

**English language learning through viewing television:  
An investigation of comprehension, incidental vocabulary acquisition,  
lexical coverage, attitudes, and captions**

by

Michael Patrick Hindley Rodgers

A thesis submitted to Victoria University of Wellington  
in fulfillment of the requirements for the degree of  
Doctor of Philosophy

Victoria University of Wellington  
2013



## ***Abstract***

In the English as a Foreign Language (EFL) setting it may be a challenge to obtain the second language input necessary for language learning. A potential source of input may be episodes of television; however, little previous research has been done indicating whether episodes are a suitable source of aural input for EFL learning. Past research has concentrated on short videos of a type that learners might not choose to learn English from. The experimental design employed in this thesis expands upon earlier methodologies by employing full-length episodes of television intended for an English-speaking audience. The thesis is comprised of five studies investigating aspects of language learning through viewing television. The first study examines comprehension gains from the first to the tenth episode viewed, comprehension across 10 episodes viewed, and the effects of vocabulary knowledge on comprehension. The results showed significant comprehension gains from the first to the final episode viewed. Comprehension scores across the eight intervening episodes were all higher than the initial episode but scores were episode-dependent. The results also showed small to moderate correlations between vocabulary knowledge and comprehension for each of the 10 episodes. The second study investigated the effects of viewing over 7 hours of television on incidental vocabulary learning, and the effects of the frequency and range of occurrence of lower frequency words within the episodes on vocabulary learning. Two tests measuring knowledge of form-meaning connection at differing sensitivities were used to assess vocabulary knowledge. Results showed vocabulary gains of approximately six words on both tests. Frequency of occurrence was found to have a medium-size correlation with vocabulary gains. No significant relationship was found between range of occurrence and acquisition. The third study examined whether increased lexical coverage leads to increased comprehension of television and greater incidental vocabulary learning. Results showed that comprehension improved with increased lexical coverage for six of the 10 episodes. In these episodes, participants with approximately 94% lexical coverage were found to have higher comprehension scores than participants with less lexical coverage. Results showed no significant relationship between incidental vocabulary acquisition and lexical coverage. In the fourth study, two surveys examined language learners' attitudes towards learning English through viewing episodes of television. One survey followed each episode and examined learners' beliefs about: their enjoyment of the episode, the usefulness of studying English through viewing the episode, their level of learning from the episode, and their comprehension of the episode. For all items, mean responses were significantly higher

following the final episode than following the first episode. On the survey that followed viewing all the episodes, participants had generally favorable attitudes towards language learning through viewing television. The fifth study investigated how the presence of captions affected the aspects of language learning examined in Studies 1 to 4. The most salient finding from this study was that the presence of captions improved comprehension for episodes early in the viewing process and for difficult episodes. Taken as a whole, this thesis shows the value of using episodes of television for language learning.

## ***Acknowledgements***

I would like to express my sincerest gratitude to my first supervisor, Dr. Stuart Webb, who was involved in every phase of this thesis. His guidance and support have been invaluable throughout every stage of completing this project.

I would also like to thank my second supervisor, Professor Paul Nation, for the help that he has supplied during the PhD process. His guidance early on in the process as well as his generosity with his time has been invaluable.

I would like also like to thank VUW for the financial support without which I would not have been able to complete my studies.

A big thanks to everyone I have met during my time in New Zealand - from flatmates to fellow students to the staff of the School of Linguistics and Applied Language Studies. I am better for having met and spent time with you.

I would like to thank my family in Japan and Canada, Miki, Michael (especially with proofreading), Noreen, Kathleen, Keith, Darcy, Liam and Raine for their support during this endeavor. Without their understanding and support I would not have been able to accomplish what I have.

Finally, I would like to thank Haruna Yokoyama for her effort and support during what seems like a long time in completing this thesis. From translation to apartment hunting, she helped me at every stage of the process. Without her I am not certain I could have completed this project.

## ***Contents***

|  |           |
|--|-----------|
| Abstract .....   | i         |
| Acknowledgements .....   | iii       |
| Contents .....   | iv        |
| List of Tables .....   | vii       |
| List of Figures .....  | xiv       |
| List of Appendices .....   | xvi       |
| <b>Chapter 1. Introduction .....</b>   | <b>1</b>  |
| 1. Why research language learning through watching television? .....                               | 1         |
| 1.1. Are episodes of television a suitable source of input for language learning? ..               | 1         |
| 1.2. The present study .....   | 4         |
| 1.3. Organization of the thesis .....  | 4         |
| <b>Chapter 2. Study 1 – Comprehension of English-language television by EFL<br/>learners .....</b> | <b>5</b>  |
| 2. Introduction .....  | 5         |
| 2.1. Comprehension of television .....   | 5         |
| 2.2. Research questions .....  | 18        |
| 2.3. Participants .....  | 18        |
| 2.4. Procedure .....   | 19        |
| 2.5. Setting .....   | 21        |
| 2.6. Pilot study .....   | 21        |
| 2.7. Materials .....   | 21        |
| 2.8. Analysis .....  | 46        |
| 2.9. Results .....   | 47        |
| 2.10. Summary of findings .....  | 53        |
| 2.11. Discussion .....   | 53        |
| 2.12. Limitations .....  | 58        |
| <b>Chapter 3. Study 2 – Incidental vocabulary learning through viewing<br/>television .....</b>    | <b>59</b> |
| 3. Introduction .....  | 59        |
| 3.1. Research investigating incidental vocabulary acquisition .....                                | 62        |
| 3.2. Research questions .....  | 76        |
| 3.3. Participants .....  | 76        |

|  |  |            |
|--|--|------------|
| 3.4.   | Procedure .....  | 79         |
| 3.5.   | Materials .....  | 81         |
| 3.6.   | Results .....  | 91         |
| 3.7.   | Summary of findings .....  | 96         |
| 3.8.   | Discussion .....   | 97         |
| 3.9.   | Limitations .....  | 102        |
| <b>Chapter 4. Study 3 – The effects of lexical coverage on aspects of viewing</b>    |  |            |
|  | <b>television: comprehension and vocabulary acquisition .....</b>          | <b>103</b> |
| 4.   | Introduction .....   | 103        |
| 4.1.   | Lexical coverage and comprehension .....                                   | 104        |
| 4.2.   | Lexical coverage and incidental vocabulary learning .....                  | 111        |
| 4.3.   | Research questions .....   | 112        |
| 4.4.   | Participants .....   | 112        |
| 4.5.   | Procedure .....  | 113        |
| 4.6.   | Materials .....  | 113        |
| 4.7.   | Results .....  | 115        |
| 4.8.   | Summary of findings .....  | 125        |
| 4.9.   | Discussion .....   | 126        |
| 4.10.  | Limitations .....  | 133        |
| <b>Chapter 5. Study 4 – EFL learners’ attitudes towards learning through viewing</b> |  |            |
|  | <b>English-language television .....</b>                                   | <b>135</b> |
| 5.   | Introduction .....   | 135        |
| 5.1.   | Learners’ attitudes towards language learning through viewing videos ..... | 135        |
| 5.2.   | Research questions .....   | 138        |
| 5.3.   | Participants .....   | 138        |
| 5.4.   | Procedure .....  | 139        |
| 5.5.   | Materials .....  | 140        |
| 5.6.   | Results .....  | 144        |
| 5.7.   | Summary of findings .....  | 156        |
| 5.8.   | Discussion .....   | 157        |
| 5.9.   | Limitations .....  | 161        |

|  |            |
|--|------------|
| <b>Chapter 6. Study 5 – Language learning through viewing television with captions .....</b> | <b>163</b> |
| 6. Introduction .....  | 163        |
| 6.1. Participants .....  | 166        |
| 6.2. Procedure .....   | 168        |
| 6.3. Materials .....   | 170        |
| 6.4. Experiment 1 – Captions and Comprehension .....   | 172        |
| 6.5. Experiment 2 – Captions and Incidental Vocabulary Acquisition .....                     | 189        |
| 6.6. Experiment 3 – Captions and Lexical Coverage .....                                      | 210        |
| 6.7. Experiment 4 – Captions and Attitudes .....   | 222        |
| 6.8. Discussion .....  | 256        |
| 6.9. Limitations .....   | 265        |
| <b>Chapter 7. Conclusion .....</b>   | <b>267</b> |
| 7. Introduction .....  | 267        |
| 7.1. Review of findings .....  | 267        |
| 7.2. Original contribution of the research .....   | 270        |
| 7.3. Pedagogical implications .....  | 271        |
| 7.4. Are episodes of television a suitable source of input for language learning? ..         | 273        |
| 7.5. Limitations and future research .....   | 275        |
| 7.5. Conclusion .....  | 276        |
| References .....   | 277        |
| Notes .....  | 291        |

## *List of Tables*

|            |  |    |
|------------|--|----|
| Table 2.1  | Participants in Study 1 prior to exclusions .....  | 18 |
| Table 2.2  | Participants excluded from the comprehension analyses for Study 1 ..   | 19 |
| Table 2.3  | Running times of the 10 episodes .....   | 23 |
| Table 2.4  | Frequency and range of modified reduced forms in the 10 episodes ...   | 24 |
| Table 2.5  | Tokens, types, word families, and cumulative coverage of the<br>combined episodes of <i>Chuck</i> .....                | 25 |
| Table 2.6  | Number and type of items on the comprehension tests for Episode 1<br>to Episode 5 .....                                | 43 |
| Table 2.7  | Number and type of items on the comprehension tests for Episode 6<br>to Episode 8, Episode A, and Episode B .....      | 43 |
| Table 2.8  | Number of comprehension ability items for each of the 10<br>comprehension tests .....                                  | 44 |
| Table 2.9  | Mean scores of raw data on comprehension tests for Episode A1 and<br>Episode B2 for Group 1 .....                      | 48 |
| Table 2.10 | Mean scores of raw data on comprehension tests for Episode B1 and<br>Episode A2 for Group 2 .....                      | 48 |
| Table 2.11 | Means of CHIPs scores on comprehension tests for Episode A1 and<br>Episode B2 for Group 1 .....                        | 48 |
| Table 2.12 | Means of CHIPs scores on comprehension tests for Episode B1 and<br>Episode A2 for Group 2 .....                        | 48 |
| Table 2.13 | Mean scores of CHIPs scores on the combined comprehension tests<br>for the Initial Episode and the Final Episode ..... | 49 |
| Table 2.14 | Comprehension test scores, expressed in CHIPs and raw scores, for<br>Episode 1 through Episode 8 .....                 | 50 |
| Table 2.15 | Scores on all comprehension tests and the combined results of the<br>VLT expressed in CHIPs .....                      | 53 |
| Table 2.16 | Pearson correlation results for the comparison of vocabulary<br>knowledge and comprehension scores for episodes .....  | 53 |
| Table 3.1  | Video types and running times of previous studies of vocabulary<br>acquisition from video .....                        | 76 |
| Table 3.2  | Experimental Group participants in Study 2 prior to exclusions .....   | 77 |
| Table 3.3  | Control Group participants in Study 2 prior to exclusions .....  | 77 |

|            |  |     |
|------------|--|-----|
| Table 3.4  | Participants excluded from the incidental vocabulary acquisition analysis of Study 2 .....                         | 78  |
| Table 3.5  | Control Group participants excluded from the incidental vocabulary acquisition analysis .....                      | 79  |
| Table 3.6  | Vocabulary test items and their range and frequency in the 10 episodes of <i>Chuck</i> .....                       | 82  |
| Table 3.7  | Vocabulary gains for Tough and Sensitive tests for Experimental Group .....  | 91  |
| Table 3.8  | Vocabulary gains for Tough and Sensitive tests for Control Group ....  | 92  |
| Table 3.9  | Relative vocabulary gains for Tough and Sensitive tests for Experimental Group .....                               | 92  |
| Table 3.10 | Relative vocabulary gains for Tough and Sensitive tests for Control Group .....                                    | 92  |
| Table 3.11 | Vocabulary gains for Tough and Sensitive tests for items .....   | 95  |
| Table 3.12 | Relative gains for Tough and Sensitive Tests for items .....   | 95  |
| Table 4.1  | Mean comprehension scores for participants in Stæhr's (2009) study by mastery of levels of the VLT .....           | 107 |
| Table 4.2  | Scores on the VLTs for Comprehension Participants .....  | 116 |
| Table 4.3  | Scores on the VLTs for Vocabulary Participants .....   | 116 |
| Table 4.4  | Mean comprehension scores for Comprehension Participants not reaching mastery of any VLT .....                     | 118 |
| Table 4.5  | Mean comprehension scores for Comprehension Participants reaching mastery of the VLT at the 2,000-word level ..... | 118 |
| Table 4.6  | Mean comprehension scores for Comprehension Participants reaching mastery of the VLT at the 3,000-word level ..... | 118 |
| Table 4.7  | Mean comprehension scores for Comprehension Participants reaching mastery of the VLT at the 5,000-word level ..... | 118 |
| Table 4.8  | Summary of ANOVA results for comprehension of episodes of <i>Chuck</i> and lexical coverage .....                  | 120 |
| Table 4.9  | LSD post hoc comparison of comprehension of episodes among the four groups of lexical coverage .....               | 120 |
| Table 4.10 | Mean comprehension scores of significantly different groups of participants with different lexical coverage .....  | 122 |

|            |   |     |
|------------|---|-----|
| Table 4.11 | Vocabulary gains for Tough and Sensitive tests for Vocabulary<br>Participants not reaching mastery of any VLT .....                                 | 123 |
| Table 4.12 | Relative vocabulary gains for Tough and Sensitive tests for<br>Vocabulary Participants not reaching mastery of any VLT .....                        | 123 |
| Table 4.13 | Vocabulary gains for Tough and Sensitive tests for Vocabulary<br>Participants reaching mastery of the VLT at the 2,000-word<br>level .....          | 123 |
| Table 4.14 | Relative vocabulary gains for Tough and Sensitive tests for<br>Vocabulary Participants reaching mastery of the VLT at the<br>2,000-word level ..... | 124 |
| Table 4.15 | Vocabulary gains for Tough and Sensitive tests for Vocabulary<br>Participants reaching mastery of the VLT at the 3,000-word<br>level .....          | 124 |
| Table 4.16 | Relative vocabulary gains for Tough and Sensitive tests for<br>Vocabulary Participants reaching mastery of the VLT at the<br>3,000-word level ..... | 124 |
| Table 4.17 | Vocabulary gains for Tough and Sensitive tests for Vocabulary<br>Participant reaching mastery of the VLT at the 5,000-word<br>level .....           | 124 |
| Table 4.18 | Relative vocabulary gains for Tough and Sensitive tests for<br>Vocabulary Participant reaching mastery of the VLT at the<br>5,000-word level .....  | 124 |
| Table 4.19 | Comparison of comprehension scores for Episode 5 and Episode 7<br>for participants at four vocabulary levels .....                                  | 130 |
| Table 4.20 | Vocabulary gains on the Tough and Sensitive tests at four different<br>lexical coverage levels .....  | 132 |
| Table 5.1  | Participants excluded from the attitudes analysis in Study 4 .....  | 139 |
| Table 5.2  | Responses to FASQs 1, 2 and 3 .....   | 150 |
| Table 5.3  | Responses to FASQ 7 .....   | 150 |
| Table 5.4  | Responses to FASQ 6 .....   | 151 |
| Table 5.5  | Responses to FASQs 4 and 12 .....   | 152 |
| Table 5.6  | Responses to FASQ 5 .....   | 153 |
| Table 5.7  | Responses to FASQs 8 and 9 .....  | 154 |

|            |  |     |
|------------|--|-----|
| Table 5.8  | Responses to FASQ 10 .....   | 154 |
| Table 5.9  | Responses to FASQ 11 .....   | 155 |
| Table 5.10 | Responses to FAS Forced-choice Items 1 to 5 .....  | 156 |
| Table 6.1  | Participants in Study 5 prior to exclusions .....  | 166 |
| Table 6.2  | Participants excluded from the comprehension, lexical coverage and<br>comprehension, and attitudes analyses for Study 5 .....  | 167 |
| Table 6.3  | Participants excluded from the incidental vocabulary acquisition, and<br>lexical coverage and vocabulary acquisition analyses for<br>Study 5 .....                   | 168 |
| Table 6.4  | Mean scores of raw data on the comprehension tests for Episode A1<br>and Episode B2 .....  | 180 |
| Table 6.5  | Means of CHIPs scores on the comprehension tests for Episode A1<br>and Episode B2 .....  | 180 |
| Table 6.6  | Comprehension test scores, expressed in CHIPs and raw scores, for<br>all episodes .....  | 181 |
| Table 6.7  | Mean CHIPs scores on the comprehension tests for all episodes in<br>Study 5 and Study 1 .....  | 183 |
| Table 6.8  | Mean gains in CHIPs from the Initial Episode to the Final Episode in<br>Study 5 and Study 1 .....  | 184 |
| Table 6.9  | Pearson correlation results for the comparisons between vocabulary<br>knowledge and comprehension scores for episodes of <i>Chuck</i> .                              | 186 |
| Table 6.10 | Pearson correlation results comparing vocabulary knowledge and<br>comprehension scores for episodes of <i>Chuck</i> for the Captions<br>and No Captions groups ..... | 187 |
| Table 6.11 | Vocabulary gains for Tough and Sensitive tests for Study 5<br>participants .....   | 203 |
| Table 6.12 | Relative vocabulary gains for Tough and Sensitive Tests for Study 5<br>participants .....  | 203 |
| Table 6.13 | Vocabulary gains for Tough and Sensitive Tests for Control Group<br>participants .....   | 204 |
| Table 6.14 | Vocabulary gains for Tough and Sensitive Tests for items .....   | 205 |
| Table 6.15 | Relative gains for Tough and Sensitive Tests for items .....   | 206 |

|            |  |     |
|------------|--|-----|
| Table 6.16 | Vocabulary gains for Tough and Sensitive Tests for participants in the Captions Group and the No Captions Group .....  | 207 |
| Table 6.17 | Relative vocabulary gains for Tough and Sensitive Tests for participants in the Captions Group and the No Captions Group .....   | 208 |
| Table 6.18 | The lexical coverage of the 10 episodes at the 2,000-, 3,000- and 5,000-word family levels .....   | 211 |
| Table 6.19 | Scores on VLT for the analysis of lexical coverage and comprehension .....   | 212 |
| Table 6.20 | Scores on VLT for the analysis of lexical coverage and incidental vocabulary learning .....  | 212 |
| Table 6.21 | Mean comprehension scores for participants not reaching mastery of any VLT .....   | 213 |
| Table 6.22 | Mean comprehension scores for participants reaching mastery of the VLT at the 2,000-word level .....   | 213 |
| Table 6.23 | Mean comprehension scores for participants reaching mastery of the VLT at the 3,000-word level .....   | 213 |
| Table 6.24 | Comparison of scores on VLT for Captions and No Captions groups .  | 215 |
| Table 6.25 | Mean comprehension scores in CHIPs for participants in the Captions and No Captions groups who failed to master the VLT at the 2,000-word level, and participants that mastered the 2,000-, 3,000- and 5,000-word levels ..... | 216 |
| Table 6.26 | Vocabulary gains for participants not reaching mastery of any VLT ..   | 218 |
| Table 6.27 | Relative vocabulary gains for participants not reaching mastery of any VLT .....   | 218 |
| Table 6.28 | Vocabulary gains for participants reaching mastery of the VLT at the 2,000-word level .....  | 218 |
| Table 6.29 | Relative vocabulary gains for participants reaching mastery of the VLT at the 2,000-word level .....   | 218 |
| Table 6.30 | Vocabulary gains for participants reaching mastery of the VLT at the 3,000-word level .....  | 218 |
| Table 6.31 | Relative vocabulary gains for participants reaching mastery of the VLT at the 3,000-word level .....   | 219 |

|            |   |     |
|------------|---|-----|
| Table 6.32 | Comparison of vocabulary gains by mastery of levels of the VLT for<br>the Captions Group and No Captions Group .....                          | 220 |
| Table 6.33 | Responses to FASQs 1, 2 and 3 .....   | 237 |
| Table 6.34 | Responses to FASQ 7 .....   | 237 |
| Table 6.35 | Responses to FASQ 6 .....   | 238 |
| Table 6.36 | Responses to FASQs 4 and 12 .....   | 239 |
| Table 6.37 | Responses to FASQ 5 .....   | 240 |
| Table 6.38 | Responses to FASQs 8 and 9 .....  | 241 |
| Table 6.39 | Responses to FASQ 10 .....  | 241 |
| Table 6.40 | Responses to FASQ 11 .....  | 242 |
| Table 6.41 | Responses to FASQs 13, 14 and 15 .....  | 243 |
| Table 6.42 | Responses to FASQ 16 .....  | 244 |
| Table 6.43 | Responses to Forced-choice Item 1 on the FAS .....  | 244 |
| Table 6.44 | Responses to FAS Forced-choice Items 2 to 6 .....   | 245 |
| Table 6.45 | Responses to FASQ 1, 2 and 3 for the Captions Group and No<br>Captions Group .....  | 247 |
| Table 6.46 | Responses to FASQ 7 for the Captions Group and No Captions<br>Group .....   | 247 |
| Table 6.47 | Responses to FASQ 6 for the Captions Group and No Captions<br>Group .....   | 248 |
| Table 6.48 | Responses to FASQs 4 and 12 for the Captions Group and No<br>Captions Group .....   | 249 |
| Table 6.49 | Responses to FASQ 5 for the Captions Group and No Captions<br>Group .....   | 249 |
| Table 6.50 | Responses to FASQs 8 and 9 for the Captions Group and No<br>Captions Group .....  | 250 |
| Table 6.51 | Responses to FASQ 10 for the Captions Group and No Captions<br>Group .....  | 251 |
| Table 6.52 | Responses to FASQ 11 for the Captions Group and No Captions<br>Group .....  | 251 |
| Table 6.53 | Comparison of responses to FAS Forced-choice Items for<br>perceptions of difficulty for the Captions Group and the No<br>Captions Group ..... | 253 |

|            |   |     |
|------------|---|-----|
| Table 6.54 | Comprehension scores, lexical coverage and effect sizes of episodes<br>where there was a significant difference between<br>participants with and without mastery of 2,000-word<br>level ..... | 261 |
|------------|---|-----|

## ***List of Figures***

|            |  |     |
|------------|--|-----|
| Figure 2.1 | Schedule for Study 1 .....   | 20  |
| Figure 2.2 | First three true/false items on the comprehension test for Episode 1 ...   | 33  |
| Figure 2.3 | Item #10 on the comprehension test for Episode 1 .....   | 34  |
| Figure 2.4 | First two multiple-choice items on the comprehension test for<br>Episode 1 .....   | 35  |
| Figure 2.5 | Sequencing items for Episode 1 .....   | 36  |
| Figure 2.6 | Examples of OMR answer fields for the comprehension tests .....  | 46  |
| Figure 2.7 | Mean CHIPs comprehension scores across all 10 episodes .....   | 51  |
| Figure 2.8 | Maximum, average and minimum comprehension scores expressed<br>as percentages across all eight successive episodes .....   | 52  |
| Figure 3.1 | Summary of previous research investigating incidental vocabulary<br>acquisition through reading, listening and viewing videos .....  | 63  |
| Figure 3.2 | Research schedule for Study 2 .....  | 80  |
| Figure 3.3 | Explanation and examples of the options for items on the Tough Test<br>using the target word ‘spy’ .....   | 85  |
| Figure 3.4 | Item #1 from the Tough Pre-Test .....  | 86  |
| Figure 3.5 | Explanation and examples of the options for items on the Sensitive<br>Test using the target word ‘spy’ .....   | 87  |
| Figure 3.6 | Item #1 from the Sensitive Pre-Test .....  | 88  |
| Figure 4.1 | First three items of the 2,000-word level of the VLT .....   | 114 |
| Figure 4.2 | Cumulative lexical coverage of the 10 episodes at the 2,000, 3,000<br>and 5,000 word-family levels .....   | 115 |
| Figure 4.3 | Mean comprehension scores in CHIPs for Comprehension<br>Participants with mastery of different levels of the VLT<br>across the 10 episodes of <i>Chuck</i> .....                     | 119 |
| Figure 4.4 | Mean comprehension scores for the participants with and without<br>mastery of the 2,000-word level for episodes where there<br>was a significant difference between the groups ..... | 126 |
| Figure 5.1 | Question 2 of the EAS .....  | 142 |
| Figure 5.2 | EASQ 1 .....   | 144 |
| Figure 5.3 | EASQ 2 and EASQ 3 .....  | 144 |
| Figure 5.4 | EASQ 4 .....   | 144 |

|             |  |     |
|-------------|--|-----|
| Figure 5.5  | Results across all 10 episodes for EASQs 1, 3 and 4 .....  | 147 |
| Figure 5.6  | Results across all 10 episodes for EASQ 2 .....  | 148 |
| Figure 6.1  | Schedule for Study 5 .....   | 169 |
| Figure 6.2  | Summary of previous research examining comprehension of video<br>with and without captions .....                 | 173 |
| Figure 6.3  | Mean CHIPs comprehension scores across all 10 episodes .....   | 182 |
| Figure 6.4  | Maximum, average and minimum comprehension scores expressed<br>as percentages across all 10 episodes .....       | 182 |
| Figure 6.5  | Mean CHIPs comprehension scores across all 10 episodes for the<br>Captions Group and the No Captions Group ..... | 185 |
| Figure 6.6  | Summary of previous research comparing vocabulary acquisition<br>through video with and without captions .....   | 190 |
| Figure 6.7  | EASQ 5 .....   | 226 |
| Figure 6.8  | Results across all 10 episodes for EASQs 1, 3 and 4 .....  | 228 |
| Figure 6.9  | Results across all 10 episodes for EASQ 2 .....  | 229 |
| Figure 6.10 | Results across all 10 episodes for EASQ 5 .....  | 230 |
| Figure 6.11 | Results across all 10 episodes for EASQ 1 for the Captions and No<br>Captions groups .....                       | 231 |
| Figure 6.12 | Results across all 10 episodes for EASQ 2 for the Captions and No<br>Captions groups .....                       | 232 |
| Figure 6.13 | Results across all 10 episodes for EASQ 3 for the Captions and No<br>Captions groups .....                       | 233 |
| Figure 6.14 | Results across all 10 episodes for EASQ 4 for the Captions and No<br>Captions groups .....                       | 234 |
| Figure 6.15 | Results for all 12 FASQs for the Captions and No Captions groups ...   | 264 |

## ***List of Appendices***

**\*\*All appendices contained on the attached CD-ROM**

### **Appendix A          *Chuck* Comprehension Tests**

|              |   |
|--------------|---|
| Appendix A1  | Episode 1 Answer Sheet                  |
| Appendix A2  | Episode 1 Comprehension Test English    |
| Appendix A3  | Episode 1 Comprehension Test Japanese   |
| Appendix A4  | Episode 2 Answer Sheet                  |
| Appendix A5  | Episode 2 Comprehension Test English    |
| Appendix A6  | Episode 2 Comprehension Test Japanese   |
| Appendix A7  | Episode 3 Answer Sheet                  |
| Appendix A8  | Episode 3 Comprehension Test English    |
| Appendix A9  | Episode 3 Comprehension Test Japanese   |
| Appendix A10 | Episode 4 Answer Sheet                  |
| Appendix A11 | Episode 4 Comprehension Test English    |
| Appendix A12 | Episode 4 Comprehension Test Japanese   |
| Appendix A13 | Episode 5 Answer Sheet                  |
| Appendix A14 | Episode 5 Comprehension Test English    |
| Appendix A15 | Episode 5 Comprehension Test Japanese   |
| Appendix A16 | Episode 6 Answer Sheet                  |
| Appendix A17 | Episode 6 Comprehension Test English    |
| Appendix A18 | Episode 6 Comprehension Test Japanese   |
| Appendix A19 | Episode 7 Answer Sheet                  |
| Appendix A20 | Episode 7 Comprehension Test English    |
| Appendix A21 | Episode 7 Comprehension Test Japanese   |
| Appendix A22 | Episode 8 Answer Sheet                  |
| Appendix A23 | Episode 8 Comprehension Test English    |
| Appendix A24 | Episode 8 Comprehension Test Japanese   |
| Appendix A25 | Episode A12 Answer Sheet                |
| Appendix A26 | Episode A12 Comprehension Test English  |
| Appendix A27 | Episode A12 Comprehension Test Japanese |
| Appendix A28 | Episode B13 Answer Sheet                |
| Appendix A29 | Episode B13 Comprehension Test English  |
| Appendix A30 | Episode B13 Comprehension Test Japanese |

|                   |   |
|-------------------|---|
| Appendix A31      | Item Types on Comprehension Tests for Episodes of <i>Chuck</i>                          |
| <b>Appendix B</b> | <b>Television Viewing Practice</b>  |
| Appendix B1       | How to Answer Comprehension Tests   |
| Appendix B2       | Introduction to Viewing Television  |
| Appendix B3       | Mr. Bean Practice Test Answer Sheet and Key   |
| Appendix B4       | Mr. Bean Practice Test Japanese   |
| Appendix B5       | Mr. Bean Comprehension Test Bilingual   |
| <b>Appendix C</b> | <b>Vocabulary Levels Test</b>   |
| Appendix C1       | Vocabulary Levels Test 2,000-word Level   |
| Appendix C2       | Vocabulary Levels Test 3,000-word Level   |
| Appendix C3       | Vocabulary Levels Test 5,000-word Level   |
| <b>Appendix D</b> | <b>Attitude Surveys</b>   |
| Appendix D1       | Study 4 Episode Attitude Survey English   |
| Appendix D2       | Study 4 Episode Attitude Survey Japanese  |
| Appendix D3       | Study 4 Final Attitude Survey Answer Sheet  |
| Appendix D4       | Study 4 Final Attitude Survey English   |
| Appendix D5       | Study 4 Final Attitude Survey Japanese  |
| Appendix D6       | Study 5 Episode Attitude Survey English   |
| Appendix D7       | Study 5 Episode Attitude Survey Japanese  |
| Appendix D8       | Study 5 Final Attitude Survey English   |
| Appendix D9       | Study 5 Final Attitude Survey Japanese  |
| Appendix D10      | Study 5 Final Survey Answer Sheet   |
| <b>Appendix E</b> | <b>Results</b>  |
| Appendix E1       | RANGE Results for Episode of <i>Chuck</i>   |
| Appendix E2       | Results of 60-item <i>Chuck</i> -specific Vocabulary Pre- and Post-Tests<br>for Study 2 |
| Appendix E3       | Results of 60-item <i>Chuck</i> -specific Vocabulary Pre- and Post-Tests<br>for Study 5 |
| <b>Appendix F</b> | <b><i>Chuck</i>-specific Vocabulary Tests</b>   |
| Appendix F1       | Vocabulary Post-Test Answer Page - Sensitive  |
| Appendix F2       | Vocabulary Post-Test Answer Page - Tough  |
| Appendix F3       | Vocabulary Post-Test Question Sheets - Sensitive  |
| Appendix F4       | Vocabulary Post-Test Question Sheets - Tough  |

|             |   |
|-------------|---|
| Appendix F5 | Vocabulary Pre-Test Answer Page - Sensitive     |
| Appendix F6 | Vocabulary Pre-Test Answer Page - Tough         |
| Appendix F7 | Vocabulary Pre-Test Question Sheets - Sensitive |
| Appendix F8 | Vocabulary Pre-Test Question Sheets - Tough     |

## **Appendix G      Miscellaneous**

|             |  |
|-------------|--|
| Appendix G1 | Ethics Approval                        |
| Appendix G2 | Participant Information Sheet English  |
| Appendix G3 | Participant Information Sheet Japanese |

---

## *Chapter 1*

---

### ***Introduction***

#### **1. Why research language learning through watching television?**

It is well established that language input is an important component of language learning. This input, in both its written and spoken forms, should be authentic (Nunan, 2002) and comprehensible (Krashen, 1985; Neuman & Koskinen, 1992). However, in the English as a Foreign Language (EFL) setting it may be a challenge to provide second language (L2) learners with sufficient input that meets these requirements. To provide L2 learners with suitable and sufficient written input, graded readers have become a popular source of authentic reading materials. As part of extensive reading programs, learners are encouraged to choose books at a comprehensible level and read as many as possible. Sources of comprehensible and authentic listening input are not so readily available. Episodes of television made for an English-speaking audience may be a potential source of L2 aural input. However, little previous research has been done that might indicate whether episodes of television are a suitable source of listening input for an EFL setting.

#### **1.1. Are episodes of television a suitable source of input for language learning?**

As part of his description of a well-designed language program, Nation (2007) outlines five conditions that must be met in order for input to be considered suitable for inclusion in a language course. By examining these conditions and their relationship to the nature of television and past research on language learning from videos it is possible to gauge the suitability of episodes of television to serve as L2 aural input.

The first of these conditions is that the input needs to be processed in large quantities. With some certainty it is possible to say that television is consumed in large quantities in the L1. Watching television is the preferred leisure activity across countries surveyed by the Organization of Economic Co-operation and Development. The amount of daily television watched per household ranges from the world leader, The U.S.A., at 8.23 hours per day to Norway at 2.39 hours a day. Other notable countries include Japan at 3.63 hours a day and New Zealand at 2.88 hours a day (OECD, 2009). If language learners were to spend even a portion of their L1 viewing time on L2 television they would be processing a large amount of input. In this way, television meets the requirement that a suitable form of input be available and consumed in large quantities.

The second condition for suitable input is that it should be familiar to the language learners (Nation, 2007). Learners should be familiar with the content of the input texts they are processing, so that they can achieve a reasonable level of comprehension. Television programs are available in many genres and cover a wide variety of subjects. Language learners intending to learn from television could choose a program that they were interested in and had some familiarity with. Learners could also build up familiarity with a program by viewing multiple and successive episodes. In doing so the learners would learn about the characters and storylines which may improve comprehension in future episodes. Previous research has indicated that language learners have been able to comprehend video (Baltova, 1999; Chung, 1999; Etemadi, 2012; Guichon & McLornan, 2008; Guillory, 1998; Huang & Eskey, 1999; Latifi, Mobalegh, & Mohammadi, 2011; Markham, Peter, & McCarthy, 2001; Markham & Peter, 2003; Taylor, 2005; Winke, Gass, & Sydorenko, 2010). However, at present, language learners' comprehension levels of episodes of television are unknown.

The third condition is that learners should be able to gain knowledge of vocabulary using context cues and background knowledge (Nation, 2007). In order for this to occur the form of input must be rich in context cues and there must be ways of building background knowledge. The imagery and dialogue in television provide a source of context cues that learners may be able to use to help them make vocabulary gains. Early on in the viewing process, learners may not have a lot of background knowledge about a television series. However, the episodic nature of television with its related episodes and recurring characters means that learners have the potential to increase background knowledge through viewing multiple episodes of the same program. If learners can make use of context cues and background knowledge when viewing episodes of television they may be able to gain knowledge of previously unknown vocabulary items as they have in past video-based research (Baltova, 1999; Huang & Eskey, 1999; Hui, 2007; Markham et al., 2001; Markham, 1999; Neuman & Koskinen, 1992; Sydorenko, 2010; Vidal, 2003, 2011; Winke et al., 2010; Yuksel & Tanriverdi, 2009).

The fourth condition is that only a small percentage of the vocabulary in the input texts is unknown to learners (Nation, 2007). It stands to reason that if a large proportion of the vocabulary were to be unknown to learners they would be impeded from comprehension or vocabulary learning. In reading and listening research, there have been studies on the percentage of known vocabulary necessary for comprehension and vocabulary learning to take place. These estimates of lexical coverage range from 90% to 99% for comprehension (Bonk, 2000; Carver, 1994; Hirsh & Nation, 1992; Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Laufer & Sim, 1985; Laufer, 1989; Nation, 2006; Schmitt, Jiang, & Grabe,

2011; Stæhr, 2009; van Zeeland & Schmitt, 2012) and 95% to 98% for vocabulary acquisition (Liu & Nation, 1985; Nation, 2001; Webb & Rodgers, 2009a, 2009b). At present, the relationship between the percentage of known vocabulary and its effects on comprehension of television and incidental vocabulary acquisition from viewing television has yet to be studied.

The fifth condition that must be met in order for input to be considered suitable for inclusion in language programs is that learners should be interested in the input and want to understand it (Nation, 2007). It has been established that watching television is a popular leisure activity and previous research has indicated that language learners are interested in learning through videos (Chung & Huang, 1998; Chung, 1999; Gruba, 2006; Wang, 2012). Viewing television is also very much a comprehension-focused activity with learners indicating that they believed comprehension of video is easier than other input modes (Brett, 1997; Chung & Huang, 1998; Gruba, 2004, 2006; Hasan, 2000). Language learners' attitudes towards learning from television indicate that episodes of television may fulfill this fifth condition; however, there has been no research conducted on attitudes towards L2 learning through viewing multiple episodes of television.

One feature of television that may improve the likelihood of it being a suitable source of L2 aural input is captions. The presence of captions when watching television may allow learners to make use of the written form of the language to better utilize episodes for language learning (Bird & Williams, 2002; Garza, 1991). EFL learners have been shown to have stronger reading skills than listening skills (Hirai, 1999) which may allow them to process more of the input than if it were received through aural means only. This may have ramifications for all five conditions for the suitability of television. The presence of captions may make language learners more eager to view large amounts of television, make episodes of television more comprehensible, increase vocabulary learning, affect the relationship between lexical coverage and comprehension and vocabulary learning, and improve learners' attitudes towards learning from television. At present, little research has been done on the effects of captions on these aspects of language learning.

Taken together, this examination of the nature of television and past video research indicates that episodes of television may fulfill the conditions outlined by Nation (2007) to be considered a suitable source of listening input for L2 learning. It also appears that captions may improve the suitability of television programs to function as a source of input. However, this is for the most part merely speculative as past research has not explicitly focused on authentic episodes of television. Previous video-based research has generally focused on short

videos and videos that learners might not normally choose to watch. There has not been any research on viewing full-length episodes of television programs on aspects of language learning.

## **1.2. The present study**

The lack of previous research and the potential of full-length episodes of television to be a source of aural input for L2 learning informs the present study which was designed to examine following aspects of language learning:

1. Comprehension of episodes of television
2. Vocabulary acquisition through viewing television
3. Lexical coverage and its effects on comprehension of television and vocabulary acquisition from viewing television
4. Attitudes towards language learning through viewing television
5. Captions and their effects on the aspects of language learning described in 1 to 4

## **1.3. Organization of the thesis**

This thesis is organized around individual studies. Each study is presented with separate literature review, research questions, methodology, results, discussion, and limitations sections. This format allows for a clearer indication of the aims and results of each study.

In general terms, each of the studies in this thesis is concerned with English language learning through viewing 10 episodes of an American television program. Study 1 (Chapter 2) investigates language learners' comprehension of the episodes of television. The relationship between vocabulary knowledge and comprehension is also addressed. Study 2 (Chapter 3) examines incidental vocabulary acquisition through viewing episodes of television. The effects of frequency and range of occurrence on vocabulary acquisition are also examined. Study 3 (Chapter 4) investigates the effects of lexical coverage on comprehension of television and on vocabulary acquisition through viewing television. The language learners' attitudes towards learning English through viewing episodes of television are examined in Study 4 (Chapter 5). Study 5 (Chapter 6) seeks to determine how the presence of captions affects the aspects of language learning that were investigated in Studies 1 through 4. The final chapter (Chapter 7) of this thesis provides a discussion of the major findings from these five studies, as well as the pedagogical implications of these findings and a consideration of potential limitations of the overall research.

### ***Study 1: Comprehension of English-language television by EFL learners***

#### **2. Introduction**

Television programs are a potential source of authentic input for EFL learners and while there are certain characteristics of television that may make it comprehensible there has not been any prior research involving comprehension of full-length episodes of television. It is unknown at present the extent to which language learners can comprehend episodes of television. There are factors, however, that might contribute to comprehension of television including: the imagery that accompanies the aural input from television programs, accumulation of background knowledge from viewing related episodes of television, and the vocabulary knowledge of the viewers. Listening comprehension studies have found that the imagery associated with video can lead to increased comprehension (e.g. Brett, 1997; Guichon & McLornan, 2008; Sueyoshi & Hardison, 2005). Past research has also indicated that knowledge about a video prior to viewing can result in increased comprehension and that background knowledge about videos may be acquired through viewing related videos (e.g. Chang & Read, 2006; Chung, 1999; Herron, York, Cole, & Linden, 1998). Reading (e.g. (Laufer & Ravenhorst-Kalovski, 2010; Schmitt et al., 2011) and listening (e.g. Bonk, 2000; Milton, Wade, & Hopkins, 2010; Stæhr, 2009) research has indicated that a language learner's vocabulary knowledge is an important factor in comprehension. The relationship of these factors to episodes of authentic television has been largely ignored yet knowledge of how these factors contribute to comprehension of television may determine the suitability of television as listening input for language learners. The present study was designed to examine (a) whether Japanese EFL learners' comprehension of episodes of authentic television changes from the first to the tenth episode viewed, (b) how comprehension of episodes of television changes across successive episodes viewed, and (c) whether language learners' vocabulary knowledge is related to their comprehension of episodes of television.

#### **2.1. Comprehension of television**

Comprehension of a television program involves the processes associated with listening comprehension but with the added support of a visual component. Fundamentally, listening comprehension is an inferential process in which the viewer constructs meaning from available knowledge sources. These knowledge sources can be differentiated as either

linguistic or non-linguistic. Linguistic sources can include phonological, lexical, syntactic, semantic, and discourse knowledge. Non-linguistic sources of knowledge can include topical, contextual and world knowledge (Buck, 2001). These knowledge types are utilized through top-down and bottom-up processing in a complex interaction the listeners use to create a mental representation of the input (Park, 2004; Rubin, 1994; Vandergrift, 2004). Bottom-up processing begins with decoding phonemes to identify individual words and construct a literal understanding of the text. Top-down processing is dependent on the background knowledge the listener brings to the text. This knowledge allows him or her to make inferences from the content (Long, 1990). These two processes are used simultaneously to construct meaning (Brindley, 1998). The contributions from top-down and bottom-up processing to the comprehension process are not constant and their relative contribution can change within different parts of a listening text (Wagner, 2002). When sufficient information has been processed through top-down and bottom-up processes comprehension can occur (Buck, 2001).

