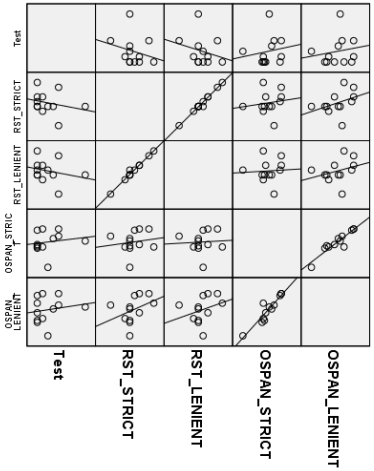
Graph 34

Scatterplot – L2 vocabulary recognition test and WM correlations for the intralingual subtitles group



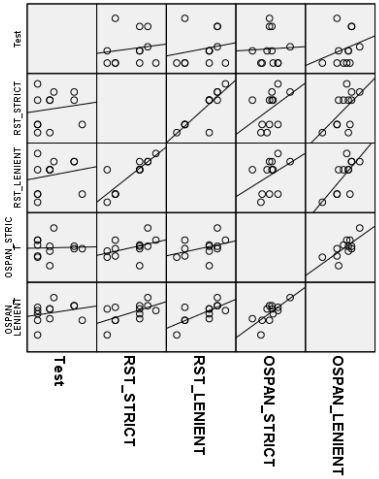
Graph 35

Scatterplot – L2 vocabulary recognition test and WM correlations for the interlingual subtitles group



Graph 36

Scatterplot – L2 vocabulary recognition test and WM correlations for the control group



The matrix scatterplots displayed in Graphs 34, 35, and 36 also indicate a lack of relationship between the participants' performance on the L2 vocabulary recognition test and their performance on the WM tests. As pointed out in the previous subsection, this can also be noticed by examining how the results are spread away from the line of best fit on the each of the WM tests' axis in relation to the L2 vocabulary recognition test's axis. Additionally, the R^2 linear values obtained signal an extremely small percentage of the influence of the variables on one another. In this study, these values have been found to vary from nearly 0 to 12%, in the intralingual subtitles group, from around 3 to 6% in the interlingual subtitles group, and from nearly 0 to 6% in the control group.

5.7.3 L2 Vocabulary Recognition Post-Test and Working Memory Correlations

Finally, this subsection will focus on the statistical results of the correlation tests performed between participants' L2 vocabulary recognition post-test and their WM tests. The results of the correlation tests performed are laid out in Table 39:

Table 39

L2 vocabulary recognition post-test and WM correlations

	Correlation Test	Post-Test x RST Strict	Post-Test x RST Lenient	Post-Test x OSPAN Strict	Post-Test x OSPAN Lenient
Intralingual Subtitles (n=12)	Spearman's Coefficient r_s	.152	-.022	-.283	-.391
	p value	.638	.945	.373	.209
	Correlation Test	Post-Test x RST Strict	Post-Test x RST Lenient	Post-Test x OSPAN Strict	Post-Test x OSPAN Lenient
Interlingual Subtitles (n=12)	Spearman's Coefficient r_s	-.256	-.235	.268	.081
	p value	.421	.462	.399	.803
	Correlation Test	Post-Test x RST Strict	Post-Test x RST Lenient	Post-Test x OSPAN Strict	Post-Test x OSPAN Lenient
Control (n=12)	Spearman's	.364	.487	.051	.276

Coefficient r_s / p value				
	.245	.109	.876	.385

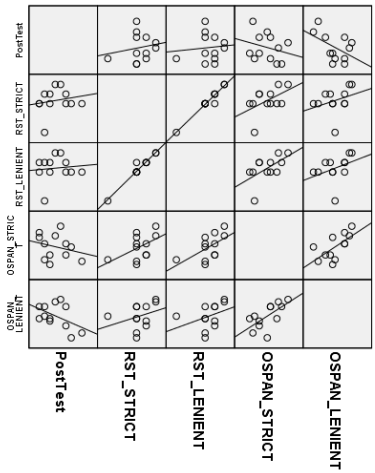
Note. n = sample size; RST = Reading Span test; OSPAN = Operation-Word Span Test; p = significance level.

Not surprisingly, the results obtained with Spearman's Rank Order Correlation Coefficient tests also revealed varied results across the groups, which range from an almost perfect absence of correlation to either a moderate negative or positive correlation. However, all of correlations obtained are not statistically significant ($p > .05$). These numbers can be understood as indicating an absence of substantial relationship between the participants' performance on the L2 vocabulary recognition post-test and on the WM tests in the data herein analyzed.

The correlations reported in Table 39 are also displayed in the following Graphs containing matrix scatterplots (with all variables) arranged by group:

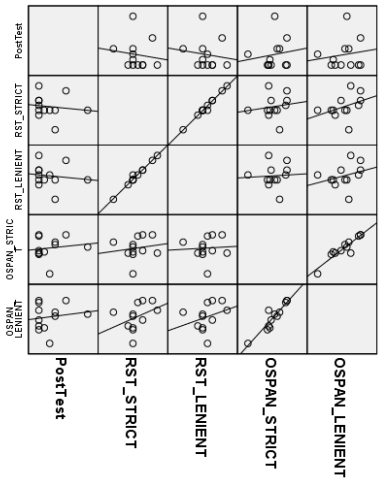
Graph 37

Scatterplot – L2 vocabulary recognition post-test and WM correlations for the intralingual subtitles group



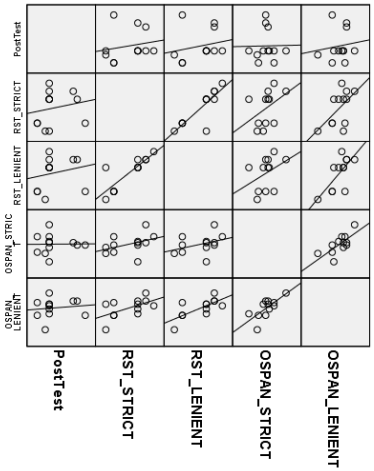
Graph 38

Scatterplot – L2 vocabulary recognition post-test and WM correlations for the interlingual subtitles group



Graph 39

Scatterplot – L2 vocabulary recognition post-test and WM correlations for the interlingual subtitles group



The matrix scatterplots displayed in Graphs 37, 38, and 39 also indicate an absence of a relationship between the participants' performance on the L2 vocabulary recognition post-test and their performance on the WM tests. This is confirmed by checking how the results are dispersed from the line of best fit on the each of the WM tests' axis in relation to the post-test's axis. Also, the R^2 linear values obtained stand for an extremely small percentage of the influence of the variables on one another. In this study, these values have been found to vary from nearly 1 to 23%, in the intralingual subtitles group, from 1 to 2% in the interlingual subtitles group, and from nearly 0 to 4% in the control group.

In sum, hypothesis 4, on whether *participants' WM capacity will interact with participants' performance in the video comprehension test and L2 vocabulary recognition test in that participants with higher WM capacity will have better scores on the video comprehension test and the L2 vocabulary recognition test,* is not supported. The results obtained with all of the correlations tests between participants' performance on the L2 vocabulary recognition test, pre-test, and post-test, and their performance on the WM tests indicate no statistically significant interaction between the sets of variables tested and analyzed. The lack of correlation between them and the results presented in this subsection will be discussed in the next Chapter.

5.8 Summary of Hypotheses and Statistical Results

This Chapter entailed the descriptive statistics and quantitative analyses performed on the data collected in order to investigate the four Research Questions and the four Hypotheses described in the Method. For now, only the hypotheses will be summarized in a tabular format since they seem to capture and translate the kernel of the statistical and quantitative nature in this study. The RQs will be properly addressed at the end of the discussion in Chapter VI.

Table 40

Hypotheses and major results

Hypotheses and Initial Predictions	Major Results
Hypothesis 1: Participants in any of the subtitling conditions – intralingual and interlingual subtitles – will outperform participants in the control condition in terms of their level of video comprehension	Partially supported.
Hypothesis 1a: Participants in the intralingual subtitles condition will outperform those in the interlingual subtitles condition in terms of their level of video comprehension.	Not Supported.
Hypothesis 2: Participants in any of the subtitling conditions – intralingual and interlingual subtitles – will outperform participants in the control condition in terms of their performance of L2 vocabulary recognition.	Not supported.
Hypothesis 2a: Participants in the intralingual subtitles condition will outperform those in the interlingual subtitles condition in terms of their performance of L2 vocabulary recognition.	Not supported.
Hypothesis 3: Scores on the participants' L2 vocabulary recognition test will vary from test to pre-test and from post-test to test, considering any of the subtitling conditions and possible gain scores in the three test moments.	Not supported.
Hypothesis 4: Participants' WM capacity will significantly interact with participants' performance in the video comprehension test and L2 vocabulary recognition test in that participants with higher WM capacity will have better scores on the video comprehension test and the L2 vocabulary recognition test.	Not supported.

Next, in Chapter VI, the results reported in the present Chapter will be discussed and possible implications for them will be considered from a qualitative oriented perspective.

CHAPTER VI
DISCUSSION

This Chapter is divided into three subsections. It has been organized as such in order to address each of the four Research Questions (RQ) posed in Chapter IV – Method. First, the results of the video comprehension test – both general and specific parts – are discussed in subsection 6.1, which relates to RQ1. Secondly, the results of the L2 vocabulary recognition tests are discussed in subsection 6.2, which provides the basis for the response to RQs 2 and 3. Finally, the results obtained with all of the correlation tests performed between and among Working Memory (WM) measures and the video comprehension test results, as well as the L2 vocabulary recognition tests results are discussed in subsection 6.3 to focus on RQ4.

6.1 Intralingual Subtitles, Interlingual Subtitles, and L2 Comprehension

As reported in Chapter II, the effects of subtitling on L2 comprehension has been, to some degree, perhaps one of the most well documented aspects in the literature, especially when it comes to European and other multilingual populations. As Vanderplank (2015) rightly put, the picture that has been built up with the literature is very informative in that the last thirty years of research into intralingual subtitles have confirmed repeatedly positive effects on (L2) language learning. Nonetheless, issues related to the contrastive nature of intralingual subtitles and interlingual subtitles, as well as long-term language development have still yet to be fully investigated.

In the present study, the results obtained with the general and specific video comprehension tests revealed that the availability of intralingual subtitles was more beneficial than the availability of interlingual subtitles or no subtitles whatsoever, though interlingual subtitles were found to aid comprehension just like intralingual subtitles were, from a statistical point of view. While the latter led participants to a significantly higher level of general and specific comprehension of the video material (content) and storyline, interlingual subtitles and control conditions were not statistically different from each other. This result is in tune with early single modalities studies (intralingual subtitles vs. soundtrack, only), such as those by Garza (1991) and Huang and Eskey (1999), who found that participants' content comprehension was larger

when intralingual subtitles were available in comparison to listening to the soundtrack without any textual aid.

The results obtained on the general and specific portions of the video comprehension test are also fairly aligned with some of those comparative (intralingual vs. interlingual subtitles) studies reported in the literature, such as Stewart and Pertusa (2004) and Hayati and Mohmedi (2011). These two pieces of research found that better performance was obtained when intralingual subtitles were available to the informants, rather than interlingual subtitles or soundtrack only. This pattern was observed in the results obtained in the present research. However, the results obtained here are not aligned with those reported in Markham, Peter, and McCarthy (2001), Markham and Peter (2003), and Latifi, Mobalegh, and Mohammadi (2011). In those studies, participants' level of content comprehension was higher when interlingual subtitles were available, rather than intralingual subtitles.

The debate around the empirical research on use of intralingual subtitles versus interlingual subtitles to foster video comprehension is far from black and white, and its grey areas are filled with sensitive issues that have not been fully investigated to date. For instance, Koskinen et al. (1996) posited that university-level students generally have better reading comprehension than listening comprehension skills and thus would benefit most if target-material was presented in the subtitles, that is, in this case, in the form of intralingual subtitles. If true, participants' frequent reliance on reading would partially explain the results found in the present investigation. Moreover, it would also confirm that the participants in the interlingual subtitles and control groups did not have the same opportunity to read the target-material in the L2, even though a translation was made available in the case of interlingual subtitles and the L2 was provided in the soundtrack.

Another sensitive issue that arises in the intralingual subtitles versus interlingual subtitles equation that could partly explain the results obtained here and elsewhere (Stewart & Pertusa, 2004) concerns the informants' attitudes towards subtitling. Not very often do researchers in the niche survey participants on how they feel about using intralingual and interlingual subtitles with a view to improving their understanding of the film content. In their study, Stewart and Pertusa comment that the surveys with their learners revealed that 65% of them prefer intralingual subtitles over interlingual ones, even if they had not been frequently watching videos with intralingual subtitles.

This particular piece of information speaks to the learners' beliefs and attitudes towards subtitling, which can have a direct impact upon their

L2 development. In the present research, data gathered with the student profile questionnaire[37] was very relevant because it outlines an interesting viewing profile. In relation to the participants in the intralingual subtitles group, 75% of them reported watching videos with Portuguese subtitles more often (twice or three times a week), whereas this number in the interlingual subtitles group amounted to around 92% and to 67% in the control group.

Although these numbers differ from each other, they would still not be enough to explain the results obtained on the general and specific parts of the video comprehension test, though they may trigger some valid insights. They may, for example, contribute to the speculation that these participants are likely to prioritize one input modality. Participants received information from two or even three competing channels or sources of input, involving two different languages. On the one hand, the audio/soundtrack in English, with challenging material to which they should attend, process, and understand; on the other hand, English or Portuguese subtitles, whose reading is also a highly cognitive demanding task, given the short period of time that subtitles last on screen (from 2-4 seconds at the most). Moreover, one cannot forget about the visual channel that certainly is attention drawing and does require some processing, after all, it is very unlikely that one watches a subtitled film and processes audio and text only.

On this very issue – the processing of images in a subtitled film – the few existing eye-tracking methodology-based studies on subtitling and L2 development so far have substantially enlarged our body of knowledge of what learners do when engaged in subtitled video watching tasks. Sydorenko (2010) reported that her beginning participants found images to be more useful than intralingual subtitles; Bisson et al (2014) found that participants tended to read the subtitles irrespective of the language in them (native or foreign) and also tended to spend more fixation time on the subtitles area when they were in a foreign language and then moved on to the images on screen.

In the present research, the fact that participants in the intralingual subtitles group outperformed those in the interlingual subtitles and control groups suggests that receiving the same input from two different competing channels (audio and subtitles), but on the same language (English) proved to be more slightly effective for content comprehension. Furthermore, it is worthwhile pointing out that the general and specific

[37] See Appendix B for the participants' profile questionnaire.

comprehension questions administered immediately after viewing were in English, which may have also facilitated or even boosted those participants' level of comprehension in the intralingual subtitles condition. As for those in the control condition, evidence from the results suggests that the video itself may have been challenging and some of its parts may have not even been processed.

Bianchi and Ciabattoni (2008) go as far as arguing that "subtitling is processed automatically and content comprehension can logically be facilitated by text in the mother tongue" (p. 78), which certainly does not seem to be the case here. In their experiment, intermediate-level learners performed better on comprehension tests using intralingual subtitles of one video (*Frantasia*) and the same participants performed better on comprehension tests using interlingual subtitles of another video (*Harry Potter)*, as compared to the control group (soundtrack only).

Drawing on these results and the ones obtained in the present research, I posit that the participants in the interlingual subtitles group may have tried to process both input modalities – audio and subtitles – and were then faced with a cognitive challenge, thus compelling them to attend to and process both sources of information containing two different languages while simultaneously paying attention to the video (images). Conversely, participants in the control group may have been overwhelmed as regards the fact that no textual aid was provided, which may have caused them to rely more on the video per se when realizing that the input provided in the audio could be difficult to attend to and process. It is important to highlight that most participants in the control group were used to watching subtitled videos with interlingual subtitles, as just reported.

This represents another sensitive issue in dealing with the use of subtitled video materials. The presence of one auditory channel and two visual channels – one being verbal and another being non-verbal – constructs a scenario in which the learner may and sometimes must choose one of the visual channels to prioritize for processing efficiency reasons, especially if that learner has not yet mastered the video's target-language. Noticeably, the choice of which source of input to focus on is probably linked to a myriad of variables, learning styles possibly being one of them. Auditory-oriented types of learners may perhaps cater for the auditory channel more than the subtitles, whereas visually-oriented types of learners may consciously or even unconsciously decide to focus more on the subtitles and the video instead.

Perego et al. (2010) hypothesized that when attention is more focused on the subtitles per se, image processing would be less effective

and that the opposite would be true as well. Their rationale was generally consistent with attentional theories that postulate early selection of information channels. The results in their experiment demonstrated that, in general, participants presented a pattern of attentional allocation, in that they willingly decided to pay more attention to the written input, rather than the visual one.

Another interesting finding is provided by Taylor (2005), who surveyed 35 learners (17 beginners and 18 intermediate Spanish learners) on their perceptions of subtitles (intralingual ones being the case in point). In the survey, 35% of the beginners stated that subtitles are distracting and confusing. The same learners also reported that they find it extremely difficult to use the three channels simultaneously. However, it is quite interesting to notice that only 11% of the intermediate learners reported similar difficulties. Another very relevant finding is that more intermediate than beginning learners of Spanish in that survey admitted that they were able to use all three channels with a view to facilitating their comprehension of the video material.

In light of Taylor's survey results combined with the present study's results, it might be that as learners progress in terms of their L2 development, and therefore develop sharper listening comprehension skills, their need to rely on subtitles can be expected to decrease. Consequently, it is also expected that learners will not need to depend on subtitles to understand the video content as much. This has been regarded in the literature as "subtitles as a crutch" (Winke, Gass & Sydorenko, 2010, p. 65), in that subtitles would facilitate or even enhance comprehension when the input provided is not very accessible or cannot be entirely attended to and processed through the auditory channel alone. With time, as one becomes more proficient in their L2 listening skills – and even more confident – their dependence on subtitles is likely to become more peripheral.

Taylor's (2005) survey also revealed that some participants (9 out of the 35 participants, 4 being beginners and 5 being intermediate-level) did mention that they completely ignored the soundtrack and focused on the subtitles alone. The implications behind such a statement are far reaching, but we must be weary given that the population in that study comprised both beginners and intermediate-level learners and what participants reported having done may not actually correspond to what they actually did on the tests.

Even though the participants in the present investigation were not specifically asked whether they attempted to process all channels

altogether, on the video retrospective questionnaire[38] administered immediately after the L2 vocabulary recognition post-test in the last session, participants were asked whether they considered the tests (video comprehension and L2 vocabulary recognition) difficult and their responses correspond to their performance obtained. In the intralingual subtitles group, 92% of the participants stated that they did not consider the test difficult; in relation to the interlingual subtitles group, 75% of them held the same opinion; finally, as for the control group, the percentage of participants who shared this opinion is drastically depleted, reaching almost 40%.

Often times, in response to the same question, some participants (30%) specifically wrote that understanding the story, the relationship between the characters, the events, the jokes, and the connections between scenes was not a problem for them because they tried to rely on all sources of information that were available. It is highly important to be mindful of the fact that out of the 36 participants, only two participants (Participants 30 and 35) in the control groups reported orally that they had already watched the episode used in the study, despite the fact that a total of 4 participants (one in the intralingual subtitles group, one in the interlingual subtitles group, and two in the control group) reported having already watched *The Big Bang Theory* before. Moreover, the participant profile questionnaire showed that *Friends* and *How I Met Your Mother* ranked amongst the most watched series by the population investigated, both of which belong to the same genre – comedy.

Going back to the results on the general and specific portions of the video comprehension test, the statistical tests and analyses performed revealed that the intralingual and interlingual subtitles groups were not statistically different from one another, even though the former outperformed the latter as regards both general and specific comprehension of the video material. It might have been the case that both groups obtained high scores on the test by relying on their preferred source of input, regardless of the language in it. For comprehension purposes, such a strategy would have sufficed for either intralingual or interlingual subtitles group, but not for the control group. It is, to some degree, safe to assume that participants in the experimental groups might have possibly relied on the reading of subtitles mostly in order to get the necessary information to make sense of what was being narrated on screen. The lower mean scores obtained by the control group would

[38] See Appendix K.

confirm this assumption because they could only have relied on their listening skills of a verbal source of input (soundtrack) of a challenging video and the images per se since no subtitles were made available in that condition.

Unfortunately, the researcher did not include a specific question to gather more data that could further enlighten this issue. Nonetheless, after the classroom data collections sessions and by the time the researcher met the participants individually for the Working Memory tests, many aspects were discussed, most of the times brought up by the participants themselves. Some of the participants informed that they usually find it very difficult to sustain attention to auditory and visual channels simultaneously. Interestingly, a few of them also mentioned that the more they get used to a certain TV series, by watching several episodes over the week, the less they tend to rely on the subtitles, be they intralingual or interlingual ones, because they mentioned that with time, they get acquainted with characters' accent, the jargons used in the series, and the storyline.

In spite of the fact that it is not possible to quantify or systematize this response given by the informants because they are solely based on the researcher's recollections, they provide some basis for the speculation that the type of video content or even its genre as well as video familiarity may have also played a role in determining learners' video comprehension of subtitled material. In fact, the relationship between content familiarity and subtitling has been a variable that has merited previous research attention in the niche. Markham, Peter, and McCarthy's (2001) revealed that participants' level of recall of the video content was much lower when no subtitles were available than when they were for non-familiar material, that is, for video materials with which participants were not so familiar in terms of content/genre. Also, Winke, Gass & Sydorenko's (2013) eye-tracking study demonstrated that video content familiarity correlated with the type of L2 being learned, which, in that case, was Chinese, a language that required more processing of subtitles with a video that was deemed more difficult for the kind of topic it portrayed, being unfamiliar to the participants.

One more aspect that deserves mentioning relates to the language used in the responses to the questions on the general portion of the video comprehension test. As previously described in the Method, participants were told that they could answer the questions in either English or Portuguese and their responses were not scrutinized by the raters. Instead, the raters were instructed to judge them as being appropriate or not as far as they could indicate comprehension. Surprisingly, more participants in

the intralingual subtitles (75%) and control (≈84%) groups answered the questions in English, whilst fewer participants (≈34%) in the interlingual subtitles group used English to respond to the questions. These numbers may also be indicative of the channel to which participants may have attended more often.