A factor that affects the degree to which top-down or bottom-up processes contribute to the comprehension process is the language proficiency of the listener. The way language learners process aural input has been found to be different for high and low proficiency learners (Wolff, 1987). Low proficiency language learners have been shown to focus more on bottom-up processes. They concentrate on trying to distinguish words in the stream of aural input. As proficiency increases, however, more attention is paid to the top-down processes (Conrad, 1985; Hansen & Jensen, 1994). When the listener can predict the spoken text of a listening passage, they do not need to rely as much on bottom-up processing. Lower proficiency EFL listeners, however, are rarely able to predict content and therefore must rely on bottom-up processing (Kelly, 1991). One characteristic of television viewing that allows language learners to utilize top-down processing is imagery. The visual elements of television enhance the top-down processing and can positively affect comprehension (Gruba, 2004; Vandergrift, 2007). This may make television a source of comprehensible aural input.

### **2.1.1. The effects of visual imagery on listening comprehension**

Whether a listening text has associated visual imagery or not has been identified as a major factor affecting listening comprehension (Rubin, 1994). The combination of aural and visual input gives viewers the opportunity to comprehend information through different channels and make connections between them (Guichon & McLornan, 2008). The theoretical foundation of this interaction is Paivio's (1990) Dual Coding Theory. In this theory, learning

is enhanced when language learners can select visual and aural information, organize the information in their working memory, build connections between the visual and aural information, and integrate the connections into their prior knowledge (Jones & Plass, 2002; Mayer & Moreno, 1998).

Studies have indicated how language learners apply the stages described by the Dual Coding Theory and make use of the images available in videos for comprehension. Gruba (2004, 2006) investigated the way that learners of Japanese interacted with television news segments. Findings indicated that learners used all aspects of a video (aural and visual) for comprehension. In the initial stages of viewing the visual elements were shown to provide primary support that shifts to secondary support as a better understanding of the video develops. However, when images were presented in a manner learners considered confusing, they indicated that the images provided no assistance to comprehension or could even impair comprehension. Hasan (2000), in a study looking at English learners' listening strategy use, found that visual support from still pictures and video helped learners understand listening texts. Participants reported that the visual clues in the form of pictures, diagrams and charts helped them understand spoken text. The use of video in listening comprehension exercises was shown to facilitate information processing. Visual support not only made the topic more comprehensible but also reduced anxiety when listeners were unfamiliar with what speakers were talking about.

A number of studies have presented findings that support the hypothesis that the visual imagery associated with videos positively affects the comprehension process (Brett, 1997; Guichon & McLornan, 2008; Jones & Plass, 2002; Maleki & Safaee Rad, 2011; Mueller, 1980; Ockey, 2007; Sueyoshi & Hardison, 2005). Three of these studies investigated the effects of still images on listening comprehension (Jones & Plass, 2002; Maleki & Safaee Rad, 2011; Mueller, 1980). A single study compared the comprehension of listening texts that were accompanied by either still images or video (Ockey, 2007). Three studies compared the comprehension of listening texts that were accompanied by video or presented in an audio-only form (Brett, 1997; Guichon & McLornan, 2008; Sueyoshi & Hardison, 2005). The following review of relevant studies focuses on how the inclusion of images with listening texts can affect language learners' comprehension.

Three studies compared the comprehension of listening texts that were presented with or without still images. In the first study, Mueller (1980) studied the effects of imagery, in the form of a line drawing, on the listening comprehension of students learning German.

Participants listened to a short interview and viewed a line drawing representing the situation before listening, after listening, or not at all. Comprehension was measured by written summaries. Less proficient participants in the treatment groups that viewed the image, performed significantly better than the participants that did not see the image (Low Before:  $M=9.78$ ; Low After:  $M=7.96$ ; Low No Image:  $M=5.25$ ). The group that viewed the image before listening had significantly higher comprehension scores than the group that viewed the image following listening. The more proficient participants did not have significantly different comprehension scores (High Before:  $M=14.37$ ; High After:  $M=11.67$ ; High No Image:  $M=11.50$ ) regardless of the treatment group.

Likewise, Jones and Plass (2002) studied the effects on comprehension of listening to a text with and without related pictures. Participants were English-speaking university students studying French. Comprehension was measured by a written recall protocol immediately after listening and then again after 3 weeks. On the immediate recall test, participants who had access to the images ( $M=9.2$  out of possible 63 idea units) had significantly higher comprehension scores than the group without access to images ( $M=3.2$  idea units). The results of the delayed posttest showed similar results for the treatment groups but with lower mean scores than on the immediate test.

Maleki and Safaee Rad (2011) obtained similar results studying the effects of still images on English language learners' performance on the listening portion of the IELTS test. Participants took two versions of the listening test, the first without any support and a second accompanied by images related to the test items. Participants were grouped by High and Low proficiency. Low proficiency participants with access to the images (High:  $M=26.73$ ; Low:  $M=19.36$ ) had significantly higher scores than when they took the test without images (High:  $M=25.53$ ; Low:  $M=15.57$ ). The high proficiency participants, however, did not have significantly different scores on the two listening tests.

Taken as a whole, the results from these three studies indicate that listening comprehension increases when language learners have access to still images related to the information presented in the audio. However, images are possibly more effective for lower level proficiency learners. These findings indicate that the more robust imagery associated with video may also provide support for listening comprehension.

A study by Ockey (2007) looked at the differences in the way language learners react to having either images or video present when listening to a lecture. English as Second Language (ESL) students at an American university listened to two lectures either accompanied by a video or still images of the lecturer. Ockey measured the time the learners

spent observing the video or still images while they listened to the lecture and completed a comprehension test. Participants reported on whether the video or still images helped or distracted them and which visual cues they used while listening. The study found that the majority of the participants spent considerably more time watching the video than looking at the stills. The responses from the participants indicated that the imagery was helpful. The learners did not report using any visual cues with the still images but reported using a variety of cues with the video. These included observing lip movements, hand motions, facial gestures, and body gestures to gain more information about the lecture. The learners' comprehension scores were not reported. This study indicates video has advantages over still images for maintaining language learners' attention and promoting the use of visual cues that may lead to increased comprehension.

Three studies compared listening comprehension of texts that were either presented as audio-only or accompanied by video. In the first study, Brett (1997) compared the comprehension of advanced English language learners who completed listening tasks based on a business English video series. The participants in the video treatment ( $M=56.9\%$ ) had higher mean comprehension scores than the audio-only treatment group ( $M=51.3\%$ ). Statistical significance of the results was not reported. Sueyoshi and Hardison (2005) studied how the presence of gestures and facial cues affected ESL learners' listening comprehension when watching or listening to a lecture. Comprehension was measured through a 20-item multiple-choice test. Low-intermediate learners had the highest mean comprehension scores when they had access to both gestures and facial cues ( $M=10.14$ ), followed by access to only facial cues ( $M=8.71$ ). The audio-only treatment had the lowest mean comprehension score ( $M=7.57$ ). For the advanced proficiency learners, the highest mean comprehension scores were associated with the facial cues treatment ( $M=13.29$ ), followed by the treatment with access to gestures and facial cues ( $M=11.14$ ), and finally the audio-only treatment ( $M=8.57$ ). Overall, the participants with access to both gestures and facial cues and only facial cues had significantly better comprehension scores than the participants in the audio-only treatment. As part of a larger study also examining L2 captions and L1 subtitles, Guichon and McLornan (2008) compared comprehension of a 3-minute news report viewed with or without video imagery. Comprehension was measured by a written summary of the report. The audio-only group made reference to an average 19.7% of the 35 possible semantic units and the video group referenced 25.1%. While the group that viewed the video had a higher mean comprehension score, the statistical significance of the results were not reported. The results

of these studies indicate the benefits that the presence of video can have on listening comprehension.

The results of previous research comparing listening comprehension with and without images indicate that comprehension is increased when learners have access to images. Still images and video were shown to lead to better comprehension than audio-only treatments. The research also indicates that images associated with listening may have a more beneficial effect on comprehension for lower proficiency learners (Maleki & Safaee Rad, 2011; Mueller, 1980). These findings indicate the potential suitability of television as a source of listening input for language learners. However, the prior research has focused on short videos and there have not been studies viewing full-length episodes of television.

### **2.1.2. The effects of background information on listening comprehension**

Similar to imagery, background knowledge is another contributor to top-down processing that may facilitate listening comprehension. When language learners are unable to fully process the vocabulary and grammar in a listening text, they rely more heavily on background knowledge for comprehension (Wolff, 1987). There are attributes of television viewing that may lead to gains in background knowledge (Herron, Cole, Corrie, & Dubriel, 1999) and potentially increase comprehension. Episodes of television are rarely viewed in isolation but rather it is more common to view multiple episodes of the same television program. These episodes generally have related storylines, recurring characters, and repeated settings. As viewers watch more episodes of the same program they build up more knowledge about the characters, their relationships to one another, and the proper names repeatedly occurring in the programs. Knowledge of proper names has been linked to increased comprehension (Kobeleva, 2012). Viewers increase knowledge about the different settings of the program and what type of events and scenes are likely to happen in these places. Viewers also learn the characters' relationships to those settings. While an episode of a television program may have a self-contained story, the individual episodes contribute to the overall story arc of a season of the program. Viewing episodes of the same program successively has the potential to build background knowledge allowing viewers to more easily use top-down processing for listening comprehension.

The role of background knowledge in L2 listening comprehension has been investigated (Chiang & Dunkel, 1992; Hasan, 2000; Long, 1990; Sadighi & Zare, 2006; Schmidt-Rinehart, 1994). Empirical research has shown that language learners who have background knowledge of a listening text have higher comprehension scores than those learners listening to

unfamiliar texts (Chiang & Dunkel, 1992; Long, 1990; Sadighi & Zare, 2006; Schmidt-Rinehart, 1994). Research has shown that language learners consider employing background knowledge an important strategy for successful comprehension (Hasan, 2000). There has, however, been little research on the effects of background knowledge on comprehension of video. Research has focused on the use of advance organizers before watching videos (Chang & Read, 2006; Chung, 1999; Herron et al., 1998) and the vocabulary in episodes of television (Rodgers & Webb, 2011; Webb, 2011).

The potential benefits of background knowledge for increasing comprehension of video are demonstrated in research on advance organizers (Chung, 1999; Herron et al., 1998). Advance organizers are generally information presented prior to the main learning material to provide a broad introduction to the activity (Chung, 1999). Herron, York, Cole, and Linden (1998) compared participants' comprehension of 10 short videos viewed with either advance organizers (information presented through declarative or interrogative sentences) or without advance organizers. Comprehension was measured through a series of short answer items for each video. The participants that used the advance organizers had significantly higher comprehension scores (Advance Organizer - Declarative:  $M = 42\%$ ; Advance Organizer - Interrogative:  $M = 41\%$ ) than the control group ( $M = 32\%$ ). Chung (1999) investigated the effects of advance organizers on comprehension of four short educational videos. Advance organizers consisted of six to eight sentences about each video presented in the participants' L1. Comprehension was measured through 10 multiple-choice items per video. Participants that had access to advance organizers before watching the video had significantly higher comprehension scores (69.8%) than participants who viewed the videos without advance organizers (66.9%). The findings indicate that the increased background knowledge through advance organizers can lead to increased comprehension.

Findings from corpus studies of television programs have also suggested that viewing multiple episodes of the same program and successive viewing of episodes may lead to increased background knowledge and increase comprehension. Webb (2011) investigated the reoccurrence of vocabulary in television in the same genre and Rodgers and Webb (2011) investigated the vocabulary that reoccurred in a season of television programs. In both studies, episodes of television related by genre or in the same season had fewer word families than episodes of random television. Fewer word families across the same amount of viewing time suggests viewers will encounter similar themes and content more regularly in related episodes. Regularly encountering thematically related content may allow viewers to build up

background knowledge more easily and apply that knowledge to following episodes. The results from these corpus studies indicate that viewing successive episodes of the same program may build background knowledge which can lead to increased comprehension.

### **2.1.3. The effects of vocabulary knowledge on listening comprehension**

Key to the utilization of bottom-up processing in aural texts is the recognition of individual words and the construction of meaning from them (Buck, 2001). This suggests that vocabulary knowledge may be an important factor in the comprehension of television. Surprisingly, there is little research on the relationship of vocabulary knowledge and comprehension of video. If this relationship is the same as it is for reading and listening, there may be implications for the use of television as comprehensible input in EFL situations.

In L2 research, there is a general consensus that vocabulary proficiency is a significant component of reading comprehension (Grabe, 1991). This has been supported by several studies that have indicated that there is a positive relationship between vocabulary knowledge and reading comprehension (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Laufer, 1989; Schmitt et al., 2011). In these studies, reported correlations between vocabulary knowledge and reading comprehension ranged between .407 ( $p < .001$ ) (Schmitt et al., 2011) and .8 ( $p$  value not reported) (Laufer & Ravenhorst-Kalovski, 2010). See Section 4.1 for more details on these studies.

There has been less research investigating the relationship between listening comprehension and vocabulary knowledge (van Zeeland & Schmitt, 2012). However, vocabulary knowledge is also thought to be a major factor in listening comprehension (Stæhr, 2009). Previous research has found a positive correlation between vocabulary knowledge and listening comprehension (Bonk, 2000; Milton et al., 2010; Stæhr, 2009) as well as a relationship between lexical coverage and comprehension (van Zeeland & Schmitt, 2012).

Three listening studies have found a positive relationship between vocabulary knowledge and comprehension. Bonk (2000) compared vocabulary knowledge and listening comprehension of four short passages with increasing lexical difficulty. The passages had equivalent word totals but included increasing amounts of low frequency vocabulary. Comprehension was measured by a written recall test in the L1 and a dictation test in the L2. He found a significant correlation ( $\tau = .45, p < .05$ ) between knowledge of the vocabulary in the passages (determined by scores on the dictation tests) and listening comprehension (determined by scores on the recall tests). Stæhr (2009) compared English language learners' vocabulary knowledge with their comprehension of a standardized listening test. Stæhr

measured vocabulary knowledge in two ways: four levels of the Vocabulary Levels Test (VLT) (2,000-, 3,000-, 5,000-, and 10,000-word levels) and a 50-item depth of vocabulary knowledge test. He found a significant correlation ( $r = .70, p < .01$ ) between the participants' combined scores on the VLT and their listening comprehension scores. Stæhr, also, found a significant correlation ( $r = .65, p < .01$ ) between the participants' combined scores on the depth of vocabulary knowledge test and their listening comprehension scores. Meara, Wade and Hopkins (2010) compared English language learners' orthographic and phonological vocabulary knowledge with their comprehension of a standardized listening test. Vocabulary knowledge was measured through the X\_Lex (orthographic) and AuralLex (phonological) tests that measure knowledge of the 5,000 most frequent words of English through 120 Yes/No items. Listening comprehension had a significant positive correlation ( $r = .52, p < .05$ ) with the orthographic test and a higher significant correlation with the phonological test ( $r = .67, p < .01$ ). The results from these three studies indicate a positive relationship between vocabulary knowledge and listening comprehension. There were, however, differences in the strengths of the correlations reported which may be due to the differing methodologies (comprehension tests, vocabulary knowledge measures, and listening texts) in the studies.

Research investigating the relationship between lexical coverage and listening comprehension also indicates a positive relationship between vocabulary knowledge and comprehension. Van Zeeland and Schmitt (2012) researched the effects of lexical coverage on second language listening comprehension. Short stories were modified with nonwords so that listeners had differing percentages of vocabulary known in each. Participants that knew 100% of the vocabulary in a story had significantly better comprehension scores than the participants that knew 98%, 95% or 90% of the vocabulary. Participants with knowledge of 98% of the vocabulary had significantly better comprehension than participants that knew 95% or 90% of the vocabulary. There was no significant difference between participants with 95% and 90% knowledge of the vocabulary though 95% group had higher mean scores. This study indicates that the more vocabulary language learners know in a text the more likely they are to have better listening comprehension

There does not appear to be any prior empirical research investigating the relationship between vocabulary knowledge and comprehension of videos. This is surprising considering the number of previous studies involving comprehension of videos. As research has indicated a relationship between vocabulary knowledge and both reading and listening comprehension, it is a reasonable hypothesis that a similar relationship exists for comprehension of video.

Information on the relationship between vocabulary knowledge and comprehension of video may prove useful for prescribing television as a source of comprehensible input for EFL learners.

#### **2.1.4. The effects of other factors on listening comprehension**

There are other factors that are believed to affect comprehension of listening texts beyond supporting imagery, background knowledge, and vocabulary knowledge. These factors are either dependent on the listening text or the listeners' individual differences. Factors that are text dependent include: the accent of the speakers, the pronunciation of the speakers, hesitations and pauses made by the speakers, the amount of reduced forms present in the speech, the prosodic nature of the speech, the speed at which the text is spoken, and the length of the listening text (Buck, 2001; Hasan, 2000; Rubin, 1994). The factors that reflect the individuality of the listener include: short term memory, ability to concentrate for extended periods of time, differences between the L1 of the listener and the language of the texts, the listener's aural experience with the target language, and the language proficiency of the listener (Goh, 2000; Rubin, 1994). While these factors may affect listening comprehension and comprehension of episodes of television it is beyond the scope of this research to control for them or examine how they affect comprehension. They do, however, need to be considered as factors that may affect language learners' comprehension of multiple episodes of television.

#### **2.1.5. Previous research involving comprehension of video**

Many studies have investigated languages learners' comprehension of L2 videos (Baltova, 1999; Chung, 1999; Etemadi, 2012; Guichon & McLornan, 2008; Guillory, 1998; Huang & Eskey, 1999; Latifi et al., 2011; Markham et al., 2001; Markham & Peter, 2003; Taylor, 2005; Winke et al., 2010). Comprehension of videos in these studies was investigated for video-only treatments as well as with other viewing treatments including: captions, keyword captions, advance organizers, captions with advance organizers, subtitles, and audio-only. A full description of each of these studies is in Section 6.4.1 in Study 5 including a summary of the viewing treatments, the comprehension test scores for the different treatment groups, and the number of participants in the treatment groups. Study 5 focuses on learning from television with and without captions and thus, this information is more relevant to that chapter. The focus of Study 1 is comprehension of video without supplementary treatments, and as such, only the relevant results from these studies are included here. Seven of these studies (Chung,

1999; Guillory, 1998; Huang & Eskey, 1999; Latifi et al., 2011; Markham & Peter, 2003; Taylor, 2005; Winke et al., 2010) used measures of comprehension that employed multiple-choice or short answer items and results are expressed as percentages. Three of the studies (Baltova, 1999; Guichon & McLornan, 2008; Markham et al., 2001) measured comprehension with tests where the results are in less comparable forms.

Four studies investigated language learners' comprehension of educational videos. Guillory (1998) examined comprehension of two textbook-related videos for learning French. Seventy English language speakers completed seven short answer items for each video. These items focused on the recall of details and inferencing from the information presented. The participants had a mean score of 7.28 out of 14 (52%). Studies by Chung (1999) and Huang and Eskey (1999) investigated language learners' comprehension of the same educational video series designed to teach English language and American culture. Participants ( $N=170$ ) in Chung's (1999) study viewed four segments (7 minutes each) of the videos under different treatment conditions including a video-only treatment. Comprehension was measured through a 10-item multiple-choice test in their L1 (Chinese). The mean score for participants viewing a segment under the video-only treatment was 6.69 out of 10 (66.9%). Huang and Eskey (1999) used a full 21-minute episode of the educational series. Fifteen participants viewed the episode twice and comprehension was measured by a 15-item multiple-choice test delivered aurally. The participants' mean score was 7.67 (51.1%). In a study of low-level Spanish-language learners viewing a 10-minute video that accompanied their textbook, Taylor (2005) measured comprehension through a free recall procedure in the participants' L1 (English) and a multiple-choice test. Immediately before viewing, participants were presented with a list of vocabulary from the video as an aid to comprehension. The mean score of the 41 participants on the multiple-choice test was 7.78 out of 15 (51.9%) and participants supplied a mean 2.41 facts about the video on the free recall test.

Two studies investigated language learners' comprehension of short documentaries. Markham and Peter (2003) measured comprehension of a 7-minute documentary with a 20-item multiple-choice listening test. Sixteen English language speakers viewed a Spanish documentary. These participants had a mean score of 7.81 out of 20 (39.1%). In another study with foreign learners of Spanish, Winke, Gass and Sydorenko (2010) tested comprehension of three documentaries that were 3 to 5 minutes long. After a brief presentation of background information, eight participants (native speakers of English and Kannada) viewed the videos

twice and completed multiple-choice items that were based on the main points of the video. Participants had a mean score of 36% on the test.

One study examined the comprehension of short sections of an animated movie. Latifi, Mobalegh and Mohammadi (2011) had 12 Persian-speaking participants view fifteen 2-minute movie clips twice each. For every clip there were 10 corresponding multiple-choice items that focused on comprehension of the main points discussed in the dialogue. Key vocabulary and idiomatic expressions were explained after the second viewing. The mean score across the 15 tests was 5.25 out of 10 (52.5%).

Three studies used comprehension tests that were less easily generalizable and comparable. Baltova (1999) conducted a study to investigate French language learners' comprehension of a 7.5-minute documentary viewed three times. Comprehension was measured by the number of idea units participants produced in response to eight open-ended questions administered immediately after viewing and again after a two-week delay. The researcher determined that there were a maximum of 22 idea units that the participants could provide. The mean score of the 29 English-speaking participants was 4.36 on the immediate comprehension test and 3.40 on the delayed test. In Markham, Peter and McCarthy's (2001) study, 68 English-speaking participants viewed a 7-minute documentary in Spanish and then completed a written summary of the video. In the summary, the participants provided an average of 8.47 idea units. In a study by Guichon and McLornan (2008), 10 French-speaking participants viewed a 3-minute English news report twice, took notes, and then had 20 minutes to write as thorough a summary as possible. The mean percentage of semantic units provided by the participants in their summaries was 25.1%.

When taken as whole, some patterns in the previous research involving comprehension of video emerge. The first of these is that the amount of video input used in these studies has been relatively short. The longest input video was the educational video used in Huang and Eskey's (1999-2000) study which was 21 minutes long but was comprised of three 7-minute chapters. The largest amount of viewing input was in Latifi, Mobalegh and Mohammadi's (2011) study with 30 minutes of viewing time. This amount of time, however, consisted of 15 short video clips from the same movie. The short clips in the previous research were also often commonly viewed twice in succession (Guichon & McLornan, 2008; Huang & Eskey, 1999; Latifi et al., 2011; Winke et al., 2010) before the comprehension tests were completed.

Another prevailing theme in the research is that the types of videos used in these are generally not representative of the types of videos that language learners might choose to watch on their own. Video types included: educational videos, documentaries, a news report,

and a movie. While a movie might be something learners might choose to watch on their own, the input videos in the study by Latifi, Mobalegh and Mohammadi (2011) were fifteen 2-minute excerpts from the movie. This is arguably not a common way to view a film.

Another common result in previous research on the comprehension of videos is relatively low mean comprehension scores. The results from the studies with somewhat comparable comprehension tests ranged from 36% to 66.9% with a mean comprehension score of 49.9%. The only study with a mean comprehension score that might be considered adequate was Chung's (1999) (66.9%) where the input video was, not surprisingly, from an educational series designed for language learners. It is, however, problematic to compare comprehension scores from the different studies as the researchers all had differing agendas for their research and differing approaches to creating comprehension items. From the studies with comprehension tests based on summaries or free recall procedures, it is possible to see the difficulties language learners may have understanding and remembering many of the important idea units after viewing a short video. Generally, the findings indicate that language learners can comprehend short videos to some degree but the amount of comprehension is dependent on the input video and the instrument used to test comprehension.

#### **2.1.6. Implications of previous research**

Previous research has indicated that the imagery present in television, accumulated background knowledge from viewing related episodes of television, and the vocabulary knowledge of the viewers are factors that may facilitate comprehension of television. However, the videos used in prior studies have generally employed short videos viewed in isolation and of a type that language learners might not typically choose to view for enjoyment. The present study attempts to build on prior research involving comprehension of video by investigating (a) language learners' comprehension gains from the first to the tenth episode viewed, (b) differences in comprehension scores across successive episodes viewed, and (c) the relationship between learners' vocabulary knowledge and comprehension of episodes of authentic television.

## 2.2. Research questions

Study 1 was designed to answer the following research questions:

1. Does comprehension of episodes of English language television change from the first episode to the tenth episode viewed?
2. Does comprehension of episodes of English language television change across successive episodes viewed?
3. Does comprehension of English language television improve with greater vocabulary knowledge?

## 2.3. Participants

There were 282 male and 133 female volunteer participants in their first and second year of university from 12 separate classes in this study. All of the participants had studied English for a minimum of seven years. The English proficiency level of the participants can be considered pre-intermediate to intermediate within the context of the university. The classes that provided the participants for Study 1 were all taught by the researcher. Details on the number of participants in each class and their university major are shown in Table 2.1.

Table 2.1 Participants in Study 1 prior to exclusions

| Class | Major       | Year of Study | Gender |     | Number of Participants |
|-------|-------------|---------------|--------|-----|------------------------|
|       |             |               | M      | F   |                        |
| 1     | Commerce    | 2             | 32     | 4   | 36                     |
| 2     | Business    | 1             | 27     | 9   | 36                     |
| 3     | Business    | 1             | 28     | 8   | 36                     |
| 4     | Engineering | 2             | 33     | 0   | 33                     |
| 5     | Law         | 2             | 31     | 7   | 38                     |
| 6     | Commerce    | 1             | 14     | 10  | 24                     |
| 7     | Law         | 2             | 26     | 11  | 37                     |
| 8     | Commerce    | 1             | 14     | 10  | 24                     |
| 9     | Engineering | 1             | 24     | 1   | 25                     |
| 10    | Language    | 1             | 30     | 12  | 42                     |
| 11    | Language    | 2             | 15     | 35  | 50                     |
| 12    | Pharmacy    | 1             | 8      | 26  | 34                     |
| Total |             |               | 282    | 133 | 415                    |

### 2.3.1. Exclusion of participants

Ninety-four participants were excluded from Study 1. Participants were excluded if they were absent from: Teaching Session 1 (VLT and Ethics Approval), Teaching Session 2 (Tough and Sensitive Vocabulary Pre-Tests), Teaching Session 3 (Viewing Episode A1 or

Episode B1), Teaching Session 12 (Viewing Episode B2 or A2), and Teaching Session 13 (Tough and Sensitive Vocabulary Post-Tests and Final Attitude Survey). Participants were also excluded from the study if they were absent from viewing more than one episode from Episode 1 to 8 (Teaching Sessions 4 to 11). It was believed that missing a single episode would not be a serious detriment to comprehension but missing two or more would have a negative effect. When participants were absent from a single teaching session for Episode 1 through Episode 8 they did not complete the comprehension test. Missing comprehension test scores were replaced using the Expectation Maximization Algorithm (Allison, 2001) which is explained in Section 2.8.2. The results from 321 participants were left for analysis after these exclusions. Table 2.2 shows the starting sizes of the 12 classes in Study 1 and the number of participants excluded from each.

Table 2.2 Participants excluded from the comprehension analyses for Study 1

| Class | Starting Size | Number of Exclusions | Final Number of Participants |
|-------|---------------|----------------------|------------------------------|
| 1     | 36            | 9                    | 27                           |
| 2     | 36            | 4                    | 32                           |
| 3     | 36            | 8                    | 28                           |
| 4     | 33            | 5                    | 28                           |
| 5     | 38            | 12                   | 26                           |
| 6     | 24            | 2                    | 22                           |
| 7     | 37            | 16                   | 21                           |
| 8     | 24            | 5                    | 19                           |
| 9     | 25            | 2                    | 23                           |
| 10    | 42            | 13                   | 29                           |
| 11    | 50            | 14                   | 36                           |
| 12    | 34            | 4                    | 30                           |
| Total | 415           | 94                   | 321                          |

### 2.3.2. Human ethics requirements

In accordance with human ethics requirements, all participants in Study 1 (and all other studies in this thesis) received a detailed explanation of the research, were given information sheets, and signed a written consent form if they agreed to participate in the research.

## 2.4. Procedure

The experimental procedure for Study 1 was repeated with 12 different university classes. Nine of the classes took place in the first semester of the Japanese university school year

which ran from April 2010 through July 2010. Three of the classes took place in the second semester which ran from September 2011 through January 2012.

#### 2.4.1. Overall schedule

Study 1 took place over 13 teaching sessions. Generally, each teaching session was separated by a week, but because of national and school holidays, there were instances where the teaching sessions were separated by two weeks. The schedule for this study is shown in Figure 2.1.

Figure 2.1 Schedule for Study 1

| Teaching Session | Study 1 Schedule  |                         |
|------------------|---|-------------------------|
|                  | Viewing Group 1   | Viewing Group 2         |
| 1                | Human Ethics Committee Ethics Form, & Vocabulary Levels Tests (2,000, 3,000, & 5,000)   |                         |
| 2                | Tough Vocabulary Pre-Test, Television Viewing Practice, & Sensitive Vocabulary Pre-Test |                         |
| 3                | <i>Chuck</i> Episode A1   | <i>Chuck</i> Episode B1 |
| 4                | <i>Chuck</i> Episode 1  |                         |
| 5                | <i>Chuck</i> Episode 2  |                         |
| 6                | <i>Chuck</i> Episode 3  |                         |
| 7                | <i>Chuck</i> Episode 4  |                         |
| 8                | <i>Chuck</i> Episode 5  |                         |
| 9                | <i>Chuck</i> Episode 6  |                         |
| 10               | <i>Chuck</i> Episode 7  |                         |
| 11               | <i>Chuck</i> Episode 8  |                         |
| 12               | <i>Chuck</i> Episode B2   | <i>Chuck</i> Episode A2 |
| 13               | Tough Vocabulary Post-Test, Final Attitude Survey, & Sensitive Vocabulary Post-Test     |                         |

#### 2.4.2. Viewing order

In Study 1, there were two viewing groups. Participants viewed either Episode A1 or Episode B1 first, followed by Episodes 1 to 8, then finally Episode B2 or Episode A2<sup>1</sup>. This was because prior to the analyses performed in this study, it was unknown if Episode A or Episode B was more difficult than the other. This might have been a factor in analyzing comprehension gains from the first to the last episode viewed. Accordingly, this counterbalanced design was implemented to control for the possibly different levels of difficulty and to allow for an analysis of comprehension gain.

<sup>1</sup> Throughout the thesis Episode A refers to Episode 12 of *Chuck* and Episode B refers to Episode 13. Episode A1 and Episode A2 are the same episode. The '1' and the '2' indicate the relative position of the episode in the viewing order of the study. The same is true of Episode B1 and Episode B2.

## **2.5. Setting**

The study took place in two separate classrooms both of which had seating for sixty-four students sitting two persons to a table. Each room had a multimedia console containing a DVD player, audio equipment, a projector, and a screen. The projector was a Panasonic PT-D400U with a resolution of the 1024 pixels by 768 pixels which was shown on the 2 meter by 1.5 meter screen at the front of the classroom. The farthest a participant sat from the screen was approximately 10 meters. In a pilot study, there were no negative comments from the participants when surveyed about the sound and picture quality. At the beginning of the study, participants were assigned to seats based on their university student numbers but any participants that were hard of hearing or had weak eyesight were encouraged to sit near to the speakers and screen.

## **2.6. Pilot study**

A pilot study was conducted with Japanese university students for whom English was a second language. The pilot had 32 participants with an English proficiency level and language learning background similar to the participants in this thesis. The purpose of this study was to determine the time requirements of the study, the suitability of the materials, and to provide data for an item analysis of the comprehension tests.

## **2.7. Materials**

### **2.7.1. The television program<sup>2</sup>**

The television program that served as the input text for this study was called *Chuck* (Schwartz, Fedak, & McG, 2007). It is a series produced by College Hill Pictures and Warner Brothers Television and was first broadcast in the U.S.A in 2007. The first season of the program contained 13 episodes which originally aired between September 24, 2007 and January 24, 2008. Prior to the broadcast of the second season of the series in September 2008, the first season of *Chuck* was released on DVD which allowed for it to be used in this study. Without access to the episodes on DVD, it would not have been possible to utilize the series in a classroom setting. The use of the episodes for educational and research purposes is covered by Victoria University of Wellington's Screenrights license and Section 48 of the Copyright Act 1994 of New Zealand.

---

<sup>2</sup> The term television program and television series are used synonymously throughout this thesis to describe *Chuck*. Because the program is of American origin, American English classification is used in this thesis. In this, a television program refers to something that people watch on television, and is usually part of a regularly occurring sequence of episodes. A run of episodes usually lasting less than a year is referred to as a season.

The series, *Chuck*, follows the life of a computer repairman, Chuck Bartowski, whose life is disrupted when government secrets are accidentally downloaded into his brain. The series follows how he adapts to life as a secret agent, cooperating with other agents assigned to protect him while preventing threats to national security. The overall story arc of the first season is concerned with Chuck learning why he has received the secrets, learning how to access the information in his head, getting accustomed to life as a spy, and coping with his romantic feelings towards one of the agents protecting him. The genre of the series is drama but it also has elements of action and comedy.

This series was chosen for five reasons. First, *Chuck* is an American drama, which was found to be a less lexically demanding genre in previous research and *Chuck* has a lexical load comparable with other American dramas (Rodgers & Webb, 2011; Webb & Rodgers, 2009a). The second reason was that the first season of *Chuck* is serial in nature. This may allow viewers to acquire background knowledge more easily than in a television series where episodes are only loosely connected. Third, this was a new series and the characters and the main story were perhaps more explicitly introduced than they might be in subsequent seasons. Fourth, *Chuck* was received favorably in the pilot study with 94% of participants rating it as very enjoyable. The fifth reason *Chuck* was chosen was that the series wasn't broadcast or available in Japan at the time of the study. This lessened the probability of participants having seen any of the episodes. If participants in the study had seen episodes, they may have acquired some knowledge of the series which may have improved their performance on the comprehension and vocabulary tests.

While the first season of *Chuck* has 13 episodes, only 10 were used in this study. This was due to the time constraints of a university semester in Japan. With vocabulary pre- and post-tests, the Introduction to Television Viewing, the Television Viewing Practice, and the Final Attitude Survey there was only enough time for viewing and testing 10 episodes. Episode 1 through Episode 8 plus Episode 12 and Episode 13 were used. The first eight episodes were used because they were successive episodes and each one was part of a general story arc. Episode 12 and Episode 13 were selected for testing purposes because they were more self-contained episodes that had less to do with the first season's story arc. Throughout this thesis, Episode 12 is referred to as Episode A and Episode 13 is referred to as Episode B to differentiate them from the eight successive episodes.

The 10 episodes used in this study have an average running time of 42 minutes and 49 seconds and range in length from 41 minutes and 15 seconds to 43 minutes and 18 seconds. With commercials, the episodes would have been originally broadcast over one hour. The

running times for each episode are shown in Table 2.3. The running time of each episode includes approximately 30 seconds of title credits and 38 seconds of closing credits.

Table 2.3 Running times of the 10 episodes

|        | Episode |        |        |        |        |        |        |        |        |        |
|--------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|        | 1       | 2      | 3      | 4      | 5      | 6      | 7      | 8      | A (12) | B (13) |
| Length | 43'18"  | 42'47" | 43'13" | 43'13" | 43'15" | 41'15" | 42'11" | 43'16" | 43'11" | 42'32" |

### 2.7.2. The vocabulary in the episodes

Analyzing the spoken vocabulary in the episodes of *Chuck* was necessary for three reasons. The first reason was to examine the suitability of the program for research purposes as it was desirable, for purposes of comparison, to find a program and series of episodes that had a vocabulary level that is typical of television as reported in Webb and Rodgers (2009a). The second reason was to provide potential target words for vocabulary tests. The procedure for using the vocabulary analysis to create vocabulary tests is presented in detail in Study 2. The third reason for analyzing the vocabulary was to provide data for the lexical coverage estimates used to examine vocabulary gains and comprehension results in relation to vocabulary knowledge. This process is presented in detail in Study 3.

To analyze the vocabulary, scripts for the 10 episodes of *Chuck* were downloaded from the Internet. These scripts are produced by fans of the television series for educational or entertainment purposes. They are available at websites such as [www.tvsubtitles.net](http://www.tvsubtitles.net) in a form intended for captioning. These caption scripts were cleaned up prior to analysis by removing time markers and media player coding. All that remained of the original scripts were the spoken words from the episodes. All the scripts were spell checked via a word processor and examined while viewing the episodes to check for any transcription errors.

Hyphenated words, reduced forms, and contractions were changed to the spellings used in Nation's (2004) British National Corpus (BNC) word lists. Reduced forms are often transcribed inconsistently and purely at the discretion of transcribers. There were 152 occurrences of reduced forms identified in the scripts of the 10 episodes. They accounted for 0.29% of the total number of tokens in the episodes used in this study. Had these not been changed to the conventional spellings, these words would have been classified as being less frequent than the most frequent 14,000 word families. A summary of the modified reduced forms can be seen in Table 2.4.

Table 2.4 Frequency and range of modified reduced forms in the 10 episodes

| Reduced Form | Corrected Form | Episode |    |    |   |    |    |    |    |        |        |       |
|--------------|----------------|---------|----|----|---|----|----|----|----|--------|--------|-------|
|              |                | 1       | 2  | 3  | 4 | 5  | 6  | 7  | 8  | A (12) | B (13) | Total |
| gonna        | going to       | 3       | 16 | 2  |   | 10 | 5  | 8  | 6  | 4      | 8      | 62    |
| wanna        | want to        |         |    |    |   |    | 1  |    | 1  |        |        | 2     |
| ya           | you            |         |    |    |   | 1  |    | 2  |    |        |        | 3     |
| gotta        | got to         |         | 1  | 3  | 1 |    | 3  |    |    |        | 1      | 9     |
| kinda        | kind of        |         | 2  |    |   | 1  | 4  |    |    |        |        | 7     |
| gotcha       | got you        |         |    |    |   | 1  |    | 1  |    |        | 1      | 3     |
| ain't        | isn't          |         | 1  |    |   |    |    |    |    |        |        | 1     |
| 'em          | them           | 2       | 2  | 1  | 3 |    |    |    | 3  | 2      | 3      | 16    |
| ol'          | old            |         |    | 1  |   |    |    |    |    |        |        | 1     |
| sucka        | sucker         |         |    |    |   |    |    |    |    | 1      |        | 1     |
| 'cause       | because        | 8       | 4  | 3  | 3 | 7  | 6  | 4  | 2  | 2      | 1      | 40    |
| -in'         | -ing           |         |    | 1  | 1 | 1  | 1  |    |    | 3      |        | 7     |
| Total        |                | 13      | 26 | 11 | 8 | 21 | 20 | 15 | 12 | 12     | 14     | 152   |

After the transcripts had been prepared for analysis, they were processed with the RANGE program (Nation & Heatley, 2002). This software sorts the vocabulary in a given text in relation to predetermined word lists. The lists used in this analysis were fourteen 1,000-word frequency lists derived by their range and frequency of occurrence in the BNC (Nation, 2006). Each list contains 1,000 word families. Each word family is rated as Level 6 according to Bauer and Nation's (1993) word family classification. Level 6 word families include inflections and more than 80 derivational affixes. Variations in the form of a word are typically through derivation, for example *pleasant* becomes *unpleasant* or *pain* becomes *painful*. All the word stems are free forms which can stand alone as opposed to bound forms which cannot occur as separate words on their own. For example, the word *reprint* consists of the affix *re-* and the free form *print* while the word *repeat* contains the bound form *peat* which cannot stand on its own.

Words in the episodes that were less frequent than the most frequent 14,000 word families are classified by the RANGE program as either *Proper Nouns*, *Marginal Words* (interjections, exclamations and hesitation procedures, for example *oh*, *uh*, *mmm*, and *ah*), or *Not in the Lists*. The proper nouns list has over 13,000 entries but did not account for all the proper nouns in the episodes so many proper nouns were classified by RANGE as *Not in the Lists*. A further 117 proper nouns were identified by the researcher in the *Not in the Lists* category, reclassified and added to the proper nouns totals. The results of this analysis with RANGE for all 10 episodes used in the study are shown in Table 2.5. Shown in the table are the number of

Table 2.5 Tokens, types, word families, and cumulative coverage of the combined episodes of *Chuck*

| Word List        | Tokens |       | Types |       | Word Families | Cumulative Coverage               |                                 |  |
|------------------|--------|-------|-------|-------|---------------|-----------------------------------|---------------------------------|--|
|                  | Raw    | %     | Raw   | %     |               | Coverage including marginal words | Coverage including proper nouns | Coverage including marginal words & proper nouns |
| 1,000            | 44367  | 85.27 | 1709  | 38.95 | 877           | 86.70                             | 88.16                           | 89.59  |
| 2,000            | 2448   | 4.70  | 851   | 19.39 | 580           | 91.40                             | 92.86                           | 94.29  |
| 3,000            | 871    | 1.67  | 423   | 9.64  | 335           | 93.07                             | 94.53                           | 95.96  |
| 4,000            | 486    | 0.93  | 232   | 5.29  | 199           | 94.00                             | 95.46                           | 96.89  |
| 5,000            | 334    | 0.64  | 171   | 3.90  | 152           | 94.64                             | 96.10                           | 97.53  |
| 6,000            | 209    | 0.40  | 114   | 2.60  | 103           | 95.04                             | 96.50                           | 97.93  |
| 7,000            | 229    | 0.44  | 80    | 1.82  | 76            | 95.48                             | 96.94                           | 98.37  |
| 8,000            | 114    | 0.22  | 71    | 1.62  | 64            | 95.70                             | 97.16                           | 98.59  |
| 9,000            | 114    | 0.22  | 55    | 1.25  | 50            | 95.92                             | 97.38                           | 98.81  |
| 10,000           | 98     | 0.19  | 36    | 0.82  | 33            | 96.11                             | 97.57                           | 99.00  |
| 11,000           | 58     | 0.11  | 44    | 1.00  | 42            | 96.22                             | 97.68                           | 99.11  |
| 12,000           | 27     | 0.05  | 22    | 0.50  | 22            | 96.27                             | 97.73                           | 99.16  |
| 13,000           | 34     | 0.07  | 25    | 0.57  | 24            | 96.34                             | 97.80                           | 99.23  |
| 14,000           | 15     | 0.03  | 13    | 0.30  | 13            | 96.37                             | 97.83                           | 99.26  |
| Proper Nouns     | 1501   | 2.89  | 258   | 5.88  | 255           | 99.26                             |                                 |  |
| Marginal Words   | 745    | 1.43  | 19    | 0.43  | 3             |                                   | 99.26                           |  |
| Not in the Lists | 380    | 0.73  | 265   | 6.04  | ?????         | 100.00                            | 100.00                          | 100.00   |
| Total            | 52030  |       | 4388  |       | 2828          |                                   |                                 |  |

tokens, types and word families for each of the fourteen 1,000-word family lists as well as the number of proper nouns, marginal words, and the low frequency words in *Not in the Lists*. Also included are the cumulative lexical coverage figures for the tokens in the television programs with and without proper nouns and marginal words. The results of this analysis, done separately for each of the 10 episodes, can be seen in Appendix E1.