A final aspect that needs to be considered is the lack of proper test announcement prior to the video watching in the present investigation. As reported in the literature, test announcement has indeed been found to correlate with participants' better performance on tests. Winke, Gass, and Sydorenko (2013) found that their participants in the test announcement condition using full captioning or keyword captioning performed better than participants who did not receive a test announcement before watching the video in those conditions. The authors argue that test announcement possibly prompted them to reanalyze the content of the subtitles more closely, thus spending additional time on the subtitles area to process it more effortfully and therefore gather more information to understand the narration.

In the present research, since no proper test announcement was included, participants were not watching the video with a pre-established goal in mind, which makes it somewhat difficult to infer with which mindset participants watched it. This means that they did not know whether they were supposed to focus attention on any specific piece of information from the video, nor did they know the nature of the upcoming tests that they would perform.

A closer look at the specific comprehension test design may raise a question as to the possibility of guessing on the participants' part, given that it was of a True or False format. However, when comparing the participants' performance across the three groups on the two parts of the test – general and specific comprehension parts – there is evidence in the scores that suggests a trend. Such a trend consists of better performance by the intralingual subtitles group on both parts of the test in comparison to the interlingual subtitles group and then the control group. The results have thus led the researcher to rule out guessing as a determining variable upon the scores obtained. Nonetheless, it is important to beware that no correlation tests were run between the results obtained on both tests because they do not essentially measure the same construct.

Considering what has been discussed in this subsection, in addressing *RQ1: How do intralingual and interlingual subtitles affect learners' video comprehension, as measured by immediate post-viewing test?*, the results and analyses thus lend support to suggest that there were more beneficial effects on general and specific video comprehension

when using intralingual subtitles over interlingual subtitles or control conditions, as suggested by the mean scores obtained on the tests. Nevertheless, participants' performance in the experimental conditions on both general and specific portions of the video comprehension test was not statistically different from one another, even though higher mean scores were actually obtained by the intralingual subtitles group.

6.2 Intralingual Subtitles, Interlingual Subtitles, and L2 Vocabulary Recognition

The effects of subtitling upon L2 vocabulary development have been explored in several studies over the last years (Markham, 1999; Bird & Williams, 2002; Stewart & Pertusa, 2004; Perego et al, 2010; Winke, Gass & Sydorenko, 2010; Matielo et al, 2013; Perez et al, 2014, to name but a few). Overall, what the literature shows is that L2 vocabulary development is possible to be aided by the use of subtitled material, but the degree to which the effectiveness of intralingual subtitles and/or interlingual subtitles is related to such development in different L2 populations still poses challenges, doubts, and inconsistencies to researchers and L2 practitioners, as became evident in the inconclusive results reported in Chapter II.

As regards the present study's participants' mean scores on the L2 vocabulary recognition pre-test, a comparison across groups was indicative of different, though not statistically significant performance. Participants in the experimental groups – intralingual and interlingual subtitles – were both outperformed by the control group on the pre-test. Nonetheless, the performance of the three groups on the L2 vocabulary pre-test was not different statistically, which signals that, from start, the groups were slightly unequal, but their unbalance on pre-test mean scores cannot be regarded as an influencing factor, as confirmed by the Kruskal-Wallis One-Way ANOVA test ($H(2) = .1.310$, $p > .05$). Moreover, participants' low mean scores on the L2 vocabulary recognition pre-test were indicative of their perception on how difficult the test was for them. This was also found elsewhere (Stewart & Pertusa, 2004; Matielo et al., 2013).

The L2 vocabulary recognition test, administered immediately after the screening of the video, was designed to measure the assumable impact of the subtitled video upon participants' ability to recognize target-words they presumably read on the subtitles and/or heard on the soundtrack. As in the pre-test, participants were required to translate the target-words into Portuguese, explain them in either English or

Portuguese or provide a synonym in English in order to show that they knew their meaning and not merely 'recognized' their form.

Participants' mean scores on the L2 vocabulary recognition test were in fact higher than on the pre-test, which confirms that the treatment must have had some effect upon their recognition ability. Despite this effect, which relates to the (un)availability of subtitles, the statistical tests revealed that the groups are not significantly different from each other ($H(2) = .916$, $p > .05$). Thus, inspecting gain scores, as shown in Table 31, in Chapter V, was very informative because it revealed that more positive gain scores in the test/pre-test comparison were obtained by the intralingual subtitles group, followed by the interlingual subtitles group and finally the control group, regardless of the fact that the control group outperformed both experimental groups in the pre-test, but not on the test.

The more positive gain scores that the participants in the intralingual subtitles condition obtained in relation to the other groups may be attributed to a large number of factors, some of which are feasible to be discussed. First of all, all participants had seen the target-words prior to video watching on the pre-test, which means that they had been provided with an untimed opportunity to visualize the word form, even if they were unsure of its meaning at that point. Yet, only participants in the intralingual subtitles group had the chance to see the target-words in written form again while watching the subtitled TV series episode, that is, in the form of English subtitles. On the one hand, participants in the interlingual subtitles would have had to attend to the auditory channel and the translation of that word in the Portuguese subtitles. Participants in the control condition, on the other hand, would have had to rely exclusively on their listening skills if they were to attend to the target-words from the video itself.

The fact that the intralingual subtitles group had more gains in L2 vocabulary recognition in the pre-test/test comparison may seem attributable to this fact, even if they did not even examine the content provided via auditory input, therefore ignoring it. As for the interlingual subtitles group, a highly cognitive demanding task would have been entailed: participants would have had to work out some kind of mental translation of the target-words to which they were listening (in English) into the words they were reading on screen (in Portuguese) while simultaneously processing the story. Finally, participants' performance in the control condition confirms that the video contained oral material to be processed that was perceived as beyond their listening skills, as evidenced by their poorer performance on the video comprehension test in comparison to the experimental groups.

The results hereby obtained are in odds with the literature in relation to L2 vocabulary development via subtitled material. The literature on subtitling and L2 vocabulary development has given rise to many conflicting results, but as far as intralingual subtitles and no subtitles are concerned, more positive effects with intralingual subtitles for L2 vocabulary development have been obtained mostly (Neuman & Koskinen, 1992; Markham, 1999), though no differences were found in Yuksel & Tanriverdi (2009). As regards comparative studies – intralingual subtitles vs. interlingual subtitles – on L2 vocabulary development, some have favored the former (Winke, Gass & Sydorenko, 2010; Zarei & Rashvand, 2011), while some other studies have shown more prominence for the latter depending on the proficiency group (Bianchi & Ciabattoni, 2008) or the association of both in the form of dual subtitles (Raine, 2013).

Winke, Gass, and Sydorenko (2010) state that the use of intralingual subtitles provides learners with an opportunity to attend to different input modalities, which can impact not only overall understanding, but also target-vocabulary recognition. It is important to keep in mind that their study did not require participants to explain the target-words in any language nor provide a synonym to them. Instead, participants were only asked to indicate whether they knew the words prior to the treatment or not, which was quantified on a vocabulary recognition scale. It is possible that participants in the present research might have purely recognized more words after the treatment in relation to the amount of target-words they 'knew' before, though this was not measured on any word recognition scale. Recognizing a word, nevertheless, could be regarded as the beginning of L2 vocabulary development (Ellis, 1997), though it does not explain the whole story.

Yuksel and Tanriverdi (2009) stress the importance of encountering words in context. In their study, participants' ability to recognize target L2 words did not differ statistically considering the intralingual subtitles group and the control group, even though the former outperformed the latter. The authors also bring up the fact that no test announcement was included in the study, so their participants also did not know on what to focus while they were watching the video. Finally, it is also important to bear in mind that word recognition was measured via self-reports scales, which are problematic in the sense that learners might not necessarily provide an accurate account of their lexical development since no translation or explanation of the lexical items was required to be produced.

Another crucial aspect that cannot go without mention regards the low scores obtained in the L2 vocabulary recognition test by the participants in all groups in the present investigation. Unlike the relationship between subtitling and L2 comprehension, in which learners may infer aspects of the story being told on screen more easily from different input sources, such as the auditory channel and the visual channels, L2 vocabulary development seems to be dependent on several meaningful and comprehensible encounters with the input provided (Nation, 1990; Horst, Cobb & Meara, 1998).

It is my contention that the performance obtained by the participants in this research might possibly be at the threshold of what those learners are capable of obtaining, considering the video adopted, the target-words, their proficiency level, and the conditions under which they performed. In other words, with such a short, limited exposure to the input – a 20-minute video – containing the target-words, across the different experimental conditions, the processing of certain parts of the input may not even have occurred at times. This argument relates directly to the Noticing Hypothesis (Schmidt, 1990, 2010), which, in a nutshell, presupposes that input does not become intake for language learning or language development unless it is noticed, that is, consciously attended, which would also count towards explaining the low mean scores obtained by the participants.

Although the Noticing Hypothesis has been harshly criticized for being vague or lacking empirical support (Schmidt, 2010), it would provide an interesting perspective into the results hereby obtained. It would make sense to assume that most target-words may not have even been noticed by the participants – let alone processed – which then would have given them no chance to recognize them or make other higher level of cognitive processing, such as inferring their meaning, providing a synonym or coming up with a translation for them based on the context of the story in which they appeared.

Conversely, it is valid to acknowledge that not everything that one learns or eventually acquires language wise is necessarily explicitly taught. This is the stance that proponents and researchers in the explicit versus implicit learning paradigm take and have devoted part of their academic life to investigate. Most scholars seem to agree that unlike the first few thousand most common words in a given language, L2 vocabulary is mainly acquired incidentally (Huckin & Coady, 1999), though the role of instruction to draw learners' awareness to L2 lexical deserves to be considered.

Nevertheless, Gass (1999) and Ellis (1994a) both criticize those who equate implicit or incidental learning with unconscious learning. Some level of consciousness is very likely to be present in any (language) learning scenario. Thus, it is possible to assume that the fact that the participants did not recognize many L2 lexical items after having watched the video could also be attributable to a lack of consciousness as a product of attention (Schmidt, 1990). Successfulness in L2 vocabulary recognition is possibly dependable on a large amount of attention directed towards lexical items.

As Ellis (1997) contends, the learning of novel vocabulary consists of the development of associative connections that are highly dependent on repetition, learning, and practice. In the present study, the tests designed might not have offered the participants with enough and meaningful chances to develop such associative set of connections because the input provided to them must have not contained enough enhancement to foster participants' more successful L2 vocabulary recognition. When watching subtitled material, following the unfolding of events can be quite daunting in itself, not to mention focus on specific lexical information while attempting to understand what is going on in the story.

When participants are provided with intralingual subtitles, they can see the target-words in writing and are provided with the opportunity to visualize or even analyze their orthography. Hence, these words can be made more discrete, thus allowing them to unpack or compartmentalize the speech in a more meaningful way. In other words, "learning to understand a language involves parsing the speech stream into chunks which reliably mark meaning." (Ellis, 2003, p. 77). When no subtitles are provided, participants do not have the chance to make that association because they solely rely on their L2 listening skills and therefore may not have reached a point in their L2 development that allows them to parse speech to facilitate its processing and understanding. With interlingual subtitles, though, the scenario changes significantly, since participants are required to establish a link between the L2 material in the soundtrack and its L1 translation in the subtitles, which constitutes a dual processing activity.

In the retrospective questionnaire, participants reported insightful information about their perceptions of the L2 vocabulary recognition tests. The most frequent aspect mentioned by the intralingual subtitles group was that, differently from the comprehension part, they claimed that they did remember reading the words on the subtitles but they could not remember exactly what they meant. In other words, they were unable

to infer their meaning. Two of them also reported that they were surprised because they remembered having seen the words both on the pre-test and on the subtitles of the video, but they were unsure of their meaning. Additionally, two of them reported that they missed some instruction as to what should constitute their focus of attention because once they started watching the episode, they became so involved with the story and its comic nature that their ability to concentrate on details was therefore compromised.

These aspects are extremely clarifying concerning the low scores obtained on the L2 vocabulary recognition tests. It might have been the case that participants could recognize several if not all words, but were unable to let it become evident on the test because a focus on meaning was also required on the L2 vocabulary recognition test, instead of a focus on form, only. Only three participants wrote that they considered the vocabulary part of the test difficult.

Interestingly, participants in the interlingual subtitles group reported different opinions about the L2 vocabulary recognition tests. On their retrospective questionnaire, three participants wrote that they had never seen the words before, which means that they could not even recall the words from their encounter on the pre-test. This confirms that the target-words were not attended by them via auditory channel or translation in the Portuguese subtitles. Two participants in this group reported that they were unsure of their meaning and were afraid of making mistakes by guessing or trying to provide an answer that would be inaccurate. Most importantly, all of the participants in this group wrote that they considered the vocabulary part of the test very difficult.

As for the control group, ten participants reported that they could not remember the target-words, but none of them specified whether they were referring to the video alone or if that inability to recall words encompassed the pre-test too. Given that a one-week gap between pre-test, test, and post-test was established, the researcher assumes that the participants referred to the video itself. In this group, two participants wrote that they could remember having heard at least a few of the words, but they were unable to attribute meaning to them, even though these very same participants had an average understanding of the episode.

The numbers displayed in Table 31 in Chapter V also revealed that some participants in all three conditions – experimental conditions and control condition – presented neutral and negative gain scores in the test/pre-test comparison. With regard to neutral gains, it is clear that more participants in the control group obtained that type of score. This means that the absence of a treatment – a subtitled video – was impactful in terms

of their L2 vocabulary recognition as much as it did the experimental groups because more neutral gains scores were obtained by this group in relation to the experimental ones.

Negative gain scores from pre-test to test, which are also displayed in Table 31, were obtained more often with the experimental groups. A possible explanation is that a few participants in these groups may have changed their answers after watching the video by the time they took the L2 vocabulary recognition test in relation to the pre-test they had taken a week before. Yet, the small negative gain scores obtained (-.33, for instance) could also be due to a slight disagreement in the ratings. Ideally, data collection for the L2 vocabulary recognition pre-test and test should be carried out on the same session, which was not done so in order to reduce disruption in the students' regularly scheduled classes.

Turning to the L2 vocabulary post-test/test gain score comparison, a few interesting issues deserve a closer look. As previously mentioned in Chapter V, more participants in the control group presented positive gain scores from post-test to test than the experimental groups. Nevertheless, responses provided to the second question on the video retrospective questionnaire[39] definitely provide solid explanation to the numbers because most of the participants who obtained more positive gain scores from post-test to test admitted that they looked up words in the dictionary or asked someone about their meaning during that week. It is expected that the participants might have looked up some of the words in the dictionary or might have asked someone about their meaning because in the third and last session they were getting in contact with the target-words words for the third time in written form. Their curiosity must have spoken louder, thus prompting them to go and check the words, even though they did not specify whether they checked them after the test or the post-test session. Participants answered the retrospective questionnaire in the third and last classroom session, that is, after the post-test.

Regardless of the fact that participants did check words in between one of the tests, which may have possibly affected the results, one can argue that the checking of unknown words could be viewed not as a downside of the research design, but as a positive aspect. Triggering their curiosity in this experiment may have planted a seed in them for the birth of a novel vocabulary development strategy: the use of the dictionary to learn new vocabulary when watching TV series in or outside the

[39] See Appendix K.

classroom setting. As pointed out in Chapter III, there has not been much research on L2 vocabulary development strategies. One of the few examples is that of Schmitt's (1997) large survey with 600 Japanese EFL learners. The survey revealed that the use of dictionary was in fact the most frequent strategy employed by the students.

Taken together, when the results of the L2 vocabulary recognition test are interpreted by looking at Jiang's (2000) L2 vocabulary development model, the tests designed and adopted in the present study would probably allow for the onset of such L2 lexical development, right at the first or even second stage, which mostly involve the mapping of L1/L2 lexical content. Jiang claims that differently from L1 vocabulary development, in L2 vocabulary development, the major and first task is to remember the lexical item.

Although many participants in the present study stated that they could remember having seen the words before, they were not able to provide an explanation, meaning, synonym or translation to most of them. Jiang also argues that little semantic, syntactic, and morphological information in the lexical entry does not necessarily equate with unavailability of such information to the learners. In order for L2 vocabulary to develop, its meaning and grammatical features may be activated by L2-L1 links, by means of translation or lexical associations, which could have happened more often when the intralingual and interlingual subtitles were made available.

Concerning the availability of intralingual subtitles, I assumed that participants in that condition would be provided with a chance to establish some relationship of the target-words with the surrounding lexical items, whose meanings would then be inferable from the context in which they occurred in the subtitles, by matching their occurrence with what was being narrated on screen. This did not happen effectively. As previously discussed, participants in any subtitling condition had to deal with the issue of time, that is, the duration of subtitles on screen. It is quite possible that the 2 seconds for one-liners and up to 4 seconds for the two-liners are not enough for participants to read, register, and carefully analyze the written input. Target lexical items were possibly mostly unattended, that is, they did not become intake partly because they were not properly processed or even noticed in the input. Furthermore, the frequency must have also been a decisive factor, since the target-words were not very salient in the input, given that they appeared in the audio/subtitles mostly twice.

As to the availability of interlingual subtitles, a different type of processing might have been involved. Perhaps participants would have

had to make an extra effort to associate the input provided in the auditory channel with the written input provided in the subtitles. This, I posit, would unavoidably require the actual processing of both auditory and visual channels. Once that is done, participants in the interlingual subtitles condition would also have had to establish a successful translation relationship between the target-words and their meanings, synonyms or explanations. This would have entailed a second piece of effort on their part, one that might be even more cognitively demanding than the noticing itself.

Lastly, in relation to the control condition, their only chance to start off their lexical development of the target-words through the video watching task would be to successfully notice and process them in the auditory input. Once that is done, I believe that they could match them onto visual cues, such as the images of the video, and/or other clues in the story to make sense of them and eventually infer their meaning. This scenario I would assume to be the less advantageous for L2 vocabulary development to occur in comparison with the other two scenarios with different input modalities. All in all, it is, to some extent, clear that L2 vocabulary recognition could be the onset of L2 vocabulary development per se in a context such as the one hereby investigated. However, in different subtitling conditions, the nature of processing would differ, as just hypothesized.

In the experimental conditions and the control condition, input was provided to the participants in different modalities and forms. Considering the low scores obtained on the L2 vocabulary recognition tests by all groups, a hypothetical explanation that could underlie the results is that the input provided might not have been very comprehensible. The Input Hypothesis, also known as the Comprehension Hypothesis put forward by Krashen (1982), postulates that learners progress in their acquisition process only when they receive L2 input that is one step beyond their current stage of linguistic development as long as it is comprehensible.

It is plausible that the lack of robust results in the L2 vocabulary recognition test may hold a relationship with input that was beyond participants' ability to comprehend it. According to Krashen, comprehensible is assumed in the sense of what the learner can understand but cannot produce yet. In the present study, involving intermediate-level students, the video chosen may have contained language that was a little difficult to process. Such a difficulty may have been aggravated when no support was provided, that is, in the control condition. Even though the results in the comprehension test were

indicative of the fact that participants in all groups did have from reasonable to great understanding of the story, with L2 vocabulary the input provided was probably more decisive in terms of their success to notice, process, and grasp novel vocabulary.

In their study, Bird and Williams (2002) found that phonological information derived from both text and sound (as is the case of intralingual subtitles) contributed to improvements in the processing of spoken words by their informants. Moreover, they claim that orthographic and phonological sources become highly integrative and interactive when one watches a subtitled video with intralingual subtitles. They also argue that "the cognitive systems dealing with auditory and visual word recognition are highly interactive and fully interconnected" (p. 17). In light of their arguments and findings, one would expect the participants in the present study in the intralingual subtitles condition to significantly outperform other participants in the other two conditions, which did not occur.

Still, it is curious to notice that in one of their experiments, Bird and Williams found that bimodal input (sound and text simultaneously presented) led to significant novel word learning only when the target-words were recycled (presented) at least three times. In the present experiment, as pointed out in the Chapter IV – Method, half of the target-words appeared twice and the other half appeared only once (in both auditory – soundtrack – and visual channels – subtitles). Hence, it is very likely that lack of enhancement in the input in terms of the frequency with which the target-words appeared did play a role in the successfulness regarding participants' ability to recognize them in the immediate L2 vocabulary recognition test.

In terms of how many exposures one needs to learn novel vocabulary, though the answer is not so straightforward, the literature does present somewhat converging insights into this matter. Nation (1990) has claimed that 5-16 exposures are needed in order to learn a word from context, whereas Meara (1997) suggested a 0.01 hypothesis – 1 uptake every 100 exposures – for L2 learners, arguing that these learners are usually unable to be exposed to large quantities of text. Horst, Cobb, and Meara (1998), in a study with low intermediate EFL learners reading a 109-page book over a ten-day period, obtained a 20% pick-up rate as regards novel lexical items. Interestingly, they also observed that words appearing over eight times in text were more likely to be picked up than those that were repeated less.

As discussed before in Chapter III, many other factors are expected to play a decisive role in the successfulness of novel vocabulary learning.

The mastery of a new word axes on many other factors such as the salience of the word in context (Brown, 1993), as well as the richness of certain contextual clues, the learner's attitudes, and possibly the size and quality of his/her existing repertoire of vocabulary (Laufer & Hadar, 1997).

In light of the discussions provided in this subsection, in relation to *(RQ2): How do intralingual and interlingual subtitles affect learners' L2 vocabulary recognition, as measured by pre-test, test, and post-test?*, the present study did not find robust L2 vocabulary recognition results by intermediate EFL learners in the experimental and control conditions herein investigated. Moreover, the present study also found no statistically significant differences between the experimental groups – intralingual and intralingual subtitles – and the control group, as regards their performance across L2 vocabulary recognition pre-test, test, and post-test.

As far as *(RQ3): How does learners' L2 vocabulary recognition, in case there is any, differ across time for any experimental condition?* is concerned, the results obtained on the L2 vocabulary recognition pre-test, test, and post-test signal a trend that reveals mostly a positive growth in performance, as evidenced by the test/pre-test and post-test/test gain score comparisons. Furthermore, the intralingual subtitles group presented more positive gains in L2 vocabulary recognition across time, even though the results of the statistical tests performed revealed no significant differences across groups and across time.

6.3 Subtitling, Video Comprehension, L2 Vocabulary Recognition, and Working Memory

To the best of my knowledge, the present investigation is the very first to empirically examine the existence of a relationship among subtitling, video comprehension, L2 vocabulary recognition, and Working Memory (WM). The motivation behind investigating this variable was based on the researcher's assumption that WM could be associated with the participants' ability to better understand the video content and to recognize L2 target-vocabulary. Because the literature on WM (see Chapter III) has repeatedly suggested that one's WM capacity correlates with one's level of L2 attainment, it might be the case that WM could be a variable that interacts with one's capacity to sustain attention to the different input channels or sources that are at play when one watches subtitled video materials.