The results from this analysis showed that the episodes, from a vocabulary perspective, were a suitable choice to serve as the input in this study. The percentage of tokens in the combined episodes of *Chuck* at the 1,000 to 5,000 word family levels were all within 0.4% of the results presented in Webb and Rodgers' (2009a) corpus analysis of television programs. These word family levels are the most relevant to this study. There were also no large differences in the relative amounts of proper nouns, marginal words, or words less frequent than the 14,000-word family list between the episodes of *Chuck* and the episodes analyzed in Webb and Rodgers.

### **2.7.3. Introduction to viewing television**

In Teaching Session 1, before the participants viewed any episodes of *Chuck*, they were given an information sheet explaining the rationale for viewing television in their English class. It was thought that this would be a unique experience for the participants and unlike the English classes they were familiar with. For these reasons, the methodology and rationale for the study were briefly explained. The information sheet was presented in both Japanese and English to ensure the participants understood it completely. The bulk of the information centered on the reasons viewing television would be a good experience for the participants. These reasons included the potential of participants to: learn the spoken form of the language through watching television and listening to the dialogue, improve their second language listening skills, learn vocabulary, increase their comprehension as they view more episodes, learn about foreign culture, and enjoy studying English. The participants read through the information sheet as a group and were encouraged to ask questions about anything they did not understand or wanted clarification on. The Introduction to Television Viewing information sheet can be seen in Appendix B2.

### **2.7.4. Television viewing practice**

In Teaching Session 2, prior to viewing episodes of *Chuck*, the participants completed the Television Viewing Practice. This was designed to expose the participants to the process of watching television and answering comprehension questions based on it. The video used was

a portion of an episode of the television series, *Mr. Bean* (Bennett-Jones, 1990) which was produced by Tiger Aspect Productions for broadcast in the United Kingdom. The video used was Act 1 of an episode called *Tee Off, Mr. Bean* (Bennett-Jones, 1995) involving the main character's adventures doing his washing at a Laundromat. The episode was first broadcast on September 20<sup>th</sup>, 1995 in the U.K. Act 1 of the episode was 13 minutes and 50 seconds in length and was a complete story. The series, in general, does not rely on spoken language but rather the actions of the characters. This video segment was chosen especially for the lack of spoken language. Any spoken language in the video was not of importance to the story but was more along the lines of background noise. Video with spoken content was avoided to prevent unintended exposure to the target vocabulary tested in Study 2. This episode of *Mr. Bean* was also chosen because its running time approximated that of two viewing sections from episodes of *Chuck* which would allow participants to get used to this viewing length. *Mr. Bean* is also a comedy and it was hoped that it would be fun and entertaining for the participants and at the same time introduce them to television viewing and the comprehension question answering process.

Before viewing the television program, the participants were introduced to the comprehension question formats and the correct way to fill out the answer sheet. The participants were given three minutes to answer the comprehension items for each viewing section and 30 seconds to preview the next set of questions. This was a similar amount of time given on the comprehension tests for *Chuck*. The first viewing section ran from 0:00 to 7:00 and the second from 7:01 to 13:50. For each viewing section there were five true/false and seven multiple-choice items. There was also an example item for each of the item types in both viewing sections. Unlike the questions on the comprehension tests created for *Chuck*, which focused on the spoken dialogue, the questions in the Television Viewing Practice concentrated on what could be seen and the actions of the characters. To familiarize the participants with the types of comprehension questions, a variety of inference, detail, and topic items were included.

Following viewing and completing the comprehension questions for both sections of *Mr. Bean*, there was a set of sequencing questions. This set of questions consisted of 12 story events from the video in a random order. The first and seventh events were provided for the participants and they were asked to provide the order of the other events in the video. Following this, the participants completed four attitude survey items. These were identical to the four Episode Attitude Survey items that followed each episode of *Chuck*. The test for *Mr. Bean* was marked by the students from an answer key provided by the researcher. The

introduction to comprehension tests, *Mr. Bean* comprehension test (in English and Japanese), answer sheet, and answer key for *Mr. Bean* are shown in Appendices B1 to B5.

### **2.7.5. Randomization**

Whenever randomization was called for during the course of creating the vocabulary and comprehension tests, the website [www.random.org](http://www.random.org) was used. This site provides true random numbers based on atmospheric noise. On the vocabulary tests, the *Integer Generator* function was used to choose the key's position for multiple-choice items and which words would serve as distractors. The *Coin Flip* function was used to choose whether an idea unit would be presented as true or false on the comprehension tests. The *Sequence Generator* was used to randomize the order of the items on the vocabulary post-tests and the comprehension tests' sequencing items. While this study aimed to use randomization wherever possible, often a quasi-random design was employed. This happened when the results of randomization clashed with the test design procedures. More details concerning randomization are included in the descriptions of the individual tests utilized in this thesis. Following randomization procedures ensured that any biases of the researcher that may manifest during the item creation process were minimized thus improving test reliability.

### **2.7.6. Episode-specific comprehension tests**

#### **2.7.6.1. Designing the comprehension tests**

To test the participants' comprehension of the episodes of *Chuck*, a comprehension test was created for each of the 10 episodes. The design process took into account:

1. How the participants would view the episodes and answer comprehension questions
2. What aspects of listening comprehension would be measured
3. What items would be used to measure these aspects of listening comprehension
4. How many items should be included on the tests

The first step in designing the comprehension tests was to decide how the participants would view each episode and complete the comprehension questions. Each episode is approximately 42 minutes in length and if participants were to answer the comprehension questions following the episode it would be a challenge for the participants to remember details from the beginning of the episode. Even with shorter listening texts, the location of the information is known to affect item difficulty (Freedle & Kostin, 1999). There was the potential for problems if participants were given all the comprehension questions prior to viewing and answered the questions as they viewed the episode. They could possibly be

distracted from viewing the episode as they tried to read and answer questions. In anticipation of these problems, each episode was divided into six viewing sections of approximately seven minutes each. Comprehension questions would be based only on a single viewing section. Participants would be given time to preview the questions before the viewing section and time to answer the questions following the viewing section. The opportunity to preview the comprehension questions was included because it has been shown to reduce anxiety towards listening tasks, and may lead to more correct answers (Chang & Read, 2006; Sherman, 1997). Viewing sections varied slightly in length because it was desirable to end sections when there was a natural break in the story. Therefore, the viewing sections were often longer or shorter than seven minutes depending where a suitable scene change was available.

The second step in designing the comprehension tests was to decide on which aspects of listening comprehension the comprehension questions would measure. To do this it was first necessary to think of listening comprehension not as a single concept, but a concept made up of different abilities (Brindley, 1998; Song, 2008). Defining the abilities that make up listening comprehension allows test designers to consider what aspects of comprehension should be included in a comprehension test (Song, 2008). Buck (2001) proposes a competency-based definition in which he identifies three required abilities in his default listening construct:

- to process extended samples of realistic spoken language, automatically and in real time,
- to understand the linguistic information that is unequivocally included in the text, and
- to make whatever inferences are unambiguously implicated by the content of the passage. (p. 114)

Buck also states that his default listening construct can be improved upon by including characteristics specific to the listening situation or specific to the listening text.

The listening comprehension tests created for the episodes of *Chuck* were designed around the tenets of Buck's construct starting with processing lengthy texts of realistic spoken language. The episodes of *Chuck* are a realistic source of spoken input that language learners might reasonably choose to learn English from. To help the participants process the language automatically and in real time yet still complete the comprehension questions, the participants were given the opportunity to preview the questions and answer them as they watched the viewing sections. Item types that are possible to answer while viewing were chosen. Participants choosing not to answer the questions as they viewed the program in real time were given time following each viewing section to complete the comprehension questions. To

measure participants' ability to understand the information included in the episodes, items designed to test the comprehension of details were included. To measure participants' ability to inference information from the content of the episodes, items designed to measure inference ability were included. The listening construct outlined by Buck was also augmented by including a characteristic important to this listening situation: understanding the topics contained in relatively lengthy viewing sections and in the episodes as a whole. Items that call for identifying the topic or main idea of a text is an aspect of listening comprehension that has been commonly featured in taxonomies of listening skills (Dunkel, Henning, & Chaudron, 1993; Field, 1998; Lund, 1991; Richards, 1983).

Measuring listening comprehension through items that test learners' ability to distinguish details, make inferences, and recognize the topic of extended portions of texts is supported by previous listening comprehension research. The inclusion of topic questions with detail and inference questions is supported by research by Song (2008) where she found that these three abilities provide a reasonably good explanation of listening comprehension. Listening for implicitly and explicitly stated information also has support from research concerning video comprehension tests with language students. Wagner (2002) found top-down and bottom-up processing skills to be correlated to listening comprehension which he believed supported Buck's (2001) default listening construct. Research on listening comprehension and item types has indicated that a mix of global and local questions should be included (Shohamy & Inbar, 1991). The use of items that measure comprehension of topic, inferencing abilities, and comprehension of details also reflects the top-down and bottom-up processing view of listening comprehension. The detail-focused items reflect bottom-up processes that involve a listener decoding specific words, clauses and sentences. The topic- and inference-focused questions reflect the top-down processes that involve a listener comprehending the themes in a text, the sequence of events, and outcomes of events (Richards, 1990).

The third step in designing the comprehension tests was to decide on the type of items to be used on the tests. There were some considerations based on the nature of the television program used in this study. First of all, when testing listening comprehension, 42 minutes is a relatively long text with a lot of content to test. Therefore, item types that could be quickly answered during and following the viewing sections were desirable. The types of items also had to function suitably as detail, inference, and topic questions. With these stipulations in mind, instead of using one single item type throughout, it was decided to use a number of item types. Listening comprehension tests that present a variety of item types better reflect the

trait of listening comprehension (Shohamy & Inbar, 1991) and are considered fairer to test takers (Spaan, 2007).

To test comprehension of each viewing section a combination of true/false and multiple-choice items was decided upon. It was believed that these item types were familiar to the participants and would need little explanation prior to the study. Both types of questions allow for gathering large amounts of information in a short period of time. True/false questions are a viable and effective format for testing knowledge quickly and produce reliable measures. They take little time for test-takers to complete so more content can be covered in a short amount of time. True/false items are also a format that is familiar to the participants so the amount of time participants need to become accustomed to the test format is reduced (Frisbie & Becker, 1991; Haladyna, 1992; Kreiter & Frisbie, 1989). Multiple-choice items provide an efficient method of covering a broad range of language knowledge with reliable scoring (Spaan, 2007). Multiple-choice items are also recommended for L2 listening tests and considered easier and more efficient than open-ended formats (In'nami & Koizumi, 2009). Multiple-choice items are also recommended for testing the ability to make inferences as they force test takers to choose the most plausible inference. Open-ended items would allow for more subjective interpretations which are difficult to score as incorrect (Brindley, 1998). Based on the positive attributes of multiple-choice and true/false items, they were chosen as the primary item types for the comprehension tests.

The number of options to use for the multiple-choice items was the next consideration. The multiple-choice format can sometimes be disadvantageous to less able listeners and result in uninformed guessing with test takers getting an item correct or incorrect for the wrong reasons (Yi'an, 1998). This, however, may be more a function of the number of options used in multiple-choice questions. Four options may lead to more confusion than three options. In this study, the multiple-choice items all had three options for three reasons. The first is that more 3-option items can be administered in the same amount of time compared with the more common 4- or 5-option items. This improves content coverage and doesn't have detrimental effects on the psychometric quality of the test (Rodriguez, 2005). The second reason for 3-option questions is that more options do little to improve an item and often result in the inclusion of implausible distractors (Haladyna & Downing, 1993; Rodriguez, 2005). The key to a good multiple-choice item is not the number of distractors but the quality of the distractors (Haladyna & Downing, 1993). The third reason to use three options is that more options expose more aspects of the text to the test takers. This can provide context clues to

other questions (Rodriguez, 2005). For these reasons, the comprehension tests for episodes of *Chuck* were created with 3-option multiple-choice items.

There was a third type of question used on the comprehension tests: sequencing items. The purpose of this item type was to test the participants' ability to process the input video as a whole. Sequencing items measure whether participants recognize the overall order of ideas in a text. They measure global comprehension where the learner attempts to understand the text as a whole (Richards, 1983). This has been described as an essential aspect of comprehension in reading (Alderson, Percsich, & Szabo, 2000) where if a reader is to comprehend a narrative text, they must be able to appreciate the order of events (Ohtsuka & Brewer, 1992). The ability to comprehend the sequence of events has also been described as an important facet of listening comprehension (Brett, 1995; J. I. Brown, 1949). Therefore, through the use of sequencing items, the ability to recognize the order of events was included in the comprehension tests for the episodes of *Chuck*.

The final step in designing the comprehension tests involved deciding on the number of items to include for each viewing section and for the sequencing task. The 90 minutes available for each teaching session and the amount of content in an episode of *Chuck* were taken into consideration when making this decision. It was desirable to have as many items as possible because of the potential for losing items during the test validation procedure that followed the pilot study. However, each viewing section only had enough content for a limited number of items. After some experimentation, the researcher determined that the roughly seven minutes of viewing time provided enough information to create 12 comprehension items. Creating more than 12 items was often difficult and in cases when more than 12 items were created there were often questionable items. Informal piloting showed that language learners with similar backgrounds to the participants in this study could answer this number of items comfortably in 2.5 to 3.5 minutes and could preview another set of 12 items in approximately 30 seconds. Consequently, test takers were given 4 minutes for each viewing section's set of comprehension items following viewing. Eleven events from an episode were used as the basis of the sequencing items. This number covered the main events of the typical episode without including events that might be considered unimportant. Piloting indicated that completing these sequencing items and the attitude survey took less than 10 minutes for the majority of participants. In total, viewing the episode and completing the comprehension questions and attitude survey took approximately 80 minutes. This left 10 minutes for administrative procedures that were a necessary component of each teaching session.

### 2.7.6.2. Creating true/false and multiple-choice items

When creating the true/false and multiple-choice items for a comprehension test, the first task was to parse each viewing section of an episode into idea units. For this process, idea units were defined as distinct events, actions, or dialogue spoken in the course of the program. While the number of idea units created for each viewing section differed, a typical result of this process is illustrated by the first viewing section from Episode 1 of *Chuck* where 44 idea units were produced for eight minutes and two seconds of video. As the comprehension questions were all based on idea units that were verbalized<sup>3</sup>, visual events that could not be linked to verbal content were removed. The remaining idea units were then sorted, by suitability, for use as true/false items and for multiple-choice items.

An example of an idea unit that was more suitable for one item type over another can be seen in the first viewing section of Episode 1. The idea unit, ‘When asked about having a girlfriend, Chuck says he had one in college named Jill.’, was better suited to being a true/false, detail-focused item (Item #3 on the comprehension test for Episode 1) while the idea unit, ‘Following the party, Morgan talks to Ellie about the party but she ignores him.’ was better suited to becoming a multiple-choice, inference-focused item (Item #10). The first idea unit can be written into a single succinct sentence with numerous words that can be substituted (*girlfriend*, *college*) to change the sentence to false if necessary. The item based on this idea unit is shown in Figure 2.2. The second idea unit is the result of considerable interaction between the characters in the episode and is important within the context of the program. This provided an opportunity to create an item measuring inference ability which is more easily accomplished through a multiple-choice item. The item based on this idea unit is shown in Figure 2.3.

Figure 2.2 First three true/false items on the comprehension test for Episode 1

1. T or F Most of the people at the party are doctors.
2. T or F Chuck’s major in university was accounting.
3. T or F Chuck had a girlfriend in university named Jill.

---

<sup>3</sup> Only verbalized idea units were used because of the research goals of another study in this thesis. Study 3 examines the relationship between lexical coverage (the percentage of words known in a text) and comprehension. If the comprehension items tested aspects of viewing other than what was said (e.g. visual elements) it would not be possible to properly explore the relationship between coverage and the participants’ comprehension of the episodes.

Figure 2.3 Item #10 on the comprehension test for Episode 1

10. How does Ellie feel about Morgan?

- A. She doesn't know who he is.
- B. She thinks he is annoying.
- C. She likes him as much as her brother.

The sorted idea units were then explored as potential comprehension questions. For the idea units to become true/false items, this meant rewriting them as succinct sentences that focused on the main point of the idea unit. For the potential multiple-choice items, this meant attempting to write a question to serve as the stem, rewriting the idea unit as the key, and determining if there were plausible distractors. Idea units that did not produce good or plausible true/false or multiple-choice items were deleted leaving the 12 idea units that the researcher believed made the best comprehension test items.

When choosing the 12 items, the listening comprehension ability (detail, inference and topic) being tested was considered but not forced. Each viewing section had differing amounts of dialogue, scene changes, characters, and action. Viewing sections also had different themes and different amounts of relevance to the episode's story and the program's overarching story. This meant that each of the viewing sections lent themselves differently to the creation of different items measuring the three listening comprehension abilities. For each viewing section between 0 and 2 topic, 2 and 7 inference, and 4 and 9 detail questions were created.

Once the five idea units that were to serve as the true/false items were decided upon, the next step for creating the items was to determine whether they would be scored as true or false. This was done quasi-randomly using the *Coin Flip* function. An idea would be left in its true (heads) state or modified to become false (tails). Modifying idea units to be false was done by either changing a key component of the sentence or negating the sentence. While the example idea unit, 'Chuck had a girlfriend in university named Jill.' became an item scored as true, it could have been made into a false item by changing *Jill* to *Janet* or *university* to *high school*. It could also have been made false by negating the sentence such as 'Chuck didn't have a girlfriend in university.' The decision on how to produce items scored as false was made by the researcher by considering what produced the most natural item within the context of the viewing section and the episode. Examples of true/false items created for Episode 1 are shown in Figure 2.2.

The procedure for creating multiple-choice items from idea units was more involved than the procedure for creating true/false items. First, the question that was to form the stem of the

item was written to be as specific to the situation as possible by making sure that it could not refer to another part of the viewing section. Next, the key and distractors for the item were created. The key was written based on the idea unit and the distractors were written to match the key in content and grammatical structure. Care was made to make the key and distractors approximately equal length. The distractors were written making sure that they could not be considered correct but were still plausible. As many distractors as possible were written and the two most plausible were chosen (Haladyna, Downing, & Rodriguez, 2002). The position of the key (A, B, or C) for each item was randomly chosen through the *Integer Generator* function. The two distractors were inserted into the remaining positions. Examples of multiple-choice items created for Episode 1 are shown in Figure 2.4.

Figure 2.4 First two multiple-choice items on the comprehension test for Episode 1

4. Why is Chuck preparing to climb out the window of his room?
  - A. He is playing a game with Morgan.
  - B. He doesn't feel comfortable at the party.
  - C. He just stole something.
5. Why does Ellie want Chuck to go out and talk to people?
  - A. She invited lots of nice girls for him to meet.
  - B. She needs his help in the kitchen.
  - C. Her boyfriend wants to meet him.

#### **2.7.6.3. Creating the sequencing items**

The sequencing items were created using the same idea units that were used to create the true/false and multiple-choice items. First, all the idea units from the six viewing sections were combined into a chronological list of events in the episode. Less important idea units were deleted and related idea units were combined. This process was repeated until eleven of the most important events in the episode remained. Each of these main points was written as a declarative sentence carefully avoiding unclear pronoun reference, leading time markers, and too much detail that may have provided contextual clues to prior or subsequent plot points. 'The episode begins.' was added as the first event to bring the total to 12 events. The order of these events was then randomized. The relative positions of the first and seventh event were indicated. The participants were required to provide the order of the rest of the story events and only the order of these 10 main events were scored. The position of seventh event was identified so as to alleviate a problem that is common in sequencing tasks: a single error can cause a score of zero. This was a problem that occurred when the sequencing items were

originally piloted. Providing the position of the seventh event gives the participants a chance to reset the ordering of their answers even if they had confused the order of the events in the first half of the task. The sequencing items for Episode 1 are shown in Figure 2.5.

Figure 2.5 Sequencing items for Episode 1

- \_\_\_\_\_ A ninja tries to steal Chuck's computer.
- \_\_\_\_\_ Casey and Sarah realize that all of their secrets of the Intersect are in Chuck's head.
- \_\_\_\_\_ Chuck and Morgan try to escape from Chuck's birthday party.
- \_\_\_\_\_ Casey starts work at Buy More.
- \_\_\_\_\_ Chuck uses a computer virus to stop a bomb.
- \_\_\_\_\_ Chuck and Sarah go onto the roof of a building.
- (Ex.) 1 \_\_\_\_\_ The episode begins.
- \_\_\_\_\_ Chuck meets Sarah again at the Buy More and they make a date.
- (Ex.) 7 \_\_\_\_\_ Chuck and Sarah go to a nightclub.
- \_\_\_\_\_ Chuck gets an email message from Bryce and images flash in his head.
- \_\_\_\_\_ Chuck takes Sarah and Casey to a hotel.
- \_\_\_\_\_ Sarah comes to the Buy More to get her phone fixed.

#### **2.7.6.4. Test layout**

Each comprehension test had the same format to reduce possible confusion that unique test formats could have caused and to increase comparability between comprehension tests of different episodes. Each viewing section's comprehension questions were printed on a single sheet of A4 paper. The true/false items were presented first, followed by the multiple-choice items. The true/false and multiple-choice items were presented in the same order their answers appeared in the viewing section. The sequencing items were also printed on a single sheet of A4 paper. Participants recorded their answers for all items on a separate answer sheet that was organized by test section.

#### **2.7.6.5. Translations**

After all the items on each comprehension test were created, they were translated into Japanese. The comprehension tests were presented in the participants' L1 for three reasons: comprehension questions presented in the L1 can make a test easier, questions in the L1 can reduce test taker anxiety, and questions presented in the L2 may measure reading comprehension as much as they are measuring listening comprehension (Shohamy, 1984). All

translations were done by a single Japanese native speaker who had viewed each episode of *Chuck*. The translations were done in consultation with the researcher to ensure that the translated items were asking the same questions intended in the original English items. English cognates were avoided in the translations whenever possible as it was believed that their inclusion would have been too leading especially as part of a multiple-choice item's key.

#### **2.7.6.6. Validation of comprehension tests**

After the translation of the 10 comprehension tests was completed and the researcher had compared the original English versions and the translated Japanese versions of the tests for inconsistencies, they were used in conjunction with the pilot study. The results from 31 participants in the pilot study were used to examine the construct validity of the individual tests.

An examination of the construct validity of a test involves investigating whether the test actually measures the intended underlying trait that the researcher aims to measure. In the case of these comprehension tests, the trait is listening comprehension ability based on the model outlined by Buck (2001) and Song (2008). The following is a description of the procedure performed on each of the comprehension tests to examine whether the items fit the proposed model. This analysis is essentially a search for items that do not fit the model and provides a rationale for their modification or exclusion. Items that do not fit the model are suspect items and their inclusion on a comprehension test may be detrimental to construct validity. Throughout the explanation of the process, examples are given using the analysis of the comprehension test created for Episode 1 of *Chuck*.

The first step in the process was to analyze the true/false and multiple-choice items from each comprehension test using the software, Winsteps (Linacre, 2010). This software was used to construct Rasch measures from each data set of persons and items. These measures are based on the Rasch Model which is best described in the words of its developer:

“...a person having a greater ability than another person should have the greater probability of solving any items of the type in question, and similarly, one item being more difficult than another means that for any person, the probability of solving the second item is the greater one.” (Rasch, 1960, p.117 quoted in Wright, 1997)

Using the test takers' responses, an estimate of person ability for each test taker and an estimate of item difficulty for each question on the test can be calculated by Rasch analysis. The estimates of person ability and item difficulty are measured on a common scale and this makes it possible to determine the probability of a test taker correctly answering any item

(Weaver, 2005). These estimates of Rasch measures produced by Winsteps and the calculations based on them were used throughout the test validation procedure.

The second step in the item analysis process was to examine the item polarity of the test items. Polarity is the direction of the responses with respect to the latent variable and a positive polarity indicates more of the latent variable. Polarity is indicated by the point measure correlation which is the correlation between the data observations and the measures of the items producing them. The Rasch Model requires a single construct underlie the items that form the hierarchical continuum. Each item should make a meaningful contribution to the underlying construct being investigated so items with negative and low point measure correlations (below 0.10) are identified. A negative point measure correlation indicates that an item is not aligned in the same direction as the latent variable and not making a contribution to the construct. From the 72 items on the comprehension test for Episode 1, 19 items were found to have negative or low point measure correlations. Before these items were flagged for removal or revision, however, the persons (participants) were examined for misfit. The removal of misfitting persons from an analysis has the potential to improve the polarity of items.

The third step in validating the comprehension tests involved looking for misfitting persons through examining the outfit mean-squares on Winsteps' person misfit table. Fit is an important tenet of the Rasch Model. The model can calculate the difference between the expected value of each person's response to each item and the actual value. From these values, improbable responses can be identified. Fit, then, is the degree of match between this pattern of observed responses and the modeled expectations. Fit can be expressed as person fit or item fit. When the data and the model do not match, this is referred to as misfit. Misfit can occur when a highly proficient person gives an incorrect response to an easy item or a low proficiency person gives a correct response to a difficult item. That is, a misfitting person's pattern of responses does not follow the pattern one would expect of him or her. Outfit is an outlier-sensitive fit statistic that indicates unexpected observations by persons on items that are relatively very easy or very hard for them. Outfit can be caused by lucky guesses or careless mistakes. The mean-square is the chi-square statistic divided by its degrees of freedom and values are expected to be close to 1.0. Values greater than 1.0 indicate underfit that degrades measurement and values less than 1.0 indicate overfit that suggests the model predicts the data too well. In general, an outfit mean-square greater than 2.0 was used as an indicator for persons to be removed but also any person with an outfit-mean-square of 1.5 or over was considered. For example, on the comprehension test for Episode 1, two persons had

outfit mean-squares that indicated misfit. One person had an outfit mean-square of 2.35 and one of 1.50. The person with the outfit mean-square of 2.35 was removed from the analysis and the data was reanalyzed. This resulted in somewhat improved point measure correlations. The person with the outfit mean-square of 1.5 was then removed and the data reanalyzed. This did not improve the point measure correlations so this person was reinstated into the data analysis.

In the fourth part of the item analysis, the items with low point measure correlations, originally identified in the second stage of the analysis, were explicitly checked. This involved reading the original English version of the item as well as the Japanese translation, examining the pattern of responses to the item from the pilot, and watching the item's viewing section again. This was in an effort to see what exactly may have been wrong with the item. Across all 10 episode comprehension tests, eight different types of errors were uncovered through this explicit examination. These included: translation problems, misentered answers, typographical errors, unclear pronoun reference, English cognate usage in questions translated to Japanese, stem problems, distractor problems, and key problems. Translation problems were uncovered by having a native speaker of Japanese (not the original translator) with a high proficiency in English read the original English wording of the items and the translations. Misentered answers (when the correct key was mislabeled in the grading software) were uncovered by looking at the response frequencies of the distractors and the key versus the correct answer. Typographical errors occurred in English which led to translation issues in Japanese where inadvertently, the wrong word was entered. Pronoun reference problems occurred in situations where it was unclear who in the stem 'he', 'she' or 'it' was referring to. In identifying stem problems, it was often not as clear what was wrong although the low point measure correlations and the frequency distribution of the options indicated that there was a problem. To remedy this, stems were rewritten to be more specific about the situation or scene that they referenced. Distractors were rewritten when they were found to be too distinct and not plausible, and therefore not considered by participants. They were also rewritten if they were too close to the correct answer and might be considered correct. The key was rewritten when it was found to be too vague, referred to an event from a different viewing section, was too detailed in describing the situation, or when it was considered too inferential and not linked to an explicit part of the dialogue.

From the 19 items identified as potentially being problematic by their low point measure correlations from the analysis of Episode 1's comprehension test, five items had obvious errors or problems and were flagged to be rewritten. Two items with low but not negative

point measure correlations were found to have expected percentages and observed percentages within 5% of each other. The observed percentage is the percentage of data points that are within 0.5 score points of their predicted value, and the expected percentage is that value predicted for this situation based on the item measures. This suggests that these two items were functioning almost as well as they could and were left in the comprehension test as is. The remaining 12 items were flagged for removal subject to the steps of the item analysis process still to be undertaken.

The fifth step was to reanalyze the data after the misfitting items (both flagged for removal and to be rewritten) were removed from the results data. Upon reanalysis, the results were inspected in the same manner as described in the second and third steps. None of the remaining items from the comprehension test for Episode 1 presented low point measure correlations and no additional persons had outfit mean-square scores large enough to be considered misfitting.

The sixth step in the item analysis process involved inspecting the outfit and infit statistics based on mean-square scores for the items remaining after items previously designated as misfitting had been removed. This is a similar process to that described in the third step of the item analysis process except instead of examining persons, items are examined. The mean-square score for an item indicates the size of its misfit. The acceptable mean-square range was set between 0.7 and 1.3 (Linacre, 2010). Both infit and outfit are fit statistics and they indicate how well data fit the model. Infit is inlier-sensitive while outfit is outlier-sensitive. Values below 1.0 indicate that the data is too predictable (infit) and values above 1.0 indicate data that is too unpredictable (outfit). The remaining items on the comprehension test for Episode 1 were found to be below the upper limit as the highest observed mean-square was 1.12. There were items with mean-squares lower than the acceptable limit of 0.7 with the lowest being 0.33. Items with low mean-squares, however, do not damage measurement although they are somewhat redundant. For tests with many items it may be advisable to remove these redundant items. The comprehension tests for *Chuck* were not considered to have an excess of items and with more participants completing the tests after the pilot study it was thought that the items with low mean-squares could be more discriminatory. No items from the comprehension test for Episode 1 were eliminated due to low mean-squares.

Using the analysis provided by Winsteps, the seventh step was to identify items with standardized fit statistic (ZSTD) scores of 2 or higher. A ZSTD is essentially a *t*-test of the hypothesis that the data fits the model perfectly. When the ZSTD is beyond 1.96 (for  $p < .05$ ) the null hypothesis (the data fits the Rasch model) is rejected (Linacre, 2010). The

comprehension test for Episode 1 had no items with ZSTD values approaching 2 for either infit or outfit.

Borderline items and persons were reinstated and the data was reanalyzed in the eighth step of the item analysis. This makes it possible to see if the borderline items and persons are now in line with the refined test. None of the items or persons that were reinstated functioned any better in this reanalysis and they were once again removed from the comprehension test for Episode 1.

The ninth and tenth steps involved comparing the hierarchal results of participant ability and item difficulty from the Rasch analysis to the outcomes predicted by the model on which the test items were based. In the ninth step, the item difficulty hierarchy was examined to see if it matched the hierarchy predicted by the model. In the tenth step, the expected participant results were compared to the participant ability hierarchy results. Essentially, these two steps examine the predictive validity of a test. The content of each episode and the relative difficulty of the items were considered by the researcher and compared to the item difficulty measures produced by the Rasch analysis. The language abilities of the participants were considered based on other educational activities in the course and familiarity with the population of Japanese university students the pilot study sample was taken from. These were compared to the hierarchy of participants' ability measures from the Rasch analysis. No items from the comprehension test for Episode 1 were flagged for removal based on these comparisons in these two stages of item analysis.

The final step in the test validation procedure was to rewrite the items that were found to be problematic. Five items in total were rewritten and included in the final version of the comprehension test for Episode 1. These included three multiple-choice and two true/false items. These items were revised as follows: Item #24 – the key was rewritten to be more specific to the situation, Item #27 – the translation was considered too ambiguous and revised, Items #28 and #56 – the stems were rewritten to be more specific to the situation, and Item #66 – the distractors were considered too similar to the key and reworded.

The 10 sequencing items on each comprehension test were not included in the item analysis carried out on the true/false and multiple-choice items. The results of the sequencing items from the pilot study were examined for items that caused the participants to answer incorrectly an inordinate number of times. These items were checked for errors in the original text, translation, and sequence within the episode. Throughout the 10 comprehension tests, outside the correction of small grammatical errors and nuanced translation revisions, very few problems were found with the sequencing items.

To sum up the test validation process for Episode 1, the comprehension test was originally created with 82 items. The test had six viewing sections with 12 items each. The test concluded with 10 sequencing items. Each viewing section's 12 items consisted of five true/false and seven multiple-choice items. The multiple-choice and true/false items on the original test included 6 topic, 43 detail, and 23 inference items. The item analysis resulted in two true/false and three multiple-choice items being flagged to be rewritten. Twelve items were flagged for removal. The items removed included five true/false and seven multiple-choice questions. One item was removed from viewing sections 2, 3 and 6, two items from sections 1 and 5, and five items from section 4. The 12 items removed included two topic questions, nine detail questions, and one inference question. The final version of the comprehension test for Episode 1 had 60 true/false and multiple-choice items and 10 sequencing items.

#### **2.7.6.7. Final versions of the comprehension tests**

Prior to the test validation procedure, each comprehension test had 30 true/false items, 42 multiple-choice items, and 10 sequencing items. Following the test validation procedure, the number of items on the comprehension tests ranged from 70 (Episode 1) to 78 (Episode 8). Not including the sequencing items, the original format of the tests with 30 true/false items and 42 multiple-choice items meant that 41.7% of the items were true/false and 58.3% of the items were multiple-choice items. On the validated tests, the ratio of true/false items to multiple-choice items ranged from 37.7% true/false items and 62.3% multiple-choice items (Episode 4) to 42.4% true/false items and 57.6% multiple-choice items (Episode 2 and Episode A). The number of true/false items in a single viewing section varied from 2 to 5 and the number of multiple-choice items ranged from 4 to 7. The number of sequencing items was 10 throughout the comprehension tests. Finalized versions of the comprehension tests (in Japanese and English) and the answer sheets for all episodes of *Chuck* are shown in Appendices A1 to A30. Tables 2.6 and 2.7 show the number and type of items in each viewing section for each of the 10 comprehension tests.

Table 2.6 Number and type of items on the comprehension tests for Episode 1 to Episode 5

|            | Episode 1 |       | Episode 2 |       | Episode 3 |       | Episode 4 |       | Episode 5 |       |
|------------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|
|            | TF        | MC    | TF        | MC    | TF        | MC    | TF        | MC    | TF        | MC    |
| Section 1  | 3         | 7     | 5         | 6     | 3         | 6     | 5         | 7     | 3         | 6     |
| Section 2  | 5         | 6     | 4         | 6     | 5         | 7     | 3         | 5     | 5         | 6     |
| Section 3  | 5         | 6     | 5         | 6     | 5         | 7     | 5         | 7     | 3         | 7     |
| Section 4  | 3         | 4     | 5         | 6     | 3         | 7     | 2         | 6     | 4         | 7     |
| Section 5  | 4         | 6     | 5         | 7     | 5         | 7     | 4         | 6     | 4         | 7     |
| Section 6  | 5         | 6     | 4         | 7     | 5         | 6     | 4         | 7     | 5         | 5     |
| Type Total | 25        | 35    | 28        | 38    | 26        | 40    | 23        | 38    | 24        | 38    |
| Type %     | 41.7%     | 58.3% | 42.4%     | 57.6% | 39.4%     | 60.6% | 37.7%     | 62.3% | 38.7%     | 61.3% |
| Sequencing | 10        |       | 10        |       | 10        |       | 10        |       | 10        |       |
| Total      | 70        |       | 76        |       | 76        |       | 71        |       | 72        |       |

Note. TF = true/false, MC = multiple-choice

Table 2.7 Number and type of items on the comprehension tests for Episode 6 to Episode 8, Episode A, and Episode B

|            | Episode 6 |       | Episode 7 |       | Episode 8 |       | Episode A |       | Episode B |       |
|------------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|
|            | TF        | MC    | TF        | MC    | TF        | MC    | TF        | MC    | TF        | MC    |
| Section 1  | 4         | 6     | 4         | 7     | 5         | 7     | 4         | 7     | 5         | 6     |
| Section 2  | 4         | 7     | 5         | 7     | 3         | 7     | 4         | 6     | 3         | 7     |
| Section 3  | 5         | 7     | 4         | 7     | 4         | 7     | 5         | 5     | 5         | 5     |
| Section 4  | 5         | 6     | 5         | 5     | 5         | 7     | 5         | 6     | 5         | 6     |
| Section 5  | 4         | 7     | 5         | 7     | 4         | 7     | 5         | 7     | 3         | 6     |
| Section 6  | 4         | 6     | 4         | 7     | 5         | 7     | 5         | 7     | 4         | 6     |
| Type Total | 26        | 39    | 27        | 40    | 26        | 42    | 28        | 38    | 25        | 36    |
| Type %     | 40.0%     | 60.0% | 40.3%     | 59.7% | 38.2%     | 61.8% | 42.4%     | 57.6% | 41.0%     | 59.0% |
| Sequencing | 10        |       | 10        |       | 10        |       | 10        |       | 10        |       |
| Total      | 75        |       | 77        |       | 78        |       | 76        |       | 71        |       |

Note. TF = true/false, MC = multiple-choice

In total, there were 742 items on the final versions of the 10 comprehension tests for an average of 74.2 items per test. The comprehension tests contained an average of 5.6 topic-based items with a maximum of 7 (Episode 6) and minimum of 4 (Episode 1). Over half the items on the comprehension tests were detail-based (50.3%) and each test averaged 37.3 detail-based items. The comprehension test for Episode 4 had the most detail items (41) and the test for Episode 1 had the least (34). There was an average of 21.3 inference questions per comprehension test with a maximum of 25 (Episode 7) and a minimum of 15 (Episode 4). The sequencing items, of which there were 10 per comprehension test, made up 13.5% of the items on the comprehension tests. The type and number of comprehension ability questions

on each of the 10 comprehension tests are shown in Table 2.8. The item types in each of the 10 episodes of *Chuck* organized by individual viewing section are shown in Appendix A31.

Table 2.8 Number of comprehension ability items for each of the 10 comprehension tests

|                | Topic | Detail | Inference | Sequencing | Total |
|----------------|-------|--------|-----------|------------|-------|
| Episode 1      | 4     | 34     | 22        | 10         | 70    |
| Episode 2      | 7     | 37     | 22        | 10         | 76    |
| Episode 3      | 5     | 38     | 23        | 10         | 76    |
| Episode 4      | 5     | 41     | 15        | 10         | 71    |
| Episode 5      | 6     | 36     | 20        | 10         | 72    |
| Episode 6      | 7     | 39     | 19        | 10         | 75    |
| Episode 7      | 5     | 37     | 25        | 10         | 77    |
| Episode 8      | 5     | 39     | 24        | 10         | 78    |
| Episode A (12) | 6     | 36     | 24        | 10         | 76    |
| Episode B (13) | 6     | 36     | 19        | 10         | 71    |
| Type Total     | 56    | 373    | 213       | 100        | 742   |
| Type %         | 7.5%  | 50.3%  | 28.7%     | 13.5%      | 100%  |

### 2.7.7. The teaching sessions

In the first teaching session of Study 1, the participants completed the ethics approval procedure. This involved reading through an introduction to the study as a group and signing the ethics approval form. Any questions about what it meant to be a participant in the study were addressed. Participants were informed that it was not mandatory to be a participant and could withdraw at any future time from the study. The participants also completed the 2,000, 3,000 and 5,000 levels of the Vocabulary Levels Test. Each participant was given as much time as needed to complete the tests at each level.

In the second teaching session, the participants completed the Tough and Sensitive Vocabulary Pre-Tests. The vocabulary test administration procedure is discussed in detail in Section 3.5.9. The participants then read through the Introduction to Television Viewing information sheet together and the instructor fielded any questions that the participants had about viewing television in class. Then, for approximately 30 minutes, the participants completed the Television Viewing Preview. This is essentially practice television viewing and comprehension question answering. Care was taken to make sure that the participants understood the format of each item type and how to fill in the answer sheets correctly.

In Teaching Session 3, participants viewed Episode A1 or B1 of *Chuck* and they completed the corresponding comprehension test and attitude survey. Viewing Group 1 viewed Episode A1 and Viewing Group 2 viewed Episode B1. Subsequently, in Teaching

Sessions 4 to 11, all participants viewed Episodes 1 to 8 and completed the corresponding comprehension tests and attitude surveys. In Teaching Session 12, Viewing Group 1 viewed Episode B2 and Viewing Group 2 viewed Episode A2 and completed their final comprehension test and attitude survey.

Teaching Sessions 3 to 12 were each 90 minutes long. The first 10 minutes of each session involved taking attendance and distributing the answer sheets for the comprehension test and attitude survey. Each participant received an answer sheet and was asked to write his or her name and student number on it. Attendance was taken to record the absent participants who were excluded from the study for that teaching session. Any participant arriving 10 minutes after the beginning of the teaching session was marked as absent from the episode. This was because he or she would not have completed all the necessary stages involved with viewing the episode and completing the comprehension test. Next, the comprehension questions for the first viewing section were distributed and the participants were given approximately 30 seconds to preview the questions. This was followed by the first viewing section. At the end of the section, the video was paused and the participants were given 3.5 minutes to complete the first set of questions. When participants had finished answering a set of comprehension questions for a viewing section, those questions were collected and they were given the next set of comprehension questions to preview. This was done so that participants could not use information from items in previous sections to answer items for that viewing section. This procedure continued for each viewing section until the episode was completed. When the comprehension questions from the final viewing section were collected, participants received the sequencing items and the Episode Attitude Survey. This was done so that the participants could not use the true/false and multiple-choice items to help answer the sequencing items. The participants were given 10 minutes to complete these two tasks. Finally, the sequencing questions, survey questions, and answer sheets were collected and the participants were dismissed.