Because this is an exploratory study – the first of its nature –, comparisons with previous studies become impossible to be drawn. Yet, the upcoming subsections intend to provide some preliminary insights into the statistical results of the correlational tests run among the variables shown in Chapter V, by firstly addressing the relationship between L2 comprehension and WM, followed by the association between L2 vocabulary recognition and WM.

6.3.1 Subtitling, L2 Comprehension, and Working Memory

Regarding the results of the RST reported in Chapter V, participants in the interlingual subtitles group outperformed participants in the intralingual subtitles group, who in turn outperformed participants in the control group, as far as the strict scoring procedure is considered, for the sake of illustration. Four participants only (Participants 13, 14, 16, and 18) were classified as higher spans, who belonged to the interlingual subtitles group. As for the OSPAN, as far as the strict scoring procedure is considered, better performance was obtained by the interlingual subtitles group, followed by the control group, and finally the intralingual subtitles group. Three participants only were classified as higher spans (Participants 13, 14, and 19), who also happen to belong to the interlingual subtitles group.

In the WM retrospective questionnaire administered at the end of the WM session (see Appendix K), most participants (85% of the respondents) stated that they considered both tests to be difficult, and this percentage also includes all of the higher spans in both RST and OSPAN tests. Most participants (70%) also stated that they perceive themselves as having a bad memory, and some of them even stated that they have a terrible memory. Such negative perception of their memory capacity may not be reflective of their actual memory skills. Instead, it may be linked to the fact that participants responded the questionnaire immediately after both WM tests had been administered and were visibly tired. Interestingly, half of those participants who reported having a bad memory described a few situations or purposes for which they believe their memory tends to work best, such as in remembering names or faces, for instance.

In relation to the correlation tests performed with the WM tests – RST and OSPAN – as reported in Chapter V, only one positive and statistically significant correlation was found, which was the RST Lenient x OSPAN Lenient for the control group ($r = .705$, $n = 12$, $p = .01$). The

only statistically significant correlation found among the WM variables amount to a strong, positive correlation between the two variables.

According to Conway et al. (2005), both WM tests are reliably considered to measure WM capacity. Engle et al. (1999) found a moderate positive correlation between the two WM tests; Daneman and Merikle (1996) carried out a meta-analysis on these two measures, and one of the results that the scholars found was that verbal processes plus storage measures of WM capacity are reliable predictors of global comprehension than are the operation-word processes plus storage measures. Moreover, the RST has been mostly associated with verbal ability, whereas the OSPAN has been mostly associated with logical and mathematical skills, as well as lexical knowledge (Conway et al., 2005).

A quick look at the pieces of research carried at *Programa de Pós-Graduação em Inglês* reveals that only one of them (Bailer, 2011) has used both measures and attempted to check whether there was a correlation between the results obtained on those WM measures. Therefore, the lack of more research with Brazilian EFL learners who have taken both WM tests makes it difficult to compare the results of the RST and OSPAN correlation tests obtained in the present investigation to further draw conclusions.

Moving on to the results between the correlation tests performed with the WM measures – RST and OSPAN tests – and the video comprehension test – general and specific parts –, it is clear that, in this study, no statistically significant relationship was found between them in any of the experimental conditions – intralingual and interlingual subtitles – and the control condition. Some factors might help explain the lack of statistically significant results obtained with those correlational tests.

Firstly, it is important to examine carefully the nature of the WM tests used in the present study. When analyzing what participants actually do when taking the RST, the common understanding is that the test taps one's ability in processing efficiency mainly as regards comprehension. Daneman and Carpenter (1980, 1983), when originally designing the test, argued that individual differences in WM capacity would reflect differences in terms of processing efficiency. They held the view that differences in the processing efficiency would be at the core of individual differences regarding language comprehension. In other words, individuals with inefficient processing would possibly have a smaller storage capacity due to the fact that they would have to allocate more of their attentional resources to the processing demands of the task itself. Hence, one's span would be likely to correlate with one's capacity to maintain information active in WM long enough to link it with new

information in order to comprehend a text, for instance. Essentially, the scholars also claimed the RST can predict comprehension because it involves many processing aspects that are involved in normal, typical reading.

As to the OSPAN test, Turner and Engle (1989) also intended to measure WM capacity through a task that could tap one's ability to store information for a brief period of time, but that is not language-specific. As demonstrated by Klein and Fiss (1999), the test has been adopted to measure WM capacity because of its high reliability and stability scores. Furthermore, the OSPAN is in tune with Engle and colleagues' view in that WM capacity is related to the processing efficiency in any given cognitive task in terms of information processing. As Prebianca (2009) explains, "capacity refers to individuals' ability to bring pieces of information from long-term memory into an active state and temporarily maintain that information for further processing by preventing other irrelevant stimuli to enter the focus of attention" (p. 34). Thus, the lack of significant relationship between the RST and OSPAN measures and the video comprehension test as well as the L2 vocabulary recognition tests in the present investigation may be related to the set of relatively different types of tasks employed.

Secondly, while watching the TV series, all three groups indeed had either two or three channels that could compete for their attentional resources. As previously explained, the intralingual subtitles group were submitted to a watching condition in which one auditory channel (soundtrack) and two visual channels (subtitles + video) were simultaneously provided, all of which shared a common language (L2); the interlingual subtitles group, however, were performing in a different condition, being provided with an auditory channel (L2 soundtrack), and two visual channels (L1 subtitles + video), thus being presented with two different linguistic sources of input; as for the control group, two channels were presented to them, that is, one auditory channel (L2 soundtrack) and one visual channel (video).

Participants' performance conditions did not entail having their attentional resources directed to any of the channels (soundtrack, subtitles and/or video). In other words, no guiding instructions were provided to them that could influence them in either prioritizing one of the input sources or even attempting to process all of them simultaneously. This may lead us to believe that participants' attentional resources may have been allocated in only one verbal channel (either soundtrack or subtitles).

As the results of the video comprehension tests demonstrated, participants in the intralingual subtitles condition outperformed the other

two groups on both general and specific comprehension portions of the test. Participants performing under that condition were provided with only one language (L2) in all verbal input sources (soundtrack + subtitles). The inference that could be made is that there were fewer attention depleting mechanisms involved in comparison to the interlingual subtitles group. This explanation, however, would not be applicable to the control group, whose performance was indeed poorer as compared to the performance of the intralingual subtitles group, but not statistically different from the interlingual subtitles group's performance.

Thirdly, although WM, measured by the RST, has been found to correlate mostly with reading comprehension, it may as well be the case that participants' processing on the video comprehension test involved a set of distinct mechanisms, especially taking into consideration the nature of the task employed in this research, involving a multimedia source, therefore not regular, typical reading. I would suspect that one's efficiency in reading subtitles (in L1 or L2) might be more closely linked with one's experience with and the amount of exposure to them. Yet, as previously reported, participants in the present research declared being more used to watching subtitled videos with interlingual subtitles than with intralingual subtitles or no subtitles whatsoever.

I would also suspect that one's capacity to process the auditory channel of the video would hold a significant relationship with one's WM capacity – or more specifically one's phonological memory – if, in fact, that channel significantly drew one's attention while processing another channel (subtitles and/or video) simultaneously. In other words, in the present investigation, participants might have not consciously attended to the auditory verbal channel. Given that participants informally stated that they tend to resort to the subtitles more often than the auditory channel, it could be the case that no seemingly major interaction would arise in this scenario. Selective attention might have played a more prominent role in determining where participants allocated more of their attentional resources to process the input provided in a given channel.

Downing (2000) has researched the relationship between selective attention and WM. The scholar contends that selective attention "reduces the load on limited-capacity cognitive systems by filtering irrelevant information from the stimulus stream" (p. 467). The author also explains that in a typical scenario containing many objects, the amount of information present exceeds the capacity of object representation systems, that is, one's capacity to process that (visual) information. Consequently, the objects can be described as "competing" for attention, and the strongest competitors – perhaps the more relevant or salient ones

– will be likely to become the focus of selective attention. As a result, they gain access to awareness and guidance of action.

When watching subtitled audiovisual material, it is possible that participants tend to focus on the most salient channel, which could also be their preferred one or the one it might seem easier for them to process or even one that, in it pure nature, is more attention drawing. In other words, visual channels – either verbal or non-verbal – could constitute their strongest competitors for attention. The problem in this scenario is that attending to the two channels would require, I assume, a considerable processing load because of the speed which with that would have to be processed (short duration on screen as regards the subtitles and the scenes) and the fact that reading is involved in processing one of the channels. If true, this would explain why participants informally reported that they tend to ignore the soundtrack of the video most of the times.

What Downing (2000) has found in his experiments is that "visual working memory and selective attention share a key functional component: The contents of working memory guide attention even when there is no explicit search task" (p. 469). As explained previously, in the present investigation participants were not provided with any specific guidance that could direct the allocation of their attentional resources'. It is then plausible to assume that when watching subtitled videos, selective attention was directed mostly towards the subtitles and the images of the video.

A somewhat similar perception about this issue is shared by Sydorenko (2010). Her assumptions relate to learners' attempts to pay attention to all three modalities. She believes that if learners did pay attention to the three channels simultaneously, this could result in cognitive overload, which "occurs even when tasks are performed in the native language and is attributed to the limits of working memory" (p. 52). These aspects are embedded with a redundancy principle of the cognitive load theory, which presupposes that redundant material slows down information processing and learning.

The contradiction surfaces when L1 and L2 studies are contrasted. Mayer, Heiser, and Lonn (2001), from a cognitive load theory perspective on multimedia learning, found that L1 speakers of English who saw an animation and listened to a parallel corresponding narration in their L1 were able to retain more information from the narration than those who also received intralingual subtitles as a third modality/channel to process. Their view is that subtitles availability is distracting when audio is also present because they essentially carry the same information, which then

follows a redundancy principle in terms of information processing capacity.

However, the literature on subtitling and L2 development has suggested quite the opposite: Subtitle availability has been found to be associated with better L2 comprehension and L2 development. Nonetheless, the preferred source of input to be attended and processed by viewers has not been extensively researched yet. To date, there have only been a handful studies that have looked at this specific aspect (Vanderplank, 1988; Taylor, 2005; Sydorenko, 2010). In general, what these studies found is that, at first, subtitles are considered distracting by the L2 viewers; with time, viewers reported that they get used to them and even start developing strategies to try to attend to all channels simultaneously.

Fourthly, another possible explanation for the lack of significant correlation between WM and video comprehension could be due to the sample size and the distribution of high spanners. Traditionally, research on the relationship between WM and language development has been carried out with much larger sample sizes (see Chapter III for the studies reviewed). In the present investigation, only 12 participants remained until the very end of the data collection in each of the three groups and the high spanners were grouped in only one experimental group, coincidentally. Therefore, it is possible that no statistically significant correlations were found in this study because of the limited number of participants and the fact that there were fewer high spanners, who were not evenly distributed across experimental and control groups.

6.3.2 Subtitling, L2 Vocabulary Recognition, and Working Memory

Now, shifting to the results between the correlation tests performed with the WM measures – RST and OSPAN tests – and the L2 vocabulary recognition tests – pre-test, test, and post-test – this study also found no statistically significant relationship between them in any of the experimental conditions – intralingual and interlingual subtitles – and the control condition. Some key elements might shed some light into the possible lack of statistically significant results obtained with those correlational tests.

Firstly, provided that WM measures, especially the OSPAN test, have mostly been found to correlate with lexical knowledge development, it would be realistic to find some significant interaction between participants' WM measures and their L2 vocabulary recognition, mostly

in the immediate L2 vocabulary recall test. This expectation in part stems from the fact that no treatment was provided in the pre-test; as for the post-test, a delayed one-week was adopted, so results would possibly interfere with long-term memory (LTM).

Still on this issue, perhaps one of the reasons underlying the lack of substantial correlations among the WM measures and the L2 vocabulary recognition tests might be associated with the nature of the tests regarding time and storage. On both RST and OSPAN tests, participants had to hold some piece of information (words) for a very brief period of time in their WM, maintaining it active somehow, and then had to recall that very same piece of information to verbalize it (saying it out loud when question marks appeared on screen). Unlike the WM tests, while watching the video in any of the experimental or control conditions, participants might not have been consciously trying to hold active any piece of information for later retrieval. Hence, the nature of the tasks involved – WM and L2 vocabulary recognition – is apparently dissimilar, particularly in terms of the processing demands required to accomplish each of them.

Secondly, as Baddeley (2009) explains, short-term (STM) and the system(s) responsible for it are part of WM, that is, integrated. The author also explains that STM then involves the capacity to store small amounts of information for very brief periods of time. Also, Schwartz and Metcalfe (1992) state that STM is considered an active memory with limited capacity to hold information for twenty to sixty seconds or so. The information received in STM is thought to be stored for a small period of time while being analyzed and interpreted. Once understood, part of the information is transmitted to LTM for permanent storage. The old information that is no longer needed may fade away from STM (Mayer & Moreno, 1998). Thus, the operations going on in STM are deemed indispensable for long-lasting storage.

In this sense, in order for information to be stored in STM and later transmitted to LTM, it has to be analyzed and possibly understood, otherwise it is lost or discarded. Again, it will only be analyzed if such piece of information is, I presume, attended. This view that links STM to consciousness is shared by a few scholars (Baars, 1986; Schmidt, 1990). In sum, their perception is that STM serves as a type of "broadcasting station" (Baars, 1983, as cited in Schmidt, 1990, p. 137) and emphasizes the role of consciousness to actuate the learning process. Although the extent to which consciousness is decisive in learning has been controversially debated in the field of SLA, many cognitive psychologists fairly agree upon the need to raise learners' consciousness to sensitize

their attention to properties of what to learn (Rutherford & Smith, 1985; Koskinen et al, 1996; Hsiao & Oxford, 2002).

In the L2 vocabulary recognition tests, more specifically on the immediate test, participants watched a 20 minute video and then were given the L2 vocabulary recognition task. It is possible to hypothesize that participants noticed some of the target lexical items, understood them, inferred their meaning or translation counterparts, but by the time they were administered the test – that is, 20 minutes later – some of those items may have not been any longer active in their STM for some reason, which could be dependent on a larger storage and processing capacity on their part. Moreover, the simple fact that participants may have been more concerned with understanding the story and enjoying the comedy – and the high scores on both general and specific comprehension portions of the video comprehension test would confirm that – would suggest that their attentional resources were not consciously allocated towards the lexical items.

Some of these lexical items, I am afraid, may have not even been successfully integrated into their STM in the first place because some of them may have not been processed, understood, and therefore registered. Furthermore, it is also plausible that lexical items were not held active in participants' STM because of a decay in memory trace due to a lack of frequency/saliency in the input. Following a similar line of reasoning, it might be the case that because very few of the participants were high spanners, in general, their capacity to hold active pieces of information about the video may have proven insufficient for later retrieval.

Thirdly, in accordance with some researchers (d'Ydewalle et al., 1991; Grimes, 1991), the redundancy produced by subtitled materials may present an influence upon consciousness and STM capacity. These theorists argue that visual channels (subtitles and video) require a certain amount of attention due to their dynamic and graphic nature. This means that the pictures that require attention can be processed with little effort, they claim. Consequently, more attention resources or capacity should be free to be used. These theorists also argue that as redundancy between the channels increases, the capacity necessary for processing the information decreases as well.

If true, this assumption would mean that if one understands the story narrated on screen more easily, more attentional resources should be available for learners to focus on other aspects of the subtitled video, e.g. vocabulary. Looking at the results of the video comprehension test and L2 vocabulary tests, it is clear that the intralingual subtitles group did perform better than the other two groups; the results also suggested that

the interlingual subtitles group outperformed the control group. Even though these performances did not always achieve statistical significance in the comparison across groups, the numbers could suggest that the availability of subtitles – and mostly English subtitles – enabled better video comprehension of the story and therefore might have made more attentional resources available to those participants performing in either experimental conditions to try to examine the linguistic content provided in the input.

A caveat surfaces amongst the issues discussed so far. Vanderplank (2015) criticizes the role of noticing and attention in the realm of subtitled film for language development. The scholar goes as far as claiming that "in learning terms, noticing is not going to be of much assistance to learners watching captioned programmes" (p. 29). For him, noticing is fleeting, that is, it cannot be recorded or properly examined. He adopts the same stance towards attention.

Vanderplank has very recently proposed a model for internalizing and using intralingual subtitles, one that suggests conscious attention on the learners' part. The problem with that model is that, as he himself admits, "the bar is set almost impossibly high for the ordinary learner-viewer to really gain in terms of genuine language acquisition from watching programmes even when there are captions available" (p. 29).

It is possible that, from a statistical point of view, no statistically significant interaction was found between WM measures and L2 vocabulary recognition because of the low scores obtained on the L2 vocabulary recognition tests, as well as the limited sample size in each experimental and control groups. More importantly, it is also conceivable to assume that no statistically significant interaction between WM measures and L2 vocabulary recognition tests was found because of the limited number of high spanners, who all happened to concentrate in one of the three groups herein investigated. Finally, it is also plausible that the WM tests and other measures used in the present study were not able to capture the complexities involved with actual processing of subtitled video and therefore significant correlations may not have surfaced due to the nature of the tests adopted.

Considering the results and the aspects discussed in this subsection, in relation to *(RQ4): Does leaners' WM capacity correlate with their performance on the video comprehension test and/or their performance on the L2 vocabulary recognition test in any subtitling condition?,* it is possible to conclude that WM capacity did not significantly interact with participants' performance on the tests (video

comprehension and L2 vocabulary recognition) in all treatment and control conditions.

It is, however, imperative to keep in mind that correlation analyses cannot be interpreted as establishing cause-and-effect relationships. It is my understanding that correlation analyses allow, with some degree of caution, to indicate the level of association among variables in a given experimental scenario. In this sense, the results herein obtained demonstrate that the variables investigated did not interact significantly since the results obtained with the correlation tests yielded no significant interactions. It is very well possible that WM could still be much related to participants' performance on the tests they performed in lesser or greater degree. Different results might be obtained with larger sample sizes and modifications in the data collection procedures and/or instruments, for instance.

This Chapter has focused exclusively on highlighting and discussing the major aspects that have emerged from the results of the present research. The next and final Chapter of this doctoral dissertation revisits the major findings of this study. It also presents limitations and suggestions for further research, as well as theoretical, methodological, and pedagogical implications of this type of investigation, taking into account the results and the findings hereby attained. Finally, it ends with a few concluding remarks on the relevance of this piece of research, which may trigger future works on issues related to the effects of subtitling upon L2 development with Brazilian EFL learners.

CHAPTER VII
CONCLUSION

This Chapter summarizes the most important findings of this doctoral dissertation as well as its major implications. In subsection 7.1, the main findings of this study are outlined. Subsection 7.2 focuses on the limitations of the present investigation and suggestions for further research. Subsection 7.3 addresses theoretical methodological implications and subsection 7.4 discusses pedagogical implications. Finally, concluding remarks are offered in subsection 7.5.

7.1 Main Findings

The present study, with a mixed design nature, focusing on qualitative and quantitative analyses, was aimed at investigating the effects of the use of intralingual and interlingual subtitles on Brazilian EFL learners' L2 development. More precisely, this study was aimed at examining whether intralingual and interlingual subtitles aid video comprehension – in terms of general and specific comprehension – as well as whether they aid L2 vocabulary recognition. Additionally, this study set out to verify whether learners' level of video comprehension and L2 vocabulary recognition correlates with their WM capacity in any subtitling condition.

As far as those objectives are concerned, the present doctoral dissertation obtained the following major findings:

- Intralingual and interlingual subtitles' effects on video comprehension: The results of the statistical tests and the analyses performed indicate that there were more beneficial effects on general and specific video comprehension when using intralingual subtitles over interlingual subtitles or control conditions. This is evidenced in the participants' mean scores obtained on the tests. Nonetheless, participants' performance in the experimental conditions – intralingual subtitles and interlingual subtitles – on both general and specific portions of the video comprehension test were not statistically different from one another. Yet, participants in the intralingual subtitles condition did outperform those in the control group and the differences between these two groups were found to be statistically significant.

- Intralingual and interlingual subtitles' effects on immediate L2 vocabulary recognition: In the present study, intralingual and interlingual subtitles did not substantially foster L2 vocabulary recognition, especially when observing the participants' performance on the L2 vocabulary recognition pre-test, test, and post-test. Moreover, no statistically significant differences between the experimental groups – intralingual and interlingual subtitles – and the control group were found as regards their performance on the L2 vocabulary recognition pre-test, test, and post-test. The lack of statistically significant results on these tests is possibly mostly linked to an absence of salient and/or comprehensible input in the instruments used for data collection.

- Intralingual and interlingual subtitles' effects on L2 vocabulary recognition across time: The results obtained on the L2 vocabulary recognition pre-test, test, and post-test point out to a trend that reveals mostly a growth in performance by the participants. This was confirmed by the test/pre-test and post-test/test gain score comparisons. Moreover, the intralingual subtitles group presented more positive gains in L2 vocabulary recognition across time, followed by the interlingual subtitles group and finally the control group. However, the results of the statistical tests performed revealed no significant differences among groups across time.

- WM capacity, subtitling, video comprehension, and L2 vocabulary recognition: In light of the statistical tests and the analyses performed, it is possible to state that WM capacity did not significantly interact with participants' performance on the tests (video comprehension and L2 vocabulary recognition) in all treatment and control conditions in the present investigation. I hypothesize that this may be in part due to different processing demands by the WM tests and the other tests (video comprehension and L2 vocabulary recognition) adopted in the present research or the types of tasks themselves. Perhaps more importantly, I also hypothesize that a lack of significant interaction among the variables may also be attributable to the small sample size of this study and a lack of even distribution of higher/lower spanner across groups.

7.2 Limitations and Suggestions for Further Research

Despite the fact that this study has been grounded in the literature in the area of SLA as an attempt to establish a conversation with cognitive and subtitling aspects, some limitations surfaced in the course of the implementation of this study, which are now properly acknowledged. First, participants were not specifically inquired about their central focus of attention while engaged in video watching activities in and outside classroom, even though some of them may have commented on that aspect when meeting up with the research for the WM data collection session.

Second, this study did not include a thorough investigation in terms of the context in which the target-words appear in the video, including factors that could either facilitate or hinder their processing and subsequent recognition. In other words, no analysis of the multisemiotic environment of the video was performed to inspect audio/video correlation.

Third, another important factor regards the instruments used for data collection, especially the one that aimed to tap participants' L2 vocabulary recognition ability. This study required participants to explain, translate or provide a synonym to the target-words, which certainly goes a little beyond pure recognition in its strictest sense. No vocabulary knowledge scales were used, so it is rather difficult to determine whether participants did recognize more words but were unable to provide an explanation, a translation or a synonym to them. Additionally, the input provided contained little enhancement or frequency, so this may have determined participants' (in)ability to recognize more L2 target-words.

Fourth, one aspect that is left undiscussed here refers to the examination of which target-words were more often recognized and which ones were less often recognized. The researcher did not carry out such an analysis due to the low scores obtained on the L2 vocabulary recognition tests.

Fifth, in dealing with quantitative research, sample sizes are very likely to play a role in determining the results, especially when correlation tests are performed. As demonstrated in the WM tests, when run together, the correlations amongst RST and OSPAN tests for all groups did indicate significant moderate positive correlations, whereas the results of separate RST and OSPAN correlation tests for each group did not. It might be possible that this study failed to find more robust results on some of the tests due to a limitation in the sample size. Larger sample sizes may

generate different results for the same instruments adopted in the present investigation. Moreover, narrow score ranges in the tests may have also influenced the absence of more robust results, specially when using correlation tests.

Sixth, the emergence of higher spans within the same group was a variable for which it was not controlled. Participants' WM capacity was measured throughout the three weeks as participants took the classroom (video) tests. Participants' WM capacity could have been tested prior to organizing the experimental and control groups to avoid having an uneven number of higher spans in relation to lower spans within the same group.

Lastly, proficiency was not systematically assessed by the researcher, although the students were all enrolled in the same English level and had previously been administered an in-house placement test at their admission in the Extracurricular non-credit Language Courses at UFSC. It is, however, extremely important to keep in mind that statistical tests run on the L2 vocabulary pre-test confirmed that the groups were not statistically different from start and participants, when enrolling the language courses, had been administered an in-house placement test.

Considering the aforementioned limitations, a number of suggestions for further research spring to mind, all of which are now mentioned. First, more studies should be carried out to investigate the continuous use of subtitled material for L2 development. As Vanderplank (2015) states, close to zero is known about the longitudinal effects of subtitled upon L2 development in any linguistic domain.

Second, more studies should be designed and implemented with the Brazilian population of different proficiency levels (beginning and advanced learners, for instance) to investigate the effects of intralingual and interlingual subtitles upon their development. As I stated in the very beginning of this doctoral dissertation, the present study, in addition to Matielo et al. (2013), represents one if not the very first attempt to provide insights into the effects of the use of subtitling for language learning purposes, taking into account Brazilian EFL learners.

Third, questionnaires or semi-structured interviews should be used to gather relevant information on what actually tends to constitute the focus of attention of learners when watching subtitled video materials. Similarly, more systematic investigations should be carried out to examine the extent to which different input modalities get processed, apart from subtitles per se. To date, very little research has been carried out to scrutinize if viewers do process the soundtrack in its entirety along with the subtitles (Bird & Williams, 2002), since reading subtitles has

been suggested to be automatic, at least by adults (d'Ydewalle & van Rensbergen, 1989; d'Ydewalle et al., 1991).

Fourth, eye-tracking methodologies might provide invaluable insights into learners' reading behavior regarding subtitles and the eye movements around the video area. As demonstrated in the very few eye-tracking studies on subtitling and language learning reported in Chapter II, examining eye movements may tell us much about the architecture of the cognitive system, especially about first fixation duration, first-pass duration (before moving to another area or looking back), regression path duration (time from entering a region until moving to another region), second-pass duration (duration of re-fixations, indicating a second reading), and total fixation, for instance. Eye-tracking methodologies applied to subtitling studies may give us decisive information about one's processing of subtitles as well as the video itself.

Fifth, future studies should consider designing, adapting, and implementing different WM tests that may provide extra relevant information to expand on the investigation of the relationship between WM capacity and learners' L2 development in the context of subtitling for language learning purposes.

Sixth, more studies need to take into consideration students' accounts concerning their experiences, strategies, attitudes, and reactions towards watching subtitled video materials, with intralingual and/or interlingual subtitles. Investigations like these should also include an analysis of learning styles, which might be insightful for this kind of research, especially considering the use of subtitled videos outside the classroom, while learners are watching videos in their leisure time and are still exposed to linguistic input that could become intake.

Seventh, future research should include test announcement prior to video watching to verify if that variable interacts with participants' performance on video comprehension and L2 vocabulary recognition tests aided by subtitling.

Finally, it is relevant to consider researching different viewing conditions for all participants in future research. The use of headphones and standard television/speaker sets should be compared and contrasted.

7.3 Methodological Implications

Three methodological implications have arisen in the course of the implementation of this doctoral investigation, taking into account the literature upon which it was based and the results herein obtained. Firstly, an important methodological implication regards how L2 vocabulary

development has been measured in the literature of subtitling and L2 development. As clearly demonstrated in Chapter II, a variety of methods have been employed to gather data on learners' L2 lexical development, some of which are functioning freely under the huge umbrella term 'L2 vocabulary development' or 'L2 vocabulary recognition' or even 'L2 vocabulary learning/acquisition'. Without having the ambition of prescribing how L2 vocabulary development in subtitling-based studies should be measured, I would argue that the use of mere immediate recognition instruments might provide us with limited information about one's L2 lexical development, which is constrained to its onset or its first step towards acquisition.

Additionally, I would suppose that L2 vocabulary development should be measured by requiring learners not only to inform whether they recognize words from the input they are provided, but also encourage them to attempt to attribute meaning to them. I imagine that form-meaning connections should be fostered even under experimental conditions because this may be informative of whether learners are indeed able to go a little beyond processing and recognizing word forms. Having learners to attribute a translation, a meaning or an explanation to target-words are far more likely to provide us with evidence that they are in the process of being learned.

Secondly, the use of vocabulary knowledge scales, as pointed out in Chapter III and again in Chapter VI, may be somewhat problematic at times depending on how their designed. However, perhaps by adapting Wesche and Paribakht's (1996) vocabulary knowledge scales and including a clear focus on meaning as regards the target-words, one may then tap learners' higher level of processing of these lexical items, which might be safer than solely relying on learners' account or their perceptions of whether they remember having seen a certain word before the treatment. Also, including statements into the scales that would aim to verify whether the target-words are processed/recognized via auditory (soundtrack) or written (subtitles) input may also be enlightening, since this would clearly indicate which input channel is mostly attended by learners and therefore more often processed.

Thirdly, a crucial methodological implication observed concerns the data collection design. In the present study, a one week-gap between pre-test, test, and post-test was adopted (for the L2 vocabulary recognition tests). Consequently, participants who missed one of those three sessions were automatically removed from the sample pool. What could be done instead is to collect participants' pre-test and test data on the very same session and their post-test data on the same week. This might ensure fewer

rates of participant mortality. Nonetheless, this would also bring up implications related to long-term memory, since the researcher wanted to verify whether participants would remember the target-words they presumably recognized one week after they had been administered the immediate test.

7.4 Pedagogical Implications

Taking into consideration the goals of this doctoral dissertation, the materials adopted, the methodological steps undertaken, the literature that circumstantiates it, and the results hereby obtained, I believe that there are at least five ways in which the present research can inform L2 pedagogy. To begin with, given the expanding use of subtitled audiovisual materials in and outside L2 classrooms and the technological advances we are currently witnessing, I find it crucial for L2 practitioners to help learners develop strategies to benefit from subtitled video materials. This means that L2 instructors should consider helping learners to cope with possible cognitive (over)load that these materials may present, especially when any textual aid (subtitles) is used. Directing learners' attention to specific features of the video material presented may be a good start to achieve that.

Secondly, a very sensitive issue that has almost been neglected in the literature regards the notion that subtitles are in fact being used as crutches. The goal behind the use of subtitled video materials is to ensure that learners who have not yet mastered a certain L2 may still benefit from that material in terms of its comprehension. However, there might come a time when subtitles may have to be abandoned. Very little is discussed in the literature in terms of whether the constant use of subtitled video materials effectively contribute to one's listening comprehension skills development or whether they cause some kind of 'dependence'.

This issue was harshly debated among Markham and Peter (2003), Stewart and Pertusa (2004), Markham (2005), and finally in Stewart and Pertusa (2005), who clearly adopt very different stances as regards the constant use of subtitled video materials for L2 instruction. Essentially, Stewart and Pertusa's argument, as opposed to Markham and Peter's, is that instead of providing learners with frequent subtitled materials with either interlingual or intralingual subtitles, L2 instructors should be more concerned with providing them with more videos that present comprehensible input. Even though I tend to agree with the scholars, it is important to keep in mind that sometimes, very/highly comprehensible input is mostly found in unauthentic video materials as opposed to more

challenging input that is found in authentic video materials. Perhaps a suggestion that is in order is that L2 practitioners may design activities or tasks that do not require the processing of the videos in their entirety, especially when more challenging input is provided.

Thirdly, a pivotal pedagogical aspect that has arisen regards the importance of instruction to foster L2 vocabulary development. As clearly demonstrated in the results obtained in the present investigation, guiding learners' attention is perhaps a key ingredient that may translate into success in their path to learn novel vocabulary. Moreover, this may prove more efficient with videos that do not provide learners with enhanced input as regards target L2 vocabulary. Considering what has been repeatedly shown in the literature in relation to the number of exposure to novel vocabulary so as to learn/acquire it, I also find it imperative that L2 teachers should design activities or tasks that can increase learners' awareness in terms of novel words. Furthermore, when using (subtitled) videos that present insidious or unenhanced input to foster L2 vocabulary pick-up, it might be strategic on the teachers' part to combine the video input with some instruction.

Similarly, I would argue that L2 practitioners should encourage learners to develop their own/new strategies to learn novel vocabulary. Among them, it might be beneficial for L2 learners to start their own glossaries instead of just looking up words in the dictionary, taking notes of their meaning or translation. Coming up with their own glossaries either in paper or electronic format could be done both in and outside learners' classroom. It is, however, extremely important that L2 teachers recycle the target-items with which they have previously worked in the classroom so learners come in contact with them again more often, even with certain lexical items that were very salient in the input provided in the subtitled video materials.

Fourthly, an important pedagogical implication derived from this research concerns learners' constraints in terms of their WM capacity. Regardless of the fact that the results obtained with the correlation tests did not show a statistically significant relationship among participants' WM capacity and their performance on other experimental tests, it is still valid to consider how many different and highly demanding cognitive processing mechanisms are at play when one watches a subtitled video material. Thus, it seems reasonably important for L2 teachers to beware of possible ways to reduce cognitive load in the scenario of multimedia learning. Mayer and Moreno (2003) lists nine ways in which one can do that, some of which are not necessarily applicable to the present context.

Nonetheless, the gist underlying their principles refers to reducing extra, non-essential information from learners' focus of attention.

In this sense, selecting material that presents information in a sequential, clear, and digestible manner might be a good idea. Also, when selecting subtitled material that includes challenging topics, it may be important for L2 teachers to verify whether the speech in the video is not too fast so learners may follow it either by attending to the auditory channel or by reading the subtitles that contain its counterpart. Another important aspect to be highlighted is that reducing extra, non-essential information from learners' focus of attention may be achieved by directing their attentional resources by means of a pre-established goal prior to video watching. In doing so, learners might concentrate their noticing capacity or attention to gather relevant information to accomplish a given task. Task/test announcement may therefore yield more beneficial effects upon learners' cognitive processing of a given subtitled video because they may deliberately free more of their attentional resources and allocate them onto the processing of pertinent sources or pieces of information, especially in the case of lower WM spanners.

Finally, even though the results of the comparisons between the experimental groups – intralingual subtitles and interlingual subtitles – were not herein found to differ from a statistical point of view, on both the video comprehension and the L2 vocabulary recognition tests, it was possible to observe more beneficial effects when intralingual subtitles were available to the participants. Based on these results and the ones reported in Chapter II, in the subtitled and language learning literature, I find it relevant for L2 instructors to consider the choices of textual aid available to be provided to their learners.

It is noteworthy to remember that the participants in the present study informed that they were more used to watching subtitled materials with interlingual subtitles rather than intralingual subtitles, but their apparent attitudes towards and comments on the retrospective questionnaires about intralingual subtitles suggest that they feel positive about them. In this vein, it may be interesting for L2 practitioners to unveil their learners' beliefs and attitudes towards the use of translational aids, asking them about their viewing habits and/or preferences. Also, it may be an interesting idea to have their learners perform activities or tasks in and outside the classroom alternating between the use of intralingual subtitles and interlingual subtitles and subsequently asking them to report how they feel about them. This might provide both L2 practitioners and researchers with some insightful information about the language learning

outcomes throughout a semester or year, which may then contribute towards a more solid understanding of possible long-term effects of subtitling exposure for L2 learning purposes.

7.5 Concluding Remarks

Having carried out the present doctoral study and written these couple of hundred pages have made me reflect thoroughly about this path, from an empirical, an instructional, and a learning point of view. As a researcher, I am more and more convinced that further investigation has yet to be carried out in order to shed light on aspects that so far have been left undiscussed here. The title of this doctoral dissertation, "Intralingual Subtitles, Interlingual Subtitles, and L2 Vocabulary Learning: An Exploratory Study with Brazilian EFL Students", encapsulates perhaps the most important and poignant aspect of this study, which stems from the fact that this research has only scratched the surface of the possible effects that subtitles may have upon L2 development as far as the Brazilian population of EFL learners are concerned, given its exploratory nature. Moreover, it is imperative to keep in mind that the present exploratory work was the very first to examine a possible relationship among WM capacity and one's performance on subtitled video materials' tasks.

I acknowledge that different choices would possibly have led to different results. Methodological modifications in the instruments used for data collection are open to be made; some of them are in fact required if one wishes to go on and continue exploring the effects of subtitles for language learning purposes, as just pointed out in the limitations and suggestions for further research in this Chapter. Other aspects revolving around the language-learning domain could and should be considered when designing instruments to collect data with this and other populations, especially if we consider the potentials that subtitled audiovisual materials can offer to both L2 instructors and L2 learners, within and outside their educational environment. On that note, I encourage fellow researchers to join me in continuing with the systematic exploration of the effects of subtitling upon L2 development with this population that has been disregarded to date.

As an L2 practitioner myself, who has been teaching EFL learners for more than ten years, I now look back on this study's results and feel more empowered to make more informed choices when designing activities and tasks or choosing subtitled audiovisual materials to be used in my classes. On the one hand, having had classroom experiences that

have translated into not so successful learning outcomes seems to make much more sense today in light of the knowledge of the body of research that I came to obtain. On the other hand, certain classroom experiences that might have gone fairly well suddenly feel much more assured and validated when looked at from an empirically-oriented point of view. Questions and issues related to successful L2 vocabulary pick-up or lack thereof, for instance, are closer to being less obscure for me.

As an L2 learner myself, who has been in contact with the English language for more than two decades now, and who also happens to very constant make use of (subtitled) video materials for his own linguistic development – especially North-American TV series –, some issues related to my ability to acquire features of the language, English being the case in point, are still very latent and will most definitely very soon constitute future avenues for my own research agenda.

To conclude, I believe that the present doctoral dissertation has been able to attain its goals in addressing relevant questions regarding the effects of intralingual and interlingual subtitles upon L2 development. I also believe that the theoretical discussions hereby presented may contribute to establishing a fruitful dialogue amongst the many fascinating areas and research niches within the disciplinary field of SLA. Finally, and perhaps most importantly, I hope that the results of the experiments that gave birth to this doctoral study may positively influence upcoming research and help L2 practitioners to reflect and make knowledgeable choices as regards their L2 instruction using subtitled video materials.

REFERENCES

Ahmed, M. O. (1989). Vocabulary learning strategies. In P. Meara (Ed.) *Beyond Words* (pp.3-14). London: BAAL, in association with Centre for Information on Language Teaching and Research.

Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics, 29,* 175-193.

Anderson, D., & Collins, P. (1988). *The impact on children's education: Television's influence on cognitive development*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Atkinson, D. (1987). The mother tongue in the classroom: a neglected resource? *English Language Teaching Journal, 41*(4), 241-247.

Atkinson, R.C.; Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In Spence, K.W.; Spence, J.T. *The psychology of learning and motivation* (Vol. 2). New York: Academic Press. pp. 89–195.

Auerbach, E. R. (1993). Reexamining English only in the ESL classroom. *TESOL Quarterly, 27*(1), 9-32.

Baars, B. J. (1983). Conscious contents provide the nervous system with coherent, global information. In R. Davidson, G. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (pp. 45–76). New York: Plenum.

Baars, B. J. (1986). *The cognitive revolution in psychology*. New York: Guilford.

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556-559.

Baddeley, A. D. (1997). *Human Memory: Theory and Practice*. 2nd ed. Bristol, UK: Psychology Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417-423.

Baddeley, A. D. (2003). Working Memory: Looking back and looking forward. *Nature Review Neuroscience, 4*, 829-839.

Baddeley, A. D. (2009). Short-term Memory. In Baddeley, A. D., Eysenck, M., & Anderson, M. C. (2009). Memory. Hove: Psychology Press.

Baddeley, A. D. (2011). Estimating the short-term component in free recall. *British Journal of Psychology, 61*(1), 13-15.

Baddeley, A. D., Gathercole, S. E. & Papagno, C., (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158-173.

Baddeley, A.D. & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), The psychology of *learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology, 8*, 485-493.

Baddeley A.D. & Logie, R. H. (1999). Working memory: the multiple-component model. In: *Models of working memory* (Miyake A, Shah P, eds), pp 28–61. Cambridge, UK: Cambridge University Press.

Bailer, C. (2011). *Working memory capacity and attention to form and meaning in EFL reading*. Dissertação (Mestrado em Língua Inglesa e Linguística Aplicada) – Pós-graduação em Letras/Inglês e Literatura Correspondente, Universidade Federal de Santa Catarina, Florianópolis.

Bailer, C., Tomitch, L. M. B., D'Ely, R. C. S. F. (2013). Working Memory Capacity and Attention to Form and Meaning in EFL Reading. *Letras de Hoje, 48*(1), 139-147.

Bianchi, F. & Ciabattoni, T. (2008). Captions and subtitles in EFL learning an investigative study in a comprehensive computer environment. Retrieved Jan 2012,

http://www.openstarts.units.it/dspace/bitstream/10077/2848/1/bianchi_ciabattoni.pdf.

Bird, S. A. & Williams, J. N (2002). The effect of bimodal input on implicit and explicit memory: an investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23*(4)*,* 509-533.

Bishop, D. (1982). *Test for reception of grammar (T. R. O. G.)* Printed for the Medical Research Council by Thomas Leach, Abindgon, Oxon, UK.

Bisson, M., Van Heuven, W. J. B., Conklin, K., Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics, 35*(2), 399-418.

Broadbent, D (1958). *Perception and Communication*. London: Pergamon Press.

Brown, C. (1993). Factors affecting the acquisition of vocabulary: Frequency and saliency of words. In T. Huckin & M. Haynes & J. Coady (Eds.), *Second language reading and vocabulary learning.* Norwood, N.J.: Ablex.

Budd, D., Whitney, P., & Turley, K. J. (1995). Individual differences in working memory strategies for reading expository text. *Memory & Cognition, 23*, 735–748.

Caimi, A. (2006). Audiovisual Translation and Language Learning. The Promotion of Intralingual Subtitles. *Journal of Specialised Translation 6*, 85-98.

Chang, S. (2006). The Interaction Between Schemata and Subtitles. Journal of National Taipei *University of Technology, 39*(1), 209-227.

Chen, Z., Cowan, N. (2009). Core verbal working memory capacity: The limit in words retained without covert articulation. *Quarterly Journal of Experimental Psychology*, *62,* 1420-1429.

Cheung, H. (1996). Nonword span a unique predictor of second language vocabulary learning. *Developmental Psychology, 32,* 867-873.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta, 49*(1), 67-77.

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research*, *20*, 445–464.

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.

Daneman, M. & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 561–584.

Daneman, M. & Merikle, P. M. (1996). Working Memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review, 3*(4), 422–433.

de Bot, K., Paribakht, S. & Wesche, M. B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition (evidence from ESL reading). *Studies in Second Language Acquisition, 3,* 309-329.

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science, 11*, 467–473.

Dunn, L. M., Dunn, L. M., Whetton, C. & Pintilie, D. (1982). *British Picture Vocabulary Scale.* Windsor: NFER-Nelson.

d'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In K. Rayner （Ed.), *Eye*

movements and visual cognition: Scene perception and reading (pp. 415–427). New York: Springer–Verlag.

d'Ydewalle, G., Praet, C., Verfaillie, K., & Van Rensbergen, J. (1991). Watching subtitled television: Automatic reading behavior. *Communication Research, 18*, 650–666.

d'Ydewalle, G., & van de Poel, M. (1999). Incidental foreign–language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research, 28,* 227–244.

d'Ydewalle, G., & Van Rensbergen, J. (1989). Developmental studies of text–picture interactions in the perception of animated cartoons with text. In H. Mandl & J. R. Levin (Eds.), Knowledge acquisition from text and pictures (pp. 233–248). Amsterdam: Elsevier Science Publishers B.V. (North–Holland).

Elliot, C. D., Murray, D. J. & Pearson, L. S. (1983). *British Ability Scales*. Windsor: NFER-Nelson.

Ellis, N. (1994a). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. *AILA Review, 11*, 37-56.

Ellis, R. (1997). *SLA Research and Language Teaching*. Oxford: Oxford University Press.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 63- 103). Malden, MA: Blackwell.

Ellis, R. (2008). *The study of second language acquisition*. 2nd Ed. Oxford: Oxford University Press.

Engle, R. W., Kane, M. J. & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In Miyake, A. & Shah, P. (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp.102-134). London: Cambridge Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., Conway, A. R. (1999). Working memory, short-term memory and general fluid intelligence: a latent variable approach. *Journal of Experimental Psychology: General, 128*(1), 309 – 331.

Erten, I. & Tekin, M. (2008). Effects of vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System, 36*, 407-422.

French, L. M. & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics, 29,* 463-487.

Friedman, N. P. & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of Experimental Psychology: General, 133*(1), 101 – 135.

Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals, 24*(3), 239-258.

Gass, S. (1999). Discussion: Incidental vocabulary acquisition. *Studies in Second Language Acquisition, 21,* 319-333.