In Teaching Session 13, the participants completed the Tough and Sensitive Vocabulary Post-Tests and the Final Attitude Survey.

### **2.7.8. Vocabulary Levels Test**

To determine whether comprehension of television improves with greater vocabulary knowledge, the results from the Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001) at the 2,000-, 3,000-, and 5,000-word levels were used. These tests were originally developed by Paul Nation (1983, 1990) and measure receptive vocabulary knowledge. A full description

of the VLT, including its format, advantages, and scoring, is included in Study 3 in Section 4.6.1. Results from piloting indicated that the three levels of the VLT used in this study were sufficient to cover the extent of the vocabulary knowledge of the participants.

## 2.8. Analysis

### 2.8.1. Optical mark recognition

All of the tests in this study were scored using the optical mark recognition (OMR) software, Remark Office OMR 6 (Gravic, 2007). This software is produced by Gravic, Inc. and recognizes customizable forms containing optical marks (bubbles and checkboxes). Test-specific answer sheets were created and the participants filled in the answer bubbles on these sheets. They were then scanned with an image scanner and the electronic image of each answer page was analyzed with the software producing spreadsheet files of the participants' responses. Answer sheets separate from the question sheets were created and used for all comprehension tests while for the VLT the questions and answer bubbles were on the same pages. Examples of portions of OMR-ready answer fields for the comprehension tests are shown in Figure 2.6. An example of a portion of the answer sheet for the VLT is shown in Study 3 in Figure 4.1. The OMR-ready answer sheets for all tests are included in the appendices.

Figure 2.6 Examples of OMR answer fields for the comprehension tests

True/False item

**1**      (T)      (F)

Multiple-choice item

**6**      (A)      (B)      (C)

Sequencing item

**69**      (2)      (3)      (4)      (5)      (6)      (8)      (9)      (10)      (11)      (12)

### 2.8.2. Expectation Maximization Algorithm

Participants that were absent from a single teaching session for Episodes 1 to 8 did not complete a comprehension test. The missing comprehension test scores for these episodes were replaced using the Expectation Maximization Algorithm. This has been shown to be an effective technique for managing missing data with favorable statistical properties (Allison, 2001). This technique replaces missing data by first obtaining expected values for the missing values using regression equations. These values are then refined through the expected log-likelihood of the data. Before completing the missing value replacement, the data was

analyzed by Little's MCAR Test to determine if the missing data occurred at random. The results of this test were all non-significant indicating that the data were missing at random and thus suitable for missing value replacement using expectation maximization. Missing comprehension test data was replaced 55 times over the eight episodes ranging from three to 12 times per episode.

### **2.8.3. Rasch measurement using the CHIPs scale**

The participants' scores on the comprehension tests and the scores of the combined VLT were expressed as CHIPs. These are units of measurement produced when test results are analyzed using the Rasch model. Measurement in the Rasch model is often expressed in logits. A logit is the unit the Rasch model produces when raw scores are transformed to interval data. This unit of measurement is the natural log of the odds of a participant successfully answering different items on a test (Smith, 2000). The value of 0 logits is usually set as the mean item difficulty. Scores for item difficulty and person ability are located along this logit scale. More able persons and more difficult items are towards the top of the scale while less able persons and less difficult items are towards the bottom (Bond & Fox, 2007). The CHIPs scale is a modified version of the logit scale. Using the CHIPs scale has many benefits. The first benefit is that the logit scale concept is not widely used so transforming the scale into the more user-friendly CHIPs scale allows people to grasp the meaning of the results more easily. The CHIPs scale has no negative numbers the way the logit scale does and has the more familiar range of 0 to 100. The next advantage of CHIPs is that the midpoint of the scale is set at 50 CHIPs. This represents the average difficulty of all the items on the test. A participant's score on a comprehension test can be easily interpreted in relation to the average difficulty of the items on that test (Weaver, 2005). Another benefit is that CHIPs scaling standardizes the results of the comprehension tests with different numbers of items. This allows for the comparison of scores on two or more comprehension tests with different numbers of items. Another advantage is the scores on the CHIPs scale are interval data which is an assumption in statistical analyses such as ANOVAs (Kerr, Hall, & Kozub, 2002).

## **2.9. Results**

### **2.9.1. First episode to tenth episode comprehension gains**

To examine how comprehension of *Chuck* changed from the first to the tenth episode viewed, the comprehension test scores for the Initial Episode and Final Episode were analyzed. To find out whether the comprehension test results from Episode A1 and B1, and

Episodes B2 and A2 could be analyzed together as the Initial and Final Episode, first the comprehension test scores of the participants in each viewing group were analyzed separately. As shown in Table 2.9, for Viewing Group 1 (Episode A1 to Episode B2) the mean raw score for Episode A1 was 41.55 out of a possible 76 (54.7%) and on Episode B2 it was 44.35 out of a possible 71 (62.5%). Table 2.10 shows the mean scores for the raw data for Viewing Group 2 (Episode B1 to Episode A2). The mean raw score for Episode B1 was 37.41 out of a possible 71 (52.7%) and on Episode A2 it was 46.19 out of a possible 76 (60.8%). Tables 2.11 and 2.12 show the mean scores for these viewing groups' first and final episode in CHIPs scores. The CHIPs scores are in line with the raw scores in that the mean score for Episode B2 (53.46) is higher than that in Episode A1 (51.14) for Viewing Group 1, and the mean score of Episode A2 (52.73) is higher than Episode B1 (50.61) for Viewing Group 2.

Table 2.9 Mean scores of raw data on comprehension tests for Episode A1 (out of 76) and Episode B2 (out of 71) for Group 1

| Comprehension Test | Mean  | Mean % | N   | SD    | Median | Minimum | Maximum |
|--------------------|-------|--------|-----|-------|--------|---------|---------|
| Episode A1         | 41.55 | 54.7%  | 208 | 6.509 | 41     | 24      | 56      |
| Episode B2         | 44.35 | 62.5%  | 208 | 6.620 | 45     | 20      | 58      |

Table 2.10 Mean scores of raw data on comprehension tests for Episode B1 (out of 71) and Episode A2 (out of 76) for Group 2

| Comprehension Test | Mean  | Mean % | N   | SD    | Median | Minimum | Maximum |
|--------------------|-------|--------|-----|-------|--------|---------|---------|
| Episode B1         | 37.41 | 52.7%  | 113 | 6.175 | 38     | 22      | 50      |
| Episode A2         | 46.19 | 60.8%  | 113 | 6.705 | 47     | 27      | 59      |

Table 2.11 Means of CHIPs scores on comprehension tests for Episode A1 and Episode B2 for Group 1

| Comprehension Test | Mean  | N   | Std. Deviation | Median | Minimum | Maximum |
|--------------------|-------|-----|----------------|--------|---------|---------|
| Episode A1         | 51.14 | 208 | 2.005          | 50.9   | 45.7    | 55.8    |
| Episode B2         | 53.46 | 208 | 2.771          | 53.6   | 43.4    | 59.9    |

Table 2.12 Means of CHIPs scores on comprehension tests for Episode B1 and Episode A2 for Group 2

| Comprehension Test | Mean  | N   | Std. Deviation | Median | Minimum | Maximum |
|--------------------|-------|-----|----------------|--------|---------|---------|
| Episode B1         | 50.61 | 113 | 2.269          | 50.8   | 44.9    | 55.5    |
| Episode A2         | 52.73 | 113 | 2.263          | 52.9   | 46.4    | 57.4    |

The difference between the comprehension scores (as measured in CHIPs) between the first episode viewed and tenth episode viewed was calculated for the participants in both viewing groups. These comprehension score gains for Viewing Group 1 ( $M=2.32$ ,  $SD=2.476$ ,

$N=208$ ) and Viewing Group 2 ( $M=2.12$ ,  $SD=2.373$ ,  $N=113$ ) were analyzed with an independent  $t$ -test. Examination of the two samples using normal Q-Q plots and a Levine test of equality of variance revealed no serious threats to the assumptions of normality or homogeneity of variance. The  $t$ -test indicated that the means of the two groups were not significantly different,  $t(319) = .701$ ,  $p = .484$ . These results suggest that the mean comprehension gains for Viewing Group 1 and Viewing Group 2 did not differ, and because the results are both expressed in CHIPs scores, the comprehension scores from Episode A1 and Episode B1 (Initial Episode), and Episode B2 and Episode A2 (Final Episode) can be analyzed collectively. The combined mean CHIPs scores for the participants' Initial and Final Episode are shown in Table 2.13.

Table 2.13 Mean scores of CHIPs scores on the combined comprehension tests for the Initial Episode (Episodes A1 & B1) and the Final Episode (Episodes B2 & A2)

| Comprehension Test | Mean  | N   | Std. Deviation | Median | Minimum | Maximum |
|--------------------|-------|-----|----------------|--------|---------|---------|
| Episode A1 & B1    | 50.95 | 321 | 2.113          | 50.9   | 44.9    | 55.8    |
| Episode B2 & A2    | 53.20 | 321 | 2.623          | 53.6   | 43.4    | 59.9    |

A paired  $t$ -test was used to compare the comprehension scores from the Initial and Final Episodes. Examination of the two samples using normal Q-Q plots and a Levine test of equality of variance revealed no serious threats to the assumptions of normality or homogeneity of variance. There was a significant difference in the comprehension scores for the Initial Episode ( $M=50.95$ ,  $SD=2.113$ ) and the Final Episode ( $M=53.20$ ,  $SD=2.623$ );  $t(320)= 16.529$ ,  $p < .001$ . The effect size as measured by  $d$  was 0.945 a value corresponding to a large treatment effect. These results suggest the viewing of the eight successive episodes between the Initial and Final Episode led to a large, reliable effect on comprehension scores for the participants in Study 1 who viewed the English-language episodes of *Chuck*.

### 2.9.2. Comprehension across eight successive episodes

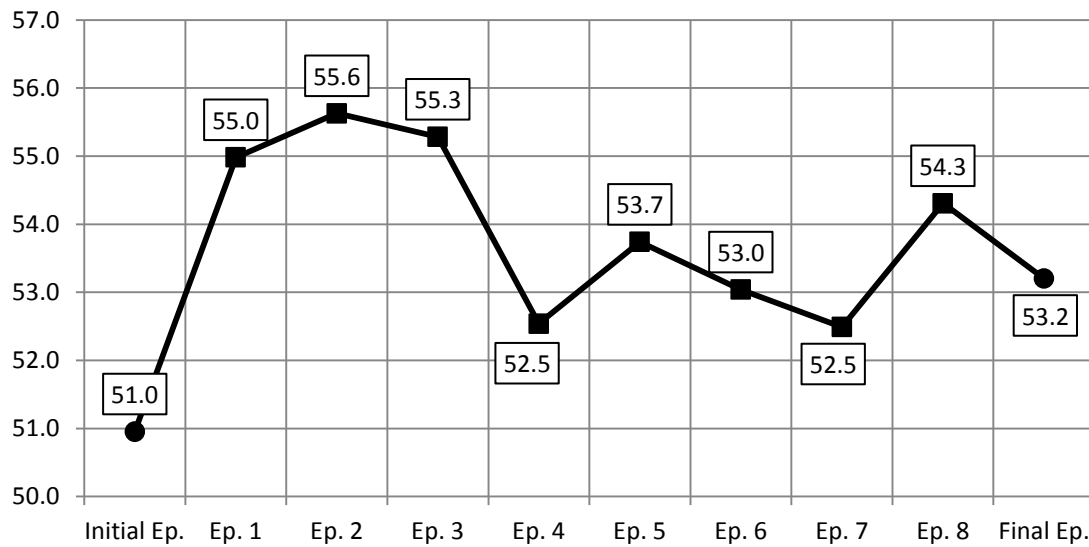
To examine how comprehension of English-language television changed across the eight successive episodes viewed, comprehension test scores were calculated for these episodes. Table 2.14 shows the results, in CHIPs and raw scores, and the descriptive statistics for each comprehension test. While the differences between comprehension tests in terms of mean CHIPs scores may seem minimal, examination of the raw score percentages shows that there was considerable difference between the results of some comprehension tests.

Table 2.14 Comprehension test scores, expressed in CHIPs and raw scores, for Episode 1 through Episode 8

|                  | Episode |       |       |       |       |       |       |       |
|------------------|---------|-------|-------|-------|-------|-------|-------|-------|
|                  | 1       | 2     | 3     | 4     | 5     | 6     | 7     | 8     |
| CHIPs Score Mean | 55.0    | 55.6  | 55.3  | 52.5  | 53.7  | 53.0  | 52.5  | 54.3  |
| CHIPs Std. Dev.  | 2.82    | 2.84  | 2.65  | 2.57  | 2.65  | 2.65  | 2.50  | 2.70  |
| CHIPs Median     | 55      | 55.9  | 55.3  | 52.7  | 54    | 53    | 52.6  | 54.3  |
| CHIPs Minimum    | 45.9    | 47.4  | 45    | 42.6  | 44.8  | 46.3  | 45.3  | 45.3  |
| CHIPs Maximum    | 62.3    | 62.5  | 66.2  | 58.9  | 61.9  | 60.1  | 60.2  | 61.4  |
| Raw Mean         | 48.4    | 53.8  | 53.4  | 43.1  | 46.1  | 46.6  | 46.4  | 51.6  |
| Raw %            | 69.2%   | 70.8% | 70.2% | 60.8% | 64.0% | 62.2% | 60.2% | 66.1% |
| Raw Std. Dev.    | 7.18    | 6.95  | 7.05  | 7.40  | 6.82  | 7.28  | 7.19  | 7.49  |
| Raw Median       | 49      | 55    | 54    | 44    | 47    | 47    | 47    | 52    |
| Raw Minimum      | 23      | 31    | 23    | 15    | 22    | 27    | 25    | 25    |
| Raw Maximum      | 63      | 67    | 72    | 59    | 63    | 63    | 65    | 68    |

The CHIPs scores show that the comprehension varied across the eight successive episodes of *Chuck*. Figure 2.7 plots the mean scores for the comprehension tests measured in CHIPs across the eight successive episodes. The results from the Initial and Final Episodes are included to better illustrate the range of comprehension scores across all episodes viewed in the study. From the Initial Episode the participants viewed to the Final Episode viewed, there was a mean increase of 2.2 CHIPs. After a mean score of 51.0 in the Initial Episode, there was a large rise in comprehension to 55.0 CHIPs for Episode 1 and then to the peak mean score in Episode 2 of 55.6 CHIPs. Comprehension test scores from Episode 3 to Episode 8 ranged from 52.5 (Episode 4 and Episode 7) to 55.3 CHIPs (Episode 3). The comprehension scores over the eight successive episodes viewed suggest that participants gained comprehension with successive viewing although there was a considerable difference in the comprehension test results for individual episodes.

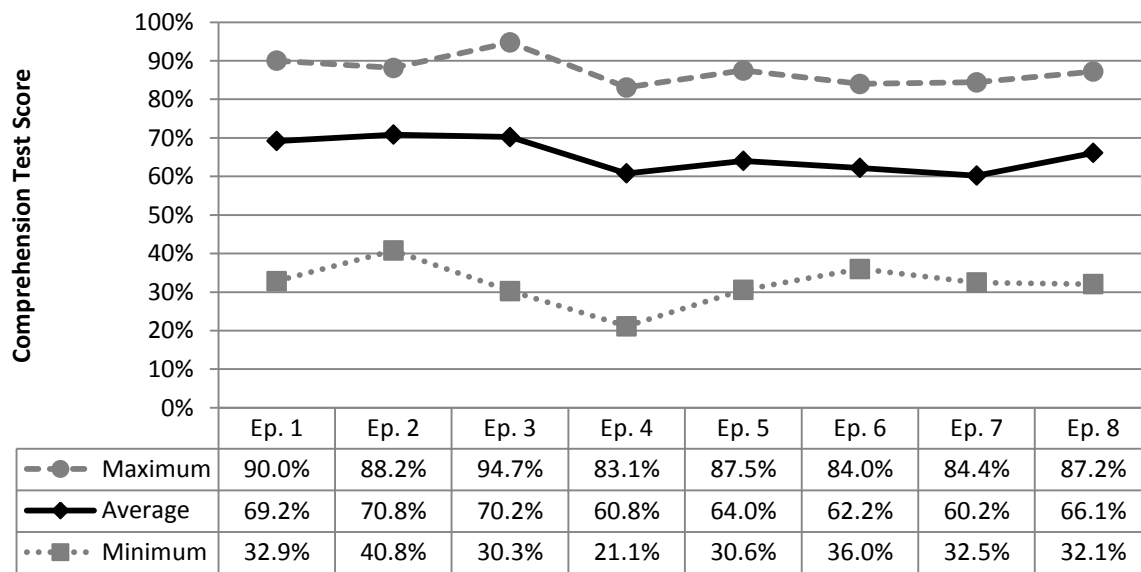
Figure 2.7 Mean CHIPs comprehension scores across all 10 episodes



A one-way repeated measures ANOVA was conducted to compare the effect of viewing time point on comprehension of the eight successive episodes of *Chuck*. The ANOVA with a Greenhouse-Geisser correction determined that the mean comprehension score differed significantly between the episodes viewed ( $F(6.607, 2114.098) = 187.2, p < 0.001$ ). Post hoc tests using the Bonferroni correction revealed that all but three pairwise comparisons of episode comprehension were significantly different ( $p < 0.05$ ). The three pairwise comparisons that were not significantly different were Episode 1 and Episode 3, Episode 2 and Episode 3, and Episode 4 and Episode 7.

Within the eight successive episodes, the large amount of variation in the participants' comprehension test results is apparent in Figure 2.8. This figure plots the maximum, mean and minimum comprehension test scores expressed as a percentage for each of the 10 episodes. While the mean comprehension scores are all over 60% (and 70% for two episodes) the average minimum comprehension score across all episodes was 32.0%. The average maximum comprehension test score was 87.4% representing an average separation of 55.4% between the maximum and minimum scores across the eight episodes. There was also variation between episodes for the minimum and maximum scores. The minimum comprehension test score ranged from 21.1% (Episode 4) to 40.8% (Episode 2) while the maximum comprehension test score ranged from 83.1% (Episode 4) to 94.7% (Episode 3). This variation between participants' comprehension test scores indicates that there were members of the sample that were able to achieve a considerable level of comprehension while others were not.

Figure 2.8 Maximum, average and minimum comprehension scores expressed as percentages across all eight successive episodes



### 2.9.3. Vocabulary knowledge and comprehension of English-language television

To investigate whether comprehension of English-language television improves with greater vocabulary knowledge a Pearson product-moment correlation coefficient was computed to assess the relationship between the amount of vocabulary knowledge participants had and their comprehension of episodes of *Chuck*. Vocabulary knowledge was quantified by combining a participant's responses on the 2,000-, 3,000-, and 5,000-word Vocabulary Levels Tests that they completed. The results were analyzed under the Rasch Model. This produces an interval measure of vocabulary ability for each participant. The mean of the vocabulary knowledge measure and the mean comprehension scores in CHIPs for the 10 episodes is shown in Table 2.15. There were small significant correlations between the two variables for nine of the 10 episodes: Initial Episode ( $r = .206$ ,  $N = 321$ ,  $p < .001$ ), Episode 1 ( $r = .248$ ,  $N = 321$ ,  $p < .001$ ), Episode 2 ( $r = .180$ ,  $N = 321$ ,  $p < .001$ ), Episode 3 ( $r = .229$ ,  $N = 321$ ,  $p < .001$ ), Episode 4 ( $r = .171$ ,  $N = 321$ ,  $p < .01$ ), Episode 6 ( $r = .261$ ,  $N = 321$ ,  $p < .001$ ), Episode 7 ( $r = .235$ ,  $N = 321$ ,  $p < .001$ ), Episode 8 ( $r = .222$ ,  $N = 321$ ,  $p < .001$ ), and Final Episode ( $r = .206$ ,  $N = 321$ ,  $p < .001$ ). There was a moderate significant correlation between the two variables for one of the 10 episodes: Episode 5 ( $r = .305$ ,  $N = 321$ ,  $p < .001$ ). The correlations for each of the 10 episodes and vocabulary knowledge can be seen in Table 2.16. These results indicate that for certain episodes of English-language television, vocabulary knowledge is a significant factor but the degree to which it is significant depends on the episode.

Table 2.15 Scores on all comprehension tests and the combined results of the VLT expressed in CHIPs ( $N=321$ )

|        | Episode |      |      |      |      |      |      |      |      |       | VLT Combined <sup>†</sup> |
|--------|---------|------|------|------|------|------|------|------|------|-------|---------------------------|
|        | Initial | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | Final |                           |
| Mean   | 51.0    | 55.0 | 55.6 | 55.3 | 52.5 | 53.7 | 53.0 | 52.5 | 54.3 | 53.2  | 51.5                      |
| SD     | 2.11    | 2.82 | 2.84 | 2.65 | 2.57 | 2.65 | 2.65 | 2.50 | 2.70 | 2.62  | 3.87                      |
| Median | 50.9    | 55   | 55.9 | 55.3 | 52.7 | 54   | 53   | 52.6 | 54.3 | 53.6  | 51.7                      |
| Min.   | 44.9    | 45.9 | 47.4 | 45   | 42.6 | 44.8 | 46.3 | 45.3 | 45.3 | 43.4  | 41.3                      |
| Max.   | 55.8    | 62.3 | 62.5 | 66.2 | 58.9 | 61.9 | 60.1 | 60.2 | 61.4 | 59.9  | 64.6                      |

Note. <sup>†</sup> VLT Combined is a product of a Rasch analysis of the combined VLT 2,000, 3,000, and 5,000 results

Table 2.16 Pearson correlation results for the comparison of vocabulary knowledge<sup>†</sup> and comprehension scores for episodes

| Initial Ep. | Ep. 1   | Ep. 2   | Ep. 3   | Ep. 4  | Ep. 5   | Ep. 6   | Ep. 7   | Ep. 8   | Final Ep. |
|-------------|---------|---------|---------|--------|---------|---------|---------|---------|-----------|
| .206***     | .248*** | .180*** | .229*** | .171** | .305*** | .261*** | .235*** | .222*** | .206***   |

Note. <sup>†</sup> vocabulary knowledge score is a product of a Rasch analysis of the combined VLT 2,000, 3,000, and 5,000 results, \*\*  $r$  is significant at  $p < .01$ , \*\*\*  $r$  is significant at  $p \leq .001$

## 2.10. Summary of findings

The main findings of Study 1 can be summarized as follows:

1. There were significant increases in comprehension scores from the first to the tenth episode of *Chuck* viewed which, on average, equated to approximately an 8% or 2.3 CHIPs score increase in comprehension.
2. There was considerable difference in the comprehension scores over the eight successive episodes viewed. Average comprehension scores ranged from 60.2% or 52.5 CHIPs to 70.8% or 55.6 CHIPs.
3. Greater vocabulary knowledge was found to have a small significant correlation with higher comprehension scores for nine out of 10 episodes and a moderate significant correlation for one of the 10 episodes.

## 2.11. Discussion

### 2.11.1. Comprehension gains from the first to the tenth episode viewed

In answer to the first research question, the results from Study 1 demonstrate gains in comprehension of English-language television from the first to the tenth episode viewed. Participants had a mean score of 50.95 CHIPs (Episode A1:  $M=54.7\%$ ; Episode B1:  $M=52.7\%$ ) on the first episode they viewed and their mean score on the tenth and final episode was 53.20 CHIPs (Episode B2:  $M=62.5\%$ ; Episode A2:  $M=60.8\%$ ). The mean

comprehension gain was 2.22 CHIPs or just over 8%. The comprehension score gains from the Initial to the Final Episode viewed were significant with a large effect size.

There does not appear to be any previous research that has compared comprehension scores from the first video viewed to the final video viewed. This makes it difficult to compare these findings to previous research on the comprehension of video. The findings, however, can be compared to previous research on the effects of background information on comprehension of videos. When language learners were offered advance organizers that supplied background information prior to viewing videos, comprehension was improved (Chung, 1999; Herron et al., 1998). Advance organizers were not a feature of this study but after viewing the first episode, participants viewed eight successive episodes from the same program. In effect, these eight episodes may function in the same way as advance organizers and increase the participants' knowledge about the television program. With each episode viewed the participants had the opportunity to learn more about the main characters and their relationships on the show. They could also learn about the personalities of the characters and how they were apt to act in different situations. The participants could gather information about the various settings of the show and the type of events that were likely to occur in these places. Participants were exposed to the overall themes in the show and how these might affect a specific story arc within a particular episode. The findings from this study suggest that the participants may have been able to make use of this gradual accumulation of background information resulting in superior comprehension of the final episode.

The results in this study support research by Rodgers and Webb (2011) and Webb (2011) that suggested that viewing related episodes of television may lead to comprehension gains. In Rodgers and Webb (2011), conclusions about possible comprehension gains were based on a corpus analysis of full seasons of television programs containing 23 or 24 episodes of television. In Webb (2011), the analysis was of programs grouped into three genres with over 40 episodes per group. In Study 1, participants made significant gains in comprehension across 10 episodes of television. Because past research has mostly involved viewing relatively short videos and rarely more than one video, viewing 10 42-minute episodes may seem like a relatively large amount of viewing time. It is, however, not a lot of time compared to the seasons or genres of television analyzed in past corpus studies. It is also not a lot of viewing time considering that in reporting OECD countries the average household television viewing time was 3.7 hours a day in 2005 (OECD, 2009). If even a quarter of this viewing time was used for viewing L2 television, the approximately 7 hours of viewing time from episodes of *Chuck* in this study could be watched in less than two weeks. It would not take a considerably

longer amount of time to view a full season of a program and increased viewing time may translate into larger gains in comprehension than those reported in this study.

### **2.11.2. Comprehension of the eight successive episodes viewed**

The second question investigated in Study 1 focused on the comprehension of English-language television across the eight successive episodes viewed. The results indicate that there is considerable variation in language learners' comprehension of authentic television. There was a range of mean comprehension scores in the episodes from 52.5 CHIPs in Episode 4 (60.8%) and Episode 7 (60.2%) to 55.6 CHIPs in Episode 2 (70.8%). The mean score across the eight successive episodes viewed was 57.9 CHIPs (65.4%). A repeated measures ANOVA indicated that there was an overall significant difference between the comprehension scores of the eight episodes. Within the sample of language learners in this study there were also substantial differences in comprehension. Across the eight episodes, the mean maximum score was 87.4% and the mean minimum score was 32.0%. The maximum score was on average 21.9% higher than the mean score and the minimum score was on average 33.4% lower.

Variation in language learners' comprehension of videos is also apparent when the results of previous video comprehension studies are compared (Chung, 1999; Guillory, 1998; Huang & Eskey, 1999; Latifi et al., 2011; Markham & Peter, 2003; Taylor, 2005; Winke et al., 2010). The range of comprehension scores reported across these studies indicates that there are considerable differences in language learners' ability to comprehend video. Comprehension scores ranged from 36% to 66.9% with a mean comprehension score of 49.9% in the earlier studies. However, it is difficult to accept direct comparisons between these studies because of the different methodologies, comprehension tests, and videos. Studies that employed multiple videos, viewed under the same conditions, unfortunately did not report the mean comprehensions scores of the videos separately. The closest comparison can be made between studies by Chung (1999) and Huang and Eskey (1999) as participants in both studies viewed video from the same educational series. In Chung's study, participants viewed approximately 28 minutes from the series while participants in Huang and Eskey's study viewed approximately 21 minutes. Both studies measured comprehension through multiple-choice items. The mean score in Chung's study was 66.9% and in Huang and Eskey's study it was 51.1%. While differing participant demographics and test items could account for comprehension differences, the results from these studies nevertheless indicate that there can be differences in comprehension of videos even from the same series.

That comprehension varies from episode to episode as language learners build up knowledge about the television series was not unexpected. Unlike the Initial and Final Episodes, where procedures were taken to control for the relative difficulty of the episodes and comprehension tests, it was never assumed that the successive episodes would be equivalently difficult for language learners. Each of the eight successive episodes can be considered a different viewing text and it is unreasonable to think that different texts, even if they are from the same television series, would be equally comprehensible to language learners. There are many factors that might affect comprehension scores from episode to episode including those based on listening comprehension: accent, pronunciation, hesitations, connected speech, prosody, speaker speed, and the length of the listening text (Buck, 2001; Hasan, 2000; Rubin, 1994). Other factors affecting comprehension are specific to viewing videos and include: visual literacy of participants, relationship of images to audio, interest in the text by the participants, and video type (Gruba, 2004; Seels, Fullerton, Berry, & Horn, 2004; Wagner, 2002). It is beyond the scope of this research to determine how these factors possibly contributed to the differing comprehension scores of the eight successive episodes. The findings from Study 1, however, are important as they establish that language learners' comprehension of authentic episodes of television can vary episode by episode and could be the impetus for future research on exactly how and to what extent comprehension of television is influenced by these factors.

### **2.11.3. The effect of vocabulary knowledge on comprehension of television**

The third question examined by Study 1 focused on the effect of vocabulary knowledge on comprehension of episodes of *Chuck*. The results indicated that there is a relationship between vocabulary knowledge and comprehension. For all 10 episodes, there were significant correlations between vocabulary knowledge, as measured by the combined scores of the 2,000-, 3,000-, and 5,000-word Vocabulary Levels Tests, and scores on the comprehension tests. The correlations ranged from .171 for Episode 4 to .305 for Episode 5 with a mean correlation over the 10 episodes of .226. These results suggest that receptive vocabulary knowledge is a factor in the comprehension of television but the degree to which vocabulary knowledge is related to comprehension depends on the episode.

The correlations in this study were smaller than those that have been reported in previous research comparing listening comprehension and vocabulary knowledge. Stæhr (2009) reported a significant correlation of .70 and Bonk (2000) reported one of .45. Milton, Wade and Hopkins (2010) reported a significant correlation of .52 between a test of orthographic

vocabulary size and listening comprehension and a correlation of .67 between phonological vocabulary size and listening comprehension. Two characteristics of television may provide some explanation for the differences in the size of correlations found in Study 1 and those reported in previous research. These are the visual component of television and the length of the episodes.

The visual support offered by television may have aided both top-down and bottom-up processing which may partially explain the lower correlations between vocabulary knowledge and comprehension found in this study. Visual support has been identified as a major factor affecting listening comprehension (Rubin, 1994) and listening comprehension has been shown to be better for videos than for audio-only forms (Brett, 1997; Guichon & McLornan, 2008; Sueyoshi & Hardison, 2005). If language learners were able to use images to facilitate top-down processing then they may have had to rely less on their ability to recognize individual words, or vocabulary knowledge, for comprehension. Conversely, the word recognition aspect of bottom-up processing may also have been aided by the presence of images. Unknown or partially-known vocabulary encountered in the episodes may have been explicitly supported by imagery. This explicit support may have allowed learners with less vocabulary knowledge to comprehend more. In these ways the reliance on the imagery may have allowed participants to compensate for lower vocabulary knowledge. This may account for the significant correlations between vocabulary knowledge and comprehension and explain the lower correlations than those reported in previous studies.

Another possible explanation for the lower correlations in this study may be the length of the input texts. The listening task in the study by Milton et al. (2010) was a module from IELTS which consists of 40 items and takes approximately 30 minutes to complete. Stæhr (2009) also used a standardized listening test. The Cambridge Certificate of Proficiency in English takes approximately 40 minutes and contains 28 items. It is important to remember that these studies involve listening to multiple short passages and the total test time includes time given to answer questions. Bonk (2000) used four short listening texts ranging from 39 to 43 seconds in length. In short listening texts, comprehension may depend on knowledge of a word that occurs only once or twice. In Study 1, however, an episode of *Chuck*, or even a viewing section within it, was much longer so vocabulary had greater potential to reoccur. Even if a participant did not originally know a word in the context of the episode, there were additional opportunities to encounter it again in the episode or even in later episodes. Participants could build knowledge of vocabulary that was previously unknown and this may have led to increased comprehension. In this way, the participants' vocabulary knowledge

measured prior to the study may have had less influence on comprehension and in turn led to lower correlations in this study than in the research with much shorter aural texts.

Results from Study 1 indicate that vocabulary knowledge is a significant factor in the comprehension of television. Vocabulary knowledge, however, does not appear to play as large a role in comprehension of authentic television programs as it does for short listening passages. The presence of images and the length of the episodes may decrease the contribution of vocabulary knowledge for comprehension. Further research is needed to investigate how the relationship between vocabulary knowledge and comprehension is affected by videos with images that provide more or less support for vocabulary and by videos with longer and shorter running times.

## **2.12. Limitations**

The language proficiency of the participants in Study 1 may limit the ability to generalize from the results of this study. Proficiency has been found to influence listening (Chang & Read, 2006; Chiang & Dunkel, 1992) and viewing (Bianchi & Ciabattini, 2008; Guillory, 1998; Maleki & Safaee Rad, 2011; Taylor, 2005) comprehension. The participants in this study were described as pre-intermediate to intermediate within the context of their university. More proficient participants may have provided a different picture of comprehension of television. A replication of this study with EFL learners across a wider range of proficiencies' may improve our understanding of how comprehension can change over successive viewing and how vocabulary knowledge affects comprehension.

The method used in this study for measuring vocabulary knowledge should also be considered when interpreting the findings. The VLT was designed to be a diagnostic test of language learners' knowledge of the 2,000-, 3,000- and 5,000-word levels. It was not designed to be used in the manner it was used in this study. There is, however, precedence for using the VLT in this way (e.g. Laufer & Ravenhorst-Kalovski, 2010; Stæhr, 2009). The VLT also measures knowledge of the written form of words rather than the spoken form which is necessary for listening comprehension. This may suggest that the correlations between vocabulary knowledge and listening comprehension were underestimated (Stæhr, 2009). For example, Milton, Wade and Hopkins (2010) found that an orthographic-based test of vocabulary knowledge had a smaller significant correlation (.52) with listening comprehension than a phonological-based vocabulary test (.67). Further research is needed to investigate how a measurement of vocabulary knowledge derived from a phonological-based vocabulary test might correlate with comprehension of video.

### *Study 2: Incidental vocabulary learning through viewing television*

#### **3. Introduction**

Vocabulary acquisition can occur either through intentional or incidental learning. Intentional learning is the acquisition of vocabulary through an activity designed to commit components of vocabulary knowledge to memory (Hulstijn, 2001; Laufer & Hulstijn, 2001). Incidental learning, however, is the by-product of an activity, usually involving comprehension, which is not explicitly designed for vocabulary learning (Gass, 1999; Huckin & Coady, 1999; Hulstijn, 2001). Acquisition occurs as learners attempt to understand new words they hear or read in context (Paribakht & Wesche, 1999). While incidental learning through reading and listening is likely responsible for the majority of first language vocabulary learning (Nation, 2001; Schmitt, 2000), it is believed that intentional or explicit learning of vocabulary may be responsible for most second language vocabulary learning (Laufer & Paribakht, 1998; Laufer, 2001; Pellicer-Sánchez & Schmitt, 2010; Webb, 2008). However, there is the potential for a certain portion of language learners' L2 vocabulary to be acquired incidentally (Gass, 1999).

While intentional learning of vocabulary may be the most effective method of building language learners' vocabulary (Folse, 2006; Laufer, 2003; Pellicer-Sánchez & Schmitt, 2010), intentional instruction of vocabulary should be supplemented with activities that promote incidental vocabulary acquisition. These activities can provide learners with exposure to vocabulary in the context-rich environments necessary to strengthen and develop lexical knowledge. Explicitly teaching vocabulary becomes problematic beyond the first 3,000 most frequent words of English where the challenge is choosing which words to teach in an efficient and effective manner. As a result of this, mid-frequency vocabulary (between the 3,000- and 9,000-word levels) is not systematically or regularly taught in ESL or EFL situations. Programs designed to encourage incidental vocabulary learning often employ materials modified for language learners such as graded readers. These materials, however, do not usually contain vocabulary beyond the 3,000 or 4,000 most frequent words. This means it is necessary to explore the use of authentic materials not designed for language learners (Schmitt & Schmitt, 2012) to encourage incidental learning of mid-frequency vocabulary. Authentic materials allow for exposure to and possible acquisition of mid-frequency

vocabulary that may not be intentionally learned or encountered in sources of input modified for language learners.

A prerequisite for incidental learning to take place is that a learner must receive a large amount of input. This means the materials being absorbed should be enjoyable so as to hold learners' attention for longer periods of time and provide repeated encounters (Day & Bamford, 1998). Novels have been suggested as a good source of authentic input where learners choose the ones that interest them the most (Pellicer-Sánchez & Schmitt, 2010). In a similar manner, viewing full episodes of television can provide a large amount of authentic input.

There has been little experimental research on incidental vocabulary learning from full length episodes of television. There have, however, been corpus-based studies (McFadden, Barret, & Horst, 2009; Rodgers & Webb, 2011; Webb & Rodgers, 2009a, 2009b; Webb, 2011) whose findings indicate that, under certain conditions, vocabulary acquisition from television may be possible. The findings from this research indicate that television programs may be a source of incidental vocabulary acquisition if language learners are able to acquire vocabulary from viewing television as they do from reading. For this to occur, learners would need to view numerous episodes of television to maximize input in the same manner they are encouraged to maximize input in extensive reading programs. Extensive reading, in coordination with language classes, has been shown to lead to vocabulary gains (Horst, 2005; Schmitt, 2008; Waring & Nation, 2004).

There are a number of characteristics of television that are likely to contribute to incidental vocabulary acquisition. The first of these is the combination of visual and aural input. Research indicates that materials which provide both forms of input simultaneously may be conducive to incidental vocabulary learning (Dubois & Vial, 2000; Duquette & Painchaud, 1996; Neuman & Koskinen, 1992). Results from incidental vocabulary acquisition studies involving reading, listening, and viewing videos have indicated that imagery may aid in vocabulary acquisition. In their study of vocabulary acquisition from reading and listening to a simplified novel, Horst, Cobb and Meara (1998) suggested that the presence of pictures in the book supported vocabulary learning. Elley (1989) found that illustrations improved vocabulary learning when children listened to stories. Neuman and Koskinen (1992) attributed vocabulary gains to encountering target words in multiple modalities. They believed that the imagery in the video the participants viewed worked together with the aural input and the printed words in the captions to foster acquisition. Koolstra and Beentjes (1999) suggested that vocabulary acquisition may be improved by using videos where the link

between the imagery and word meaning are clearly linked. In addition, findings from vocabulary learning studies have indicated that imagery can support vocabulary acquisition. Avila and Sadoski (1996) found that language learners studying vocabulary through the keyword technique that involves the use of mnemonic pictures had superior vocabulary learning over those studying through rote repetition or translation. Language learners have also been shown to learn more vocabulary when glosses that accompanied listening tasks were supplemented with pictures (Jones & Plass, 2002). Glosses supplemented with video clips were found to be even more effective than glosses accompanied by pictures for vocabulary learning through reading (Al-Seghayer, 2001).

The second characteristic of television that may aid incidental vocabulary learning concerns the idea that learners focus on understanding the programs they are viewing. Incidental vocabulary learning is thought to occur during comprehension-focused processing (McFadden et al., 2009) which has been shown to occur in reading studies (Horst, 2005; Nation, 2001; Waring & Takaki, 2003). Viewing episodes of television can be considered a comprehension-focused activity.

The third characteristic of viewing television that may be advantageous to learning vocabulary is the serial nature of the programs themselves. Viewing an episode of a program is rarely done in isolation. Many viewers choose to view episodes of one program on a regular basis. A learning approach which involves regular viewing of one program could lead to large incidental vocabulary learning gains (Webb & Rodgers, 2009a). Studies on the occurrence of vocabulary in episodes of the same television program indicate that it would be advantageous for learning vocabulary to view related episodes of programs rather than random television programs (Rodgers & Webb, 2011; Webb, 2011). Texts on the same theme have been shown to provide favourable learning conditions and allow for greater repetition of unknown vocabulary (Gardner, 2008; Nation, 2001). One way that regular viewing of a single program may aid vocabulary learning is by increasing encounters with low frequency vocabulary. Themes, characters and plots in successive episodes of the same program have much in common and low frequency vocabulary related to the program reoccurs more often across multiple episodes of one show than across random episodes (Rodgers & Webb, 2011; Webb, 2011). Words that are unknown to learners and are encountered repeatedly in context are more likely to be learned (Rodgers & Webb, 2011; Webb & Rodgers, 2009a; Webb, 2008). Overall, the combination of aural and visual input, the comprehension-focused nature of viewing television, and the serial nature of episodes of television are characteristics that suggest that viewing television may lead to incidental vocabulary acquisition.

The potential of full-length episodes of television as a source for incidental vocabulary acquisition informs the present study which was designed to examine (a) whether viewing English-language television leads to gains in vocabulary knowledge for EFL learners, (b) whether language learners with greater vocabulary knowledge learn more vocabulary than those with less vocabulary knowledge, and (c) how the frequency and range of occurrence of vocabulary in the episodes of television affects the rate at which vocabulary is learned.

### **3.1. Research investigating incidental vocabulary acquisition**

Numerous studies have examined L2<sup>4</sup> learners' incidental vocabulary acquisition through reading, listening, and viewing videos (R. Brown, Waring, & Donkaewbua, 2008; Day, Omura, & Hiramatsu, 1992; Horst et al., 1998; Horst, 2005; Pellicer-Sánchez & Schmitt, 2010; Vidal, 2003, 2011; Waring & Takaki, 2003; Webb, 2007, 2008). These studies are summarized in Figure 3.1 which provides the number of participants, the L1 and target language of the participants, the type of vocabulary test, a summary of the treatment conditions, the number of target words tested, and the mean vocabulary gains found in each study. Seven of these studies investigated language learners' incidental vocabulary acquisition through reading (Day et al., 1992; Horst et al., 1998; Horst, 2005; Pellicer-Sánchez & Schmitt, 2010; Waring & Takaki, 2003; Webb, 2007, 2008) and one (R. Brown et al., 2008) compared vocabulary acquisition from reading, listening, and reading and listening. The pair of studies by Vidal (2003, 2011) tested vocabulary acquisition through viewing videos of academic lectures with the latter study comparing vocabulary acquisition from video to vocabulary acquisition from reading. Other studies have compared incidental vocabulary learning from video and videos with captions (Baltova, 1999; Huang & Eskey, 1999; Hui, 2007; Markham et al., 2001; Markham, 1999; Sydorenko, 2010; Winke et al., 2010; Yuksel & Tanriverdi, 2009). The following survey of the most relevant studies focuses on examples of incidental vocabulary acquisition through different input modes as well as the amount of vocabulary learning that took place.