Gass, S. & Selinker, L. (1994). *Second language acquisition: An introductory course*. Hillsdale, NJ: Lawrence Erlbaum.

Gathercole, S. E. & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81*, 439-454.

Grimes, T. (1991). Mild auditory-visual dissonance in television news may exceed viewer attentional capacity. *Human Communication Research, 18*(2), 268–298.

Harrington, M. & Sawyer, M. (1992). L2 Working Memory Capacity and L2 Reading Skill. *Studies in Second Language Acquisition, 14*(1), 25-38.

Hayati, A. & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of ESL learners. *British Journal of Educational Technology 42*(1), 181-192.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*, 207-223.

Horst, M. & Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *The Canadian Modern Language Review, 56*, 308-328.

Hsiao, T. Y., & Oxford, R. (2002). Comparing Theories of Language Learning Strategies: A confirmatory Factor Analysis. *The Modern Language Journal, 86*, 368-383.

Huang, H., & Eskey, D. E. (1999). The effects of closed-captioned television on the listening comprehension of intermediate English as a second language (ESL) students. *Journal of Educational Technology Systems, 28,* 75-96.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition, 21*, 181-193.

Hummel, K. M. (2010). Translation and short-term L2 vocabulary retention: Hindrance or help? *Language Teaching Research, 14*(1)*,* 61-74.

Jiang, N. (2000). Lexical representation and development in second language. *Applied Psycholinguistics, 21*(1)*,* 47-77.

Kane, M. J., Beckley, M. K., Conway, A. R., Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology, General, 130*(2), 169-183.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working-memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, J. Towse (Eds.), (2007). *Variation in Working Memory*. Oxford University Press.

Kempen, G. & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science, 11*(2), 201-258.

Kempen G, Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14,* 185–209.

Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments, & Computers, 31*, 429-432.

Koosltra, C. M. & Beentjes, J. W. J. (1999). Children vocabulary acquisition in a foreign language through watching subtitled television programs at home. *ETR&D, 47*(1), 51-60.

Koskinen, P. S., Knable, J. E., Markham, P. L., Jensema, C. J., & Kane, K. W. (1996). Captioned Television and the Vocabulary Acquisition of Adult Second Language Correctional Facility Residents. *Journal of Educational Technology Systems, 24*(4), 359-373.

Krashen, S. (1982/2009). *Principles and Practice in Second Language Acquisition.* Pergamon Press.

Kruger, J. & Steyn, F. (2013). Subtitles and Eye Tracking: Reading and Performance. *Reading Research Quarterly, 49*(1), 105-120.

Lado, R. (1956). Patterns of difficulty in vocabulary. *Language and Learning, 6*(1), 23-41.

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication 50*(1), 46-70.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS.* New York: Routledge.

Latifi, M., Mobalegh, A. & Mohammadi, M. (2011). Movie Subtitles and the Improvement of Listening Comprehension Ability: Does it help?. *Journal of Language Teaching and Learning, 1*(2), 18-29.

Laufer, B. (1997). What's in a word that makes it hard or easy? Intralexical factors affecting the difficulty of vocabulary acquisition. In *Vocabulary Description, Acquisition and Pedagogy,* eds. M. McCarthy and N. Schmitt. Cambridge University Press. pp. 140-155.

Laufer, B., & Hadar, L. (1997). Assessing the effectiveness of monolingual, bilingual, and "bilingualised" dictionaries in the comprehension and production of new words. *The Modern Language Journal, 81*, 189-196.

Lucindo, E. S. (2006). Tradução e Ensino de Línguas Estrangeiras. *Scientia Traductionis 3,* 1-11.

Markham, P. L. (2005). More on caption use: A responde to Stewart and Pertusa. *Foreign Language Annals, 38*(2), p. 283.

Markham, P. L. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals, 32*(3), 321-328.

Markham, P. L., & Peter, L. (2003). The influence of English language and Spanish language captions on foreign language listening/reading comprehension. *Journal of Educational Technology Systems, 31*(3), 331-341.

Markham, P. L., & Peter, L., & McCarthy, T. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals, 34*(5)*,* 439-445.

Martin, K. I. & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*(3), 379-413.

Masoura, E. & Gathercole, S. E. (1999). Phonological short-term memory and foreign vocabulary learning. *International Journal of Psychology, 34*, 383-388.

Matielo, R., Collet, T., D'Ely, R. C. S. F. (2013). The effects of interlingual and intralingual subtitles on vocabulary learning by

Brazilian EFL learners: An exploratory study. *Revista Intercâmbio, v. 27*, 83-99.

Matielo, R., D'Ely, R. C. S. F., Baretta, L. (2015). The effects of interlingual and intralingual subtitles on second language learning/acquisition: a state-of-the-art review. *Trabalhos em Linguística Aplicada, v. 54*, 161-182.

Mayer, R., & Moreno, R. (1998). A Split-Attention Effect in Multimedia Learning: Evidence for Dual Processing Systems in Working Memory. *Journal of Educational Psychology, 90(2)*, 312-320.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*, 187–198.

McGeoch, J. A. & McDonald, W. T. (1931). Meaningful relation and retroactive inhibition. *American Journal of Psychology, 43,* 579-588.

Meara, P. (1997). Towards a new approach to modelling vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 109-121). Cambridge: Cambridge University Press.

Melton, A.W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behaviour, 2,* 1-21.

Milton, J. (2008). Vocabulary uptake from informal learning tasks. *Language Learning Journal, 36*(2)*,* 227-237.

Miyake, A. & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, Massachusetts: Heinle & Heinle Publishers.

Nation, I. S. P. (1994). *New Ways in Teaching Vocabulary*. Alexandria, VA: TESOL.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1), 59-82.

Nation, I.S.P. and Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Neuman, S.B. (1989). The impact of different media on children's story comprehension. *Reading Research and Instruction, 28,* 38-47.

Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly, 27*, 94-106.

Norman, D. A. & Shallice, T. (1980). *Attention to action: Willed and automatic control of behaviour*. Center for Human Information Processing (Technical Report No. 99). University of California, San Diego.

Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology*, *13(3)*, 243–272.

Perez, M. M., Peters, E., Desmet, P. (2013). Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL, 26*(1), 21-43.

Perez, M. M., Peters, E., Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal, 99*(2), 308-328.

Prebianca, G. V. V. (2009) *Working memory capacity, lexical access and proficiency level in L2 speech production*. Tese (Doutorado em Língua Inglesa e Linguística Aplicada) – Pós-graduação em Letras/Inglês e Literatura Correspondente, Universidade Federal de Santa Catarina, Florianópolis.

Price, K. (1983). Closed-captioned TV: An untapped resource. *MAT-SOL Newsletter, 12*(2)*,* 1-8.

Raine, P. (2013). Incidental increase in depth of vocabulary knowledge through the viewing of subtitled, authentic videos. In N. Sonda & A. Krause (Eds.), *JALT 2012 Conference Proceedings. Tokyo: JALT*, 492-505.

Raven, J. C. (1984). Manual for the Coloured Progressive Matrices (Revised). Windsor, UK: NFER-Nelson.

Read, J. (2000). *Assessing vocabulary.* Cambridge, MA: Cambridge University Press.

Repovš, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience, 139,* 5-21.

Robinson, P. & Gilabert, R. (2007). Task complexity, the cognition hypothesis and second language learning and performance. *International Review of Applied Linguistics, 45*(3), 161-176.

Rokni, S. J. A. & Ataee, A. J. (2014). The Effect of Movie Subtitles on EFL Learners' Oral Performance. *International Journal of English Language, Literature and Humanities 1*(5), 201-215.

Romanelli, S. (2006). Traduzir ou não traduzir em sala de aula? Eis a questão. *Inventário - Revista dos Estudantes do Programa de Pós-graduação em Letras e Lingüística da UFBA*, *5,* 1-10.

Romanelli, S. (2009). O uso da tradução no ensino-aprendizagem das línguas estrangeiras. *Revista Horizontes de Linguística Aplicada, v. 8*(2), 200-219.

Rutherford, W., & Smith, M. S. (1985). Consciousness-raising and universal grammar. *Applied Linguistics 6,* 274-82.

Sanaoui, R. (1995). Adult learners' approaches to learning vocabulary in second languages. *The Modern Language Journal, 79,* 15-28.

Schmidt, R. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics, 11*, 129-158.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker, *Proceedings of CLaSIC 2010, Singapore, December 2-4* (pp. 721-737). Singapore: National University of Singapore, Centre for Language Studies.

Schmitt, N. (1997). Vocabulary learning strategies. In Schmitt, N. and McCarthy, M. (eds.), *Vocabulary: Description, Acquisition, and Pedagogy*. Cambridge University Press.

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1074-1083.

Schweers, C.W. (1999). Using L1 in the L2 classroom. *English Teaching Forum, 37*(2), 6-13.

Segalowitz, N., & Segalowitz, S. J. (1993). Skilled performance, practice and the differentiation of speedup from automatization effects. *Applied Linguistics, 14*, 369–385.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology, 45A*, 21-50.service

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society, London B, 298*, 199-209.

Sharif, M . R. & Ebrahimian, M. (2013). A Comparative Study of 'Subtitled' vs. 'Auditory' Documentaries Comprehension among Persian Speakers of English. *International Journal of Applied Linguistics & English Literature, 2*(2), 185-192.

Souza, J. P. (1999). Tradução e Ensino de Línguas. *Revista do GELNE, 1*(1), 141-151.

Stewart, M.A. & Pertusa, I. (2004). Gains to language learners from viewing target language closed-captioned films. *Foreign Language Annals, 37*(3)*,* 438-447.

Stewart, M.A. & Pertusa, I. (2005). Responding to Markham's comments. *Foreign Language Annals,38*(2)*,* 284.

Sydorenko, T. (2010). Modality of Input and Vocabulary Acquisition. *Language Learning & Technology 14*(2), 50-73.

Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals, 38*(3), 422-427.

Tomitch, L.M.B. (2003a) *Reading: text organization perception and working memory capacity*. Florianópolis, SC: PGI/UFSC, ARES - Advanced Research in English Series.

Torres, A. C. G. (2003) *Working memory capacity and reader's performance on main idea construction in L1 and L2*. Tese (Doutorado em Língua Inglesa e Linguística Aplicada) – Pós-graduação em Letras/Inglês e Literatura Correspondente, Universidade Federal de Santa Catarina, Florianópolis.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127 – 154.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498–505.

Unsworth, N., & Spillers, G.J. (2010). Working memory capacity: Attention, Memory, or Both? A direct test of the dual-component model. *Journal of Memory and Language, 62,* 392-406.

Van Lommel, S., A. Laenen & G. D'Ydewalle (2006). Foreign grammar acquisition while watching subtitled television programs. *British Journal of Educational Psychology, 76*, 243-258.

Vanderplank, R. (1988). The value of teletext subtitles in language learning. *ELT Journal, 42*(4), 272–281.

Vanderplank, R. (2015). Thirty Years of Research Into Captions/Same Language Subtitles and Second/Foreign Language Learning: Distinguishing between 'Effects of' Subtitles and 'Effects with' Subtitles for Future Research. In Gambier, Y., Caimi, A., & Mariotti, C. (2015). *Subtitles and Language Learning: Principles, Strategies and Practical Experiences*. Bern, Switzerland: Peter Lang.

Verhagen, J., & Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology, 141,* 65-82.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth vs. breadth. *Canadian Modern Language Review, 53*, 13-39.

Willis, M., & Ohashi, Y. (2012). A model of L2 vocabulary learning and retention. *The Language Learning Journal, 40*(1), 125-137.

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning and Technology, 14*(1), 65-86.

Winke, P., Gass, S., & Sydorenko, T. (2013). Factors Influencing the Use of Captions by Foreign Language Learners: An Eye-Tracking Study. *The Modern Language Journal, 97*(1), 254-275.

You, Y. (2011). Factors in vocabulary acquisition through reading. *Indiana Teachers of English to Speakers of Other Languages Journal, 8*(1)*,* 43-57.

Yuksel, D. & Tanriverdi, B. (2009). Effects of Watching Captioned Movie Clip on Vocabulary Development of EFL Learners. *Turkish Online Journal of Educational Technology, 8*(2), 48-54.

Zanon, N.T. (2006). Using subtitles to enhance foreign language learning. *Porta Linguarum, 6,* 41-52.

Zarei, A. A., & Gilanian, M. (2013). The effect of multimedia modes on L2 vocabulary learning. *International Journal of Management and Humanity Sciences, 2*(S), 1011-1020.

Zarei, A. A. & Rashvand, Z. (2011). The Effect of Interlingual and Intralingual, Verbatim and Nonverbatim Subtitles on L2 Vocabulary Comprehension and Production. *Journal of Language Teaching and Research, 2*(3), 618-625.

Zhou, S. (2004). Effects of visual intensity and audiovisual redundancy in bad news. *Media Psychology, 6*, 237–256.

APPENDIX A

CONSENT FORM

UNIVERSIDADE DO ESTADO DE SANTA CATARINA - UDESC
GABINETE DO REITOR
COMITÊ DE ÉTICA EM PESQUIS
ENVOLVENDO SERES HUMANOS – CEPSH

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

O(a) senhor(a) está sendo convidado(a) a participar de uma pesquisa de doutorado intitulada "Ensino e Aprendizagem de Inglês", que busca estudar os mecanismos envolvidos na aprendizagem de inglês, mais especificamente de vocabulário de inglês como língua estrangeira. Este estudo visa contribuir para o entendimento dos processos envolvidos na retenção de vocabulário em condições específicas de aprendizagem.

Se aceitar participar da pesquisa, O(a) senhor(a) (i) responderá a um questionário no início da pesquisa em sala de aula; (ii) assistirá a um vídeo de 20min de duração em sala de aula; (iii) responderá perguntas de conhecimento de vocabulário em sala de aula em três momentos distintos, com intervalo de uma semana em cada; (iv) realizará dois testes de memória fora da sala de aula em horário a ser acordado com o pesquisador. Ao ser concluída, esta tese será defendida até fevereiro de 2017, e o estudo tornar-se-á público.

Nesse tipo de pesquisa quantitativo-qualitativa não há riscos físicos, entretanto o participante poderá se sentir desconfortável durante a realização dos testes ou poderá sentir cansaço mental durante dos testes de memória. É importante, portanto, salientar que você não precisa

responder a nenhuma questão ao longo da pesquisa que lhe cause desconforto ou qualquer tipo de constrangimento.

Os benefícios e vantagens em participar deste estudo serão obtenção de resultados de testes de memória de trabalho, como bem eventuais ganhos em retenção de vocabulário em inglês ao longo do período de coleta de dados em sala de aula, que terá a duração de três semanas (um encontro cada semana).

As pessoas que estarão acompanhando os procedimentos serão os pesquisadores Me. Rafael Matielo, Dra. Raquel Carolina Souza Ferraz D'Ely (DLLE/UFSC) e Dra. Luciane Baretta (UNICENTRO), que poderão ser contatados a qualquer momento durante o processo de coleta de dados.

Além disso, o(a) senhor(a) poderá se retirar do estudo a qualquer momento, sem qualquer tipo de constrangimento.

Solicitamos a vossa autorização para o uso de seus dados para a produção de artigos técnicos e científicos. As informações fornecidas e o material coletado serão absolutamente confidenciais e não haverá identificação nominal dos participantes, nem divulgação de quaisquer informações que podem revelar sua identidade.

Agradecemos a vossa participação e colaboração.

Assinatura do Doutorando

Assinatura da Orientadora

Assinatura da Coorientadora

Florianópolis, ____ / ____/ _____

Contatos: Rafael Matielo: rafaelmatielo@yahoo.com.br (48 – 9986-9463) – R. Lauro Linhares, 897 – 402C – Porto da Trindade, Trindade, Florianópolis-SC.
Raquel Carolina Souza Ferraz D'Ely: raqueldely@gmail.com (48 – 9989-5806) - Departamento de Língua e Literatura Estrangeiras (DLLE), UFSC, Trindade, Florianópolis-SC.
Luciane Baretta: barettaluciane@gmail.com (42-9957-1241) – Setor de Ciências Humanas, Letras e Artes, UNICENTRO - Campus Santa Cruz: Rua Padre Salvador, 875, Guarapuava-PR.

TERMO DE CONSENTIMENTO

Declaro que fui informado sobre todos os procedimentos da pesquisa e, que recebi de forma clara e objetiva todas as explicações pertinentes ao projeto e, que todos os dados a meu respeito serão sigilosos. Eu compreendo que neste estudo, as medições dos experimentos/procedimentos de tratamento serão feitas em mim, e que fui informado que posso me retirar do estudo a qualquer momento.

Nome por extenso
_____
_____

Assinatura _____ Local: _____ Data: ____/____/____ .

APPENDIX B

STUDENT PROFILE QUESTIONNAIRE

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE COMUNICAÇÃO E EXPRESSÃO – CCE
DEPARTAMENTO DE LÍNGUA E LITERATURA ESTRANGEIRAS – DLLE

Nome: _____
Idade:
Curso de Graduação/Pós-Graduação/Nível de Escolaridade:
Nacionalidade:
E-mail:
Turma:

QUESTIONÁRIO

1. Indique há quanto tempo você estuda inglês:

    a. ( ) Há menos de 02 anos;
    b. ( ) De 2 a 3 anos;
    c. ( ) De 3 a 4 anos;
    d. ( ) Há mais de 4 anos.

2. Você tem contato com inglês (você poderá marcar mais de uma opção, marcando 1 para a alternativa que mais compete à você, 2 para a segunda mais relacionada, e assim sucessivamente):

    a. ( ) Somente na sala de aula, nas aulas de inglês;
    b. ( ) Escutando música;
    c. ( ) Assistindo a filmes/seriados de TV;
    d. ( ) Praticando online, através de jogos, chat, etc;
    e. ( ) Lendo trabalhos/livros acadêmicos;
    f. ( ) Em momentos de lazer/diversão;
    g. ( ) Outros: _____.

3. Você já teve alguma experiência em algum país falante de inglês?

    a. ( ) Não;
    b. ( ) Sim.

4. Em caso afirmativo, por quanto tempo?

    a. ( ) Menos de 1 mês;
    b. ( ) De 1 a 3 meses;
    c. ( ) De 3 a 6 meses;
    d. ( ) De 6 meses a 1 ano;
    e. ( ) Mais de um ano.

5. Você estuda inglês com a finalidade de (você poderá marcar mais de uma opção, marcando 1 para a alternativa que mais compete à você, 2 para a segunda mais relacionada, e assim sucessivamente):

    a. ( ) Desenvolver-se profissional/academicamente;
    b. ( ) Viajar;
    c. ( ) Sem objetivos específicos;
    d. ( ) Outros: _____.

6. Com que frequência você assiste a filmes/seriados de TV:

    a. ( ) Raramente;
    b. ( ) Pelo menos 1 vez por semana;
    c. ( ) De 2 a 3 vezes por semana;
    d. ( ) Todos os dias.

7. Dentre os seriados de TV a seguir, a quais deles você costuma assistir (você poderá marcar mais de uma opção):

    a. ( ) American Horror Story CSI;
    b. ( ) Breaking Bad;
    c. ( ) Friends;
    d. ( ) Game of Thrones;
    e. ( ) How I met your mother;
    f. ( ) Pretty Little Liars;
    g. ( ) The Big Bang Theory;
    h. ( ) The Walking Dead;
    i. ( ) True Blood;
    j. ( ) Two and a Half Men;
    k. ( ) Under the Dome;
    l. ( ) Outros:

8. Você assiste **com mais frequência** a seriados/filmes (marque **APENAS UMA** opção):

    a. ( ) Dublados;
    b. ( ) Com legendas em inglês;
    c. ( ) Com legendas em português;
    d. ( ) Sem legendas.

9. Nas aulas de inglês, você costuma assistir a filmes/seriados de TV:

    a. ( ) Nunca;
    b. ( ) Raramente;
    c. ( ) Às vezes;
    d. ( ) Frequentemente.

10. As atividades feitas pelo seu professor/a na sala de aula, relativas a filmes/seriados têm o objetivo de (você poderá marcar mais de uma opção, marcando 1 para a alternativa que ocorre com mais frequência, 2 para a segunda mais aplicada, e assim sucessivamente):

    a. ( ) introduzir o tópico da unidade;
    b. ( ) fazer perguntas de compreensão oral;
    c. ( ) explorar questões culturais;
    d. ( ) trabalhar gramática;
    e. ( ) trabalhar pronúncia;
    f. ( ) Outros: _____.

11. Você acredita que assistindo a filmes/seriados de TV seja possível aprender ou incrementar seus conhecimentos na língua estrangeira?

    a. ( ) Não;
    b. ( ) Sim. Dê exemplos do que você acredita que seja possível aprender/melhorar:_____
    _____
    _____

b. Você teria algo a dizer sobre os filmes/seriados que você mais assiste e seu papel no aprendizado de inglês?
_____
_____
_____

APPENDIX C

L2 VOCABULARY RECOGNITION PRE-TEST

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE COMUNICAÇÃO E EXPRESSÃO
DEPARTAMENTO DE LÍNGUA E LITERATURA ESTRANGEIRAS
CURSOS EXTRACURRICULARES DE LÍNGUA ESTRANGEIRA



Course/Level/Class: English 5 **Semester:** 2014.2
Schedule:
Teacher:
Researcher: M.A. Rafael Matielo

Student's Name:_____.

ACTIVITY 1

A) Write down the meaning, a synonym or an explanation using your own words in English or Portuguese for the words given below. It is important that you only answer the ones of which you are sure:
Ex. House: casa;

Blueberry: fruta de cor azulada, da família da amora, comum nos EUA.

1. **Mint** _____
2. **Embodiment**_____
3. **Bookshelf** _____
4. **Slot** _____

5. **Award** _____
6. **Broadband** _____
7. **Pointless** _____
8. **Membership** _____
9. **Liability** _____
10. **Pushy** _____
11. **Affection** _____
12. **Guinea pigs** _____
13. **Mice** _____
14. **Moisturizer** _____
15. **Bachelor** _____
16. **Showdown** _____
17. **Obnoxious** _____
18. **Mustache** _____
19. **Resemblance** _____
20. **Wrath** _____

APPENDIX D

L2 VOCABULARY RECOGNITION TEST

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE COMUNICAÇÃO E EXPRESSÃO
DEPARTAMENTO DE LÍNGUA E LITERATURA ESTRANGEIRAS
CURSOS EXTRACURRICULARES DE LÍNGUA ESTRANGEIRA



Course/Level/Class: English 5 **Semester:** 2014.2
Schedule:
Teacher:
Researcher: M.A. Rafael Matielo

Student's Name:_____.

ACTIVITY 2

The Big Bang Theory is an American comedy series that takes place in Pasadena, California. The show focuses on five characters: two roommates who are scientists and researchers, Leonard and Sheldon, their neighbor Penny who is a waitress, and their friends Howard and Rajesh, who are also researchers.

A) After watching Episode 8, Season One of the sitcom "The Big Bang Theory", answer the questions below:

1. Why are Raj's parents insisting on him to meet the Indian girl, Lalita Gupta?

_____
_____
_____
_____

2. In the episode, why does Penny make cocktail for the boys?

_____
_____
_____
_____

3. At the end of the episode, why is Sheldon singing and playing?

_____
_____
_____
_____

B) Analyze the statements below and put T when you think it is true or F when you think it is false, according to the episode you have just watched:

( ) Raj is used to drinking.
( ) Leonard and Raj are roommates.
( ) Sheldon enjoys drinking alcohol.
( ) Raj does not seem to have any problems when talking to women.
( ) Raj drank a Grasshopper cocktail, which is a green drink.

C) Write down the meaning, a synonym or an explanation using your own words in English or Portuguese for the words given below. It is important that you only answer the ones of which you are sure:

1. **Embodiment** _____
2. **Slot** _____
3. **Pointless** _____

4. **Membership** _____
5. **Pushy** _____
6. **Guinea pigs** _____
7. **Showdown** _____
8. **Obnoxious** _____
9. **Resemblance** _____
10. **Wrath** _____

APPENDIX E

L2 VOCABULARY RECOGNITION POST-TEST

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE COMUNICAÇÃO E EXPRESSÃO
DEPARTAMENTO DE LÍNGUA E LITERATURA ESTRANGEIRAS
CURSOS EXTRACURRICULARES DE LÍNGUA ESTRANGEIRA



Course/Level/Class: English 5 **Semester:** 2014.2
Schedule:
Teacher:
Researcher: M.A. Rafael Matielo
Student's Name:_____.

ACTIVITY 3

A) Write down the meaning, a synonym or an explanation using your own words in English or Portuguese for the words given below. It is important that you only answer the ones of which you are sure:

1. **Embodiment** _____
2. **Slot** _____
3. **Pointless** _____
4. **Membership** _____
5. **Pushy** _____
6. **Guinea pigs** _____

7. **Showdown** _____
8. **Obnoxious** _____
9. **Resemblance** _____
10. **Wrath** _____

APPENDIX F

READING SPAN TEST

Reading Span Test – Practice Session

Set 1

1 - Caiu o número de profissionais que diziam querer ficar por muito tempo no atual **emprego**. (15 palavras, *Você S/A*, fevereiro de 2011, p.51)

2 - O consumo de proteínas estimula a produção de células dos tecidos ósseos e musculares, acelerando o **crescimento**. (17 palavras, *Superinteressante*, agosto de 2000, versão online)

Set 2

3 - Adotar uma postura ética eleva tanto o nível de felicidade quanto ganhar um **aumento**. (14 palavras, *Superinteressante*, dezembro de 2010, versão online)

4 - De modo geral, os imigrantes vindos do Terceiro Mundo têm famílias mais numerosas que os **europeus**. (16 palavras, *Veja*, 24 de outubro de 2007, p.120)

5 - Descobriu-se que o grau de identificação com a equipe não tinha relação com as vitórias ou **derrotas**. (17 palavras, *Mente e Cérebro*, maio de 2011, p.41)

Set 3

6 - Para construir a trama os atores passaram, durante dois meses, por um processo diretamente influenciado pelo **cinema**. (17 palavras, *Mente e cérebro*, maio de 2010, p.11)

7 - O açúcar é uma parte natural da vida humana desde os primórdios de nossa **existência**. (15 palavras, *Veja*, 24 de outubro de 2007, p.11-12)

8 - O consumo isolado de farinha de linhaça não vai baixar os tão desejados pontinhos da **balança**. (16 palavras, *Women's Health*, abril de 2010, p.46)

9 - Não se esqueça de incluir a cidade de onde escreve e telefone para **contato**. (14 palavras, *Mente e cérebro*, maio de 2010, p.7)

Reading Span Test – Testing Session

Set 1

7- Tornou-se cada vez mais comum ver adultos usando aparelhos nos dentes.

8- Em quase cinquenta anos de política, ACM acumulou fortuna e poder.

9- A crise de energia de falta só é não brasileira chuvas.

10- Cientistas analisam imagens das galáxias para traçar o mapa evolutivo do cosmo.

11- O maior símbolo da resistência uma foi colonização a contra africana mulher.

12- Ao chegar ao campo de batalha, os soldados da Grécia antiga bebiam vinho.

Set 2

13- Os bebês nascem programados a com emocional vínculo um formar para mãe.

14- Acreditava-se que a função biológica e natural da mulher era ter filhos.

15- Em vários países, é autorização sem pessoais informações arquivar crime.

16- Diz a ancestral sabedoria que quem controla a respiração controla a mente.

17- A ciência está comprovando a eficácia de base a populares receitas de ervas.

12- Algumas soluções para nossos problemas nossos de diante bem florescendo estar podem olhos.

13- A ciência e a tecnologia tornaram-se o aspecto dominante da guerra.

14- Um estudo indica que do dependentes são americanos médicos dos 20% ópio.

15- Os medicamentos passaram a para chave a como vistos ser cura.

Set 3

16- O leite materno em leite o que do melhor sendo continua pó.

17- Mais importante do que entender é sentir a palavra de Deus.

18- Mobilização social deve de subida contra sociedade da luta incluir preços.

19- Os títulos das reportagens estão dentre os elementos mais importantes do jornal.

20- Galileu foi o primeiro a apontar o telescópio para os céus.

21- Segundo Aristóteles, a Lua, os de feitos seriam estrelas as e planetas éter.

22- A humanidade só venceu e descobriu coisas novas pelo aperto e por crises.

23- A política monetária tem sido do alta a conter para eficaz pouco dólar.

24- Deus criou o universo, mas são os designers que estão repaginando o mundo.
25- Funcionários que fumam apresentam baixa produtividade porque perdem tempo com o vício.
26- A construção de grandes obras cinco a quatro de leva geração de anos.
27- O desmantelamento da na cubana economia a jogou Soviética União lona.

Set 4

28- Não há dúvida de que mais prisões devem ser construídas em todo país.
29- Os óleos de peixe podem evitar o câncer de mama e pulmão.
30- O alho ajuda a evitar os altos níveis de colesterol no sangue.
31- Em qualquer empresa há apenas a e cultura a importantes coisas duas marca.
32- Ao longo da história as pessoas foram segregadas por castas e classes.
33- Nos motores de carros, o do oxigênio o com se combina hidrogênio ar.
34- Covas mostra que, com honestidade, a valer pode política fazer caráter e pena.
35- Mário Covas cumpriu seu trabalho, e agora pode descansar em paz.
36- A função dos a para fartura a garantir é pajés tribo.
37- Na área energética, está faltando uma visão estratégica de longo prazo.

38- Os homens a sobre comida botar para escritório em trabalham mesa.

39- E provável que o nossas mudará que existência uma traga progresso vidas.

40- O yoga trata da relação entre a mente e o corpo.

41- O estresse, a ansiedade ou os desencadear podem fortes muito emoções mesmo tiques.

42- Pesquisas centradas na genética procuram a causa da agressividade em características dos genes.

APPENDIX G

READING SPAN TEST INSTRUCTIONS

READING SPAN TEST: INSTRUÇÕES

1. Sentenças em português aparecerão na tela. À medida que você lê em voz alta as sentenças, você deverá julgar se elas estão corretas ou incorretas, isto é, se elas fazem sentido.

2. Após cada sentença, você deve falar em voz alta se ela está correta ou incorreta.

3. Ao final de cada sequência de sentenças, quando uma tela com pontos de interrogação aparecer, você deverá tentar lembrar da última palavra de cada sentença que você leu e dizer as palavras que lembra em voz alta na sequência em que elas apareceram para você anteriormente.

4. Dois pontos de interrogação (??) no slide significam que você deve tentar lembrar e falar em voz alta as últimas palavras de duas sentenças, três pontos de interrogação (???) significam que você deve tentar lembrar e falar em voz alta as últimas palavras de três sentenças e assim por diante. Portanto, o número de pontos de interrogação indica o número de palavras que você tem que recordar e verbalizar.

5. Você precisa tentar lembrar as palavras e dizê-las em voz alta preferencialmente na mesma ordem que elas apareceram na tela. Se você não conseguir lembrar da ordem em que as palavras apareceram, você poderá dizê-las em voz alta mesmo assim.

6. Não diga em voz alta as palavras antes da tela com os sinais de interrogação aparecer.

7. Procure não tossir, hesitar, repetir-se e/ou interagir com o pesquisador. Você poderá repetir o treinamento se desejar. Seu teste será gravado a à gravação somente o pesquisador terá acesso.

APPENDIX H

OPERATION-WORD SPAN TEST

Operation-Word Span Test – Practice Session

Mathematical Operations	Sets
(9 ÷ 3) -2 = 2 ?	Lábio
(8 ÷ 4) -1 = 1 ?	Ficha
(6 ÷ 2) + 1 = 4 ?	Joia
(6 x 3) -2 = 11 ?	Grito
(4 x 2) +1 = 9 ?	Saia
(10 ÷ 2) + 4 = 9 ?	Cofre
(2 + 3) + 3 = 8 ?	Lenda
(7 + 3) - 2 = 8 ?	Pilha
(3 - 1) + 1 = 1 ?	Noite
(9 - 1) ÷ 2 = 4 ?	Perna
(3 x 5) - 2 = 12 ?	Classe
(4 x 3) - 3 = 10 ?	Granja

(2 + 7) + 4 = 12 ?	Loja
(10 – 4) ÷ 2 = 4?	Carne

Operation-Word Span Test – Testing Session

Mathematical Operations	**Words**
Set 1	
(10 ÷ 2) – 3 = 2 ?	carta
(10 ÷ 10) – 1 = 2 ?	lençol
(7 ÷ 1) + 2 = 7 ?	terra
(3 ÷ 1) - 2 = 3 ?	papel
(2 x 1) - 1 = 1 ?	avó
(10 ÷ 1) + 3 = 13 ?	tinta
(9 x 2) + 1 = 18 ?	guerra
(9 ÷ 1) - 7 = 4 ?	chuva
(8 x 4) - 2 = 32 ?	fila
(9 x 3) - 3 = 24 ?	água
(4 ÷ 1) + 1 = 4 ?	maçã
(10 ÷ 1) - 1 = 9 ?	ferro

(8 x 4) + 2 = 34 ?	jornal
Set 2	
(6 x 3) + 2 = 17 ?	feira
(6 ÷ 3) + 2 = 5 ?	lago
(6 x 2) - 3 = 10 ?	fogão
(8 ÷ 2) + 4 = 2 ?	lixo
(8 ÷ 2) - 1 = 3 ?	dedo
(9 ÷ 1) - 5 = 4 ?	balde
(6 ÷ 2) - 2 = 2 ?	ladrão
(7 x 2) - 1 = 14 ?	rocha
(6 x 2) - 2 = 10 ?	padre
(2 x 2) + 1 = 4 ?	jardim
(7 x 1) + 6 = 13 ?	leite
(3 ÷ 1) + 3 = 6 ?	braço
(10 ÷ 1) + 1 = 10 ?	cobra
(4 x 4) + 1 = 17 ?	fita
(3 x 3) - 1 = 8 ?	irmão

Set 3	
(3 x 1) + 2 = 2 ?	telha
(4 ÷ 2) + 1 = 6 ?	vinho
(5 ÷ 5) + 1 = 2 ?	foto
(2 x 3) + 1 = 4 ?	mala
(9 ÷ 3) - 2 = 1 ?	bruxa
(10 ÷ 2) - 4 = 3 ?	álbum
(5 ÷ 1) + 4 = 9 ?	dente
(10 x 2) + 3 = 23 ?	vidro
(7 ÷ 1) + 6 = 12 ?	trilha
(3 x 2) + 1 = 6 ?	feijão
(6 x 4) + 1 = 25 ?	nuvem
(9 ÷ 3) - 1 = 2 ?	calça
(8 ÷ 1) - 6 = 4 ?	pato
(9 x 1) + 9 = 1 ?	festa

APPENDIX I

OPERATION-WORD SPAN TEST INSTRUCTIONS

OPERATION-WORD SPAN TEST: INSTRUÇÕES

Nesta tarefa você tentará memorizar palavras que você verá na tela do computador. Entre as palavras que serão apresentadas na tela, você terá que resolver operações matemáticas simples.

Você verá na tela uma equação seguida de uma palavra. Sua tarefa é ler a equação em voz alta e verificar se o resultado da mesma está ou não correto dizendo **SIM** ou **NÃO** no microfone. Imediatamente após dar sua resposta, você lerá a palavra também em voz alta. Você poderá pensar por alguns instantes na resposta, mas não poderá esperar para ler a palavra.

Vamos ver um exemplo:

(2 + 1) ÷ 3 = 1 ? GATO

Nesse caso você leria em voz alta: "Dois mais um, dividido por 3 é igual a um? Em seguida, você responderia **SIM** porque 1 é o resultado correto da equação. Imediatamente após dizer **SIM**, você leria em voz alta a palavra **GATO**. Você deve tentar memorizar esta palavra.

Você verá 3 blocos com 2, 3, 4 e/ou 5 pares de equações e palavras. Quando todos os pares de equações e palavras de um determinado conjunto forem apresentados, você verá uma tela com pontos de interrogação que indicarão o número de palavras que você viu naquele conjunto. Este será o sinal para você começar a dizer em voz alta as palavras que você conseguiu memorizar, respeitando a ordem em que elas apareceram na tela.

Em seguida, você verá outro conjunto de equações e palavras e repetirá os mesmos procedimentos até o fim do experimento.

Procure se concentrar na tarefa e prestar bastante atenção durante a apresentação das equações e das palavras, pois elas permanecerão por apenas alguns segundos na tela do computador.

Além de tentar memorizar a palavra apresentada após a equação, é muito importante que você também se esforce para acertar o resultado das equações. Você precisa tentar lembrar as palavras e dizê-las em voz alta preferencialmente na mesma ordem que elas apareceram na tela. Se você não conseguir lembrar da ordem em que as palavras apareceram, você poderá dizê-las em voz alta mesmo assim.

Não diga em voz alta as palavras antes da tela com os sinais de interrogação aparecer. Procure não tossir, hesitar, repetir-se e/ou interagir com o pesquisador. Você poderá repetir o treinamento se desejar. Seu teste será gravado a à gravação somente o pesquisador terá acesso.

APPENDIX J

TEACHERS PROFILE QUESTIONNAIRE

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE COMUNICAÇÃO E EXPRESSÃO – CCE
DEPARTAMENTO DE LÍNGUA E LITERATURA ESTRANGEIRAS – DLLE

Nome: _____
Idade: _____
Nacionalidade: _____
E-mail:

QUESTIONÁRIO

1. Qual a sua formação acadêmica?

   _____
   _____
   _____
   _____

2. Há quanto tempo você leciona inglês e em quais contextos você atua?

   _____
   _____
   _____
   _____
   _____

3. Resumidamente, comente como você percebe o processo de aprender e ensinar uma língua estrangeira?

4. Nas suas aulas, você costuma trabalhar com vídeos? Em caso afirmativo, com quais objetivos? Você usa legendas em português ou em inglês?

5. Caso você utilize vídeos em suas aulas, você costuma disponibilizar legendas? Em caso afirmativo, de que tipo e por quê?

APPENDIX K

VIDEO RETROSPECTIVE QUESTIONNAIRE

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE COMUNICAÇÃO E EXPRESSÃO – CCE
DEPARTAMENTO DE LÍNGUA E LITERATURA ESTRANGEIRAS – DLLE

Nome: _____
Idade: _____
Nacionalidade: _____
E-mail:

QUESTIONÁRIO

1. Você considerou os testes difíceis? Comente algum aspecto que lhe tenha chamado a atenção.

2. Após alguma das etapas anteriores da pesquisa, você ficou com alguma palavra na cabeça? Se sim, qual(is)?

3. Você chegou a checar alguma palavra no dicionário ou com o seu/sua professor(a)? Em caso afirmativo, você se recorda qual(is) dela(s) você checou ou perguntou?

4. Você acredita que seriados de TV são uma boa fonte ou forma de aprendizagem em língua estrangeira, como vocabulário, por exemplo? Comente brevemente sua opinião sobre isso e diga que tipo de legendas (em inglês, em português ou sem legendas) você acha que funcionam melhor com seu estilo de aprendizagem.

WORKING MEMORY RETROSPECTIVE QUESTIONNAIRE

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE COMUNICAÇÃO E EXPRESSÃO – CCE
DEPARTAMENTO DE LÍNGUA E LITERATURA ESTRANGEIRAS – DLLE

Nome: _____
Idade: _____
Nacionalidade: _____
E-mail:

QUESTIONÁRIO

1. Você considerou os testes difíceis? Comente algum aspecto que lhe tenha chamado a atenção.

2. Você se considera uma pessoa que tem uma boa memória? Por quê (não)?

3. No seu cotidiano, você consegue prestar atenção simultânea em mais de uma tarefa? Dê algum exemplo que lhe venha à mente.

4. Você se sentiu cansado ou incomodado durante os testes com alguma coisa? Comente algum aspecto que desejar.

APPENDIX L

TESTS OF NORMALITY FOR THE VIDEO COMPREHENSION TEST (GENERAL AND SPECIFIC PARTS)

Table 3

Tests of normality: General video comprehension

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Intralingual Subtitles	.417	12	.000	.608	12	.000
Interlingual Subtitles	.307	12	.003	.764	12	.004
Control	.307	12	.003	.764	12	.004

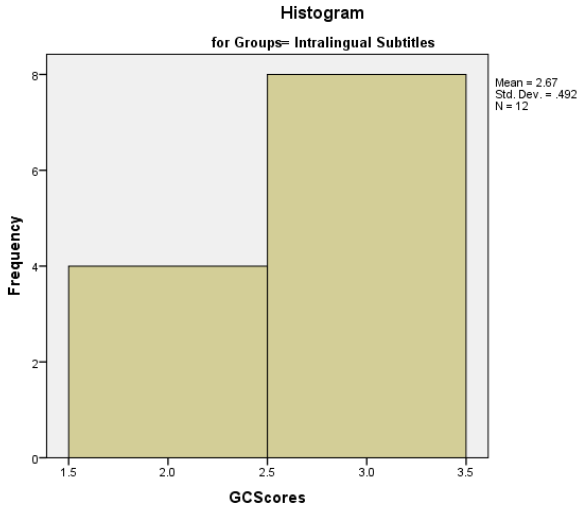
Table 4

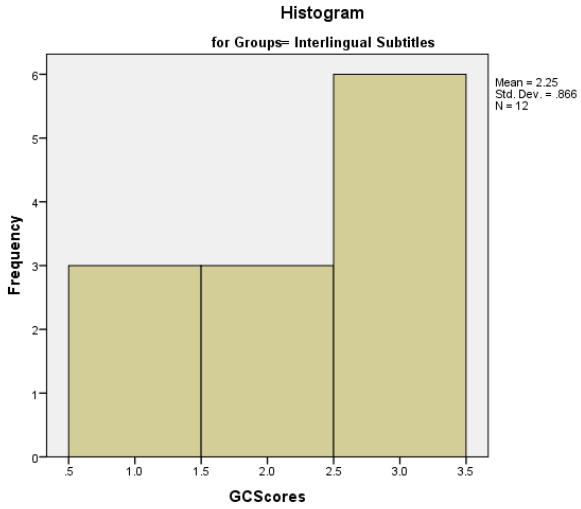
Tests of normality: Specific video comprehension

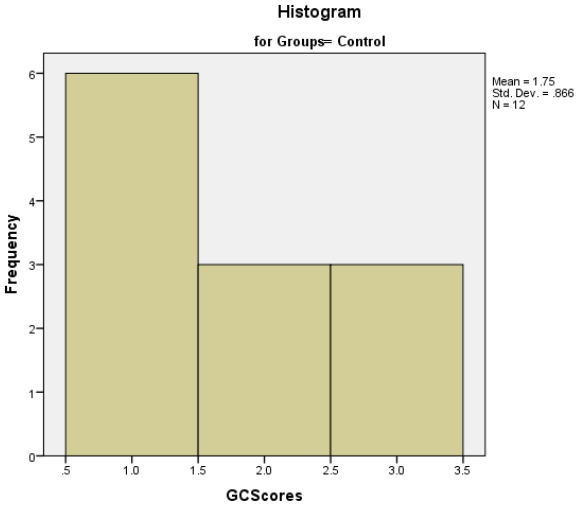
Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Intralingual Subtitles	.446	12	.000	.592	12	.000
Interlingual Subtitles	.357	12	.000	.710	12	.001
Control	.258	12	.027	.802	12	.010

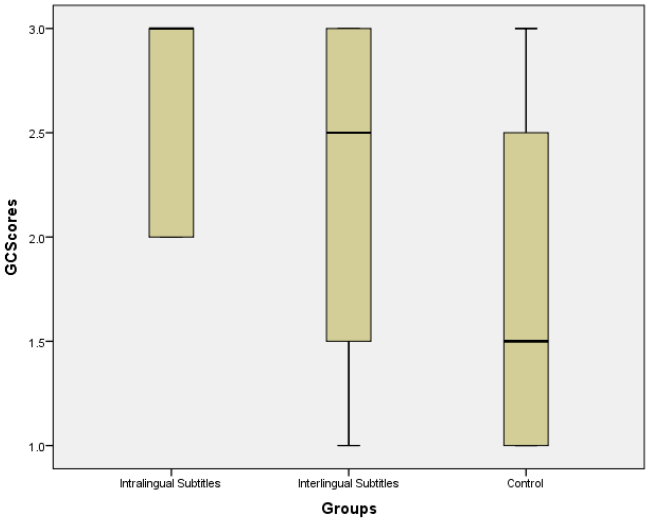
APPENDIX M

HISTOGRAMS AND BOX PLOTS FOR THE VIDEO COMPREHENSION TEST (GENERAL COMPREHENSION)