---

<sup>4</sup> There have been many studies that have investigated incidental vocabulary learning through reading, listening, and viewing videos in the L1 (e.g. Elley, 1989; Jenkins, Stein, & Wysocki, 1984; Oetting, Rice, & Swank, 1995; Rice & Woodsmall, 1988). While these studies are worth considering they are outside the scope of this thesis.

Figure 3.1 Summary of previous research investigating incidental vocabulary acquisition through reading, listening and viewing videos

| Study                           | Participants                              | L1       | Target Language | Vocabulary Test   | Treatment   | Words Tested    | Mean Vocabulary Gains   |
|---------------------------------|---|----------|-----------------|---|---|-----------------|---|
| Day, Omura and Hiramatsu (1992) | 191 high school & 397 university students | Japanese | English         | Meaning-focused multiple-choice test  | Participants read a 1,032-word story  | 17 target words | 5.2 words (high school)<br>9.3 words (university)   |
| Horst, Cobb and Meara (1998)    | 34 university students                    | n/a      | English         | Pre- and post-test versions of a meaning-focused multiple-choice test and a 13-item word association test   | Participants read a 21,232-word simplified novel.   | 45 target words | Multiple-choice test: 4.62 words<br>Word association test: 1.18 words                                       |
| Vidal (2003)                    | 116 university students                   | Spanish  | English         | A modified VKS in pre-, post- and delayed post-tests  | Participants viewed 3 academic lectures of 1,738, 1,837 and 1,812 words each                    | 36 target words | Pre-test rating of 1.41 for vocabulary knowledge increased to 16.38 on the post-test                        |
| Waring and Takaki (2003)        | 15 university students                    | Japanese | English         | Knowledge of target words was measured by a word form recognition test, a multiple-choice test measuring form and meaning, and a receptive translation test | Participants read a graded reader (5,872 words) with target words replaced with disguised forms | 25 target words | Word form recognition test: 15.3 words,<br>Multiple-choice test: 10.6 words,<br>Translation test: 4.6 words |

Figure 3.1 Summary of previous research investigating incidental vocabulary acquisition through reading, listening and viewing videos (cont.)

| Study                               | Participants           | L1 Language | Target Language | Vocabulary Test   | Treatment  | Words Tested   | Mean Vocabulary Gains   |
|-------------------------------------|------------------------|-------------|-----------------|---|--|--|---|
| Horst (2005)                        | 17 adult ESL learners  | Various     | English         | Yes/No pre-test of vocabulary from the extensive reading corpus and a Yes/No post-test of vocabulary from the books participants read | Participants read graded readers in an extensive reading program. The mean number of books read was 10.52 but the individual number of books read varied | 100 target words from the book corpus and 100 target words from the books read | 16.88 words (self-reported)   |
| Brown, Waring and Donkaewbua (1999) | 35 university students | Japanese    | English         | Target words were tested through a multiple-choice test and a translation test  | Participants in 3 groups read, read and listened to, or listened to three graded readers with target words replaced with disguised forms                 | 28 target words  | Reading: 12.54 words (multiple-choice), 4.10 words (translation), Reading-Listening: 13.31 words (multiple-choice), 4.39 words (translation), Listening: 8.20 words (multiple-choice), 0.56 words (translation) |

Figure 3.1 Summary of previous research investigating incidental vocabulary acquisition through reading, listening and viewing videos (cont.)

| Study                               | Participants            | L1 Language | Target Language | Vocabulary Test  | Treatment  | Words Tested    | Mean Vocabulary Gains  |
|-------------------------------------|-------------------------|-------------|-----------------|--|--|-----------------|--|
| Webb (2008)                         | 50 university students  | Japanese    | English         | Target words were tested through tests of recall of meaning, recognition of meaning, recall of form, and recognition of form                           | Participants read three sets of 10 sentences with each sentence containing 1 of 10 disguised forms. After the 1 <sup>st</sup> set of sentences, the 2 <sup>nd</sup> and 3 <sup>rd</sup> sets of sentences were more informative for the experimental group | 10 target words | Recall of meaning: M=1.31 words, Recognition of meaning: M=6.77 words, Recall of form: M=5.96 words, Recognition of form: M=8.00 words |
| Pellicer-Sánchez and Schmitt (2010) | 20 university students  | Spanish     | Nigerian        | Target words were tested via multiple-choice spelling and word class tests. Knowledge of meaning was tested by an interview and a multiple-choice test | Participants read a 67,000-word novel over a month and then were tested on knowledge of Nigerian words that occurred in the English-language story   | 34 target words | Spelling: 11.65 words, Word Class: 6.65 words, Receptive Meaning: 14.45 words, Productive Meaning: 4.80 words                          |
| Vidal (2011)                        | 230 university students | Spanish     | English         | Target words were tested by a modified VKS in pre-, post- and delayed post-tests   | Participants viewed or read 3 academic lectures of 1,738, 1,837 and 1,812 words each   | 36 target words | Viewing Group: Pre-test rating of 1.08 for vocabulary knowledge increased to 28.94 on the post-test<br>Reading Group: 1.82 to 42.67    |

### **3.1.1. Incidental vocabulary learning from reading and listening**

Incidental vocabulary acquisition has been shown to occur through reading (Day et al., 1992; Horst et al., 1998; Horst, 2005; Pellicer-Sánchez & Schmitt, 2010; Waring & Takaki, 2003; Webb, 2008) and reading and listening (R. Brown et al., 2008). These studies had differing methodologies making direct comparison of results impractical. However, the underlying trend from these studies is that incidental vocabulary learning has been shown to occur through reading.

In a study by Day, Omura and Hiramatsu (1992), participants in high school and university were tested on knowledge of 17 target words after reading a short story. The participants in high school had a mean score of 5.2 (out of 17). A corresponding control group, who only completed the vocabulary test, had a mean score of 4.1. Participants in university had a higher mean score of 9.3 words which was compared to their control group's mean score of 6.3 words. Both treatment groups had significantly higher numbers of words known than their corresponding control groups.

Horst, Cobb and Meara (1998) investigated incidental vocabulary learning by language learners reading a simplified novel. Forty-five target words were tested in pre- and post-tests through a meaning-focused multiple-choice test and a 13-item word association test. Participants had a pre-test mean of 21.64 on the meaning-focused test and a post-test average of 26.26. This represented a significant mean gain of 4.62 words. On the item recognition test, there was a significant difference between the pre-test mean ( $M=5.53$ ) and the post-test mean ( $M=6.71$ ).

Waring and Takaki (2003) measured knowledge of 25 disguised form target words appearing in a graded reader. Vocabulary knowledge was measured through three tests: Test 1 – a word form recognition test, Test 2 – a multiple-choice test measuring form and meaning, and Test 3 – a test of receptive recall of meaning. The vocabulary tests were administered immediately following reading the book and after two delay periods. On the immediate post-test, Test 1 had the highest mean score ( $M=15.3$  of the 25 target words), followed by Test 2 ( $M=10.6$ ), and then Test 3 ( $M=4.6$ ). There was considerable decay in knowledge of the vocabulary on both delayed post-tests.

In an examination of incidental vocabulary learning through extensive reading, Horst (2005) studied participants in an ESL environment who read graded readers across a six-week period. Participants' vocabulary gain was measured through 100-item pre- and post-tests. Statistical analysis indicated a significant increase from the pre- to the post-test. Participants

had a mean increase in knowledge of 6.59 words from the 50 items from the 1,001 to 2,000 most frequent words of English and a mean increase of 10.29 words from off-list items.

Webb (2008) examined how varying amounts of context affected incidental vocabulary learning from reading. In this study, participants in an experimental group and a control group completed a reading task where they read three sets of 10 sentences with each sentence containing 1 of 10 disguised form target words. While the first set of 10 sentences was the same for all participants, the second and third sets for the experimental group were more informative than those for the comparison group. The participants in the experimental group had significantly higher scores on a test of recall of meaning (Experimental Group:  $M=1.31$ , Comparison Group:  $M=0.13$ ) and a test of recognition of meaning (Experimental Group:  $M=6.77$ , Comparison Group:  $M=4.38$ ).

Pellicer-Sánchez and Schmitt (2010) studied the vocabulary gains through reading a novel that was in English but the 34 target words were chosen from naturally occurring Nigerian words. Knowledge of the target words was measured through a series of three tests and an interview. In the interview, knowledge of meaning was assessed by the participants reading a target word and then saying everything they knew about that word. The mean scores were 11.65 for the spelling test, 6.65 for the part of speech test, and 14.45 on the meaning recognition test. The mean score for the meaning recall test was 4.80.

One study examined incidental vocabulary acquisition from three input modes: reading, reading and listening, and listening-only. Brown, Waring and Donkaewbua (1999) had participants complete each of the treatments with one of three graded readers. Disguised forms were substituted for twenty-eight target words. There were two vocabulary tests measuring form and meaning that were administered immediately after the treatments and two delay periods. On the immediate multiple-choice test, the mean scores were: reading: 12.54, reading and listening: 13.31, and listening-only: 8.20. On the translation test the mean scores were: reading: 4.10, reading and listening: 4.39, and listening-only: 0.56. There were significant differences between the reading treatment and the listening treatment, and between the reading and listening treatment and the listening treatment on both post-tests. There was no significant difference between the reading treatment and reading and listening treatment. On the delayed post-tests there was little decay in the number of target words known for the multiple-choice test but considerably more decay for the translation test.

The results from these studies indicate the extent of incidental vocabulary acquisition possible through reading and point to the levels of vocabulary learning that may be possible through viewing videos. Some of the findings from the reading research have particular

relevance to the potential of incidental video learning through video. The first of these was that the more informative the context that language learners encountered vocabulary the more apt they were to learn it (Webb, 2008). This could be especially important in a context rich input mode like watching television. Another finding with particular relevance to learning through videos is that incidental vocabulary learning was found to occur through reading multiple texts (Horst, 2005). Reading books in an extensive reading program is similar to watching multiple episodes of television programs. One finding from the studies indicates that the rate of incidental vocabulary learning through video could be lower than for reading. Brown, Waring and Donkaewbua (2008) found that there was greater vocabulary acquisition through reading than there was through listening. However, the combination of aural input with the imagery in video may ameliorate this. Overall, previous research on incidental vocabulary acquisition through reading may indicate the levels of vocabulary learning that may take place through viewing episodes of television.

### **3.1.2. Incidental vocabulary learning from video**

A pair of studies explored incidental vocabulary acquisition through viewing videos featuring academic lectures. Vidal (2003, 2011) investigated the effects of viewing lectures on vocabulary acquisition. In the first study, students from a Spanish university viewed videos of three lectures on economic, sociocultural and environmental issues affecting tourism. For each lecture there were 12 target words, classed as either low frequency, academic, or technical and occurring 6, 5, 4, 3, 2, or 1 times in the specially-made videos. Vocabulary gain was measured by a modified Vocabulary Knowledge Scale (VKS) test prior to, immediately after, and after a one month delay. There was a significant difference between the pre- and post-test results indicating a positive effect of viewing lectures on vocabulary acquisition. Comparison of the post-test and delayed post-test results showed that participants retained between 43% and 54% of their vocabulary gains.

Vidal (2011) used the same lectures and measurement instrument to compare the effects of viewing the three lectures and reading three academic texts on incidental vocabulary acquisition. Three groups of participants viewed the lectures, read the academic texts, or only completed the vocabulary measures. The written texts were based on the information presented in the lectures. There was a significant difference between the mean scores of treatment groups with the Reading Group having the highest mean on both the immediate and the delayed post-test followed by the Viewing Group and the Control Group. A comparison of

the immediate post-test and the delayed post-test showed that the Reading Group retained an average of 46.9% of their vocabulary gain and the Viewing Group retained an average 50.4%.

The results from these two studies indicate that incidental vocabulary learning occurs through viewing academic lectures although it is less than through reading similar texts. The results are somewhat tempered, however, by the use of a modified version of the VKS to investigate vocabulary gain and the analysis of the results from this rating scale as if they were interval data. While the gains in vocabulary knowledge were attributed to encountering the words in the context of the lecture it may have been that the participants had been primed for the target words in the pre-test and encountering the words in the lecture may not have been the only impetus for the development of vocabulary knowledge. The ordinal data obtained from the modified VKS should not be added together and analyzed as if it were interval data. While the ability to generalize from these studies may be limited, the research points to a gain in vocabulary knowledge even when viewing a video of a lecture that has very little support from imagery. The results also indicate that while there is vocabulary learning from viewing video, it is significantly less than vocabulary learning from reading.

### **3.1.2.1. Incidental vocabulary learning from video as a comparison to vocabulary learning from video with captions**

Numerous studies (Baltova, 1999; Huang & Eskey, 1999; Hui, 2007; Markham et al., 2001; Markham, 1999; Sydorenko, 2010; Winke et al., 2010; Yuksel & Tanriverdi, 2009) have compared incidental vocabulary acquisition through viewing video with captions to viewing without captions. Study 5 in this thesis carries out a similar comparison and therefore examines the results and methodologies from these studies in detail. A full report on these studies is presented in Section 6.5.2. However, by focusing on the results from the viewing groups that did not have access to captions it is possible to see the vocabulary gains participants in previous research have obtained through viewing video. While these studies all have different methodologies which makes direct comparison difficult, it is possible to see the vocabulary gain trends that have been achieved. In these studies gains in knowledge of form and meaning, recognition of form, production of form, and growth in vocabulary knowledge as measured by the VKS were reported.

Most of the studies investigating vocabulary acquisition through viewing videos have measured increase in knowledge of meaning. The majority of these studies reported increased knowledge of meaning as raw gains or as the number of words learned through the course of the study. In one such study, Baltova (1999) investigated learning and retention of French

vocabulary through viewing a documentary. Twenty-nine participants viewed the video three times and then were tested on the meaning of 30 target words. A cloze test was administered immediately after the final viewing and again after two weeks. On the immediate test, the mean gain was 11.66 words and on the delayed test the mean score was 12.16 words. In Huang and Eskey's (1999) study, 15 participants viewed a 7-minute video twice and knowledge of target words believed to be unknown was measured via a meaning-focused multiple-choice test. Participants knew an average of 3.47 of the 8 target words following viewing. Fifteen participants in Markham, Peter and McCarthy's (2001) study viewed a short science video and completed a 10-item multiple-choice vocabulary test. Participants knew an average of 4.67 target words after viewing the documentary. Participants in a study by Hui (2007) viewed a video and were tested on knowledge of 10 target words. Participants were divided into two subgroups of high and low English proficiency. Knowledge of the target words was measured in three ways: recognition of spoken phonological form, production of orthographic form, and knowledge of meaning. High proficiency participants were able to recognize an average of 5.17 of the target words on the listening test. Low proficiency participants recognized a mean 4.37 target words. The high proficiency learners had a mean score of 4.37 on the orthographic test and the low proficiency learners had a mean score of 3.47. On the meaning focused tests, the high proficiency participants ( $M=4.07$ ) again had higher mean scores than the low proficiency participants ( $M=3.30$ ).

Two studies calculated the incidental vocabulary acquisition of their participants in terms of percent gain of target words determined to be unknown to the participants prior to viewing videos. Target words in both studies were tested in both their aural and written forms. In Sydorenko's (2010) study, nine participants viewed three short segments of a television program. On an L2 to L1 translation test, participants showed a 25% gain in knowledge of the meaning of target words presented in the written form and 18% of the vocabulary presented aurally. On a recognition test, the participants had gains of 63% of the written items and 69% of the items presented aurally. In Winke, Gass and Sydorenko's (2010) study, participants had gains in knowledge of 4% of target words tested after viewing three short documentaries twice. Vocabulary knowledge was measured via an L2 to L1 translation test.

There were two studies that measured aspects of vocabulary knowledge other than form and meaning. Markham (1999) examined acquisition of 100 target words through viewing two educational videos. Recognition of the aural form of target words occurring an average of 3 times per video was measured. The 33 participants correctly identified an average of 64.1% of the target words for Video 1 and the 24 participants viewing Video 2 had a mean score of

55.5%. In a study that used the VKS to measure development of vocabulary knowledge, Yuksel and Tanriverdi (2009) investigated incidental vocabulary learning from watching a portion of a television program. The 51 participants that viewed the videos without captions had a pre-test average on the 5-point rating scale of 1.74 ( $SD=0.37$ ). This mean score is between Level 1 (participants had never seen the word before) and Level 2 (participants recognized the form of the word but did not know its meaning) on the VKS. Following viewing the segment of an episode of a situation comedy, the post-test average was 2.57 ( $SD=0.33$ ). This mean response indicates that the participants believed that for most of the target words they either had seen the word before but did not know what it meant (Level 2) or they thought they recognized the word and thought they knew what it meant (Level 3). There was a significant difference between the pre- and post-test scores.

The results from these studies indicate that incidental vocabulary learning can occur through viewing videos. It may be difficult, however, to make a definitive estimate of the amount of vocabulary that can be learned through viewing videos because each study had different videos, different amounts of viewing time, and a different number of target words. What can be surmised from the results from the studies is an indication that language learners can incidentally gain knowledge of vocabulary through watching short videos.

### **3.1.3. How does a language learner's vocabulary knowledge affect incidental vocabulary acquisition?**

A learner's prior vocabulary knowledge might be a factor that influences the amount of incidental vocabulary learning they achieve. It is believed that more words are likely to be learned if other related words are known (Gass, 1999). The greater a learner's vocabulary proficiency, the fewer number of words in a text that should be unknown. Fewer unknown words means a learner can assign more attention to these unknown words in the course of processing the text. It stands to reason that more attention given to fewer words could increase the rate of acquisition. Another reason that prior vocabulary knowledge may affect incidental vocabulary acquisition is that to learn a previously unknown word a learner must rely on knowledge of the context the word appears in. Understanding that context and knowing the surrounding words is essential for learning to successfully take place (Huckin & Coady, 1999). Greater vocabulary knowledge would increase the likelihood of comprehending the context and knowing more words in a text.

There have been relatively few studies investigating the relationship between a learner's prior vocabulary knowledge and the amount of vocabulary they incidentally acquire but the

studies that have been conducted indicate a relationship between proficiency and learning. Day, Omura and Hiramatsu (1992) found that university students had superior incidental vocabulary gains from reading a short story compared to high school students. That the university students with more formal English language instruction, and presumably higher vocabulary knowledge, learned more vocabulary from the story suggests that there was a relationship between prior vocabulary knowledge and incidental vocabulary learning. Oetting, Rice and Swank (1995) examined L1 vocabulary learning from viewing a video. The participants in the study were 6- to 8-year-old children who were either normally developed or had specific language impairments. While both groups were found to have significant vocabulary gains, the amount of gain was related to age and impairment. The fact that the language impaired participants and younger participants made the smallest gains indicates that vocabulary proficiency is related to vocabulary gain. Horst, Cobb and Meara (1998) found medium correlations between language learners' gains from reading a simplified novel and their scores on the 2,000- and 5,000-word level Vocabulary Levels Test. Zahar, Cobb and Spada (2001) investigated the incidental vocabulary acquisition of middle-school ESL learners through reading. Vocabulary proficiency was determined by the participants' results on five levels of the VLT. Learners with higher prior knowledge had greater gains from the pre- to post-test except for learners with the highest level of vocabulary proficiency. This was attributed to a ceiling effect for these high proficiency learners. Past research indicating that vocabulary proficiency influences incidental vocabulary acquisition and the lack of research investigating this with vocabulary learning through viewing videos suggests that further research is warranted.

#### **3.1.4. How is incidental vocabulary acquisition affected by the frequency and range of occurrence in text?**

Incidental vocabulary acquisition is generally accepted to be a gradual process where gains in knowledge are a function of repeated encounters over time (Ellis, 2002; Nation, 2001; Schmitt, 2008). The frequency at which words are encountered in texts and the propensity for these words to be learned is of particular interest to researchers. Studies have repeatedly indicated that the more a language learner encounters a word the more likely they are to gain knowledge of it. However, over the years no definite number of encounters that can guarantee vocabulary learning has been agreed upon. This indicates that different number of exposures may be necessary for different types of vocabulary knowledge and from different modes of input (Laufer & Ravenhorst-Kalovski, 2010). Previous research has

generally employed different testing procedures and different treatments making it challenging to compare the number of encounters necessary for vocabulary learning from one study to another. Generally however, the results of prior research give a general idea of the trend of improved acquisition with more encounters.

Research investigating the number of encounters necessary for vocabulary learning to take place has mostly occurred in the vocabulary acquisition through reading context. In the previously described study, Horst, Cobb and Meara (1998) also examined the correlation between the number of times words occurred in their treatment book and the target words' relative gain by the participants. Target words in their study occurred 2 to 17 times and a correlation of 0.49 was found between rate of occurrence and relative gain. They concluded that with eight or more repetitions of a target word in a text vocabulary learning was likely to occur.

Other studies have found different numbers of encounters lead to acquisition of vocabulary. Saragi, Nation and Meister (1978) found considerably more vocabulary gain when participants encountered words 6 or more times. Nagy, Herman and Anderson (1985) found that a single exposure to a word provided enough information to facilitate acquisition but the rate at which that happens was quite low. Rott (1999) found that 6 exposures was better for vocabulary learning than 2 or 4 exposures. Waring and Takaki (2003) found that there was not a clear number of encounters that led to higher acquisition rates for language learners reading a graded reader but found a general increase in vocabulary acquisition as exposure rates increased. Still, the authors concluded that for many words, 20 to 30 exposures in context might be necessary for learning to take place. In an extensive reading study that examined a single French-language learner's vocabulary acquisition, Pigada and Schmitt (2006) did not find a single exposure figure that indicated acquisition of meaning but found that over 10 exposures, there was an increase in acquisition rate. There were, however, words that were still not acquired with over 20 exposures. Webb (2007) found that 10 encounters led to gains in vocabulary knowledge measured across a number of components of vocabulary knowledge. Pellicer-Sánchez and Schmitt (2010) found that learning began with 5 to 8 occurrences, and increased with 10 to 17 exposures.

In Brown, Waring and Donkaewbua's (2008) study of vocabulary learning from reading, reading and listening, and listening to graded readers, they investigated the retention of 28 words that occurred 2 to 20 times and were grouped in four frequency bands. For the reading treatment and the reading and listening treatment, the more frequently words were encountered, the better they were learnt. Unfortunately for the listening treatment a

relationship between frequency and vocabulary acquisition was not apparent which was attributed to floor effects. The authors suggest that the findings support previous research indicating that the number of encounters necessary for long-term retention through reading and listening is likely to be higher than 7 to 9 times.

Two studies investigated frequency and vocabulary learning through viewing videos. Vidal (2003) saw learning occur at just one encounter but generally found a linear rise in vocabulary acquisition as a word was encountered (encounters ranged from 1 to 6 times). Similarly, Vidal (2011) found that vocabulary acquisition generally increased with more repetitions but that frequency of occurrence was a better indicator of acquisition in the reading treatment than in the viewing videos treatment. Participants in the viewing videos treatment needed 5 or 6 encounters for vocabulary acquisition to become substantial.

As shown, previous research into the relationship between the frequency of vocabulary occurrence and acquisition has produced a range of results for the number of encounters necessary for learning to take place. Across the reading, listening, and viewing videos modes the trend seems to be incidental vocabulary acquisition is more likely to take place when a word is encountered more often. It appears, however, that more encounters are necessary for incidental vocabulary acquisition to take place from listening or from viewing than through reading.

In language learning situations like extensive reading, where vocabulary may be encountered in not just one book but in an array of books, another factor that may affect incidental vocabulary learning is the range of texts in which the vocabulary is encountered. That is, assuming a constant number of exposures, is a language learner more apt to learn a word that has appeared in a wider or narrower range of texts? There appears to be no research examining how vocabulary acquisition is affected by the range of texts a target word appears in. While there have been studies (R. Brown et al., 2008; Horst, 2005; Kweon & Kim, 2008; Pigada & Schmitt, 2006) that have examined vocabulary acquisition through reading a range of texts, the effect of range of occurrence on vocabulary acquisition was not investigated. Investigating the range of vocabulary occurrence may be valuable when considering vocabulary learning in programs where words might be learned through repeated exposures but words learned may not have appeared in all the texts a language learner reads, listens to, or views.

### **3.1.5. Implications for previous research**

Overall, research examining incidental vocabulary acquisition through reading and listening (R. Brown et al., 2008; Day et al., 1992; Horst et al., 1998; Horst, 2005; Pellicer-Sánchez & Schmitt, 2010; Waring & Takaki, 2003; Webb, 2007, 2008) and studies examining incidental vocabulary acquisition through viewing videos (Baltova, 1999; Huang & Eskey, 1999; Hui, 2007; Markham et al., 2001; Markham, 1999; Sydorenko, 2010; Vidal, 2003, 2011; Winke et al., 2010; Yuksel & Tanriverdi, 2009) indicate the potential for language learners to acquire vocabulary incidentally. However, the amount of learning that occurred differed by mode of input and by study. Overall, the studies indicated small but significant vocabulary gains regardless of the input mode. Language learners at a range of proficiency levels from a range of L1 backgrounds experienced incidental vocabulary gains. The frequency of occurrence of the vocabulary in the input texts was also shown to be related to vocabulary acquisition but the number of encounters needed for vocabulary learning to occur is likely situation-specific.

With regard to vocabulary learning through viewing videos, the following types of videos served as the input in previous studies: segments of a comedy series, documentaries, academic lectures, educational videos, and segments of a movie. All of these forms of video were shown to promote incidental vocabulary acquisition. These video types, however, are not a representative sample of what a language learner might choose to learn English from over a prolonged period of time. Because an episode of an American comedy series is 22 minutes and a drama is 44 minutes and the length of the individual videos in the studies cited here averaged much less than these, typical viewing would likely involve much greater exposure to language through television. The video types and running times of these studies are shown in Table 3.1. This highlights the need for further research on incidental vocabulary learning from viewing multiple videos that are longer, promote multiple encounters with vocabulary, and are characteristic of typical television viewing.

Table 3.1 Video types and running times of previous studies of vocabulary acquisition from video

| Study                              | Video Type                                 | Video Length      |
|------------------------------------|--|-------------------|
| Baltova (1999)                     | 1 documentary                              | 7.5 minutes       |
| Markham (1999)                     | 2 excerpts from educational programs       | 12 and 13 minutes |
| Huang and Eskey (1999)             | 1 episode from an educational video series | 21 minutes        |
| Markham, Peter and McCarthy (2001) | 1 documentary                              | 7 minutes         |
| Vidal (2003)                       | 3 academic lectures                        | 14 to 15 minutes  |
| Hui (2007)                         | 1 documentary                              | 16 minutes        |
| Yuksel and Tanriverdi (2009)       | 1 segment of a situation comedy            | 9.2 minutes       |
| Sydorenko (2010)                   | 3 segments from a comedy series            | 2 to 3 minutes    |
| Winke, Gass and Sydorenko (2010)   | 3 documentaries                            | 3 to 5 minutes    |
| Vidal (2003)                       | 3 academic lectures                        | 14 to 15 minutes  |
|                                    | Average                                    | 10.1 minutes      |

### 3.2. Research questions

Study 2 was designed to answer the following research questions:

1. Does viewing English-language television lead to increased word knowledge for Japanese EFL learners?
2. Do English-language learners with greater vocabulary knowledge incidentally learn more vocabulary through television than learners with less vocabulary knowledge?
3. Are unknown words that occur more frequently in television programs more likely to be learned?
4. Are unknown words that occur across a greater range of episodes of a television program more likely to be learned?

### 3.3. Participants

There were 229 male and 60 female volunteer participants in their first and second year of university from nine separate classes in this study. All of the participants had studied English for a minimum of seven years. The English proficiency level of the participants can be considered pre-intermediate to intermediate within the context of the university. The classes were all taught by the researcher. The participants described here make up the Experimental

Group for this study. Details on the number of participants in each class and their university major are shown in Table 3.2.

Table 3.2 Experimental Group participants in Study 2 prior to exclusions

| Class | Major       | Year of Study | Gender |    | Number of Participants |
|-------|-------------|---------------|--------|----|------------------------|
|       |             |               | M      | F  |                        |
| 1     | Commerce    | 2             | 32     | 4  | 36                     |
| 2     | Business    | 1             | 27     | 9  | 36                     |
| 3     | Business    | 1             | 28     | 8  | 36                     |
| 4     | Engineering | 2             | 33     | 0  | 33                     |
| 5     | Law         | 2             | 31     | 7  | 38                     |
| 6     | Commerce    | 1             | 14     | 10 | 24                     |
| 7     | Law         | 2             | 26     | 11 | 37                     |
| 8     | Commerce    | 1             | 14     | 10 | 24                     |
| 9     | Engineering | 1             | 24     | 1  | 25                     |
| Total |             |               | 229    | 60 | 289                    |

### 3.3.1. Control Group

To serve as a Control Group, there were a further 60 male and 26 female volunteer participants in their first and second year of study in three separate classes at the same university. The classes were all taught by a single instructor (not the researcher). All of the participants had studied English for a minimum of seven years. Like the participants in the Experimental Group, the English proficiency level of the Control Group participants was considered pre-intermediate to intermediate within the context of the university. The number of participants in each class and their university major are presented in Table 3.3.

Table 3.3 Control Group participants in Study 2 prior to exclusions

| Class | Major       | Year of Study | Gender |    | Number of Participants |
|-------|-------------|---------------|--------|----|------------------------|
|       |             |               | M      | F  |                        |
| 1     | Law         | 2             | 15     | 7  | 22                     |
| 2     | Commerce    | 2             | 23     | 17 | 40                     |
| 3     | Engineering | 1             | 22     | 2  | 24                     |
| Total |             |               | 60     | 26 | 86                     |

### 3.3.2. Human ethics requirements

In accordance with human ethics requirements, all participants in Study 2 received a detailed explanation of the research, were given information sheets, and signed a written consent form.

### 3.3.3. Exclusion of Experimental Group participants

One hundred and two participants were excluded from the Experimental Group from Study 2. Participants were excluded if they were absent from any of the treatment and testing sessions. They were excluded if they missed viewing an episode of *Chuck* because they would not have the opportunity to encounter episode-specific vocabulary and would have reduced encounters with some target words. These exclusions left 187 participants for the analysis of incidental vocabulary learning. The exclusion process is illustrated in Table 3.4 which shows the starting sizes of the nine classes and the number of participants excluded from each.

Table 3.4 Participants excluded from the incidental vocabulary acquisition analysis of Study 2

| Class | Starting Size | Number of Exclusions | Final Number of Participants |
|-------|---------------|----------------------|------------------------------|
| 1     | 36            | 17                   | 19                           |
| 2     | 36            | 7                    | 29                           |
| 3     | 36            | 14                   | 22                           |
| 4     | 33            | 8                    | 25                           |
| 5     | 38            | 17                   | 21                           |
| 6     | 24            | 3                    | 21                           |
| 7     | 37            | 23                   | 14                           |
| 8     | 24            | 8                    | 16                           |
| 9     | 25            | 5                    | 20                           |
| Total | 289           | 102                  | 187                          |

### 3.3.4. Exclusion of Control Group participants

From the Control Group, 13 participants were excluded from the study. The basis for their exclusion was based on the exclusion procedure for the Experimental Group. Participants in the Control Group were excluded from the analysis if they missed two or more teaching sessions between the vocabulary pre- and post-tests. There were no specific activities these participants missed but rather it was assumed that repeated absenteeism indicated a lack of effort toward their coursework and their results would not reflect the potential vocabulary learning possible in their English class. The exclusion process is illustrated in Table 3.5 which shows the starting sizes of the three classes and the number of participants excluded from each.

Table 3.5 Control Group participants excluded from the incidental vocabulary acquisition analysis

| Class | Starting Size | Number of Exclusions | Final Number of Participants |
|-------|---------------|----------------------|------------------------------|
| 1     | 22            | 5                    | 17                           |
| 2     | 40            | 5                    | 35                           |
| 3     | 24            | 3                    | 21                           |
| Total | 86            | 13                   | 73                           |

### 3.4. Procedure

The overall schedule and the in-class procedures for Study 2 were identical to those explained in the procedure section of Study 1 (Section 2.4). The experimental procedure for Study 2 was repeated with nine different university classes and three additional university classes that served as the Control Group. These courses took place in the first semester of the Japanese university school year which ran from April 2010 through July 2010.

#### 3.4.1. Overall schedule

Study 2 took place over thirteen teaching sessions in one university semester. Generally, each teaching session was separated by a week but, because of national and school holidays, there were instances where the teaching sessions were separated by two weeks. The schedule for this study is shown in Figure 3.2.

Figure 3.2 Research schedule for Study 2

| Teaching Session | Study 2 Schedule   |                 |   |
|------------------|--|-----------------|---|
|                  | Viewing Group 1  | Viewing Group 2 | Control Group   |
| 1                | Human Ethics Committee Ethics Form & Vocabulary Levels Tests (2000, 3000, & 5000)  |                 |   |
| 2                | <i>Chuck</i> -specific 60-Item Tough Vocabulary Pre-Test, Television Viewing Practice & <i>Chuck</i> -specific 60-Item Sensitive Vocabulary Pre-Test |                 | Human Ethics Committee Ethics Form, <i>Chuck</i> -specific 60-Item Tough Vocabulary Pre-Test & <i>Chuck</i> -specific 60-Item Sensitive Vocabulary Pre-Test |
| 3                | Episode A1   | Episode B1      |   |
| 4                | Episode 1  |                 |   |
| 5                | Episode 2  |                 |   |
| 5                | Episode 2  |                 |   |
| 6                | Episode 3  |                 |   |
| 7                | Episode 4  |                 |   |
| 8                | Episode 5  |                 |   |
| 9                | Episode 6  |                 |   |
| 10               | Episode 7  |                 |   |
| 11               | Episode 8  |                 |   |
| 12               | Episode B2   | Episode A2      |   |
| 13               | <i>Chuck</i> -specific 60-Item Tough Vocabulary Post-Test, Final Attitude Survey & <i>Chuck</i> -specific 60-Item Sensitive Vocabulary Post-Test     |                 | <i>Chuck</i> -specific 60-Item Tough Vocabulary Post-Test & <i>Chuck</i> -specific 60-Item Sensitive Vocabulary Post-Test                                   |

### 3.4.2. Viewing order

In Study 2, there were two viewing groups. Participants viewed either Episode A1 or Episode B1 first, followed by Episodes 1 to 8, then finally Episode B2 or Episode A2. This was because participants in this study were also participants in Study 1 which investigated comprehension of the episodes of *Chuck*. Prior to the analyses in Study 1, it was unknown if Episode A or Episode B was more difficult than the other. This might have been a factor in analyzing Initial to Final Episode comprehension gains. Analysis of the comprehension scores in Study 1 revealed no significant difference between the comprehension scores of the participants who viewed Episode A1 or Episode B1 first. Because participants in Study 2 viewed all the same episodes and Episodes A1 and B1 were found to be statistically equivalent, the results of all participants were analyzed together.

### 3.4.3. Control Group schedule

The Control Group had only two teaching sessions that were part of Study 2. In Teaching Session 2 and 13, the participants completed the *Chuck*-specific 60-item Vocabulary Pre- and Post-Tests. There were some minor differences in the administration procedures for these two teaching sessions compared to those of the Experimental Group. For the Control Group, the Human Ethics Committee Ethics Form and explanation was completed prior to the Tough Pre-Test in Teaching Session 2. The Tough and Sensitive tests were still separated by 30 to 40 minutes. The content in this time was left to the discretion of the instructor of the Control Group. The instructor chose to complete portions of the classes' textbooks.

### 3.5. Materials

The materials used in Study 2 were the same as those used in Study 1 but included the *Chuck*-specific vocabulary tests before and after viewing the 10 episodes of the television series. The materials from Study 1 are described in full in Section 2.7.

#### 3.5.1. Target vocabulary

To test for incidental vocabulary learning through viewing television, two vocabulary tests were created. The target words for these tests were chosen using the results from the analysis of the vocabulary in the 10 episodes of *Chuck* detailed in Section 2.7.2 of Study 1. In this analysis, the vocabulary in the episodes was sorted by occurrence in the fourteen 1,000-word frequency lists based on the BNC (Nation, 2006).

Words occurring in the episodes of *Chuck* from the third through fourteenth word lists were considered to be potential target words. Words that were from the first and second 1,000-word lists were discounted as it was thought that there was a good chance that the majority of the participants in this study would know or have encountered most of these words. Words occurring in the episodes from the third through fourteenth word family lists were then sorted by frequency. Word families occurring five or more times throughout the 10 episodes were considered for inclusion as target vocabulary. This produced a list of 96 word families. From this list proper nouns were excluded. These included proper nouns such as *Harry*, *Tang* and *Victor*. Next, words whose translations in Japanese were loanwords were excluded. An example of this was *helicopter* which translates as ヘリコプター (*herikoputā*). Words like the interjection, *whoa*, were also excluded because they would be too ambiguous when translated. This process left 58 word families to which two frequently occurring words from the *Not in the Lists* category (words less frequent than the 14,000 most frequently

occurring word families) of the analysis were added. The most commonly occurring word type from the word families in this list of 60 were used as the vocabulary test items. See Table 3.6 for all the target words and their frequency and range of occurrence in the episodes.

Table 3.6 Vocabulary test items and their range and frequency in the 10 episodes of *Chuck*

| Word Family  | Word List | Range | Family Frequency | Word Family          | Word List | Range | Family Frequency |
|--------------|-----------|-------|------------------|----------------------|-----------|-------|------------------|
| spy          | 5         | 10    | 54               | flight               | 3         | 2     | 7                |
| buddy        | 6         | 9     | 43               | nuclear              | 3         | 2     | 7                |
| steal        | 3         | 8     | 25               | sweat                | 3         | 5     | 7                |
| flash        | 3         | 7     | 22               | asset                | 4         | 3     | 6                |
| mission      | 3         | 8     | 21               | auction              | 5         | 1     | 6                |
| professor    | 4         | 1     | 17               | defect               | 5         | 1     | 6                |
| bug          | 3         | 4     | 17               | anniversary          | 3         | 3     | 6                |
| poison       | 3         | 2     | 17               | cheat                | 3         | 2     | 6                |
| kidding      | 3         | 8     | 15               | congratulate         | 3         | 4     | 6                |
| costume      | 4         | 2     | 13               | fry                  | 3         | 3     | 6                |
| receiver     | 4         | 1     | 12               | ruin                 | 3         | 4     | 6                |
| worm         | 5         | 1     | 12               | bullet               | 4         | 4     | 5                |
| freak        | 5         | 5     | 11               | former               | 4         | 3     | 5                |
| shrimp       | 7         | 1     | 11               | hazard               | 4         | 2     | 5                |
| undercover   | 7         | 5     | 11               | intimate             | 4         | 3     | 5                |
| compromise   | 3         | 7     | 11               | pants                | 4         | 2     | 5                |
| idiot        | 3         | 6     | 11               | cute                 | 5         | 3     | 5                |
| remote       | 3         | 4     | 11               | hug                  | 5         | 5     | 5                |
| surveillance | 8         | 7     | 10               | violate              | 8         | 2     | 5                |
| weapon       | 3         | 5     | 10               | assassin             | 10        | 3     | 5                |
| sizzle       | 7         | 1     | 9                | breathe              | 3         | 2     | 5                |
| rescue       | 3         | 2     | 9                | desert               | 3         | 1     | 5                |
| cage         | 4         | 2     | 8                | hero                 | 3         | 5     | 5                |
| fake         | 5         | 3     | 8                | inspire              | 3         | 5     | 5                |
| cop          | 3         | 4     | 8                | install              | 3         | 4     | 5                |
| torture      | 5         | 4     | 7                | interrupt            | 3         | 5     | 5                |
| antidote     | 10        | 1     | 7                | mystery              | 3         | 1     | 5                |
| buck         | 3         | 6     | 7                | nervous              | 3         | 2     | 5                |
| crazy        | 3         | 6     | 7                | nerd <sup>†</sup>    | X         | 7     | 17               |
| extract      | 3         | 3     | 7                | crisper <sup>†</sup> | X         | 1     | 7                |

Note. † Target words selected from the *Not in the Lists* category from the RANGE results; Range refers to the number of episodes that a word family occurred in.

### 3.5.2. Vocabulary test design

The tests of incidental vocabulary learning in Study 2 were based on the item creation procedures outlined in a study by Nagy, Herman, and Anderson (1985). To assess vocabulary learning from a reading task, they created a multiple-choice test where each target word was tested at three levels of difficulty. Similarity in meaning between the target word and the

distractors was the basis for the different levels of difficulty. Distractors at the highest level of difficulty had meanings or concepts closely associated with the target word, while the distractors at the middle level of difficulty were mostly the same part of speech but semantically varied. The distractors at the lowest difficulty level were chosen to be dissimilar in both meaning and part of speech. Instead of separate tests for each level of difficulty, the three types of multiple-choice items at different sensitivity levels were administered in one test. Each multiple-choice item had six options. One option was the correct answer (key) while '*don't know*' was always the final option. The remaining options were filled by three different types of distractors. One distractor type was the correct answer for another target word. Another distractor was the meaning of a non-target word encountered in the reading text. The final type of distractor was a word or a definition of a concept closely related to or similar to the target word's meaning but which did not occur in the text. These key concepts of including distractors at different difficulty levels, distractors that were the same or different parts of speech, distractors that were non-target words that occurred in the text, and distractors that were the answers to other target words informed the method by which the vocabulary tests in Study 2 were created.