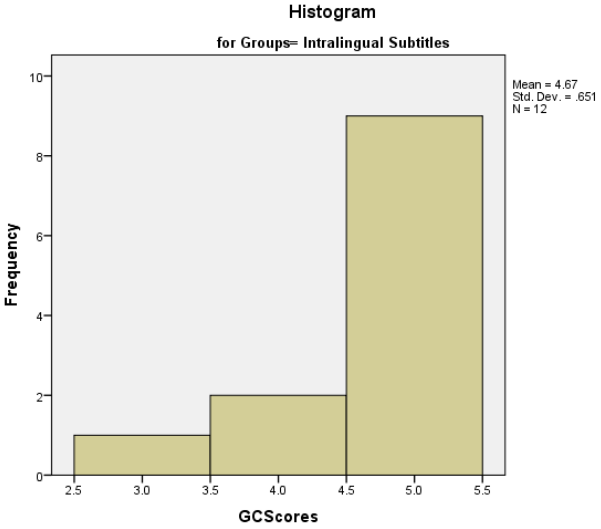


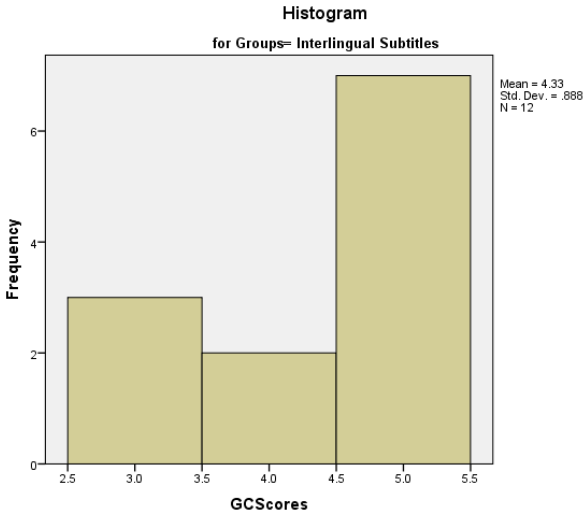

Histogram
for Groups= Control
Mean = 1.75
Std. Dev. = .866
N = 12
Frequency
GCScores

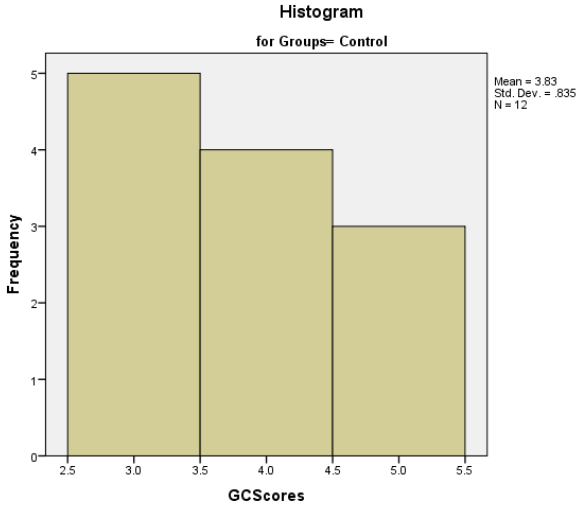

GCScores
Intralingual Subtitles
Interlingual Subtitles
Control
Groups

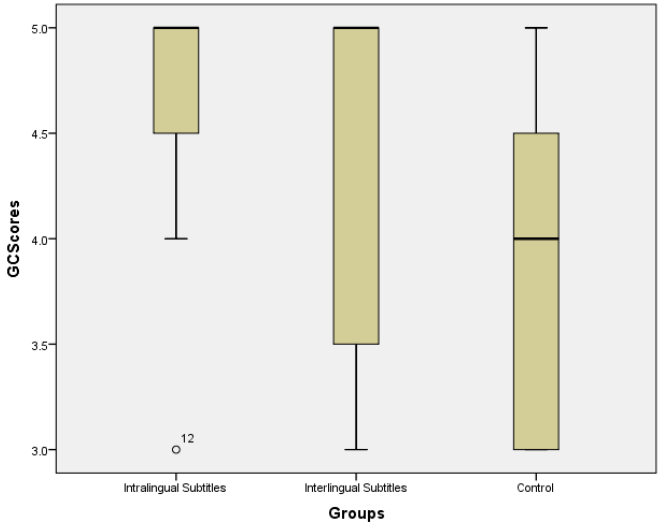
APPENDIX N

HISTOGRAMS AND BOX PLOTS FOR THE VIDEO COMPREHENSION TEST (SPECIFIC COMPREHENSION)






Histogram
for Groups= Control
Mean = 3.83
Std. Dev. = .835
N = 12
Frequency
GCScores


GCScores
12
Intralingual Subtitles
Interlingual Subtitles
Control
Groups

APPENDIX O

INTER-RATER RELIABILITY FOR THE L2 VOCABULARY RECOGNITION – PRE-TEST

Table 5

Inter-rater reliability for the L2 vocabulary recognition – pre-test: Intralingual subtitles group

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Raters
.85	.84	3

Table 6

Inter-rater reliability for the L2 vocabulary recognition – pre-test: Interlingual subtitles group

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Raters
.75	.76	3

Table 7

Inter-rater reliability for the L2 vocabulary recognition – pre-test: Control group

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Raters
.98	.98	3

APPENDIX P

INTER-RATER RELIABILITY FOR THE L2 VOCABULARY RECOGNITION – TEST

Table 8

Inter-rater reliability for the L2 vocabulary recognition - test: Intralingual subtitles group

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Raters
.92	.93	3

Table 9

Inter-rater reliability for the L2 vocabulary recognition - test: Interlingual subtitles group

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Raters
.95	.96	3

Table 10

Inter-rater reliability for the L2 vocabulary recognition - test: Control group

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Raters
.99	.99	3

APPENDIX Q

INTER-RATER RELIABILITY FOR THE L2 VOCABULARY RECOGNITION – POST-TEST

Table 11

Inter-rater reliability for the L2 vocabulary recognition – post-test: Intralingual subtitles group

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Raters
.88	.89	3

Table 12

Inter-rater reliability for the L2 vocabulary recognition – post-test: Interlingual subtitles group

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Raters
.97	.97	3

Table 13

Inter-rater reliability for the L2 vocabulary recognition – post-test: Control group

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	Number of Raters
.98	.99	3

APPENDIX R

TESTS OF NORMALITY FOR THE L2 VOCABULARY RECOGNITION TESTS

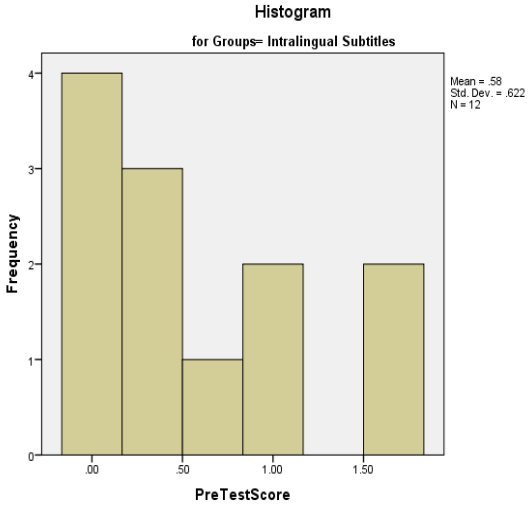
Table 14

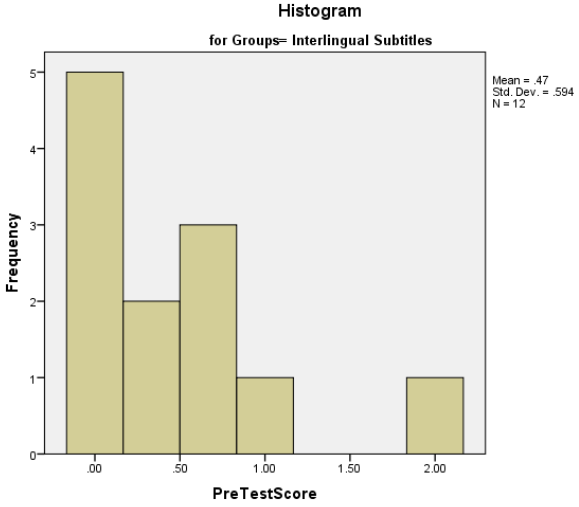
Tests of normality: L2 vocabulary recognition tests

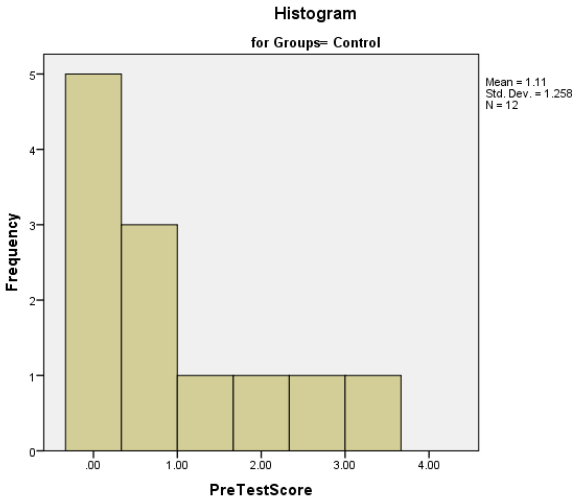
	Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Pre-Test	Intralingual Subtitles	.240	12	.056	.841	12	.029
	Interlingual Subtitles	.213	12	.138	.791	12	.007
	Control	.228	12	.085	.843	12	.030
Test	Intralingual Subtitles	.152	12	.200*	.918	12	.271
	Interlingual Subtitles	.229	12	.082	.771	12	.005
	Control	.224	12	.099	.810	12	.012
Post-Test	Intralingual Subtitles	.165	12	.200*	.940	12	.495
	Interlingual Subtitles	.255	12	.030	.751	12	.003
	Control	.376	12	.000	.801	12	.010

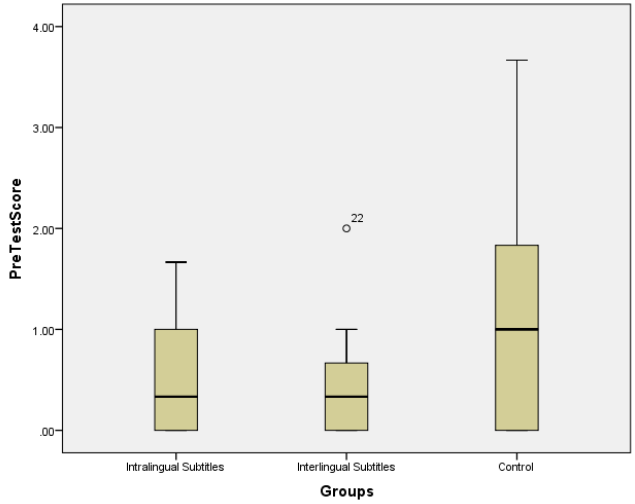
APPENDIX S

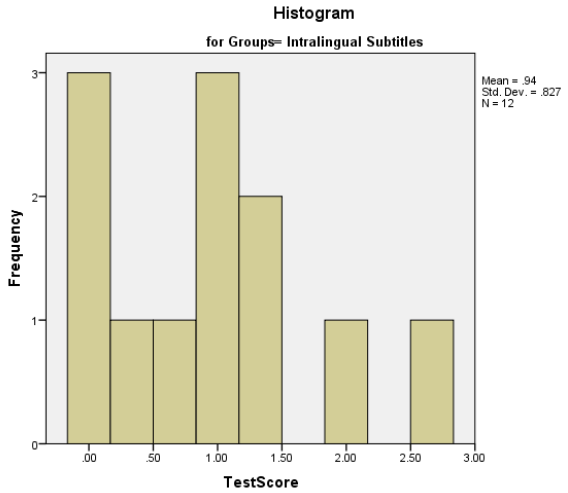
HISTOGRAMS AND BOX PLOTS FOR THE L2 VOCABULARY RECOGNITION TESTS

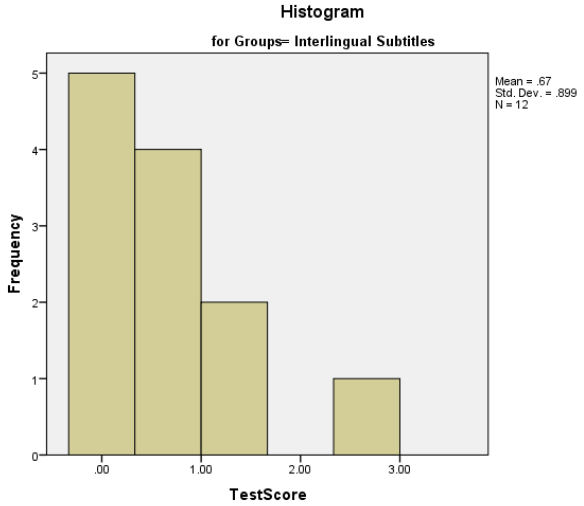


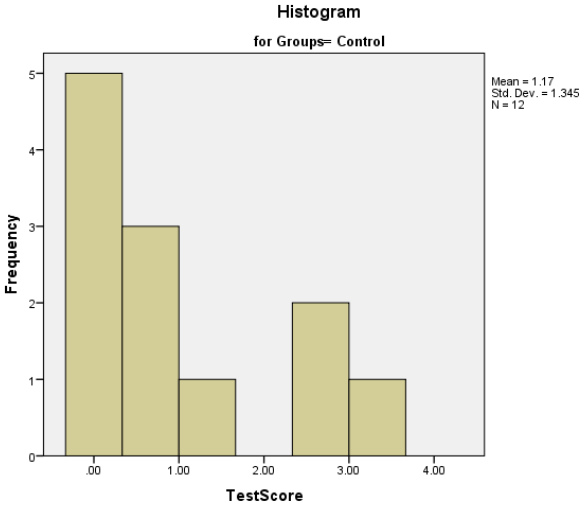


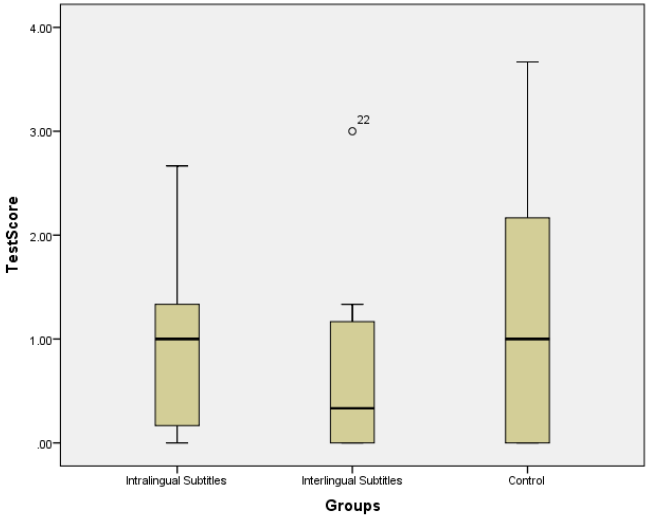

Histogram
for Groups= Control
Mean = 1.11
Std. Dev. = 1.258
N = 12
Frequency
PreTestScore
0
1
2
3
4
5
.00
1.00
2.00
3.00
4.00

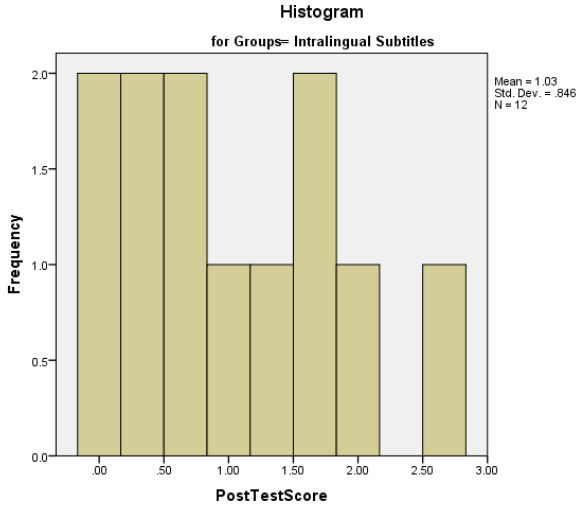

PreTestScore
Groups
.00
1.00
2.00
3.00
4.00
22
Intralingual Subtitles
Interlingual Subtitles
Control

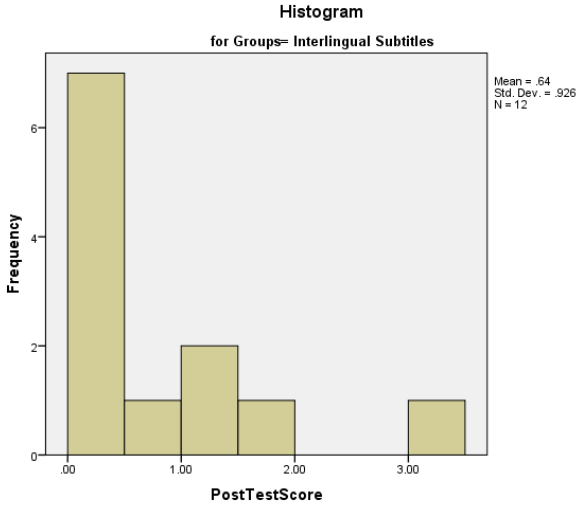

Histogram
for Groups= Intralingual Subtitles
Mean = .94
Std. Dev. = .827
N = 12
Frequency
TestScore
0
1
2
3
.00
.50
1.00
1.50
2.00
2.50
3.00

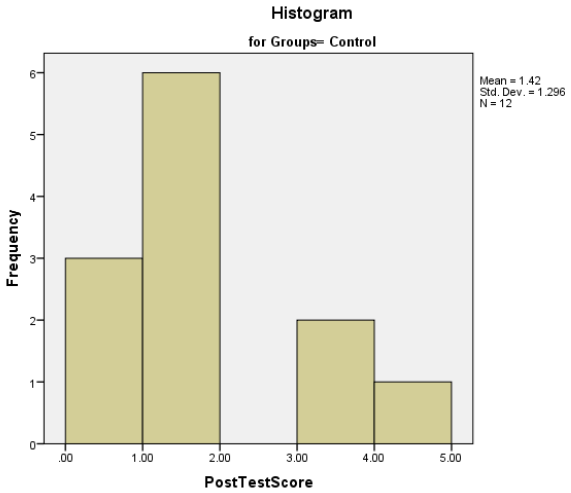

Histogram
for Groups= Interlingual Subtitles
Mean = .67
Std. Dev. = .899
N = 12
Frequency
TestScore
0
1
2
3
4
5
.00
1.00
2.00
3.00

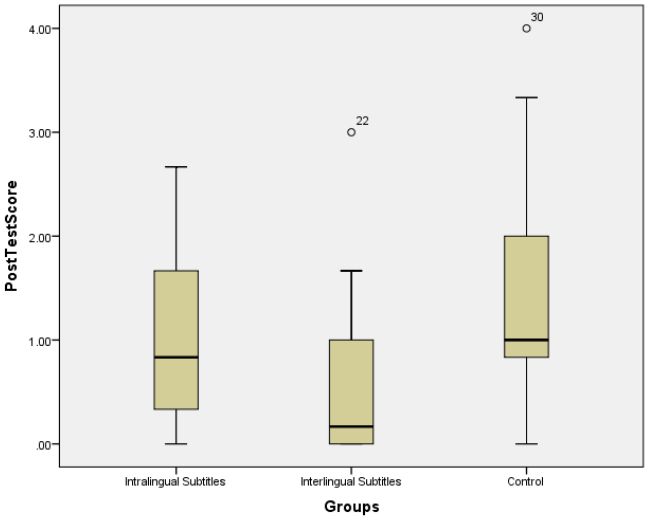

Histogram
for Groups= Control
Mean = 1.17
Std. Dev. = 1.345
N = 12
Frequency
TestScore
0
1
2
3
4
5
.00
1.00
2.00
3.00
4.00


TestScore
Groups
.00
1.00
2.00
3.00
4.00
22
Intralingual Subtitles
Interlingual Subtitles
Control


Histogram
for Groups= Intralingual Subtitles
Mean = 1.03
Std. Dev. = .846
N = 12
Frequency
PostTestScore


Histogram
for Groups= Interlingual Subtitles
Mean = .64
Std. Dev. = .926
N = 12
Frequency
PostTestScore


Histogram
for Groups= Control
Mean = 1.42
Std. Dev. = 1.296
N = 12
Frequency
PostTestScore
0
1
2
3
4
5
6
.00
1.00
2.00
3.00
4.00
5.00


PostTestScore
Groups
.00
1.00
2.00
3.00
4.00
22
30
Intralingual Subtitles
Interlingual Subtitles
Control

APPENDIX T

TESTS OF NORMALITY FOR THE WORKING MEMORY TESTS

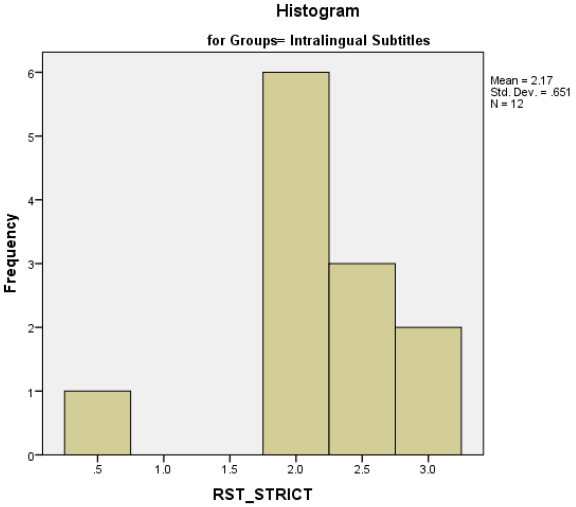
Table 15

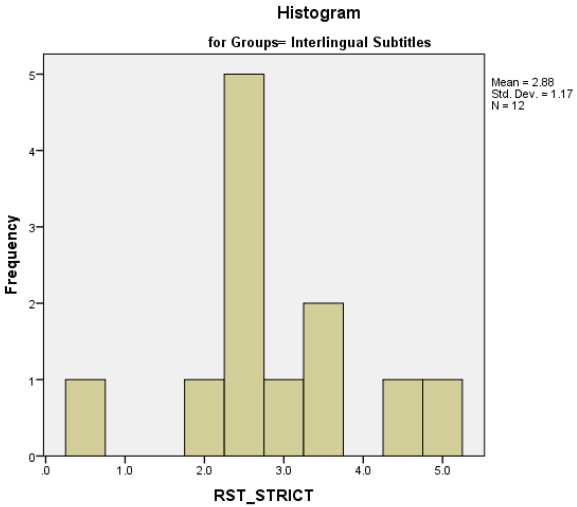
Tests of normality: Working memory tests (RST and OSPAN)

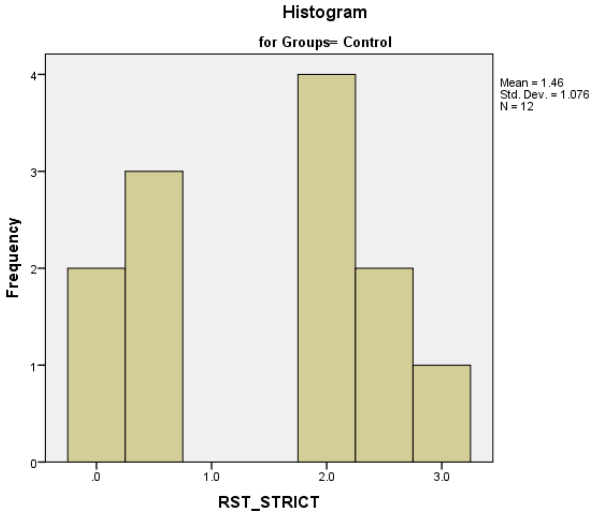
WM Test	Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
RST Strict	Intralingual Subtitles	.316	12	.002	.810	12	.012
	Interlingual Subtitles	.209	12	.155	.926	12	.339
	Control	.276	12	.012	.866	12	.058
RST Lenient	Intralingual Subtitles	.292	12	.006	.812	12	.013
	Interlingual Subtitles	.194	12	.200	.943	12	.532
	Control	.296	12	.005	.857	12	.045
OSPAN Strict	Intralingual Subtitles	.161	12	.200	.956	12	.725
	Interlingual Subtitles	.199	12	.200	.896	12	.140
	Control	.138	12	.200	.969	12	.903
OSPAN Lenient	Intralingual Subtitles	.196	12	.200	.944	12	.558
	Interlingual Subtitles	.115	12	.200	.948	12	.613
	Control	.172	12	.200	.930	12	.385

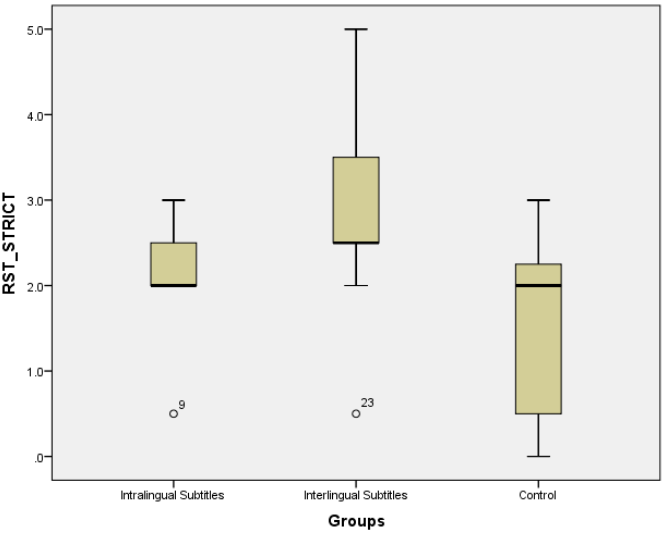
APPENDIX U

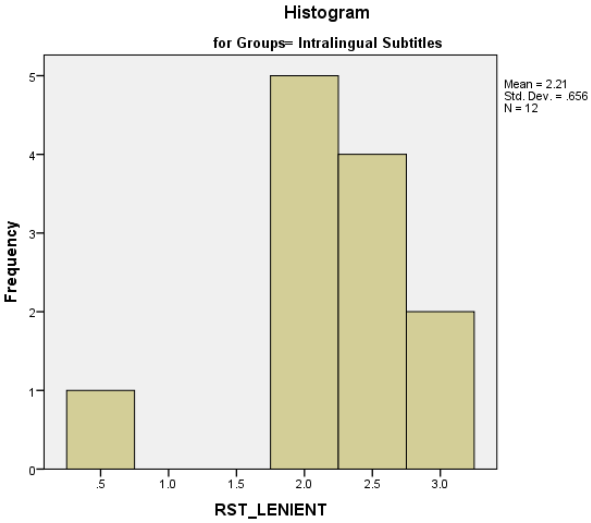
HISTOGRAMS AND BOX PLOTS FOR THE WORKING MEMORY TESTS

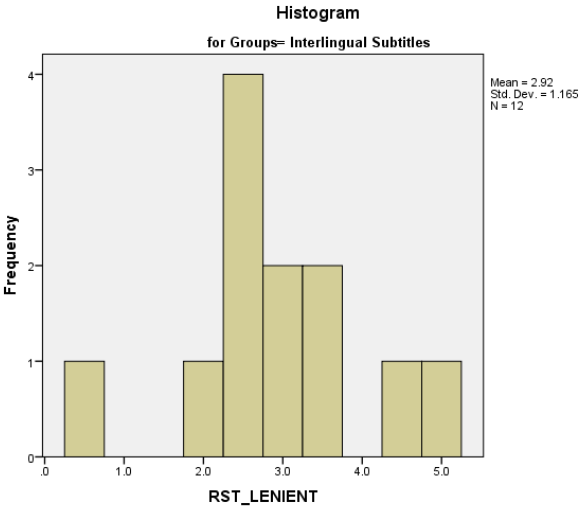


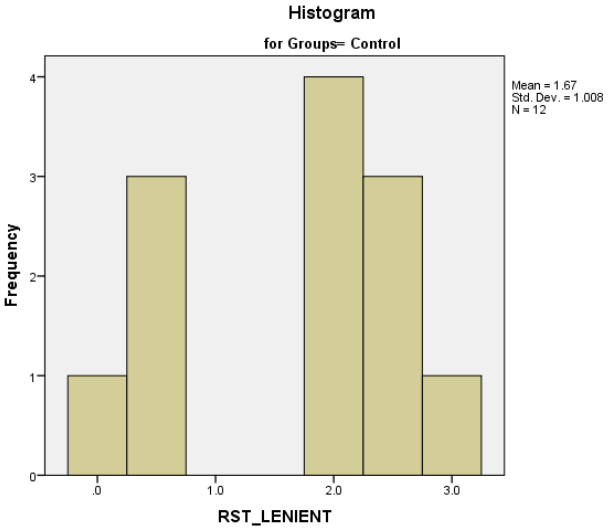


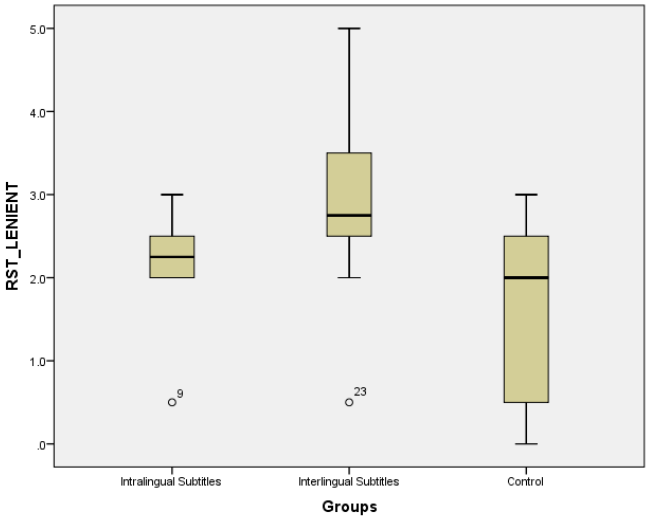

Histogram
for Groups= Control
Mean = 1.46
Std. Dev. = 1.076
N = 12
Frequency
RST_STRICT

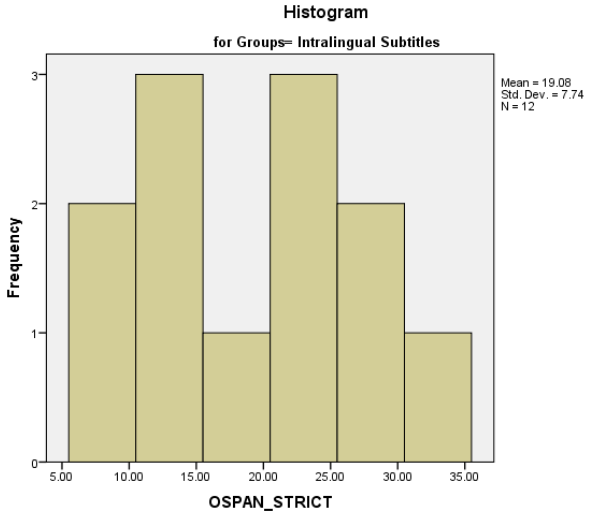

RST_STRICT
Intralingual Subtitles
Interlingual Subtitles
Control
Groups

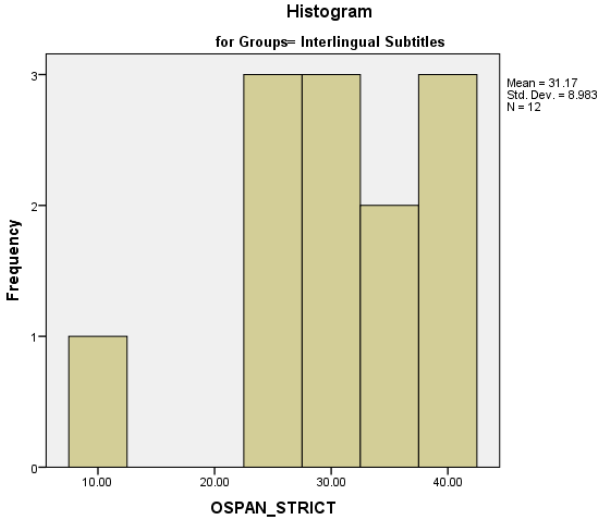

Histogram
for Groups= Intralingual Subtitles
Mean = 2.21
Std. Dev. = .656
N = 12
Frequency
RST_LENIENT

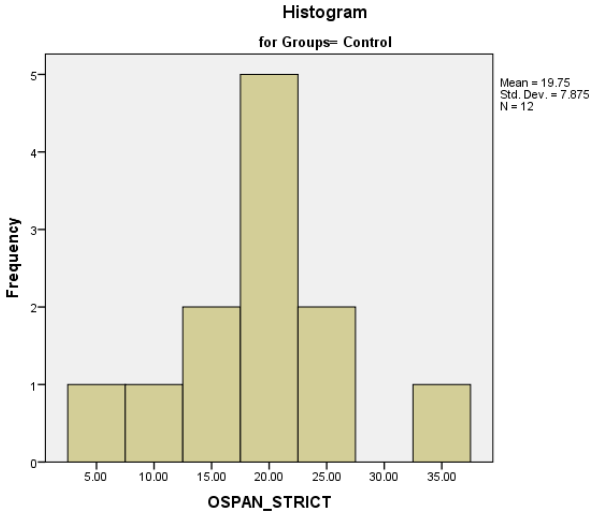

Histogram
for Groups= Interlingual Subtitles
Mean = 2.92
Std. Dev. = 1.165
N = 12
Frequency
RST_LENIENT

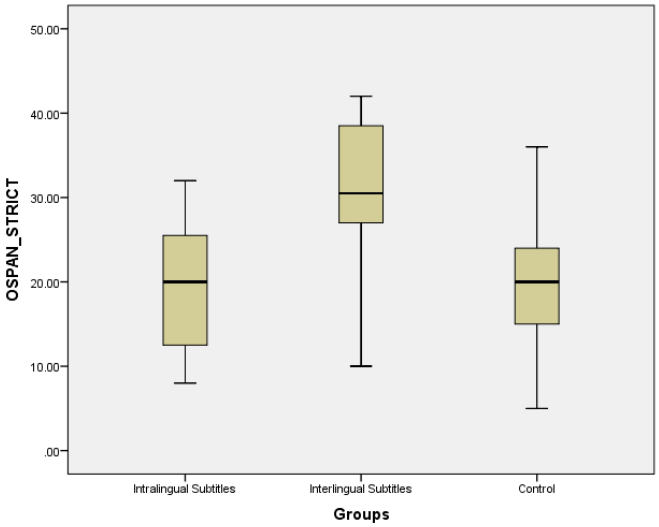

Histogram
for Groups= Control
Mean = 1.67
Std. Dev. = 1.008
N = 12
Frequency
RST_LENIENT

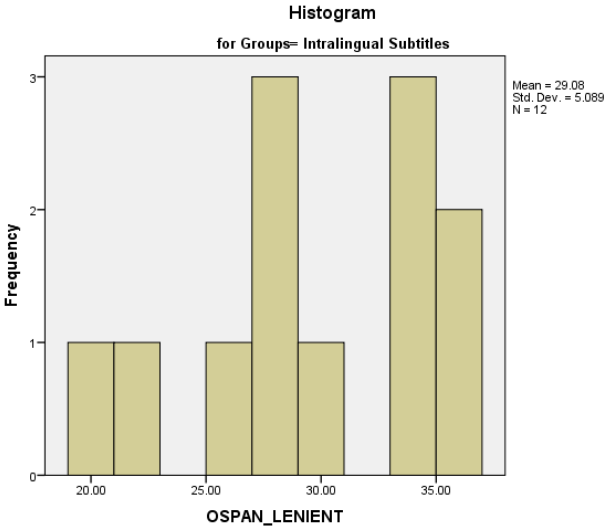

RST_LENIENT
9
23
Intralingual Subtitles
Interlingual Subtitles
Control
Groups

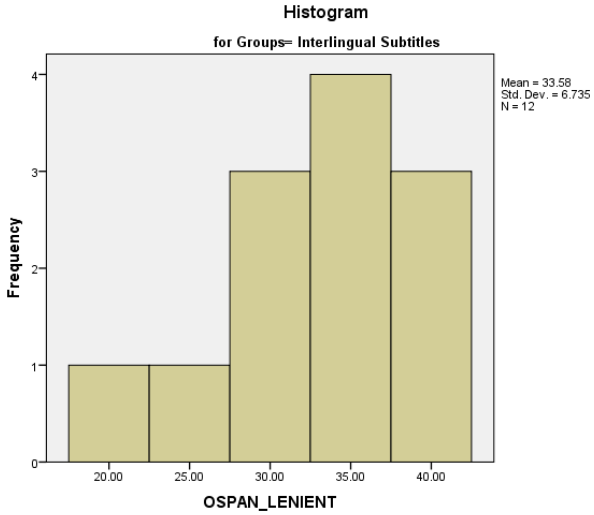

Histogram
for Groups= Intralingual Subtitles
Mean = 19.08
Std. Dev. = 7.74
N = 12
Frequency
OSPAN_STRICT
0
1
2
3
5.00
10.00
15.00
20.00
25.00
30.00
35.00

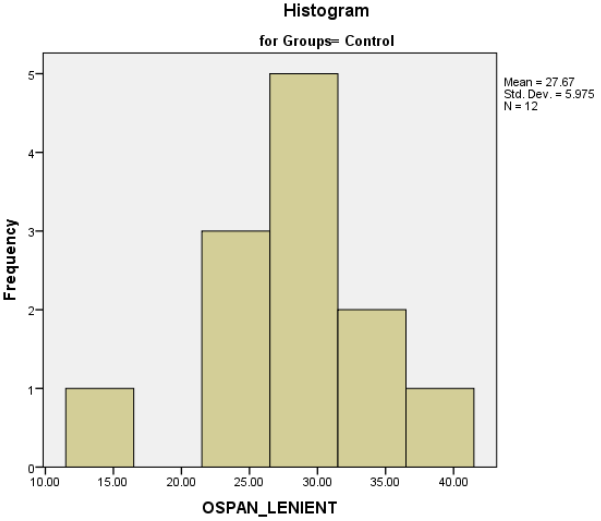

Histogram
for Groups= Interlingual Subtitles
Mean = 31.17
Std. Dev. = 8.983
N = 12
Frequency
OSPAN_STRICT
0
1
2
3
10.00
20.00
30.00
40.00

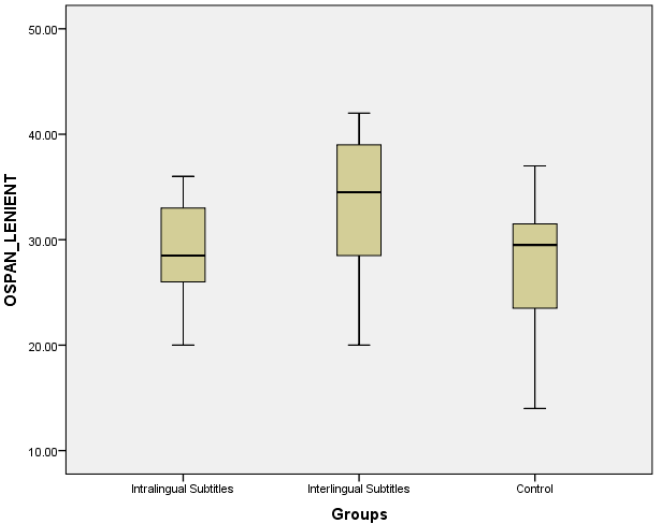

Histogram
for Groups= Control
Mean = 19.75
Std. Dev. = 7.875
N = 12
Frequency
OSPAN_STRICT
5.00
10.00
15.00
20.00
25.00
30.00
35.00


OSPAN_STRICT
50.00
40.00
30.00
20.00
10.00
.00
Intralingual Subtitles
Interlingual Subtitles
Control
Groups


Histogram
for Groups= Intralingual Subtitles
Mean = 29.08
Std. Dev. = 5.089
N = 12
Frequency
0
1
2
3
20.00
25.00
30.00
35.00
OSPAN_LENIENT


Histogram
for Groups= Interlingual Subtitles
Mean = 33.58
Std. Dev. = 6.735
N = 12
Frequency
0
1
2
3
4
20.00
25.00
30.00
35.00
40.00
OSPAN_LENIENT


Histogram
for Groups= Control
Mean = 27.67
Std. Dev. = 5.975
N = 12
Frequency
OSPAN_LENIENT


OSPAN_LENIENT
Intralingual Subtitles
Interlingual Subtitles
Control
Groups

APPENDIX V

PARTICIPANTS' GENERAL COMPREHENSION RAW SCORES

Table 41

General comprehension raw scores

	Participant	GCScore
Intralingual Subtitles (n=12)	P1	3
	P2	3
	P3	3
	P4	2
	P5	3
	P6	3
	P7	3
	P8	3
	P9	2
	P10	2
	P11	3
	P12	2
	Mean	2.67
	SD	.492
	Minimum	2
	Maximum	3
Interlingual Subtitles (n=12)	P13	3
	P14	2
	P15	3
	P16	3
	P17	2
	P18	3
	P19	3
	P20	2
	P21	1
	P22	3
	P23	1
	P24	1
	Mean	2.25
	SD	.866
	Minimum	1
	Maximum	3

Control (n=12)	P25	3
	P26	1
	P27	1
	P28	2
	P29	1
	P30	3
	P31	2
	P32	3
	P33	1
	P34	2
	P35	1
	P36	1
	Mean	1.75
	SD	.866
	Minimum	1
	Maximum	3

Note: n = sample size; SD = standard deviation; GCScores = general comprehension scores

APPENDIX W

PARTICIPANTS' SPECIFIC COMPREHENSION RAW SCORES

Table 42

Specific comprehension raw scores

	Participant	SCScores
Intralingual Subtitles (n=12)	P1	5
	P2	5
	P3	5
	P4	5
	P5	5
	P6	5
	P7	5
	P8	4
	P9	5
	P10	4
	P11	5
	P12	3
	Mean	4.67
	SD	.651
	Minimum	3
	Maximum	5
Interlingual Subtitles (n=12)	P13	3
	P14	5
	P15	5
	P16	5
	P17	3
	P18	5
	P19	4
	P20	5
	P21	5
	P22	5
	P23	4
	P24	3
	Mean	4.33
	SD	.888
	Minimum	3
	Maximum	5

Control (n=12)	P25	5
	P26	3
	P27	3
	P28	3
	P29	4
	P30	5
	P31	5
	P32	4
	P33	3
	P34	3
	P35	4
	P36	4
	Mean	3.83
	SD	.697
	Minimum	3
	Maximum	5

Note: n = sample size; SD = standard deviation; SCScores = specific comprehension scores

APPENDIX X
PARTICIPANTS' L2 VOCABULARY RECOGNITION RAW SCORES

Table 43

L2 vocabulary recognition raw scores

	Participant	Pre-Test	Test	Post-Test
Intralingual Subtitles (n=12)	P1	1.67	2.67	2.67
	P2	.00	.00	.00
	P3	1.00	2.00	2.00
	P4	1.00	1.00	.67
	P5	.00	.00	.67
	P6	.67	1.33	1.33
	P7	.33	.00	.00
	P8	1.67	1.33	1.67
	P9	.33	1.00	.33
	P10	.00	1.00	1.00
	P11	.00	.33	1.67
	P12	.33	.67	.33
	Mean	.58	.94	1.02
	SD	.621	.826	.846
	Minimum	.00	.00	.00
	Maximum	1.67	2.67	2.67
Interlingual Subtitles (n=12)	P13	.33	.33	.00
	P14	1.00	1.33	1.67
	P15	.00	.00	.33
	P16	.00	.00	.00
	P17	.67	.67	.67
	P18	.33	.00	.00
	P19	.00	.00	.00
	P20	.67	.33	.00
	P21	.00	1.00	1.00
	P22	2.00	3.00	3.00
	P23	.67	1.33	1.00
	P24	.00	.00	.00
	Mean	.47	.66	.63
	SD	.593	.898	.926
	Minimum	.00	.00	.00
	Maximum	2.00	3.00	3.00

Control (n=12)	P25	1.00	1.00	1.00
	P26	.00	.00	.00
	P27	1.00	1.00	1.00
	P28	1.67	1.00	1.00
	P29	.00	.00	.67
	P30	3.67	3.67	4.00
	P31	1.00	1.33	1.00
	P32	3.00	3.00	3.00
	P33	.00	.00	1.00
	P34	2.00	3.00	3.33
	P35	.00	.00	.00
	P36	.00	.00	1.00
	Mean	1.11	1.16	1.41
	SD	1.25	1.34	1.29
	Minimum	.00	.00	.00
	Maximum	3.67	3.67	4.00

Note: n = sample size; SD = standard deviation; GCScores = general comprehension scores