While individual items were based on aspects of Nagy, Herman and Anderson's design, Study 2 used two different vocabulary tests to measure gains in vocabulary knowledge from viewing episodes of *Chuck*. The two tests were designed at different levels of sensitivity as outlined by Nation and Webb (2011). While at different sensitivities, each test was still designed to measure the same aspect of word knowledge: form and meaning. The first test is referred to as the Tough Test and was less sensitive to knowledge of word meaning and designed to be the more difficult test. The Sensitive Test, which is considered the easier of the two tests, was designed to be more sensitive to gains in vocabulary knowledge. The different levels of sensitivity were achieved through item design. On the Tough Test, the multiple-choice items contained distractors that shared aspects of form or meaning with the correct answer. These were considered more difficult to discern from the correct option. On the Sensitive Test, the multiple-choice items contained distractors that did not share aspects of form and meaning with the target word. The Tough Test was administered before the Sensitive Test to avoid having the easier items on the Sensitive Test provide clues to the answers of items on the following test.

### 3.5.3. Tough Pre-Test

Items measuring knowledge of the *Chuck*-specific vocabulary on the 60-item Tough Test consisted of a stem, the key, five distractors, and an ‘*I don’t know this word.*’ option. Similar to Nagy, Herman, and Anderson’s (1985) study, the ‘*I don’t know this word.*’ option was included to reduce the amount of guessing. Two of the five distractors for each item were semantically-related words that were taken from the same or a lower frequency BNC word list as that of the target word. Each of these distractors was chosen to be the same part of speech (P.O.S.) as the target word. Three methods were used to choose suitable semantically-related distractors. The first method was through the use of a thesaurus (Kipfer, 2003). The entries in the thesaurus for certain target words were analyzed with the RANGE software to see if they were in the same or a lower frequency word frequency list as the target word. This excluded many possible distractors. When a potential distractor was found, it was translated to Japanese. If the possible distractor was too similar in Japanese meaning, it was excluded. This meant that the words chosen as distractors had to be close in meaning but not so close as to be the same in Japanese. An example of a distractor chosen through this procedure is Item #3 on the pre-test. The target word for this item is *flight* and the distractor from the thesaurus method was *departure*.

The second method for choosing distractors involved the researcher reading through the BNC word list that the target word was from and looking for words believed to be semantically related. If there were no suitable words in the same word list as the target word then the next less frequent word list was examined. An example of a distractor created via this method is Item #6 on the pre-test for the target word *professor*. The distractor *pastor* was chosen from a less frequent word list and *professor* and *pastor* are both occupations.

The third process for selecting a distractor involved searching for words that had both a semantic and orthographic relationship with the Japanese translation of the target word. In conjunction with the people who did the translation for Study 1, words in Japanese that had similar or the same Chinese characters as the translation of the target word were identified. Japanese words are often made up of two or more of these logographic units (kanji). Japanese words that contained the same kanji as the translation of the target word, had a similar meaning, and whose English translation was from a suitable low frequency word list were used as distractors. In Item #10 from the pre-test, the distractor *ornament* (装飾品) and the target word *costume* (衣装) share a common kanji (装) and are semantically related.

There was a third type of distractor for each item on the Tough Test. This was a distractor for the same target word on the Sensitive Test. The distractor was quasi-randomly chosen from the three distractors from that test. These distractors were a different P.O.S from the target word.

The fourth type of distractor on the Tough Test was a key from another item. These distractors were quasi-randomly chosen but care was taken so that they were not semantically related to the target word and not from the same item that supplied this distractor on the Sensitive Test. The fifth type of distractor on the Tough Test was selected from the words that occurred less than five times in the episodes of *Chuck* viewed in this study. These distractors were chosen at random but if a selection was semantically similar to the target word, another random choice was used. Figure 3.3 shows an example and explanation of the options for Item #1 on the Tough Test.

Figure 3.3 Explanation and examples of the options for items on the Tough Test using the target word ‘spy’ (Item #1)

| Option Description   | Test Option | English Translation          |
|--|-------------|------------------------------|
| <b>Key</b>   | 密偵          | spy                          |
| <b>Semantically Related/Same P.O.S. Distractor</b><br>• chosen from low frequency word family lists  | 消防士         | fireman                      |
| <b>Semantically Related/Same P.O.S. Distractor</b><br>• chosen from low frequency word family lists  | 仲介業者        | broker                       |
| <b>Sensitive Test Distractor</b><br>• quasi-randomly chosen and different P.O.S.   | やめる         | quit                         |
| <b>Key from Different Item Distractor</b><br>• quasi-randomly chosen, not semantically related, and different from Sensitive Test                        | ジュージュー      | sizzling ( <i>Item #21</i> ) |
| <b>Non-target Word Distractor</b><br>• quasi-randomly chosen from vocabulary encountered less than 5 times in all episodes, and not semantically related | 降伏する        | surrender                    |
| <b>‘I don’t know this word.’ Distractor</b>  | この単語を知らない。  | I don’t know this word.      |

To determine the order that the options appeared in each item, first the position of the key was established. This was done quasi-randomly with the other distractors then inserted around it. The first distractor inserted was the distractor that was a key from another test item. Starting with the first item, this distractor was inserted into option position A. For the subsequent test items, this distractor was rotated, in order, through the option letters (A to F). Next, the two semantically related distractors were inserted into the available options on either side of the key. When the key was in option position E, the second distractor was put in the first available location starting with option position A. Then, the distractor that was a non-target word occurring in the episodes of *Chuck* was inserted into the first available option position following the key. The distractor from the Sensitive Test from the same target word was chosen by rotating through the noun, verb, adjective, and adverb options in that order. This distractor was put into the last remaining option position. It was hoped that this repeated occurrence of item distractors and keys would make it more difficult for participants to pick up the association between the target word and its definition from the test alone. This procedure was repeated for each item on the Tough Test. Figure 3.4 shows an example (Item #1) from the Tough Test and its stem, key, five distractors, and ‘*I don’t know this word.*’ option.

Figure 3.4 Item #1 from the Tough Pre-Test

|    |            |
|----|------------|
| 1. | <b>spy</b> |
| A  | ジュージュー     |
| B  | 密偵         |
| C  | 消防士        |
| D  | やめる        |
| E  | 降伏する       |
| F  | 仲介業者       |
| G  | この単語を知らない。 |

#### 3.5.4. Sensitive Pre-Test

Items testing knowledge of the *Chuck*-specific vocabulary on the 60-item Sensitive Test consisted of a stem, the key, five distractors, and an ‘*I don’t know this word.*’ option. Three of the five distractors on the Sensitive Test were chosen from the same BNC word list as the target word. They were, however, a different part of speech from, and not semantically related to the stem. These three distractors were chosen quasi-randomly by first selecting random numbers between 1 and 1,000. These numbers corresponded to the position of a headword in the appropriate BNC word list. If the headword at that position was semantically related to the

target word the next word in the list was used. If the headword, or any of its associated word family, was not the correct part of speech the successive word in the word list was examined until a suitable distractor was found.

The fourth distractor for each item was chosen from words occurring in the episodes but not occurring frequently enough to be considered as a target word. This meant that over the course of the 10 episodes used in this study, these words were encountered less than five times. From a list of these non-target words, distractors were chosen quasi-randomly. If the word selected was close semantically to the target word, the next word on the list was used. The fifth type of distractor used on the Sensitive Test was a key from a different item on the test. This item was quasi-randomly chosen. If, however, the key from the different item was semantically related, or if the item was too proximal then a different item was chosen. It was hoped that this repeated occurrence of item keys would make it more difficult for participants to pick up the association between the target word and its definition from the test alone. Figure 3.5 shows an example and explanation of the options for Item #1 on the Sensitive Test with the Japanese options and their corresponding English translations.

Figure 3.5 Explanation and examples of the options for items on the Sensitive Test using the target word ‘spy’ (Item #1)

| Option Description   | Test Option | English Translation           |
|--|-------------|-------------------------------|
| <b>Key</b>   | 密偵          | spy                           |
| <b>Different P.O.S. Distractor</b><br>• from same word family list, and not semantically related   | 先天的な        | inherent ( <i>adjective</i> ) |
| <b>Different P.O.S. Distractor</b><br>• from same word family list and not semantically related  | 謙虚に         | humbly ( <i>adverb</i> )      |
| <b>Different P.O.S. Distractor</b><br>• from same word family list and not semantically related  | やめる         | quit ( <i>verb</i> )          |
| <b>Key from Different Item Distractor</b><br>• quasi-randomly chosen, not semantically related, and different from Tough Test                            | かわいい        | cute ( <i>item 47</i> )       |
| <b>Non-target Word Distractor</b><br>• quasi-randomly chosen from vocabulary encountered less than 5 times in all episodes, and not semantically related | 焼却炉         | incinerator                   |
| <b>‘I don’t know this word.’ Distractor</b>  | この単語を知らない。  | I don’t know this word.       |

The same procedure for each item was performed to determine the positions of the options. First, the position of the key was determined randomly. The first distractor inserted was the distractor that was a key from another test item. Starting with the first item, this distractor was inserted into option position A. For the subsequent test items, this distractor was rotated, in order, through the option letters (A to F). If the designated space was filled by the key then the next available space was used. The distractors that were determined by their part of speech were inserted starting with the first one being placed in the first option position available after the key. The other two distractors in this category were inserted in the subsequent open positions. The order these distractors were inserted was noun, verb, adjective, and adverb. The distractor that was a non-target word that occurred in the episodes was inserted into the last remaining option position. Figure 3.6 shows an example (Item #1) from the Sensitive Test and its stem, key, five distractors, and ‘*I don’t know this word.*’ option.

Figure 3.6 Item #1 from the Sensitive Pre-Test

|    |            |
|----|------------|
| 1. | <b>spy</b> |
| A  | 息をする       |
| B  | 先天的な       |
| C  | 謙虚に        |
| D  | 素晴らしい      |
| E  | 密偵         |
| F  | やめる        |
| G  | この単語を知らない。 |

### 3.5.5. Translation to Japanese

To produce the final version of the vocabulary tests, the target words and the distractors were translated into Japanese. This involved a pair of translators working separately to produce a list of translations. To provide context for the target words, transcripts of *Chuck* were given to the translators to guarantee the most suitable translation. The lists of translations were compared and any translations that were different were discussed until a consensus on the best translation was reached. Particular attention was paid to ensure that the part of speech of the English word and the Japanese translation was the same. In the course of making the vocabulary tests, some translating of the target words and corresponding distractors had been done by the researcher. The agreed upon translations were substituted into the test framework to produce preliminary versions of the tests. Finally, both the tough and sensitive versions of these vocabulary tests were trialed with Japanese colleagues of the researcher who provided feedback on the physical structure of the test, the translations used,

and the choice of distractors. Based on their advice, small modifications were made to finalize the tests.

### **3.5.6. Aural presentation of the stem**

When participants in Study 2 completed the Tough and Sensitive tests, the stems for each item were presented in both their written and aural forms. The participants were instructed to answer the items on these tests in coordination with an audio track that accompanied each test. The aural form of each stem was recorded by the researcher. Each stem was spoken twice within a five second span followed by 10 seconds of silence. The fifteen seconds provided to answer each item was tested in the pilot study and was found to be enough time for participants to answer items. Time was also given at the end of each test for participants to check their answers and complete any remaining items. The aural form was included to make sure those participants who acquired knowledge of a target word through watching and listening to the television program had the chance to activate that knowledge to answer the item on the post-tests.

### **3.5.7. Vocabulary post-tests**

The post-test versions of the Tough and Sensitive tests used the same items as the pre-tests but presented them in a different order. To do this, the items from the pre-test were first put into a random order and examined for items with a key and a distractor in common that were too close together. When these items were near to each other, one of them was moved. The goal was to not have these two items presented on the same page (the 60 test items were presented over 3 pages). The item order on the post-tests, like the pre-tests, was the same for both the tough and sensitive versions.

### **3.5.8. Format of the vocabulary tests**

Both the Tough and Sensitive tests in the pre- and post-test situations were formatted in the same manner. For each test, there was an answer page that included spaces to answer all 60 items. The items were distributed in three columns of 20 items each on one side of the page. The opposite side of the page had space for the participant's name and student number. For each item, the seven letters (A to G) that corresponded with item options were presented horizontally. Each letter was printed in the Optical Mark Recognition bubble font. Participants were instructed to shade in the letter of the most appropriate option for each item. Each test was presented in a four page test booklet. The first page of the booklet presented the instructions in the participants' L1 and explained the format of the test and how to complete it.

The remaining three pages of the booklet presented 20 items each. The items were presented in four columns of five items each. Each item presented the English target word's stem with the seven options arranged vertically beneath it. The format of the tests and the individual items can be seen in Appendices F1 to F8 which present the answer pages for all vocabulary tests and the pre- and post-test versions of the Tough and Sensitive tests.

### **3.5.9. Vocabulary test administration procedure**

The Tough and Sensitive tests, in both the pre- and post-test situations, were administered to the participants in the same teaching session. The Tough Test was administered first. To begin the test administration procedure the participants were told that they were about to take a vocabulary test. They were told that the test was a measure of how much vocabulary they knew and not for the purposes of determining their grade in the course. Each participant received an answer sheet and was asked to write their name and student number on it. As a group, the format of the vocabulary tests was reviewed on the OHC and included information on how to answer the items, the audio track that accompanied the tests, and the purpose of the '*I don't know this word.*' option on the test. Any questions that the participants had about the tests were addressed at this time. The test sheets were then handed out and the participants were encouraged to review the test procedures that were printed in their L1 on the first page of the test booklet. All participants began the test together in synchronization with the audio track. They were given 15 seconds to answer each item on the test which meant a total of 15 minutes for the complete test. The participants were given a further 5 minutes to check their answers, answer any items they had not answered earlier, and to make sure that they had answered all the items. Answer sheets and test booklets were then collected. Participants then completed approximately forty minutes of unrelated material which differed depending on whether it was the pre-test or the post-test. Participants then completed the Sensitive Test in the same manner as that described for the Tough Test.

### **3.5.10. Scoring the vocabulary tests**

The tough and sensitive versions of the vocabulary pre- and post-tests were scored with Optical Mark Recognition software and those results were analyzed in a spreadsheet. Each word for each test was scored as *known*, *learned*, or *not learned*. If a participant responded correctly on the pre- and post-test then the word was scored as *known*. If a participant responded incorrectly or chose the '*I don't know this word.*' option on the pre-test and responded correctly on the post-test, the item was scored as *learned*. If the participant

answered incorrectly or chose the ‘*I don’t know this word.*’ option on the pre- and the post-test, the item was scored as *not learned*. An item was also scored as *not learned* if the participant answered correctly on the pre-test but incorrectly on the post-test.

### 3.6. Results

#### 3.6.1. Analysis of the results from the Tough and Sensitive tests

The effect of viewing 10 episodes on vocabulary knowledge of the 187 participants from Study 2 was examined by measuring knowledge of the *Chuck*-specific target words at two sensitivities before and after viewing. The mean number of words known on the Tough Test was 32.1 while on the Sensitive Test it was 37.2. For a target word to be considered known by a participant, it had to be answered correctly on both the pre- and post-test. The mean number of words known indicates that the participants knew between 53.5% and 62.0% of the target words before viewing episodes of *Chuck* depending on the sensitivity of the test. Individuals, however, differed by which items they knew and how many. On the Tough Test, the maximum number of target words known was 45 and the minimum number known was 2. On the Sensitive Test, the maximum known was 53 target words and the minimum was 16. Therefore, there was a lot of variation in the amount of vocabulary learning possible between individuals.

Vocabulary gain was determined by subtracting the number of target words known in both the pre- and post-test from the number of targets words correct on the post-test. The mean vocabulary gains on the Tough and Sensitive tests were 6.36 and 6.78 words respectively<sup>i</sup>. There was considerable variance in gain as the minimum gains, maximum gains, and standard deviations shown in Table 3.7 indicate.

Table 3.7 Vocabulary gains for Tough and Sensitive tests for Experimental Group (N=187)

| Vocabulary Test | Mean Real Gain | SD    | Median | Minimum | Maximum |
|-----------------|----------------|-------|--------|---------|---------|
| Tough           | 6.36           | 2.741 | 6      | 1       | 20      |
| Sensitive       | 6.78           | 3.214 | 6      | 1       | 23      |

The 73 participants in the Control Group also completed the two tests of the *Chuck*-specific target words at two sensitivities. The results for the Control Group are presented in Table 3.8. The mean gain on the Tough Test was 5.32, and on the Sensitive Test it was 5.37. The minimum gain was 0 words and the maximum gain was 9 words for both sensitivities of the vocabulary test.

Table 3.8 Vocabulary gains for Tough and Sensitive tests for Control Group ( $N=73$ )

| Vocabulary Test | Mean Real Gain | SD    | Median | Minimum | Maximum |
|-----------------|----------------|-------|--------|---------|---------|
| Tough           | 5.32           | 1.914 | 5      | 0       | 9       |
| Sensitive       | 5.37           | 2.118 | 5      | 0       | 9       |

In addition to calculating the raw gains on the two vocabulary tests it was also necessary to determine the relative vocabulary gains of the participants. Because participants who knew more of the target words had less room for improvement than participants who knew fewer target words, absolute word gains may not give a complete picture of vocabulary learning. Relative vocabulary gain was calculated to take into consideration these varying opportunities for gain. Relative gain percentage was developed by Shefelbine (1990) and was explained and used by Horst, Cobb and Meara (1998) in their study of vocabulary acquisition through reading. Relative gain for participants used the following formula:

$$\text{Relative Gain for Participants} = \frac{\text{Number of Target Words Learned}}{\text{Number of Items on Test} - \text{Number of Target Words Known}} \times 100$$

The results for relative gain on the Tough and Sensitive tests for the Experimental Group are shown in Table 3.9 and in Table 3.10 for the Control Group. For the Experimental Group, the mean relative gain for the Tough Test was 23.03% and 29.61% for the Sensitive Test. The mean relative gain for the Control Group on the Tough Test was 20.86% and the mean was 25.42% for the Sensitive Test. There were large differences between the minimum and maximum relative gains for both tests and in both treatment groups.

Table 3.9 Relative vocabulary gains for Tough and Sensitive tests for Experimental Group ( $N=187$ )

| Vocabulary Test | Mean Relative Gain | Median | Minimum | Maximum |
|-----------------|--------------------|--------|---------|---------|
| Tough           | 23.03%             | 22.73% | 2.94%   | 46.67%  |
| Sensitive       | 29.61%             | 28.57% | 4.35%   | 70.00%  |

Table 3.10 Relative vocabulary gains for Tough and Sensitive tests for Control Group ( $N=73$ )

| Vocabulary Test | Mean Relative Gain | Median | Minimum | Maximum |
|-----------------|--------------------|--------|---------|---------|
| Tough           | 20.86%             | 20.69% | 0.00%   | 40.00%  |
| Sensitive       | 25.42%             | 23.53% | 0.00%   | 50.00%  |

### 3.6.2. The effects of viewing English-language television on incidental vocabulary acquisition

To determine whether the raw vocabulary gains on the Tough and Sensitive tests were significant, a series of *t*-tests were carried out. Prior to performing *t*-tests on the data, an informal analysis of the distribution of the gain scores on the Tough and Sensitive Test using a histogram and normal Q-Q plot revealed no serious threats to the assumption of normality. Paired samples *t*-tests showed that the difference between mean gains of the target words for the Experimental Group on the Tough,  $t(186) = 30.448, p < .001$ , and Sensitive,  $t(186) = 26.288, p < .001$ , tests were significant. There were large treatment effects for both the Tough ( $d=1.04$ ) and the Sensitive Test ( $d=1.18$ ). These results indicate that the Experimental Group in Study 2 made significant, large gains in the number of target words they knew through viewing episodes of English-language television.

To determine whether the raw vocabulary gains by the Control Group on the Tough and Sensitive tests were significant, two *t*-tests were performed. Prior to performing *t*-tests on the data, an informal analysis of the distribution of the gain scores on the Tough and Sensitive tests using a histogram and normal Q-Q plot revealed no serious threats to the assumption of normality. Paired samples *t*-tests showed that the difference between mean gains of the target words for the Control Group on the Tough,  $t(72) = 23.727, p < .001$ , and Sensitive,  $t(72) = 21.661, p < .001$ , Tests were significant. There were large treatment effect for the Tough Test ( $d=0.90$ ) and a medium treatment effect for the Sensitive Test ( $d=0.71$ ).

To compare the vocabulary gains for the participants in the Experimental Group and the Control Group, a series of *t*-tests were undertaken. Prior to performing *t*-tests on the data, an informal analysis of the distribution of the gain scores on the Tough and Sensitive Test using a histogram and normal Q-Q plot revealed no serious threats to the assumption of normality. The results of the independent samples *t*-tests showed that the difference between gains of the target words for the Experimental Group and the Control Group on the Tough,  $t(258) = 2.890, p < .01$ , and Sensitive,  $t(258) = 3.200, p < .01$ , tests were significant. There was a small treatment effect for the Tough Test ( $d= 0.49$ ) and a small treatment effect for the Sensitive Test ( $d= 0.43$ ). These results indicate that viewing the 10 episodes of English-language television between the Tough and Sensitive Pre- and Post-Tests had a small significant impact on vocabulary gain when compared to the participants (Control Group) who did not view any episodes of *Chuck* between the vocabulary pre- and post-tests and studied English in a more conventional setting for the same period of time.

### **3.6.3. The effects of vocabulary knowledge on relative vocabulary gain through viewing English-language television**

To investigate whether participants with greater vocabulary knowledge incidentally gain more vocabulary through viewing television than those with less vocabulary knowledge, the relative vocabulary gains were compared to a measure of their vocabulary knowledge. Vocabulary knowledge was calculated by combining the participants' results on the three VLT at the 2,000-, 3,000-, and 5,000-word levels and analyzing this using the Rasch Model to obtain an interval measure. With this measure of vocabulary knowledge it was possible to investigate whether those participants who had more vocabulary knowledge made greater relative gains in knowledge of the target vocabulary than participants with less vocabulary knowledge.

A Pearson product-moment correlation coefficient was used to assess the relationship between the relative vocabulary gains from viewing 10 episodes of English-language television and the vocabulary knowledge of the participants. There were small non-significant correlations between the relative vocabulary gain on the Tough Test,  $r = .021$ ,  $N=187$ ,  $p = .773$ , and the Sensitive Test,  $r = .080$ ,  $N=187$ ,  $p = .277$ , and vocabulary knowledge. These results indicate that participants with more vocabulary knowledge did not increase their relative vocabulary learning through watching television compared to those with less vocabulary knowledge.

### **3.6.4. The effects of frequency and range on incidental vocabulary acquisition through viewing English-language television**

To investigate whether words that occur more frequently and words that occurred across a greater range of television episodes were more likely to be learned, gains in the number of participants who learned the target words were analyzed. The mean raw gains for items on the Tough and Sensitive tests were 19.83 and 21.12 people respectively. That is, across the 60 items on the vocabulary tests, an average of approximately 20 of the 187 participants learned each item on the Tough Test. An average of approximately 21 participants learned each item on the Sensitive Test. There was, however, considerable variation in the number of participants who learned items. On both the Tough and Sensitive Test, the minimum number of participants gaining knowledge of an item was 0. On the Tough Test, the maximum increase was 46 participants and on the Sensitive Test it was 59 participants. Table 3.11 presents a summary of these results.

Table 3.11 Vocabulary gains for Tough and Sensitive tests for items ( $N=60$ )

| Vocabulary Test | Mean Real Gain | SD    | Median | Minimum | Maximum |
|-----------------|----------------|-------|--------|---------|---------|
| Tough           | 19.83          | 12.49 | 19     | 0       | 46      |
| Sensitive       | 21.12          | 16.15 | 20     | 0       | 59      |

To take into account the differing numbers of participants who knew items and therefore differing amounts of possible gain the relative gain of the items was calculated through the following formula:

$$\text{Relative Gain for Items} = \frac{\text{Number of People Who Learned the Item}}{\text{Number of People Answering Item} - \text{Number of People Knew the Item}} \times 100$$

The mean relative gain for items on the Tough Test was 33.02% and for the Sensitive Test, it was 39.78%. There were large differences between the minimum and maximum relative gains for both tests. The descriptive statistics for this analysis are described in Table 3.12. The results for each item, including the number of participants who knew each target word, learned a target word, did not learn the target word, and the relative gain, on both the Tough and Sensitive tests are presented in Appendix E2.

Table 3.12 Relative gains for Tough and Sensitive tests for items ( $N=60$ )

| Vocabulary Test | Mean Relative Gain | Median | Minimum | Maximum |
|-----------------|--------------------|--------|---------|---------|
| Tough           | 33.02%             | 29.60% | 0%      | 69.23%  |
| Sensitive       | 39.78%             | 34.60% | 0%      | 100%    |

The 60 target words that the Tough and Sensitive test items were based on occurred from 5 to 54 times throughout the 10 episodes of *Chuck*. The target words had an average frequency of occurrence of 10.2 times across these episodes. A Pearson product-moment correlation coefficient was used to assess the relationship between the relative gain for the target words and their frequency of occurrence across the episodes. There was a medium significant correlation between the relative gain for items on the Tough Test and frequency of target word occurrence,  $r = .30$ ,  $N=60$ ,  $p < 0.05$ . The relative gains for the Sensitive Test had a small non-significant correlation,  $r = .18$ ,  $N=60$ ,  $p = .162$ , with the frequency of target word occurrence. These results indicate that there was a significant relationship between how frequently a target word occurred in the 10 episodes and the relative gain of items on the Tough Test but this relationship was not significant for the Sensitive Test.

The range of occurrence for the 60 target words on the Tough and Sensitive tests was from one episode to 10 episodes with an average range of occurrence of 3.7 episodes. A Pearson product-moment correlation coefficient was used to assess the relationship between

the relative gain for the target words and their range of occurrence across the episodes of *Chuck*. There was a small non-significant negative correlation,  $r = -.048$   $N=60$ ,  $p = .770$ , between the relative item gains for the Tough Test and the range of occurrence of the target words and a small non-significant negative correlation for the Sensitive Test,  $r = -.007$ ,  $N=60$ ,  $p = .956$ . These results indicate no relationship between the range of episodes in which a target word was encountered and the amount of relative gain for the word on either the Tough or Sensitive Test.

It may be more worthwhile to consider range and frequency together and analyze a target word's relative frequency of occurrence across the 10 episodes rather than range alone. Relative frequency is the overall frequency of occurrence of a target word in the episodes divided by the number of episodes (range) that the target word occurred in. This indicates how concentrated the occurrences of a target word are in single episodes. If the concentration of a target word in a single episode is higher it may make the word more salient and increase the amount of learning. The mean relative frequency of the 60 target words in the episodes of *Chuck* was 3.6 with a minimum relative frequency of 1.0 and a maximum relative frequency of 17.0. A Pearson product-moment correlation coefficient was used to assess the relationship between the relative gains of the target words by the participants on the Tough and Sensitive tests and the target words' relative frequency in the episodes. There was a small significant correlation,  $r = .25$ ,  $N=57$ ,  $p < 0.05$ , between the relative item gains for the Tough Test and the relative frequency of the target words. There was a small non-significant correlation for the Sensitive Test,  $r = .11$ ,  $N=58$ ,  $p < 0.196$ . These results indicate that there was a significant relationship between the relative frequency of target words and the relative gain of items on the Tough Test but this relationship was not significant for the Sensitive Test. This analysis of relative frequency and vocabulary gains indicates that learning is more apt to take place when vocabulary is encountered more often in a single episode.

### **3.7. Summary of findings**

The main findings for Study 2 can be summarized as follows:

1. Incidental vocabulary learning from viewing 10 episodes of English-language television was on average a gain of over six words regardless of the sensitivity of the test. The learners' mean uptake of approximately six words means that they learned almost a quarter of the vocabulary that they could possibly learn. These gains were significantly higher than those of the Control Group.

2. Greater vocabulary knowledge was not found to have a significant correlation with greater vocabulary gains made through watching television.
3. The frequency with which the target words occurred in the 10 episodes of *Chuck* and the relative gains for these target words was found to have a medium-size correlation for the Tough version of the vocabulary test. There was no significant correlation between frequency and vocabulary gains on the Sensitive Test.
4. The range of episodes in which target words occurred had no significant relationship between the relative gains for these target words on either vocabulary test.

### **3.8. Discussion**

#### **3.8.1. Incidental vocabulary learning from viewing television**

Study 2 has expanded on the designs of previous studies of vocabulary acquisition through viewing video in four ways. First, gains in vocabulary knowledge were measured with tests at differing sensitivities. Knowledge of 60 target words was measured with a sensitive and tough test before and after viewing the 10 episodes of *Chuck*. Second, this study examined the effects of a large amount of viewing time. The largest amount of viewing time participants in previous studies had was 45 minutes (Vidal, 2003, 2011). In Study 2 viewing time was over seven hours. Third, the videos used in Study 2 were episodes of authentic television, the type of which language learners may be more likely to choose to watch on their own. Unlike the majority of the types of videos cited in previous research (educational videos, segments of television or film, academic lectures, and documentaries), these episodes of television were less designed for learning from. Finally, participants viewed successive episodes of the same program. Unlike the 10 episodes of television viewed in Study 2, when multiple videos in the previous research were viewed they had little relationship to one another and provided little opportunity for increasing exposures to low frequency vocabulary. Overall, the experimental design of Study 2 expanded on earlier methodologies in a number of significant ways.

The results from Study 2 demonstrate incidental learning of vocabulary does occur through watching television. Participants had mean vocabulary gains of 6.4 words on the Tough Test and 6.8 words on the Sensitive Test through viewing 10 successive episodes of *Chuck*. The Experimental Group's gains from the pre- to post-test were significantly greater than those of the Control Group. The results support the earlier findings of Baltova (1999), Huang and Eskey (1999), Hui (2007), Markham, Peter and McCarthy (2001), Sydorenko

(2010), Vidal (2003, 2011), and Winke, Gass and Sydorenko (2010) who found incidental vocabulary learning occurred through viewing videos.

It is difficult to make direct comparisons between the mean gains from the results of Study 2 and previous research because of the differing number of test items and different treatments in the studies. The overarching finding from this research and previous research, however, is that vocabulary can be incidentally learned from viewing videos. In Study 2 the mean number of words learned was over 6 for each vocabulary test but this amount may be a conservative estimate of the amount of vocabulary learning that occurred. There are several ways that vocabulary learning may have been underestimated. Firstly, there may have been gains for words that were not tested. Only words with five or more occurrences in the episodes of *Chuck* were considered as target words. As the number of occurrences decreases the likelihood of acquisition decreases but words that were particularly salient in an episode, were supported well with onscreen visual images, or were explicitly explained may potentially have been learned. Words from the 2,000 most frequent words of English were also not tested. The results from the VLT at the 2,000-word level (described in detail in Study 3 as they are more relevant for the later chapter) indicated that 77.5% of participants did not have knowledge of a proportion of words at this frequency level. However, because the 2,000 level vocabulary occurs very frequently, there were many opportunities for participants to make gains of previously unknown words from this frequency level.

Through viewing the episodes of television there were also opportunities for participants to increase their depth of vocabulary knowledge but this was not measured. Participants likely increased their depth of knowledge for words that they had prior knowledge of before viewing, especially words in the 1,000- and 2,000-word frequency lists. Gains in knowledge of these words would be expected to come in the form of collocation, aural form, and multiple meanings of words. Encountering these known words in the context-rich episodes would likely have increased these components of knowledge (Webb & Rodgers, 2009b), because learning vocabulary in a contextualized situation such as viewing television gives learners as strong sense of a word's meaning and use (Huckin & Coady, 1999).

Another reason that the results reported in Study 2 might be considered conservative has to do with the nature of the vocabulary testing procedure. The Tough and Sensitive Post-Tests came a week after the final episode was viewed. This meant that there were 10 weeks between the first episode viewed and the post-tests. It is conceivable that gains in knowledge made from earlier episodes may have been lost to decay. Previous research (R. Brown et al., 2008; Elley, 1989; Vidal, 2003, 2011; Waring & Takaki, 2003) has shown that gains in

knowledge demonstrated in immediate post-tests experienced considerable decay when target words were tested again after a delay. It is conceivable that gains in vocabulary knowledge would have been more apparent if participants were measured for knowledge of the target words immediately following each episode.

It is worthwhile to compare the incidental vocabulary acquisition results from Study 2 not only to previous research involving viewing videos, but also research involving reading. The results from Study 2 indicate that incidental vocabulary learning from television is comparable to incidental learning from reading. One reading study with a methodology similar enough to make comparison possible was by Horst, Cobb and Meara (1998). Through reading a novel, they found that participants gained a mean of 4.6 words from 45 target words. From 60 target words, participants in Study 2 were found to have gained a mean 6.4 words on the Tough Test and 6.8 words on the Sensitive Test. In the study by Horst, Cobb and Meara one word was learned for every 9.78 target words tested and in Study 2 approximately one word was learned for every 9.09 target words tested. This indicates the similarity of the incidental vocabulary gains from reading compared with those from viewing television.

While the vocabulary gains may appear similar, viewing television may actually be a more efficient method of incidentally learning vocabulary compared with reading. In the Horst, Cobb and Meara study participants read the *Mayor of Casterbridge* which contained 21,232 tokens and took 6 hours to read. The 10 episodes of *Chuck* in this study took approximately 7 hours to view and contained 52,030 tokens. It may appear quicker to read the book but people have been shown to read less than they watch TV. In Japan, where the participants from Study 2 are from, the average television viewing time in 2007 was 3.6 hours per day (OECD, 2009) while Japanese students reported spending one to two hours a day reading for enjoyment (OECD, 2011). If language learners in Japan were to spend a third of the amount of time they normally watch television watching English-language television they would watch L2 television 1.2 hours a day. Similarly, if they read English-language books one third the amount they spend reading in their L1 for pleasure, they would read 0.5 hours a day. This would translate to 438 hours watching English-language television and 182.5 hours reading English-language books a year. Given the rates of acquisition found in Study 2 and in the Horst, Cobb and Meara (1998) study, this indicates that language learners could potentially learn approximately 401 words through watching television while learning approximately 140 words through reading. Even if the learners only spend 30 minutes a day watching English-language television and 30 minutes a day reading English-language books,

they could potentially learn more vocabulary through viewing television (approximately 167 words) than through reading (approximately 140 words).

The comparability of viewing television to reading in terms of the potential for incidental vocabulary acquisition makes television potentially suitable to form the basis of extensive viewing programs. While books provide the authentic input in extensive reading, episodes of television would provide the input in extensive viewing. The growth in popularity of extensive reading programs was born out of an effort to provide language learners with authentic input, which, among other qualities, provides learners with opportunities to encounter, in context, unknown vocabulary which can lead to incidental vocabulary learning. As the amount of authentic L2 exposure increases, the likelihood that words will be learned increases (Nagy et al., 1985). Television is similarly suited to provide significant amounts of authentic input to language learners. In addition to the potential for building vocabulary knowledge, television has the potential to improve listening skills, increase comprehension skills, provide exposure to varieties of English, and be a source of motivation (Vandergrift, 2007). Episodes of television could be utilized to provide a source of L2 input in EFL environments where such input is rare.

### **3.8.2. Frequency and range of vocabulary occurrence and incidental vocabulary acquisition**

The results of Study 2 indicated that as the number of exposures to target words increases the likelihood the words are learned increases. This supports the findings of previous investigations involving reading (Horst et al., 1998; Neuman & Koskinen, 1992; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Rott, 1999; Saragi et al., 1978; Waring & Takaki, 2003; Zahar et al., 2001). Horst, Cobb and Meara (1998) found a correlation of .49 between the frequency of target word occurrence and gain, Zahar, Cobb and Spada (2001) found a correlation of .36, and Saragi, Nation and Meister (1978) found a correlation of .36. In Study 2, there was a significant correlation of .30 between the number of times each word occurred in the episodes of television and the relative gains of the items on the Tough Test. While the correlation for this study is the lowest of those reported it is still medium-sized which is generally believed to be the relationship between the vocabulary repetition and acquisition. This is because frequency is but one of many factors affecting vocabulary learning (Nation, 2001).

In Study 2, vocabulary learning occurred with relatively few exposures to the target words. Seventeen target words occurred the minimum 5 times throughout the 10 episodes of *Chuck*.

There were gains by participants on all but one these 17 target words on the Tough Test. The target word that did not show any gain, *cute*, was known by 182 of 187 participants prior to viewing any episodes meaning that there were few participants that could have gained knowledge of it. The fact that a high proportion of the target words occurring 5 times had substantial gains indicates that even with relatively few exposures to a word, there was enough information presented in context, both visually and aurally, in the television episodes for learning to occur (Webb, 2008).

There are some features of Study 2 that may have led to the smaller correlation between frequency and vocabulary gains than those reported in previous studies. The majority of the previous research investigated frequency of occurrence and vocabulary acquisition from single texts. In this study, however, there were 10 individual episodes that the target words occurred in. In the previous research, the vocabulary tests were generally immediately following exposure to the input text. In Study 2, the minimum amount of separation from the tests was one week when the participants viewed the Final Episode and then completed the Tough and Sensitive Post-Tests the following week. There was also the possibility of 10 weeks between encountering a target word for the first time and the post-tests. This delay between first encountering a target word and testing could have caused decay in knowledge. This occurred in Vidal's (2011) study involving vocabulary learning from viewing videos and reading. When participants were tested a month after their immediate post-tests, their delayed post-test scores were almost half what their immediate post-test scores were for both treatments. In Study 2, had post-tests come immediately following the individual episodes of *Chuck*, more pronounced vocabulary learning might have been evident, and thereby a stronger relationship with frequency of occurrence.

Study 2 also investigated the effects of the target words' range of occurrence across the episodes and the target words' acquisition by the participants. This analysis revealed no significant correlation between the two variables. While it was hypothesized that encountering the vocabulary in different contexts, i.e. different episodes, could possibly improve acquisition, it appears that there was no relationship between range and acquisition. It may be more worthwhile to consider range and frequency together and analyze a target word's relative frequency which is the overall frequency of occurrence of a target word in the episodes divided by the number of episodes that the target word occurred in. Results indicated a significant relationship between the relative frequency of the target words and the relative gain of items on the Tough Test but this relationship was not significant for the Sensitive Test. These findings indicate that incidental vocabulary learning is more apt to take place

when vocabulary is encountered more often in a single episode which may be a result of words being more salient when encountered more often in a single episode rather than across many episodes.

The findings in Study 2 indicating a relationship between the frequency of vocabulary occurrence and incidental acquisition, as well as the relationship between relative frequency of the vocabulary in the episodes and incidental acquisition, are valuable. They add evidence to support the role that frequency of occurrence has in incidental learning of L2 vocabulary. Repetition of vocabulary in different contexts is thought to be vital to learning words (Schmitt, 2008; Webb, 2008) and by viewing successive episodes of the same series the repetitions of low frequency words increase, thus improving the chance of their being learned (Rodgers & Webb, 2011). Overall, the results support the notion of viewing successive episodes of a single television program to increase the frequency at which vocabulary occurs thereby both increasing the opportunities for, and improving the chances of incidental vocabulary acquisition.

### **3.9. Limitations**

The target words tested by the Tough and Sensitive tests are a limitation of Study 2. The target words represented words taken from the 3,000- to 14,000-word level BNC lists (and two off-list target words) that occurred more than 5 times across the 10 episodes. It is very likely that the participants made gains in knowledge of vocabulary that occurred less than 5 times. Previous video-based research has shown gains to target words occurring less than 5 times (Vidal, 2003). Also in Study 2, results from the VLT indicated that the majority of participants did not have mastery of the 2,000-word level. There are many frequently occurring words at this level that were not tested that participants could have learned. If this research were to be replicated with participants at a similar proficiency level, including target words from the 1,000- and 2,000-word frequency lists would be recommended. It might also be advisable to measure knowledge of a large number of words from the episodes of television in a pre-test and create personalized lists of target words for participants in a manner similar to what was done in studies by Winke, Gass and Sydorenko (2010) and Sydorenko (2010).

### ***Study 3: The effects of lexical coverage on aspects of viewing television: comprehension and vocabulary acquisition***

#### **4. Introduction**

The previous studies in this thesis have shown the potential for language learning from viewing episodes of television. Study 1 demonstrated that, on average, language learners were able to comprehend episodes of L2 television to what might be considered an adequate level. The participants in Study 2 were shown to be able to make significant gains in their vocabulary knowledge through watching television. However, these studies also showed that there were considerable differences between the results of the participants. One possible explanation for these individual differences may be the vocabulary knowledge of the participants. The better a language learner's vocabulary knowledge, the less unknown vocabulary there is in episodes of television. Less unknown vocabulary may lead to greater comprehension and a better chance of incidental vocabulary learning.

While there are many factors that influence comprehension, vocabulary knowledge is believed to be the most influential factor (Laufer & Sim, 1985; Stæhr, 2009; Stahl, Jacobson, Davis, & Davis, 1989). Lexical coverage of a text is the percentage of words known by a language learner at a particular vocabulary level and points to how much vocabulary learners need to know for comprehension to take place. Once the lexical coverage of a given text or series of texts is ascertained it is a valuable measurement for language learners because it provides a target vocabulary size for comprehension. Upon reaching this target, the learners should be better able to understand the text in question. While it is unquestionably a factor in comprehension, lexical coverage's role in incidental vocabulary learning is less substantiated. However, without sufficient lexical coverage language learners may be unable to acquire knowledge of the meaning of words presented in context.

The bulk of the research on the effects of lexical coverage has examined reading and listening. Similar research involving lexical coverage and viewing television may prove equally useful. If language learners are able to comprehend television and learn vocabulary through viewing television in the way they are thought to from reading and listening, viewing television may be a valuable source of authentic input. There are also relatively few studies examining the relationship between coverage and comprehension and how this relationship varies from text to text (Nation & Webb, 2011). This is particularly important for language

learning through viewing television as learners may be encouraged to watch multiple episodes of the same television series and lexical coverage and comprehension may vary from episode to episode. Overall, there is a lack of research concerning television and the relationships between lexical coverage and comprehension, and lexical coverage and incidental vocabulary acquisition.

The previous research into the relationship between lexical coverage and comprehension and between lexical coverage and incidental vocabulary learning informs the present study which was designed to examine (a) whether increased lexical coverage of English-language television leads to increased comprehension by EFL learners, and (b) whether learners with greater lexical coverage of television have more incidental vocabulary learning than those learners with less lexical coverage.

#### **4.1. Lexical coverage and comprehension**

Numerous studies have examined the relationship between lexical coverage and comprehension (Bonk, 2000; Carver, 1994; Hirsh & Nation, 1992; Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Laufer & Sim, 1985; Laufer, 1989; Nation, 2006; Rodgers & Webb, 2011; Schmitt et al., 2011; Stæhr, 2009; van Zeeland & Schmitt, 2012; Webb & Rodgers, 2009a; Webb, 2011). Six of these studies investigated the relationship between lexical coverage and reading comprehension (Carver, 1994; Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Laufer & Sim, 1985; Laufer, 1989; Schmitt et al., 2011). Three studies looked at the relationship between lexical coverage and listening comprehension (Bonk, 2000; Stæhr, 2009; van Zeeland & Schmitt, 2012). There were four studies that analyzed the vocabulary in texts to determine the vocabulary size, and corresponding lexical coverage, necessary to understand different types of discourse (Hirsh & Nation, 1992; Nation, 2006; Rodgers & Webb, 2011; Webb & Rodgers, 2009a). The following survey of the most relevant studies focuses on the percentage of vocabulary that needs to be known in a text for comprehension to occur.

The basis for the presupposition that lexical coverage is a principal determiner of comprehension comes from Laufer and Sim's (1985) study into the threshold of L2 reading competence for academic texts. In attempting to identify the nature of this threshold, the researchers looked at whether it is dependent on semantic, syntactic, discourse or subject-matter based knowledge. In the first stage of the research, participants that were believed to be at the threshold of L2 reading competence were identified by their performance on a series of comprehension tests. Participants in this group were then given the reading section of a

standardized English proficiency test to measure the threshold level in terms of an objective external instrument. The participants' scores on this test were between 65% and 70%. This range of scores was considered appropriate for minimal reading ability for academic purposes. In a follow-up study with a larger sample size, the researchers determined that vocabulary was the most important factor for comprehension by observing how the participants handled known language elements to interpret meaning. They also observed how supplying the meanings to unknown language elements improved the participants' ability to interpret meaning of the text. Building on this early research, later studies attempted to identify the percentage of lexical coverage at which this threshold for comprehension begins.

Results from studies investigating the lexical coverage figure necessary for reading comprehension have produced varying coverage estimates. Hu and Nation (2000) studied the effects of differing coverage levels on reading comprehension. Low frequency words were replaced with nonwords to create versions of an easy fiction story with 95%, 90% and 80% coverage levels. The original story with no replacements was the 100% coverage version. At a coverage level of 80%, no participants attained adequate comprehension. At 90% lexical coverage, only a small number of participants achieved adequate comprehension. At 95% coverage, a minority of participants gained adequate comprehension. At 100% coverage, most learners were able to comprehend the text. A lexical coverage of 98% was determined through regression analysis to be necessary for adequate comprehension.

Hu and Nation's (2000) findings are supported by Carver's (1994) L1 study. Carver measured native speakers' reading levels and had the participants read passages that had been analyzed for lexical difficulty. This was to determine the relationship between the number of unknown words in a passage and the relative difficulty of the passage. Two groups of participants (elementary school students and graduate school students) underwent similar treatments in which they read passages and underlined unknown words. The research indicated that lexical coverage was related to the difficulty of a text and 98% to 99% coverage of a text provided sufficient L1 comprehension.

Investigating L2 learners of English, Laufer (1989) attempted to determine the percent coverage needed to ensure reasonable reading comprehension of an academic text. A reasonable level of comprehension was defined as a score of 55% (the lowest passing grade at the university at which the research took place). Two tests were used to measure reading comprehension: a standardized test employing multiple-choice questions and a test employing open-ended questions designed by the researchers. As the participants answered the questions they also underlined any words they did not know in the reading texts. Next, the participants

completed a vocabulary test where they translated or paraphrased words from the texts. The participants' lexical coverage was determined by subtracting unknown words on the vocabulary test and words indicated as unknown in the reading texts from the total number of words in the reading passages. Analysis showed that the group with 95% lexical coverage and above had a significantly higher number of readers reach reasonable comprehension of the academic texts than the group with lexical coverage below 95%. Subsequent research by Laufer and Ravenhorst-Kalovski (2010) went on to suggest that 95% lexical coverage is necessary for minimal comprehension while 98% lexical coverage is necessary for optimal reading comprehension.

Schmitt, Jiang and Grabe (2011) investigated how each percentage point from 90% to 100% coverage affects comprehension. Participants with 12 different L1s indicated whether they knew 120 words taken from the reading texts allowing the researchers to estimate each participant's coverage of the texts. The participants then read the texts and completed comprehension tests. The participants' scores on the comprehension tests were plotted against their coverage figures. The results revealed a relatively linear relationship between the percentage of lexical coverage and reading comprehension. Participants with 90% coverage had comprehension scores just over 50% while participants with 100% coverage had comprehension scores just over 60% or 75% depending on the text. The authors conclude that if 60% comprehension is necessary then 95% coverage is sufficient. However, if 70% comprehension is necessary 98% to 99% coverage is necessary.

This review of lexical coverage and reading comprehension research shows a range of figures have been proposed. The lowest figure is Laufer's (1989) suggestion that 95% coverage is necessary for reasonable comprehension of a text. Hu and Nation (2000) suggested that 98% coverage is likely necessary, while Carver (1994) proposed an even more conservative estimate of 98% to 99% coverage as being necessary for comprehension. Schmitt, Jiang, and Grabe (2011) suggested that different amounts of coverage are necessary for different levels of reading comprehension. They proposed that 95% coverage is sufficient for 60% comprehension but for 70% comprehension, 98% to 99% coverage is necessary. Laufer and Ravenhorst-Kalovski (2010) also suggested different levels of lexical coverage were necessary for different levels of comprehension. They proposed that 95% lexical coverage is necessary for minimal comprehension and for optimal reading comprehension 98% lexical coverage is necessary.

Three studies examining the relationship between lexical coverage and listening comprehension found a range of lexical coverage figures at which comprehension was said to

take place. Bonk (2000) studied the effects of coverage on listening comprehension using four short passages with increasing lexical difficulty. The passages had equivalent word totals but included increasing amounts of low frequency vocabulary. Comprehension was measured by a written recall test in the L1 and a dictation test in the L2. While there were outliers (participants who had low lexical recognition but had what was considered good comprehension and the inverse) coverage levels of 90% were found to be present in 87% of the cases of good comprehension. Bonk goes on to claim that a language learner may be able to operate at an even lower level of lexical coverage in a real-world situation where there can be support from images, context, or background knowledge.

Stæhr (2009) investigated the relationship between lexical coverage and listening comprehension for English language learners. Stæhr used the Cambridge Certificate of Proficiency in English where the listening texts are described as a short dialogue, short monologues, a radio broadcast, a radio interview, and an informal discussion. Stæhr estimated the lexical coverage of the participants on the listening test using scores on four levels of the Vocabulary Levels Test (VLT). Table 4.1 shows the mean comprehension scores of participants achieving mastery of the different levels of the VLT and the lexical coverage that each level provided on the listening tests. There was a significant difference between the comprehension scores of the participants with mastery of the 3,000 level and 5,000 level. Stæhr concludes that 98% lexical coverage might be sufficient for adequate listening comprehension.

Table 4.1 Mean comprehension scores for participants in Stæhr's (2009) study by mastery of levels of the VLT

| VLT    | <i>n</i> | Lexical Coverage | Mean Comprehension Score |
|--------|----------|------------------|--------------------------|
| 2,000  | 22       | 90.43%           | 54.1%                    |
| 3,000  | 34       | 93.94%           | 59.1%                    |
| 5,000  | 48       | 98.12%           | 72.9%                    |
| 10,000 | 8        | 99.27%           | 80.0%                    |

Van Zeeland and Schmitt (2012) expanded the research on lexical coverage and listening comprehension with a study investigating both L1 and L2 listening. Native and non-native speakers of English listened to four short stories. The stories were modified with nonwords so that each story had a lexical coverage of 100%, 98%, 95%, or 90%. Comprehension of each story was measured with a multiple-choice test. Vocabulary knowledge of the participants was measured with the 2,000 level VLT. For the native speakers, comprehension was

significantly better for the 100% coverage story than the stories with less coverage. Comprehension of the 98% coverage story was significantly higher than the story with 95% coverage. There was no significant difference between the story with 95% coverage and the story with 90% coverage. Surprisingly, comprehension of the 98% coverage story was not significantly greater than that of the story with 90% coverage. For non-native speakers, comprehension was significantly higher for the 100% coverage story than the stories with less coverage. Comprehension of the 98% coverage story was significantly higher than the 95% and 90% coverage stories. There was, however, no significant difference in comprehension between the stories with 95% coverage and 90% coverage. The results indicate that lexical coverage contributes to listening comprehension in both the L1 and L2. There was more variation in the scores for the non-native speakers suggesting that coverage is but one of many factors affecting L2 listening comprehension. Van Zeeland and Schmitt suggest 98% lexical coverage for high comprehension, but in situations where more lenient comprehension levels are acceptable, 95% and 90% coverage will suffice. Important for low-level language learners was the finding that with 90% coverage of the stories, 75% of L2 listeners had comprehension scores of 70% or over.

For listening comprehension, the majority of the coverage figures suggested have been similar to those in reading research. Stæhr (2009) found that 98% coverage was necessary for listening to a variety of texts. Bonk (2000), however, found that L2 learners with coverage of 90% or above had good listening comprehension. Van Zeeland and Schmitt (2012) suggested that for a high level of listening comprehension 98% coverage is necessary but 95% or 90% coverage is sufficient for lower but satisfactory levels of comprehension.

Four studies analyzed the vocabulary in different types of texts to estimate the lexical coverage necessary for comprehension. Hirsh and Nation (1992) looked at the vocabulary demands of three short unsimplified novels. These novels were thought to have a light vocabulary load because they were written for young native speakers of English. The researchers concluded that to read for pleasure or with general ease, a learner would need to have a vocabulary of around 5,000 word-families which corresponded to a lexical coverage of approximately 97% to 98%.

The vocabulary size needed for comprehension of different types of written discourse (newspapers, a novel, and a graded reader), as well as different types of spoken discourse (listening to conversations and watching a children's movie) was investigated by Nation (2006). Based on previous research into comprehension and lexical coverage, Nation examined the vocabulary size necessary to reach 95% and 98% coverage. Knowledge of the

2,000 most frequent word-families of English was shown to provide over 95% coverage of a graded reader. However, the most frequent 4,000 word-families were necessary to reach 95% coverage for newspapers and a novel. Knowledge of the most frequent 3,000 word-families accounted for over 98% coverage of graded readers while 8,000 word-families was necessary for 98% coverage of newspapers and 9,000 word-families for a novel. Nation proposed 98% coverage as ideal for unsupported comprehension of most texts. While spoken language was found to make use of more high frequency words than most written text, Nation suggested that lexical coverage greater than 98% may be needed for listening comprehension due in part to the temporary nature of spoken language.

Webb and Rodgers (2009a) examined the lexical coverage of television programs. This study was corpus-driven so there was no experimental investigation of coverage and comprehension but it lays the groundwork for research into the relationship between coverage and comprehension of television. The vocabulary in 88 English-language television programs in six genres was analyzed. The results indicate that a vocabulary size of 3,000 word-families plus knowledge of proper nouns and marginal words accounted for 95.45% coverage. However, the vocabulary size necessary to gain 95% coverage of the different genres ranged from 2,000 to 4,000 word-families plus proper nouns and marginal words. Overall, 7,000 word-families plus proper nouns and marginal words provided 98.27% coverage but again there was variation between genres. Knowledge of 5,000 to 9,000 word-families plus proper nouns and marginal words were needed to gain 98% coverage depending on the genre. Individual episodes of the programs also showed considerable variation in coverage. These findings were supported by Rodgers and Webb's (2011) analysis of a much larger corpus of television programs. Because the combination of visual and aural input is believed to make comprehension of television programs easier, Webb and Rodgers (2009a) suggest that 95% lexical coverage might be sufficient for comprehension.

The estimates of lexical coverage based on analyses of different texts provided relatively similar figures for what is considered necessary for comprehension. For analyses based on written texts, Hirsh and Nation (1992) suggested a lexical coverage of 97% to 98% for short novels. Nation (2006) similarly suggested coverage of 98% would be likely necessary for unaided reading comprehension. Nation also suggested that a lexical coverage greater than 98% may be needed for comprehension of spoken text. Webb and Rodgers (2009a) and Rodgers and Webb (2011) suggested 95% lexical coverage may be sufficient for adequate comprehension of television because of the added support of the images inherent in television.

While these lexical coverage figures provide valuable insight into the percentage of vocabulary in a text a language learner needs to know in order for comprehension to occur, direct comparison of these lexical coverage figures is difficult. This is because of the differing benchmarks at which the researchers consider comprehension to have taken place. Generally, these benchmarks have been set by the researchers based on study-specific rationales, and the concept of sufficient comprehension assuredly varies from study to study (Nation & Webb, 2011). For example, Stæhr (2009) examined three different levels of adequate comprehension, 60%, 68%, and 75%. These were based on 60% being a pass, 75% being a B- grade and 68% falling between the two. Hu and Nation (2000) assumed that 14 out of 14 on their comprehension test would obviously be adequate but, allowing for the possibility of human error, a score of 12 out of 14 (approximately 85%) was also considered to be an adequate comprehension level. These two examples highlight the lack of commonality for definitions of comprehension which makes comparing results challenging.

Variability between the comprehension tests and the input texts used in previous studies also potentially affects the relationship between lexical coverage and comprehension (Nation & Webb, 2011). Measuring comprehension of reading texts, for example, was done through open ended questions by Laufer and Sim (1985), while Schmitt, Jiang and Grabe (2011) used a combination of multiple-choice items focused on making inferences from the text and a graphic organizer that required recognition of the organizational pattern of the text. Within research on lexical coverage and listening comprehension, differences between input texts have also been pronounced. There were four listening passages used in Bonk's (2000) study. They each had approximately 85 tokens and were expository texts describing African customs. The four texts in van Zeeland and Schmitt's (2012) study, on the other hand, were first person narratives of approximately 450 tokens each.

Together, the findings from these studies suggest that for comprehension to occur, language learners need between 90% and 99% lexical coverage of a text. The coverage figure necessary, however, appears to depend on the input mode of the text as well as the level of comprehension desired. While there is no previous research that has examined the relationship between lexical coverage and comprehension of television, there have been corpus-driven studies that suggest that the coverage figure necessary for comprehension to take place may be lower than that of reading. This is an effect of the combined aural and visual input inherent in television.

#### **4.2. Lexical coverage and incidental vocabulary learning**

There have been very few studies that have investigated the relationship between lexical coverage and incidental vocabulary learning. One study investigated incidental vocabulary learning through reading texts with differing amounts of lexical coverage (Liu & Nation, 1985). Nation (2001) and Webb and Rodgers (2009a, 2009b) have suggested coverage figures at which incidental vocabulary learning may take place. Other studies have identified factors that may affect the lexical coverage necessary for incidental vocabulary learning to occur (Elley, 1989; Gruba, 2004; Horst et al., 1998; Huckin & Coady, 1999; Wang, 2012).

Liu and Nation (1985) carried out a study with 59 teachers of ESL across a wide range of English proficiency levels. The participants read texts where 1 in 10 or 1 in 25 low frequency words were replaced by nonwords. Vocabulary learning was measured by having the participants provide translations or synonyms for the nonwords. Participants learned 34.25% of the nonwords in the texts with 1 in 25 words replaced and 30.50% of the nonwords in the texts with 1 in 10 words replaced. This indicates that the participants were able to provide meanings for the nonwords more easily in the texts that had the lower percentage of unknown words. Verbs were the easiest learned, followed by nouns, adverbs and adjectives. Liu and Nation hypothesized that this is because verbs and nouns have more contextual support within written passages. From a lexical coverage perspective, the language learners were able to learn words more easily when they had 96% coverage (1 in 25 words replaced) than when they had 90% coverage (1 in 10 words replaced). The findings from this study indicate that the more words a language learner knows in a text the more likely they are to learn unknown words occurring in that text.

Lexical coverage figures at which incidental vocabulary learning is believed to take place have been proposed. Nation (2001) suggested that 98% lexical coverage may be the ideal level of lexical coverage for incidental vocabulary learning from reading to occur. Two studies analyzed the vocabulary in movies and television to determine the amount of vocabulary knowledge necessary to obtain certain levels of lexical coverage. From their corpus-driven research, Webb and Rodgers (2009a, 2009b) suggested that 95% coverage, or knowledge of the most frequent 3,000 word-families plus proper nouns and marginal words, may be sufficient for incidental vocabulary learning to occur.

The lower coverage figure proposed by Webb and Rodgers is based on the idea that the combination of visual imagery and aural input in video may make it easier to learn words incidentally. This is supported by previous research that has found that imagery accompanying reading texts is linked to increased incidental vocabulary learning (Elley,

1989; Horst et al., 1998). When viewing videos, language learners have been shown to use the images to clarify the meaning of unknown vocabulary (Gruba, 2004). There are a number of other factors that may affect the relationship between lexical coverage and vocabulary learning through viewing videos. First, texts that learners find interesting are more conducive to learning than texts that are less interesting (Dörnyei & Ushioda, 2011; Huckin & Coady, 1999). Furthermore, learners have been shown to find language learning through viewing videos to be an interesting and enjoyable experience (Wang, 2012). Second, prior knowledge of the subject matter is linked to improved incidental vocabulary acquisition (Huckin & Coady, 1999). Viewing successive episodes of a television series builds up knowledge about the program's story arcs, the characters, and their relationships to one another. Knowledge gained in early episodes can be utilized in later episodes. However, there is a characteristic of the listening and viewing process that may negatively affect the amount of lexical coverage necessary for vocabulary learning from video to take place. This is the temporal nature of listening and video texts. In reading situations, time pressures are not as much a factor because readers in most settings can control their reading speed. Viewers do not usually have the luxury of viewing at their own speed. Moreover, in contrast to reading, there is not usually an opportunity to easily review the text when viewing television (Huckin & Coady, 1999). Further research, examining vocabulary learning through viewing television is necessary to help clarify the relationship between lexical coverage and incidental vocabulary learning.

#### **4.3. Research questions**

Study 3 was designed to answer the following research questions:

1. Does comprehension of English-language television improve with increased lexical coverage?
2. Do English language learners with greater lexical coverage incidentally learn more vocabulary through viewing television than those learners with less lexical coverage?

#### **4.4. Participants**

The participants in Study 3 were the same as those described in Study 1 and Study 2. Data from the 321 participants from Study 1 was used for the analysis of the effects of lexical coverage on comprehension. Data from the 187 participants from Study 2 was used for the analysis of the effects of lexical coverage on incidental vocabulary acquisition. See Sections 2.3 and 3.3 for a full description of the participants from these studies. For the purposes of

Study 3, the participants from Study 1 are referred to as the Comprehension Participants and the participants from Study 2 as the Vocabulary Participants.

#### **4.5. Procedure**

Because the participants in Study 3 were the same as those in the previous studies in this thesis, the overall schedule and the in-class procedures were identical to those explained in the procedure sections from Studies 1 and 2. The experimental procedure and setting for these studies are described in Sections 2.4 and 3.4.

#### **4.6. Materials**

The materials used in Study 3 were the same as those used in Study 1 and Study 2. The VLT, which was only used peripherally in the previous studies, was used in this study to estimate vocabulary knowledge of the 2,000-, 3,000- and 5,000-word levels. Results from the three levels of the VLT were analyzed with the comprehension and vocabulary acquisition results from Study 1 and Study 2 to investigate how these aspects of language learning are affected by different levels of lexical coverage. The materials from Studies 1 and 2 are described in full in Sections 2.7 and 3.5.

##### **4.6.1. Vocabulary Levels Test**

In order to estimate the vocabulary knowledge of the participants, the VLT was used at the 2,000-, 3,000-, and 5,000-word levels. These tests were originally developed by Paul Nation (1983, 1990) and measure receptive vocabulary knowledge. There are two additional VLT levels (10,000 and the Academic Word List levels), but these were not used in this study as piloting showed the participants would be unlikely to have mastered them. Mastery of the 2,000- and the 3,000-word levels indicates that a language learner has the knowledge necessary to function effectively in English, while the 5,000-word level is the upper boundary of high-frequency vocabulary worth studying explicitly in a classroom setting (Nation, 1990).

The VLT is a matching test where respondents match three words or phrases with their definitions or synonyms from a list of six options. There are 30 items on each test with the items divided into 10 independent sections each containing six options and three items. The results of the test indicate the proportion of the words at each frequency level a learner knows. Two versions of the VLT were produced and published by Schmitt (2000) (Version 1) and Schmitt, Schmitt and Clapham (2001) (Version 2). In this study, Version 1 (Schmitt, 2000, pp. 192-200) of the VLT was used but its format was slightly modified from the published version so that Optical Mark Recognition could be used for scoring. While respondents were

asked to write the target word's number next to the definition in the original version, respondents in this study were asked to fill in the appropriate bubble for the target word. An example of the first set of questions at the 2,000-word level is shown in Figure 4.1. All levels of the VLT in this study used this format. The instructions for each level of the VLT were translated into Japanese (the L1 of the participants). This was to assist participants in completing the tests which may have had an unfamiliar format. The three levels of the VLT used in this thesis are shown in Appendices C1, C2, and C3.

Figure 4.1 First three items of the 2,000-word level of the VLT

| WORDS |           | MEANING | ANSWERS    |   |     |     |     |     |     |     |
|-------|-----------|---------|------------|---|-----|-----|-----|-----|-----|-----|
| A     | birth     | 1       | game       | 1 | (A) | (B) | (C) | (D) | (E) | (F) |
| B     | dust      | 2       | winning    | 2 | (A) | (B) | (C) | (D) | (E) | (F) |
| C     | operation |         |            |   |     |     |     |     |     |     |
| D     | row       | 3       | being born | 3 | (A) | (B) | (C) | (D) | (E) | (F) |
| E     | sport     |         |            |   |     |     |     |     |     |     |
| F     | victory   |         |            |   |     |     |     |     |     |     |

Source: adapted from Schmitt (2000)

Using the VLT has several advantages. The VLT is relatively simple to administer with a format that requires very little reading. This allowed for the three levels used in this study to be administered easily within one teaching session. The VLT is also a well-established test that has been rigorously validated (Read, 2000) and has been used in previous studies where an estimate of lexical coverage or vocabulary level was needed. This suggests that the VLT has both good content and face validity. It is assumed for this research that a participant with mastery of the 2,000-word VLT has a lexical coverage at the 2,000-word level for episodes of *Chuck*.

When scoring the VLT, a score of 24 out of 30<sup>5</sup> was used as the cut point for showing mastery of a level. This cut score is admittedly not very conservative but it has precedence as an indicator of mastery of a level. This 80% cut score was suggested by Schmitt (as cited in Xing & Fulcher, 2007, p. 184) and used in Xing and Fulcher's (2007) study of the reliability of two versions of the VLT. It is also comparable to the cut score (14 out of 18) used by Hu and Nation (2000) with an earlier version of the VLT at the 2,000-word level in their study of lexical coverage and reading comprehension.

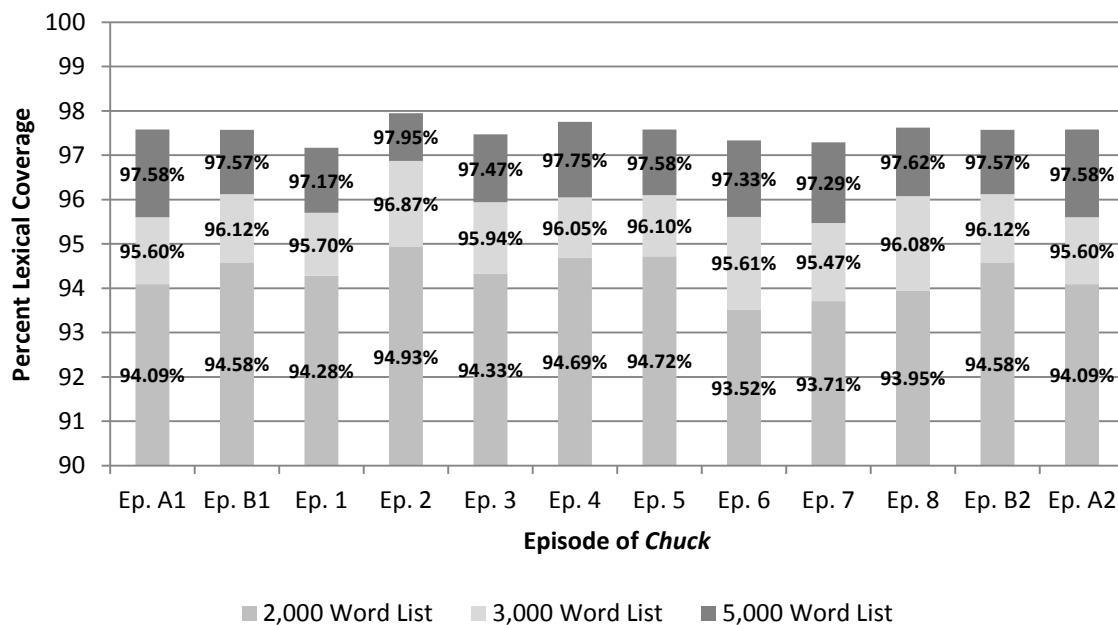
<sup>5</sup> The data in this study was also analyzed using the more conservative cut scores of 27 and 29 out of 30. The results were not greatly affected so the cut score of 24 out of 30 was used to allow for the inclusion of more participants at differing levels of lexical coverage.

## 4.7. Results

### 4.7.1. Vocabulary Levels Tests and the lexical coverage of episodes of *Chuck*

To determine whether aspects of viewing television: comprehension and vocabulary acquisition, improved with increased lexical coverage it was first necessary to determine the participants' lexical coverage for each of the 10 episodes. Results of the VLT were used to estimate the participants' lexical coverage of the episodes. To do this, first the vocabulary in the episodes was analyzed using the fourteen 1,000-word BNC lists (Nation, 2006). The procedure and full results of this analysis are described in Section 2.72. The participants' results on the VLT were then compared to this analysis. It is assumed that a participant with mastery of a level of the VLT has command of the vocabulary in the corresponding BNC word lists. Because the VLT is not based on the same word lists that are used to analyze the vocabulary in *Chuck*, it is only possible to approximate the lexical coverage of participants. Figure 4.2 presents the cumulative lexical coverage of each episode at the 2,000-, 3,000-, and 5,000-word levels. These estimates of lexical coverage include proper nouns and marginal words as these words may have a lower learning burden (Nation, 2006) and based on the proficiency level of the participants, it was reasonable to believe they would be able to understand the meanings of those words.

Figure 4.2 Cumulative lexical coverage (including proper nouns and marginal words) of the 10 episodes at the 2,000, 3,000 and 5,000 word-family levels



Three hundred and twenty-one Comprehension Participants were included in the analysis of lexical coverage and comprehension. To determine whether comprehension of episodes of

*Chuck* improved with increased lexical coverage it was first necessary to determine the participants' lexical coverage of the episodes. This was done by first analyzing the results of the three levels of the VLT. The mean score on the 2,000-word level test for the Comprehension Participants was 20.7 out of 30 with 95 participants achieving the mastery score of 24 out of 30. On the 3,000-word level test the mean score was 14.8 out of 30 with 14 participants achieving the mastery score of 24. Three participants achieved mastery of the 5,000-word level test where the mean score was 13.5 out of 30. There were 209 participants who failed to obtain mastery of the VLT at the 2,000-word level. Table 4.2 shows the results of the three levels of the VLT including the mean scores and the number of participants achieving mastery of each level.

Table 4.2 Scores on the VLTs for Comprehension Participants

| Levels Test | N   | Mean | SD  | Median | Minimum | Maximum | Mastery† |
|-------------|-----|------|-----|--------|---------|---------|----------|
| VLT 2000    | 321 | 20.7 | 5.4 | 21     | 6       | 30      | 95       |
| VLT 3000    | 321 | 14.8 | 5.0 | 15     | 4       | 29      | 14       |
| VLT 5000    | 321 | 13.5 | 4.6 | 14     | 2       | 26      | 3        |

Note. †Mastery of VLT set at 24 out of 30.

Two hundred and twenty-six Vocabulary Participants were included in the analysis of lexical coverage and vocabulary acquisition. Their mean score on the 2,000-word level test was 19.0 out of 30 with 39 participants achieving the mastery score of 24 out of 30. On the 3,000-word level test, the mean score was 13.4 out of 30 with two participants achieving the mastery score of 24. One participant achieved mastery of the 5,000-word level test. The mean score on that test was 12.0 out of 30. There were 145 participants who failed to obtain mastery of the VLT at the 2,000-word level. Table 4.3 displays the results of the three levels of the VLT including the mean scores and the number of participants achieving mastery of each level.

Table 4.3 Scores on the VLTs for Vocabulary Participants

| Levels Test | N   | Mean | SD  | Median | Minimum | Maximum | Mastery† |
|-------------|-----|------|-----|--------|---------|---------|----------|
| VLT 2000    | 187 | 19.0 | 5.5 | 19     | 6       | 30      | 39       |
| VLT 3000    | 187 | 13.4 | 4.2 | 13     | 4       | 26      | 2        |
| VLT 5000    | 187 | 12.0 | 4.3 | 12     | 2       | 26      | 1        |

Note. †Mastery of VLT set at 24 out of 30.

The 95 Comprehension Participants and 39 Vocabulary Participants who achieved mastery of the VLT at the 2,000-word level had coverage of the episodes of *Chuck* ranging

from 93.52% (Episode 6) to 94.93% (Episode 2). Their average lexical coverage across the 10 episodes was 94.28%. The 14 Comprehension Participants and two Vocabulary Participants with mastery of the VLT at the 3,000-word level had an average coverage of 95.95% for the 10 episodes viewed. Episode 7, with 95.47% coverage, was the lowest for these participants, while Episode 2 was the highest at 96.87%. The three Comprehension Participants and one Vocabulary Participant who achieved mastery of the VLT at the 5,000-word level had an average lexical coverage of 97.53%, a maximum coverage of 97.95% in Episode 2, and a minimum of 97.17% in Episode 1. It is not possible to estimate the coverage of the participants who did not obtain a mastery score of any of the VLTs but it can be assumed that it is less than the lexical coverage at the 2,000-word level.

#### **4.7.2. The effects of lexical coverage on comprehension**

The Comprehension Participants and their scores on the 10 comprehension tests for the 10 episodes of *Chuck* were grouped by their performance on the VLTs. Table 4.4 shows the mean comprehension test scores for the episodes for the 209 participants who failed to master the VLT at the 2,000-word level. Table 4.5 shows the comprehension scores of the 95 participants who achieved mastery of the 2,000-word level. Table 4.6 shows the comprehension scores of the 14 participants who achieved mastery at the 3,000-word level, and Table 4.7 shows the comprehension scores of the three participants who achieved mastery of the 5,000-word level. In these tables, the average raw score and average raw percentage for the Initial and Final Episodes are not presented because the participants had different first and last episodes based on their viewing group (different viewing groups are explained in Section 2.4.2). The mean comprehension scores for these episodes are also expressed in terms of ability measures known as CHIP scores which are raw test scores analyzed by the Rasch Model to obtain interval data in a range from 1 to 100. For a full explanation of CHIPs see Section 2.8.3. The mean CHIP score for participants not reaching mastery of the 2,000-word level across all 10 episodes viewed was 53.3. The mean CHIP score of participants that reached the 2,000-word level was 54.1, while the mean for the group reaching the 3,000-word level was 54.2. The mean CHIP score for participants reaching mastery of the 5,000-word level was 56.0. The highest mean CHIP score for any single episode was by the group of participants reaching mastery of the 5,000-word level and was 59.3 (Episode 2). The lowest mean CHIP score for any single episode was by the group of participants not reaching mastery of the 2,000-word level and was 50.8 (Initial Episode). Figure 4.3 shows the variation

in mean comprehension scores between the 10 episodes for participants grouped by their mastery of the different levels of the VLT.

Table 4.4 Mean comprehension scores for Comprehension Participants not reaching mastery of any VLT ( $n=209$ )

| Mean Score | Initial | Ep. 1 | Ep. 2 | Ep. 3 | Ep. 4 | Ep. 5 | Ep. 6 | Ep. 7 | Ep. 8 | Final |
|------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| CHIP       | 50.8    | 54.6  | 55.4  | 55.0  | 52.4  | 53.3  | 52.7  | 52.3  | 54.0  | 52.9  |
| Raw        |         | 48.0  | 53.4  | 52.9  | 42.8  | 45.7  | 46.3  | 46.0  | 51.1  |       |
| Raw %      |         | 68.6% | 70.2% | 69.6% | 60.2% | 63.5% | 61.7% | 59.7% | 65.5% |       |

Table 4.5 Mean comprehension scores for Comprehension Participants reaching mastery of the VLT at the 2,000-word level ( $n=95$ )

| Mean Score | Initial | Ep. 1 | Ep. 2 | Ep. 3 | Ep. 4 | Ep. 5 | Ep. 6 | Ep. 7 | Ep. 8 | Final |
|------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| CHIP       | 51.2    | 55.8  | 55.9  | 55.8  | 52.7  | 54.3  | 53.6  | 52.8  | 54.8  | 53.6  |
| Raw        |         | 50.3  | 54.4  | 54.7  | 43.6  | 47.6  | 48.1  | 47.1  | 52.8  |       |
| Raw %      |         | 71.8% | 71.6% | 72.0% | 61.5% | 66.1% | 64.2% | 61.2% | 67.7% |       |

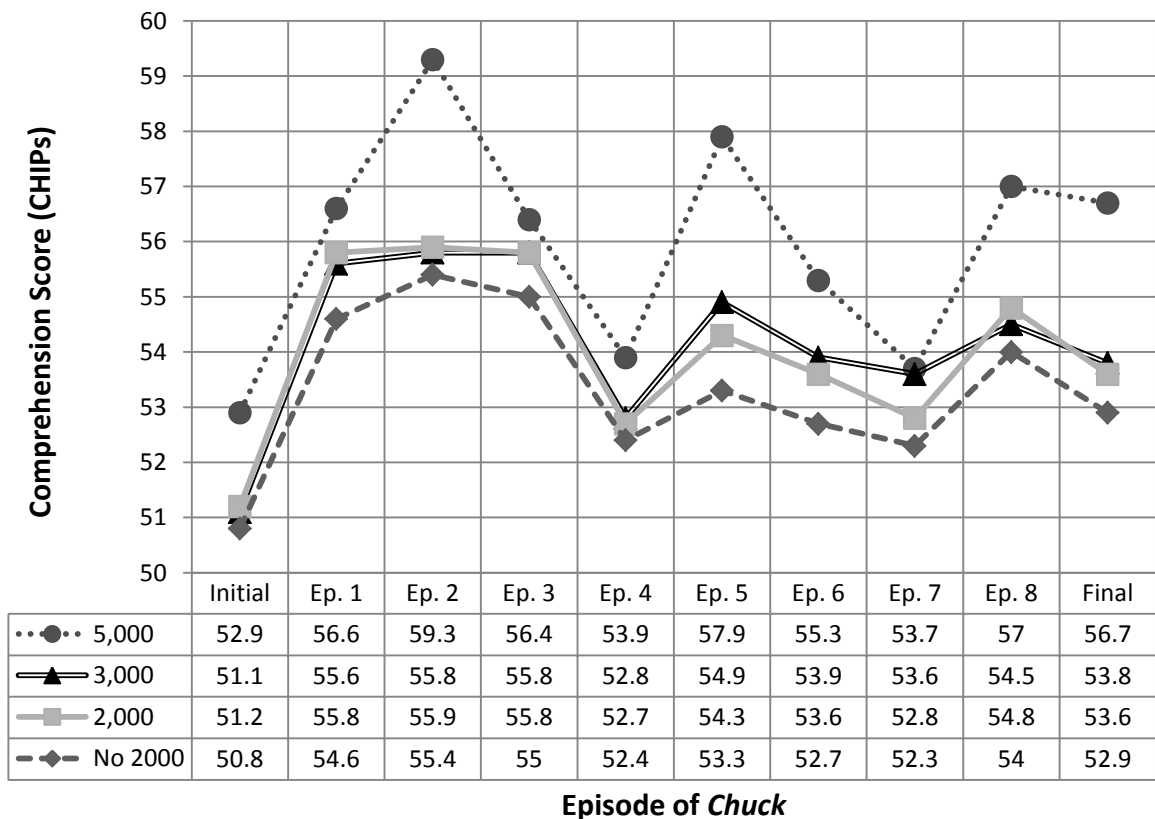
Table 4.6 Mean comprehension scores for Comprehension Participants reaching mastery of the VLT at the 3,000-word level ( $n=14$ )

| Mean Score | Initial | Ep. 1 | Ep. 2 | Ep. 3 | Ep. 4 | Ep. 5 | Ep. 6 | Ep. 7 | Ep. 8 | Final |
|------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| CHIP       | 51.1    | 55.6  | 55.8  | 55.8  | 52.8  | 54.9  | 53.9  | 53.6  | 54.5  | 53.8  |
| Raw        |         | 49.9  | 54.1  | 54.9  | 44.1  | 49.1  | 48.7  | 49.5  | 52.2  |       |
| Raw %      |         | 71.3% | 71.2% | 72.2% | 62.1% | 68.2% | 65.0% | 64.3% | 66.9% |       |

Table 4.7 Mean comprehension scores for Comprehension Participants reaching mastery of the VLT at the 5,000-word level ( $n=3$ )

| Mean Score | Initial | Ep. 1 | Ep. 2 | Ep. 3 | Ep. 4 | Ep. 5 | Ep. 6 | Ep. 7 | Ep. 8 | Final |
|------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| CHIP       | 52.9    | 56.6  | 59.3  | 56.4  | 53.9  | 57.9  | 55.3  | 53.7  | 57.0  | 56.7  |
| Raw        |         | 51.7  | 62.0  | 56.0  | 46.3  | 56.0  | 52.3  | 49.7  | 58.7  |       |
| Raw %      |         | 73.8% | 81.6% | 73.7% | 65.3% | 77.8% | 69.8% | 64.5% | 75.2% |       |

Figure 4.3 Mean comprehension scores in CHIPs for Comprehension Participants with mastery of different levels of the VLT across the 10 episodes of *Chuck*



Differences among the mean comprehension scores for each episode by participants with the different lexical coverage levels were assessed using a series of one-way ANOVAs. A Levene test of homogeneity of variance conducted prior to the ANOVAs did not indicate the assumption of homogeneity of variance was significantly violated for any episode, ( $p > .05$ ). There was a significant but small effect (as indicated by  $\eta^2$  values) of lexical coverage, as represented by mastery of the 2,000-, 3,000- and 5,000-word levels, on comprehension for six of the 10 episodes: Episode 1, Episode 3, Episode 5, Episode 6, Episode 8, and the Final Episode. A summary of these analyses is shown in Table 4.8. Post hoc comparisons (LSD) of the four groups indicated that for these six episodes the comprehension scores for the participants with mastery of the 2,000-word level were significantly higher than the comprehension scores for the participants without mastery of any level of the VLT. For Episode 5, the post hoc comparison revealed that the comprehension scores for the participants with mastery of the 3,000-word level were significantly higher than the comprehension scores for the participants without mastery of any level of the VLT. Also in Episode 5, the comprehension scores for the participants with mastery of the 5,000-word level were significantly higher than the comprehension scores for the participants without mastery

of any level of the VLT and the participants with mastery of the 2,000-word level. For the Final Episode, the post hoc comparison also revealed that the comprehension scores for the participants with mastery of the 5,000-word level were significantly higher than the comprehension scores for the participants without mastery of any level of the VLT and the participants with mastery of the 2,000-word level. A summary of the post hoc analyses is shown in Table 4.9.

Table 4.8 Summary of ANOVA results for comprehension of episodes of *Chuck* and lexical coverage

|                 | <i>Df</i> | <i>F</i> | $\eta^2$ | <i>p</i> |
|-----------------|-----------|----------|----------|----------|
| Initial Episode | 3, 317    | 1.55     | .01      | .201     |
| Episode 1       | 3, 317    | 4.52     | .04      | .004**   |
| Episode 2       | 3, 317    | 2.27     | .02      | .081     |
| Episode 3       | 3, 317    | 2.78     | .03      | .041*    |
| Episode 4       | 3, 317    | 0.70     | .01      | .549     |
| Episode 5       | 3, 317    | 6.84     | .06      | .000***  |
| Episode 6       | 3, 317    | 3.68     | .03      | .012*    |
| Episode 7       | 3, 317    | 1.94     | .02      | .123     |
| Episode 8       | 3, 317    | 2.66     | .02      | .048*    |
| Final Episode   | 3, 317    | 3.81     | .03      | .011*    |

Note. \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

Table 4.9 LSD post hoc comparison of comprehension of episodes among the four groups of lexical coverage

|               | Finding      | <i>Sig.</i> |
|---------------|--------------|-------------|
| Episode 1     | 2K > Less 2K | .001*       |
| Episode 3     | 2K > Less 2K | .008*       |
| Episode 5     | 2K > Less 2K | .002*       |
|               | 3K > Less 2K | .028*       |
|               | 5K > Less 2K | .003*       |
|               | 5K > 2K      | .003*       |
| Episode 6     | 2K > Less 2K | .007*       |
| Episode 8     | 2K > Less 2K | .030*       |
| Final Episode | 2K > Less 2K | .030*       |
|               | 5K > Less 2K | .011*       |
|               | 5K > 2K      | .039*       |

Note. \* $p < .05$ , Less 2K=no mastery at any word level, 2K=mastery at 2,000-word level, 3K=mastery at 3,000-word level, 5K=mastery of the 5,000-word level

Across the six episodes where significant differences were identified, the difference between participants with mastery of the 2,000-word level and participants without mastery of

2,000-word level was most commonly identified as significant. The mean coverage for participants with mastery of the 2,000-word level was 94.2% (including proper nouns and marginal words). For the participants without mastery of the 2,000-word level, coverage is considered to be less than 94.2%. The average comprehension score for participants with mastery of the 2,000-word level was 66.7% while the average comprehension score for participants without mastery of the 2,000-word level was 64.1%. For these six episodes, participants with mastery of the 2,000-word level had a range of comprehension scores from 62.5% in the Final Episode to 72.0% in Episode 3. The participants without mastery of the 2,000-word level had a range of raw comprehension scores from 61.4% in the Final Episode to 69.2% in Episode 3.

In Episode 5, there was a significant difference between participants with mastery of the 3,000-word level and participants without mastery of the 2,000-word level. The mean comprehension score for the participants with mastery of the 3,000 level was 68.2% with a lexical coverage of 96.10%. The mean comprehension score for the participants without mastery of the 2,000 level was 62.6% with a lexical coverage of less than 94.72%.

In Episode 5 and the Final Episode, there were significant differences between the participants with mastery of the 5,000-word level and participants with and without mastery of the 2,000-word level. For both episodes, the lexical coverage for the participants with mastery of the 5,000-word level was 97.58%. The mean comprehension score for this group of participants was 77.8% for Episode 5 and 73.9% for the Final Episode. Lexical coverage for the participants with mastery of the 2,000-word level was 94.27% for Episode 5 and 94.34% for the Final Episode. Mean comprehension for these participants was 66.1% for Episode 5 and 63.2% for the Final Episode. Participants without mastery of the 2,000-word level had a mean comprehension score of 62.6% for Episode 5 and 61.0% for the Final Episode. For the six episodes where significant differences were found, the lexical coverage and mean comprehension scores for the groups of participants are presented in Table 4.10.

Table 4.10 Mean comprehension scores of significantly different groups of participants with different lexical coverage

|                        | Episode 1    |          | Episode 3    |          | Episode 5      |          |
|------------------------|--------------|----------|--------------|----------|----------------|----------|
| Significant Difference | 2K > Less 2K |          | 2K > Less 2K |          | 2K > Less 2K   |          |
| Comprehension Score    | 71.8%        | 67.8%    | 72.0%        | 69.2%    | 66.1%          | 62.6%    |
| Lexical Coverage       | 94.93%       | < 94.93% | 94.33%       | < 94.33% | 94.72%         | < 94.72% |
| Significant Difference |              |          |              |          | 3K > Less 2K   |          |
| Comprehension Score    |              |          |              |          | 68.2%          | 62.6%    |
| Lexical Coverage       |              |          |              |          | 96.10%         | < 94.72% |
| Significant Difference |              |          |              |          | 5K > Less 2K   |          |
| Comprehension Score    |              |          |              |          | 77.8%          | 62.6%    |
| Lexical Coverage       |              |          |              |          | 97.58%         | < 94.72% |
| Significant Difference |              |          |              |          | 5K > 2K        |          |
| Comprehension Score    |              |          |              |          | 77.8%          | 66.1%    |
| Lexical Coverage       |              |          |              |          | 97.58%         | 94.72%   |
|                        | Episode 6    |          | Episode 8    |          | Final Episode† |          |
| Significant Difference | 2K > Less 2K |          | 2K > Less 2K |          | 2K > Less 2K   |          |
| Comprehension Score    | 71.8%        | 67.8%    | 72.0%        | 69.2%    | 63.2%          | 61.0%    |
| Lexical Coverage       | 93.52%       | < 93.52% | 93.95%       | < 93.95% | 94.34%         | < 94.34% |
| Significant Difference |              |          |              |          | 5K > Less 2K   |          |
| Comprehension Score    |              |          |              |          | 73.9%          | 61.0%    |
| Lexical Coverage       |              |          |              |          | 97.58%         | < 94.34% |
| Significant Difference |              |          |              |          | 5K > Less 2K   |          |
| Comprehension Score    |              |          |              |          | 73.9%          | 61.0%    |
| Lexical Coverage       |              |          |              |          | 97.58%         | < 94.34% |

Note. †Mean comprehension score and mean lexical coverage of the two viewing groups (Episode A2 and B2)

These analyses indicate that for English-language television, increased lexical coverage can lead to increased comprehension but comprehension can also be episode dependent. For episodes where differences in lexical coverage were not found to be significantly linked to comprehension other factors such as the content of the episode, the relation of the episode to previous episodes viewed, or composition of the comprehension tests may be interfering with the comprehension and coverage relationship.

#### 4.7.3. The effects of lexical coverage on incidental vocabulary acquisition

The Vocabulary Participants and their vocabulary gain scores on the Tough and Sensitive Tests were grouped by their performance on the three levels of the VLT. Table 4.11 shows the mean real gain scores for the 145 participants who failed to master the VLT at the 2,000-word

level and Table 4.12 shows their mean relative gain scores. Table 4.13 shows the mean real gain scores of the 39 participants who achieved mastery of the 2,000-word level and Table 4.14 shows their mean relative gain scores. Table 4.15 shows the mean real gain scores of the two participants who achieved mastery at the 3,000-word level and Table 4.16 shows their mean relative gain scores. Table 4.17 shows the mean real gain scores of the one participant who achieved mastery at the 5,000-word level and Table 4.18 shows their mean relative gain score.

The mean real gain of the participants not reaching mastery of the 2,000-word level across all 10 episodes viewed was 6.70 words on the Tough Test and 6.99 words on the Sensitive Test. This group had mean relative gains scores of 23.0% and 29.2% respectively on the two vocabulary tests. The mean real gain of participants that had mastery of the 2,000-word level was 5.15 words (22.4% relative gain) on the Tough Test and 6.23 words (31.7% relative gain) on the Sensitive Test. The mean real gain of two participants that reached the 3,000-word level was 5.0 words (28.8% relative gain) on the Tough Test and 2.0 words (15.7% relative gain) on the Sensitive Test. The mean real gain of the one participant that reached the 5,000-word level was 7.0 words (46.7% relative gain) on the Tough Test and 7.0 words (41.2% relative gain) on the Sensitive Test.

Table 4.11 Vocabulary gains for Tough and Sensitive tests for Vocabulary Participants not reaching mastery of any VLT ( $n=145$ )

| Vocabulary Test | Mean Real Gain | SD   | Median Real Gain | Minimum Real Gain | Maximum Real Gain |
|-----------------|----------------|------|------------------|-------------------|-------------------|
| Tough           | 6.70           | 2.97 | 7                | 1                 | 20                |
| Sensitive       | 6.99           | 3.47 | 7                | 1                 | 23                |

Table 4.12 Relative vocabulary gains for Tough and Sensitive tests for Vocabulary Participants not reaching mastery of any VLT ( $n=145$ )

| Vocabulary Test | Mean Relative Gain | Median Relative Gain | Minimum Relative Gain | Maximum Relative Gain |
|-----------------|--------------------|----------------------|-----------------------|-----------------------|
| Tough           | 23.0%              | 23.3%                | 2.9%                  | 45.7%                 |
| Sensitive       | 29.2%              | 27.8%                | 4.3%                  | 58.1%                 |

Table 4.13 Vocabulary gains for Tough and Sensitive tests for Vocabulary Participants reaching mastery of the VLT at the 2,000-word level ( $n=39$ )

| Vocabulary Test | Mean Real Gain | SD   | Median Real Gain | Minimum Real Gain | Maximum Real Gain |
|-----------------|----------------|------|------------------|-------------------|-------------------|
| Tough           | 5.15           | 2.10 | 5                | 1                 | 12                |
| Sensitive       | 6.23           | 3.66 | 5                | 1                 | 15                |

Table 4.14 Relative vocabulary gains for Tough and Sensitive tests for Vocabulary Participants reaching mastery of the VLT at the 2,000-word level ( $n=39$ )

| Vocabulary Test | Mean Relative Gain | Median Relative Gain | Minimum Relative Gain | Maximum Relative Gain |
|-----------------|--------------------|----------------------|-----------------------|-----------------------|
| Tough           | 22.4%              | 22.2%                | 4.3%                  | 44.4%                 |
| Sensitive       | 31.7%              | 30.0%                | 5.7%                  | 70.0%                 |

Table 4.15 Vocabulary gains for Tough and Sensitive tests for Vocabulary Participants reaching mastery of the VLT at the 3,000-word level ( $n=2$ )

| Vocabulary Test | Mean Real Gain | SD   | Median Real Gain | Minimum Real Gain | Maximum Real Gain |
|-----------------|----------------|------|------------------|-------------------|-------------------|
| Tough           | 5.00           | 0    | 5                | 5                 | 5                 |
| Sensitive       | 2.00           | 1.41 | 2                | 1                 | 3                 |

Table 4.16 Relative vocabulary gains for Tough and Sensitive tests for Vocabulary Participants reaching mastery of the VLT at the 3,000-word level ( $n=2$ )

| Vocabulary Test | Mean Relative Gain | Median Relative Gain | Minimum Relative Gain | Maximum Relative Gain |
|-----------------|--------------------|----------------------|-----------------------|-----------------------|
| Tough           | 28.8%              | 28.8%                | 26.3%                 | 31.3%                 |
| Sensitive       | 15.7%              | 15.7%                | 8.3%                  | 23.1%                 |

Table 4.17 Vocabulary gains for Tough and Sensitive tests for Vocabulary Participant reaching mastery of the VLT at the 5,000-word level ( $n=1$ )

| Vocabulary Test | Mean Real Gain | SD | Median Real Gain | Minimum Real Gain | Maximum Real Gain |
|-----------------|----------------|----|------------------|-------------------|-------------------|
| Tough           | 7.00           | -- | 7                | 7                 | 7                 |
| Sensitive       | 7.00           | -- | 7                | 7                 | 7                 |

Table 4.18 Relative vocabulary gains for Tough and Sensitive tests for Vocabulary Participant reaching mastery of the VLT at the 5,000-word level ( $n=1$ )

| Vocabulary Test | Mean Relative Gain | Median Relative Gain | Minimum Relative Gain | Maximum Relative Gain |
|-----------------|--------------------|----------------------|-----------------------|-----------------------|
| Tough           | 46.7%              | 46.7%                | 46.7%                 | 46.7%                 |
| Sensitive       | 41.2%              | 41.2%                | 41.2%                 | 41.2%                 |

To investigate whether there was a difference between participants with different vocabulary levels and their relative vocabulary gains on the Tough and Sensitive Tests through watching 10 episodes of television four one-way ANOVAs were run. The first two ANOVAs compared the relative vocabulary gains on both vocabulary tests of those participants who did not have mastery of the 2000-word level VLT, those who had mastery of the 2,000-word VLT, those who had mastery of the 3000-word VLT, and those who had

mastery of the 5,000-word VLT. Neither ANOVA found any significant difference between the four groups of participants: Tough Test [ $F(3,183) = 2.584, p = .055$ ], and Sensitive Test [ $F(3,183) = 1.553, p = .202$ ]. Because there were only two participants with mastery of the 3,000-word level vocabulary test and one participant with mastery of the 5,000-word level, they may not have been functioning well as a comparison groups. With this considered, the third and fourth ANOVAs compared the difference between the relative vocabulary gains on the Tough and Sensitive Tests of the participants who did not have mastery of the 2000-word VLT, and the combined group of the participants who had mastery of the 2,000-word level, 3000-word level, and 5000-word level. Neither ANOVA found any significant difference between the two groups: Tough Test [ $F(1,185) = 0.039, p = .844$ ], and Sensitive Test [ $F(1,185) = 0.830, p = .363$ ]. Results indicate no significant relationship between relative gain and mastery of the VLT at the 2,000-, 3,000- or 5,000-word level.

#### **4.8. Summary of findings**

The main findings for Study 3 can be summarized as follows:

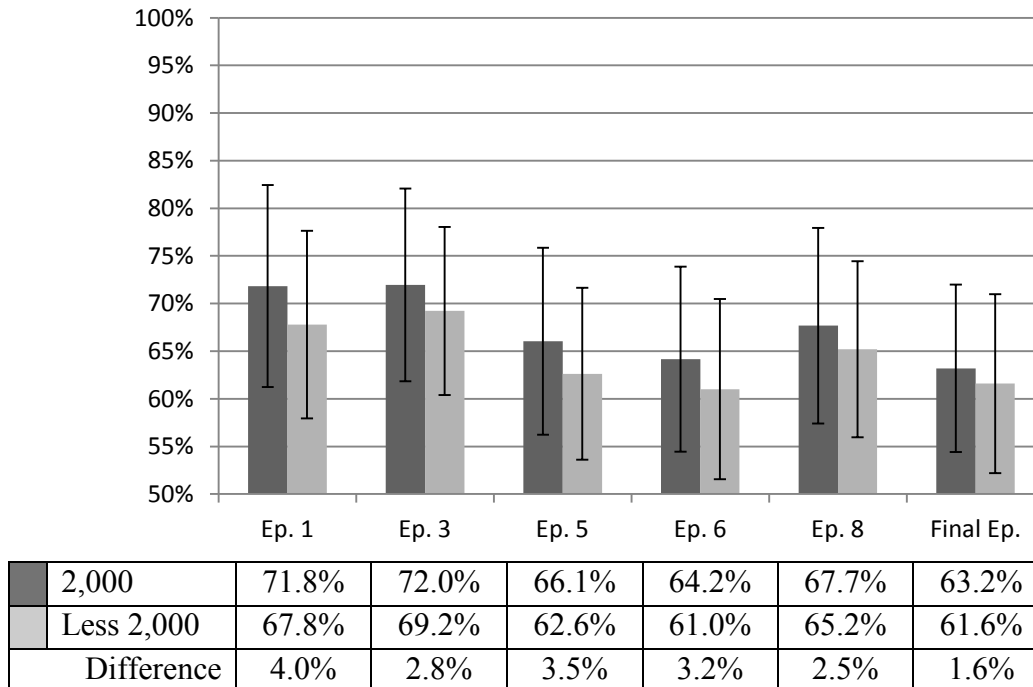
1. Comprehension of English-language television was shown to improve with increased lexical coverage for some but not all episodes of *Chuck*. Participants with an average lexical coverage of 94.28% (as indicated by mastery of the 2,000-word level) were found to have significantly higher comprehension scores than those participants with an average lexical coverage less than 94.28% (as indicated by failure to master the 2,000-word level) for six of the 10 of episodes. For these six episodes, the participants with an average of 94.28% lexical coverage had a mean comprehension score of 66.7% while the participants with less than 94.28% lexical coverage had an average comprehension score of 64.1%. For Episode 5, the participants with mastery of the 3,000 word-level (96.10% lexical coverage) had significantly higher comprehension scores than those participants who failed to master the 2,000-word level. For Episode 5 and the Final Episode, participants with mastery of the 5,000-word level were found to have significantly higher comprehension than those participants who had mastery of the 2,000-word level and participants who failed to master the 2,000-word level. The mean lexical coverage for participants with mastery of the 5,000-word level for these two episodes was 97.58%.
2. Incidental vocabulary acquisition from English-language television was not shown to improve with increased lexical coverage.

## 4.9. Discussion

### 4.9.1. Lexical coverage and comprehension

The results from Study 3 demonstrate that lexical knowledge contributes to comprehension of episodes of television viewed in the L2. The mean lexical coverage of the six episodes where comprehension scores were significantly higher for the participants with and without mastery of the 2,000-word level may indicate the coverage level needed for adequate comprehension of television. The lexical coverage afforded by mastery of the 2,000-word level across the six episodes averaged over 94% but had a coverage ranging from 93.52% (Episode 6) to 94.72% (Episode 5). This coverage figure is lower than the 95% figure suggested by Webb and Rodgers (2009a) for the lower boundary of adequate comprehension. Webb and Rodgers do not define what constitutes adequate comprehension. Figure 4.4 shows the mean comprehension scores for the participants with mastery of the 2,000-word level and the participants without mastery of the 2,000-word level.

Figure 4.4 Mean comprehension scores for the participants with and without mastery of the 2,000-word level for episodes where there was a significant difference between the groups. Error bars show the mean plus or minus one standard deviation



The 2,000-word level group had a mean comprehension score of 67.5% for the six episodes shown in Figure 4.4, but the mean comprehension scores ranged from 63.2% (Final Episode) to 72.0% (Episode 3). Adequate comprehension is obviously a subjective term with

the level of comprehension that is considered adequate situationally dependent. If the television program were viewed in a situation like it was – successive episodes over a relatively long period of time – adequate comprehension is surely the amount of comprehension necessary so that viewing was sufficiently enjoyable for the language learners to continue viewing. This is similar to extensive reading programs where reading should be a pleasurable experience where learners build fluency in reading and have opportunities to learn vocabulary through multiple exposures. It is perhaps more appropriate to say that for these six episodes mastery of the 2,000-word level afforded participants the minimum amount of comprehension for viewing. But even that assertion is difficult to make as the error bars in Figure 4.4 indicate. They represent plus or minus one standard deviation from the mean comprehension score and show that there was a good deal of variation in comprehension for participants with and without mastery of the 2,000-word level. Many participants without mastery of the 2,000-word level had higher comprehension scores than the mean comprehension score for participants with mastery of the 2,000-word level. It is perhaps more appropriate to state that the lexical coverage afforded by mastery of the 2,000-word level is an indicator of when comprehension is likely but this can be dependent on the television program and the language learner.

The lexical coverage necessary for the upper boundary of adequate comprehension as defined by Webb and Rodgers (2009) was 98%. No participant in Study 3 reached that level of coverage of an episode of *Chuck*. The 97.95% lexical coverage of Episode 2 by the participants with mastery of the 5,000-word level was the highest coverage level across the 10 episodes. While the mean comprehension score (81.6%) for this group was the highest of any lexical coverage group in any episode viewed in this study there was no statistical difference between the groups of participants with different coverage levels for Episode 2. This may be in part due to the lower coverage groups having very similar mean comprehension scores (70.2%, 71.6% and 71.2%) and the small number of participants with mastery of the 5,000-word level. Unfortunately the results from Study 3 do not provide conclusive support for 98% lexical coverage being the upper boundary of adequate comprehension. The results, however, do indicate that a relationship between increased coverage and increased comprehension exists.

The findings from Study 3 are comparable to the findings from previous lexical coverage and reading comprehension research. While participants in this study did not reach the 98% or higher lexical coverage figure suggested by some as necessary for ideal comprehension (Carver, 1994; Hirsh & Nation, 1992; Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski,

2010; Nation, 2006; Schmitt et al., 2011), the 94% coverage figure identified in this study is similar to the lower coverage figures identified in some studies. The 94% lexical coverage figure, acknowledged as the level at which comprehension of television is likely, is comparable to Hu and Nation's (2000) study where a minority of participants with 95% coverage had adequate coverage, Laufer's (1989) study where reasonable comprehension began at 95%, Laufer and Ravenhorst-Kalovski's (2010) study that found 95% coverage to be necessary for minimal comprehension, and Schmitt, Jiang and Grabe's (2011) study where 95% coverage was considered sufficient for 60% comprehension.

The 94% coverage figure proposed in this study for when comprehension of television is likely is also comparable to prior lexical coverage studies involving listening. While some listening studies have posited that 98% lexical coverage is necessary for comprehension (Nation, 2006; Stæhr, 2009), other studies suggested that the lexical coverage necessary for listening comprehension is lower than for reading comprehension. Bonk (2000) found that a coverage of 90% led to good listening comprehension of short listening passages. Van Zeeland and Schmitt (2012) suggested 98% coverage for high comprehension, but in situations where more lenient comprehension levels are acceptable, 95% and 90% coverage is sufficient. In their study on comprehension of television, Webb and Rodgers' (2009a) suggested that the combination of visual and aural input may allow a coverage level of 95% to be sufficient for comprehension.

An interesting finding from this study was the difference in comprehension scores between participants with mastery of the 2,000-word level and those with mastery of 3,000-word level. For six of the 10 episodes of *Chuck*, participants with 2,000-word level mastery had significantly higher comprehension scores than those participants without mastery of the 2,000-word level. Only in Episode 5 did participants with mastery of the 3,000-word level have significantly better comprehension results than the participants without mastery of the 2,000-word level. In Figure 4.3 which shows the mean comprehension scores for participants with mastery of the different levels of the VLT, the mean scores of the participants with mastery of the 3,000-word level are only higher than participants with mastery of the 2,000-word level for four of 10 episodes and never significantly higher. A possible explanation for this might be the relatively small amount of lexical coverage that is provided by knowing the 3,000-word level compared with knowing the 2,000-word level. Across the 10 episodes, the mean increase in lexical coverage is only 1.67%. Considering that knowledge of the 2,000 most frequent words of English plus proper nouns and marginal words provides 94.29% coverage the increase in lexical coverage from knowing 1,000 more words is relatively small.

While the results from Study 3 indicate that knowing more vocabulary leads to better comprehension, an increase of 1.67% appears to not be large enough to consistently make a significant difference in understanding.

This finding adds more support for the importance of language learners knowing and explicitly studying the first 2,000 words of English. This should include not just the form and meaning aspect of knowledge but also developing depth of vocabulary knowledge. If one of the goals of a language learner is to view authentic television, then development of knowledge of the aural form of the vocabulary in the first 2,000 words is even more important. While it should be a goal of language learners after mastering the 2,000-word level to master the 3,000-word level the relatively small amount of added coverage that this 1,000 words represents means that it may be more efficient to deal with frequently occurring vocabulary beyond the first 2,000 words through glossing or preteaching.

Something salient from the findings in Study 3 is that the relationship between lexical coverage and comprehension differs between episodes of the same television program. There was a relationship between increased coverage and comprehension for only six of the 10 episodes viewed. This contrasts with findings from van Zeeland and Schmitt's (2012) study of lexical coverage and listening comprehension where there were similar results regarding the relationship between coverage and comprehension across the four texts used in the study. The lack of a relationship between coverage and comprehension may be somewhat expected for the Initial Episode. Participants were encountering a television program which they knew nothing about for the first time and they had yet to build up background information about the series which could be used to aid comprehension. Conceivably, this is a viewing situation that language learners would have a difficult time with regardless of their lexical coverage of the episode. The inconsistency between subsequent episodes, however, is somewhat more puzzling. For example, the comprehension scores from Episode 5 indicate a strong relationship between comprehension and coverage while the scores from Episode 7 do not. In Episode 5, mean comprehension scores increased with increased coverage and there were significant differences between groups of participants. In Episode 7, while there is a pattern of increasing comprehension with higher lexical coverage, the difference between comprehension scores at the different levels is very slight. The comprehension scores for Episode 5 and Episode 7 for each vocabulary level are shown in Table 4.19.

Table 4.19 Comparison of comprehension scores for Episode 5 and Episode 7 for participants at four vocabulary levels

| Vocabulary Level           | Episode 5 | Episode 7 |
|----------------------------|-----------|-----------|
| Less than 2,000-word level | 63.5%     | 59.7%     |
| 2,000-word level           | 66.1%     | 61.2%     |
| 3,000-word level           | 68.2%     | 64.3%     |
| 5,000-word level           | 77.8%     | 64.5%     |

There are several possible explanations for why there is a lack of consistency between comprehension and coverage of texts in Study 3 that was demonstrated in van Zeeland and Schmitt's study. The first explanation is that differences between the input texts affect results. Study 3 employed episodes of authentic television to which estimates of the learners' vocabulary size were compared to estimate lexical coverage. In van Zeeland and Schmitt's study, known vocabulary was controlled to produce different coverage levels by replacing words outside the most frequent 2,000 words of English with nonwords. The use of authentic video texts in Study 3 means that it would have been impossible to control the known and unknown vocabulary in this manner. While the texts in van Zeeland and Schmitt's study were chosen to be authentic listening texts they were also chosen to be as similar to one another as possible and their difficulty was controlled. While there was some similarity between the episodes used in Study 3 as they were all from the same season of the same program, the relative difficulty of the episodes was something that could not be controlled, nor would be controllable, under typical viewing conditions.

Another possible explanation for the inconsistent relationship between coverage and comprehension for episodes in Study 3 is the length of these episodes. Previous research involving listening and lexical coverage has generally used very short listening tasks. The four passages used by Bonk (2000) ranged from 39 seconds to 43 seconds and averaged 85 running words. There were seven different listening passages in Stæhr's (2009) study. There were four short passages ranging from 45 seconds to 1 minute (142 to 195 tokens) and three longer passages ranging from 3.15 minutes to 3.30 minutes (558 to 665 tokens) in length. The four texts in the study by van Zeeland and Schmitt (2012) ran for between 1.58 minutes and 2.08 minutes (471 to 484 tokens). In contrast, the episodes of *Chuck* were an average of 42 minutes and 49 seconds in length and averaged 5,203 tokens with 7 minute viewing periods between which comprehension responses might be completed. The texts used in the listening comprehension studies were more of the type that might be encountered in a listening comprehension test but perhaps not in a real world situation. The longer videos viewed in this

study represent something that people might regularly encounter and view for enjoyment. It is unrealistic to expect that people viewing episodes of *Chuck* try and remember everything that happened or even understand everything that occurred. Because of the relatively long viewing times for episodes of television, comprehension may have a less consistent relationship with lexical coverage than with much shorter listening passages.

One more feature that may have contributed to the inconsistent relationship between increased lexical coverage and improved comprehension is the presence of images that are a significant part of the viewing experience. Participants in the previous reading and listening studies did not have access to images accompanying the input. Viewers of television programs obviously have access to and rely on the images that accompany the aural input to support comprehension. While the comprehension questions in Study 3 were all based on the spoken dialogue from *Chuck*, the images present in different episodes may have provided varying amounts of support for comprehension. In situations where more support was provided by the imagery, participants with lower lexical coverage may have been able to use the images to ameliorate their comprehension. This may have reduced the effects of lexical coverage on comprehension for at least some of the comprehension test items. For other items where the imagery did not explicitly support comprehension, participants with higher lexical coverage could have used their vocabulary knowledge to better understand the episode. This may have led to the significantly different amounts of comprehension for different lexical coverage groups.

The television program that participants viewed in Study 3 can be categorized simply as a spy drama/comedy. The findings for this program in this genre indicated that the relationship between comprehension and lexical coverage can be episode-specific. There might be a different relationship found for different programs and different genres of television. For example, the amount of coverage necessary to make comprehension likely may change if a language learner were to watch a situation comedy, a documentary, or science fiction drama. Both Nation and Webb (2011) and van Zeeland and Schmitt (2012) have theorized that the relationship between lexical coverage and comprehension is likely to differ between genres and text types. The results from Study 3 also highlight the fact that vocabulary, while an important factor, is not the only factor affecting comprehension (these factors were reviewed in Study 1 [see Section 2.1]). Different factors are likely to have influenced comprehension in the episodes. More research needs to be done investigating the factors involved in comprehension of video and the relationship between lexical coverage and these factors.

#### 4.9.2. Lexical coverage and incidental vocabulary learning

The results of Study 3 revealed that there was no significant relationship between lexical coverage and incidental vocabulary gains. This finding contrasts the earlier research by Liu and Nation (1985) that indicated that higher lexical coverage leads to superior incidental vocabulary learning. In Study 3, vocabulary learning occurred at a range of levels of lexical coverage. Table 4.20 shows the mean real and relative gains for the Vocabulary Participants. Results for the participants are divided by their mastery of the three levels of the VLT used in this study. It is apparent from this table that participants at four different lexical coverage levels (as represented by mastery of levels of the VLT) were able to make vocabulary gains. The vocabulary gains for language learners with greater lexical coverage do not appear to be very different from the participants with less lexical coverage. However, there are too few participants with mastery of 3,000- and 5,000-word levels to fully support this claim.

Table 4.20 Vocabulary gains on the Tough and Sensitive tests at four different lexical coverage levels

| Vocabulary Test |                    | < 2,000 word level ( <i>n</i> =145) | 2,000 word level ( <i>n</i> =39) | 3,000 word level ( <i>n</i> =2) | 5,000 word level ( <i>n</i> =1) |
|-----------------|--------------------|-------------------------------------|----------------------------------|---------------------------------|---------------------------------|
| Tough           | Mean Real Gain     | 6.70                                | 5.15                             | 5.00                            | 7.00                            |
|                 | Mean Relative Gain | 23.0%                               | 22.4%                            | 28.8%                           | 46.7%                           |
| Sensitive       | Mean Real Gain     | 6.99                                | 6.23                             | 2.00                            | 7.00                            |
|                 | Mean Relative Gain | 29.2%                               | 31.7%                            | 15.7%                           | 41.7%                           |

The findings also indicate that incidental vocabulary learning can occur at relatively low levels of lexical coverage. Webb and Rodgers (2009a, 2009b) suggested that 95% coverage, or knowledge of the most frequent 3,000 word-families plus proper nouns and marginal words, may be sufficient for incidental vocabulary learning to occur. In Study 3, participants were found to have vocabulary gains at 94.28% lexical coverage (mastery of the 2,000-word level including proper nouns and marginal words) and below 94.28% lexical coverage (without mastery of the 2,000-word level including proper nouns and marginal words). This indicates that vocabulary acquisition through viewing television can occur for learners with lower lexical coverage than previously believed.

While these findings may be contrary to what was expected, this may be due in part to the participants' vocabulary size. Only 39 of the participants had mastery of the 2,000-word level, two had mastery of the 3,000-word level, and one had mastery of the 5,000-word level. Additional participants with a greater range of lexical coverage may have more accurately shown the effects of increased coverage on vocabulary learning. Webb and Rodgers (2009a)

suggested that 95% coverage may be sufficient for incidental learning to take place but only three participants had at least this mean coverage level (provided by knowledge of the 3,000 most frequent words of English and proper nouns and marginal words) across the 10 episodes. However, even if more participants had mastery of the 3,000- and 5,000-word level, their coverage of the episodes still may have been insufficient to show a significant difference between vocabulary gains at different coverage levels. Nation (2001) suggested that 98% coverage may be necessary to incidentally learn words through reading. No participants in this study reached this level of coverage. Across the 10 episodes, a language learner would need to know the 7,000 most frequent words of English plus the proper nouns and marginal words in the episodes to have 98% coverage. To find a significant relationship between increased lexical coverage and incidental vocabulary learning, a number of participants may have to reach 98% coverage.

The results from Study 3 also suggest that lexical coverage may not be as important a factor for incidental vocabulary learning from viewing television as it may be for reading. When language learners are reading they must use knowledge of proximal vocabulary to learn words. When watching television, however, contextual knowledge is supported by images. Learners can make use of these images to learn the meaning of unknown vocabulary, and in turn lessen the impact of lexical coverage on incidental vocabulary learning. Interest in language learning through viewing television and background knowledge of the series built up through viewing multiple episodes are additional factors that may lessen the influence of lexical coverage on incidental vocabulary learning through viewing episodes of television.

#### **4.10. Limitations**

The number of participants reaching mastery of the 3,000- and 5,000-word levels is a limitation of Study 3. Although the number of participants at these levels was small, there were some indications that this amount of vocabulary knowledge could have an effect on comprehension. The results indicated that when lexical coverage approaches 98%, L2 television viewers may have significantly greater comprehension. However, considerably more participants at these vocabulary levels are necessary to confirm this. There were even fewer participants reaching mastery of the 3,000- and 5,000-word levels for the analysis of the effects of lexical coverage on vocabulary learning. The relative gain scores of the participant with mastery of the 5000-word level indicated that incidental vocabulary learning may be superior when lexical coverage approaches 98%. However, further research with a larger sample of participants at this level of vocabulary knowledge is warranted.

Another limitation of Study 3 is the instrument used to measure knowledge of the most frequent words in English of the participants. The VLT measures knowledge of written form and meaning. A test similar to the VLT, that measures knowledge of the spoken form, may provide a more precise measurement of the lexical coverage of spoken texts. However, at present there does not appear to be a standardized test similar to the VLT that can provide a measurement of vocabulary knowledge in this manner. A fourth limitation is that the lists used to create the 2,000-, 3,000- and 5,000-word levels of the VLT were based on early word frequency counts (Kucera & Francis, 1967; Thorndike & Lorge, 1944; West, 1953). However, the lists that were used to analyze the vocabulary in the episodes of *Chuck* were based on BNC word lists developed by Nation (2004, 2006). A test analogous to the VLT based on the BNC word lists would allow for more precise measurement of lexical coverage.

### ***Study 4: EFL learners' attitudes towards learning through viewing English-language television***

#### **5. Introduction**

Previous studies in this thesis have attempted to measure aspects of language learning from viewing multiple episodes of a television program. In addition, to these quantitative indicators of the effectiveness of viewing television for language learning, it is also useful to gauge how language learners feel about the experience of learning through television. For many of the participants in these studies, learning through television would have been a unique experience and they were likely encountering this form of language learning for the first time. There has not been any research that has looked at language learners' reactions to learning from longer videos over time. The present study was designed to examine Japanese EFL learners' (a) attitudes towards comprehension of television across episodes, (b) attitudes towards vocabulary learning through viewing television, (c) attitudes towards aspects of language learning through viewing television, and (d) beliefs about sources of difficulty encountered when viewing English-language television.

#### **5.1. Learners' attitudes towards language learning through viewing videos**

There have been many studies that have surveyed learners' attitudes towards language learning through viewing videos (Baltova, 1994; Brett, 1997; Chung & Huang, 1998; Chung, 1999; Gruba, 2004, 2006; Hasan, 2000; Sueyoshi & Hardison, 2005; Vanderplank, 1988; Wang, 2012). These studies investigated learners' beliefs about their comprehension of videos (Brett, 1997; Chung & Huang, 1998; Gruba, 2004, 2006; Hasan, 2000; Sueyoshi & Hardison, 2005; Wang, 2012), vocabulary learning through watching television (Wang, 2012), their affective attitude towards learning through viewing videos (Chung & Huang, 1998; Chung, 1999; Gruba, 2006; Wang, 2012), and their general language learning experience through watching videos (Baltova, 1994; Chung, 1999; Sueyoshi & Hardison, 2005). There has also been research on the difficulties that learners perceive to be part of language learning from videos (Chung & Huang, 1998; Gruba, 2006; Kobeleva, 2012; Vanderplank, 1988; Wang, 2012).

Previous research has indicated that language learners' attitudes towards comprehension of video have been positive. Language learners consider video-use beneficial to comprehension (Brett, 1997; Chung & Huang, 1998; Gruba, 2004, 2006; Hasan, 2000). As

reported by Hasan (2000) and Sueyoshi and Hardison (2005), learners believe the benefits to comprehension originate from being able to see the speakers' body language, gestures and facial expressions in video. Wang (2012) reported that learners believed that images provide opportunities for better comprehension. Learners' beliefs about the ability of video to aid comprehension are important because comprehension may not always just be a matter of language proficiency but also a positive attitude towards the task may be a factor (Gruba, 2006). Gruba (2006) observed that when learners were able to understand an L2 video they felt it was a pleasurable experience that made them want to learn more and try harder.

Language learners' beliefs about the efficacy of viewing videos for learning vocabulary have not been widely surveyed. Wang (2012) studied 28 participants' attitudes towards learning vocabulary through viewing 12- to 13-minute portions of American situation comedies. The vocabulary learning in this study was not incidental but part of explicit vocabulary instruction. The attitudes of the participants towards vocabulary learning through video were surveyed using 4-point Likert-scales and interviews. Over 50% of participants *agreed* or *slightly agreed* that the television program helped them to learn vocabulary. Participants noted that learning vocabulary was aided by the images in the video, repetition of the vocabulary, the plot of the program, and the emotions of the characters. Over 50% of participants believed the level of the vocabulary in the programs was appropriate for them and over 70% of the participants *agreed* or *slightly agreed* that they preferred learning vocabulary through viewing television over reading. Learners did frequently comment, however, that concentrating on learning vocabulary distracted them from following the story at times. Overall, the results from this study indicate that learners believe that viewing videos can be a valuable part of the vocabulary learning process.

Viewing videos has been likened to a form of play where learners can enjoy the process of learning English (Gruba, 2006). This is important because positive attitudes towards learning tasks can result in increased motivation (Dörnyei & Ushioda, 2011) and comprehension (Gruba, 2006). A number of studies have reported that language learners enjoy viewing videos to learn English. In Chung and Huang's (1998) study, 93% of 155 participants stated that they enjoyed learning from an educational video and Chung (1999) reported that 85% of 156 participants stated that they liked being taught English through watching videos. A majority of the participants in Wang's (2012) study also reported that viewing television programs to learn English was a pleasant experience.

In addition to a belief that learning through viewing videos is enjoyable, research has indicated that learners also believe there are benefits to language learning in general terms

from viewing videos. Learners have been shown to prefer learning through video over aural-only formats especially when they have become accustomed to and comfortable with video (Baltova, 1994). Chung (1999) found that 80% of learners in her study thought that viewing videos was helpful for learning English. Language learners have also indicated that they believe watching videos is beneficial for their listening skills. Participants in Sueyoshi and Hardison's (2005) study rated watching videos as the best way to improve English listening ability. The ESL participants in Sueyoshi and Hardison's study also indicated that watching television was the third most frequent way, after homework and in class study, that they learned English. Overall, learners' feedback on learning through viewing videos indicates that they consider it to have positive effects on several aspects of language learning.

Five characteristics of videos that learners have difficulty with have been identified in previous survey-based research: the vocabulary, the speed at which the dialogue is delivered, the pronunciation of the dialogue, the content, and the names of people featured in a video. Over 66% of participants in Chung and Huang's (1998) study ranked vocabulary as their number one or number two problem when viewing educational videos. Learners in this study claimed that they would have had very little understanding of the videos had they not been pretaught vocabulary prior to viewing. Gruba (2006) also observed that when participants were unfamiliar with the vocabulary in a news clip they found the viewing process a frustrating experience. Chung and Huang (1998) reported that 79% of their participants rated the speed of the speech as the number one or two problem they encountered when watching a video. Similarly, when participants viewed segments of authentic television in Wang's (2012) study they noted that the speed of the dialogue was problematic. Learners in Wang's study reported in interviews that they thought that the pronunciation of the English in the videos was also a source of difficulty. Learners linked this to the English native speakers' frequent use of connected speech as opposed to the type of pronunciation they were used to. The content of the texts themselves has also been reported as a source of difficulty for language learners. Vanderplank (1988) reported that his participants believed they had difficulties with programs when they were unfamiliar with the culture described in them. Gruba (2006) also reported that in situations where learners don't understand the culture or context of a video, they can become discouraged with the viewing process. When viewing episodes of television where little background to the characters was presented, Vanderplank (1988) observed that learners had more difficulties. This has been supported in listening comprehension research by Kobeleva (2012). She found that language learners who were pre-taught the proper names occurring in a news text had significantly higher listening comprehension scores than those

learners who did not have knowledge of the proper names. Kobeleva also found that learners with knowledge of the proper names rated the comprehension tasks as easier than learners without knowledge of the proper names.

Although these earlier studies are valuable, they are also limited to some degree. For example, the previous research on language learners' attitudes towards viewing videos has utilized videos with short running times (Baltova, 1994; Brett, 1997; Chung, 1999; Gruba, 2004, 2006; Sueyoshi & Hardison, 2005; Wang, 2012), videos viewed in isolation (Chung & Huang, 1998; Wang, 2012), types of videos more conducive to language learning (educational, documentaries, news clips), or segments of longer narrative videos (Baltova, 1994; Brett, 1997; Chung & Huang, 1998; Chung, 1999; Gruba, 2004, 2006; Sueyoshi & Hardison, 2005; Wang, 2012). Further research needs to investigate whether findings from these studies are consistent with language learners' attitudes towards viewing multiple full-length episodes of television programs intended for an English-speaking audience as this may indicate the potential efficacy of authentic television for use in language learning.

## **5.2. Research questions**

Study 4 was designed to answer the following research questions:

1. Do EFL learners' beliefs about learning through English-language television change through viewing multiple episodes of a program?
2. What are EFL learners' beliefs about language learning through viewing English-language television?
3. Do EFL learners' beliefs about their comprehension of English-language television change through viewing multiple episodes of a program?
4. What are EFL learners' beliefs about their comprehension of English-language television?
5. What are EFL learners' beliefs about vocabulary learning through viewing English-language television?
6. What do EFL learners believe are sources of difficulty when they view English-language television?

## **5.3. Participants**

There were 229 male and 60 female volunteer participants in their first and second year of university from nine separate classes in this study. Details about these participants are provided in Section 3.3 of Study 2.

### 5.3.1. Exclusion of participants

Sixty-three participants were excluded from Study 4. Participants were excluded if they were absent from: Teaching Session 1 (VLT and Ethics Approval), Teaching Session 2 (Tough and Sensitive Vocabulary Pre-Tests), Teaching Session 3 (Viewing Episode A1), Teaching Session 12 (Viewing Episode B2), and Teaching Session 13 (Tough and Sensitive Vocabulary Post-Tests and Final Attitude Survey). Participants were also excluded from the study if they were absent from viewing more than one episode from Episode 1 to 8 (Teaching Sessions 4 to 11). Data for participants who failed to complete an Episode Attitude Survey accompanying a comprehension test or the Final Attitude Survey was treated as missing and not included in the analyses. The results from 226 participants were left for analysis after these exclusions. The exclusion process is illustrated in Table 5.1 which shows the starting sizes of the nine classes in Study 4 and the number of participants excluded from each.

Table 5.1 Participants excluded from the attitudes analysis in Study 4

| Class | Starting Size | Number of Exclusions | Final Number of Participants |
|-------|---------------|----------------------|------------------------------|
| 1     | 36            | 9                    | 27                           |
| 2     | 36            | 4                    | 32                           |
| 3     | 36            | 8                    | 28                           |
| 4     | 33            | 5                    | 28                           |
| 5     | 38            | 12                   | 26                           |
| 6     | 24            | 2                    | 22                           |
| 7     | 37            | 16                   | 21                           |
| 8     | 24            | 5                    | 19                           |
| 9     | 25            | 2                    | 23                           |
| Total | 289           | 63                   | 226                          |

### 5.4. Procedure

The procedure for Study 4 follows the procedure outlined for Study 1. The overall schedule and in-class procedure for Study 1 are described in Chapter 2. Two attitude surveys were administered to the participants in Study 4. The Episode Attitude Survey was administered following the viewing of each episode. It was administered 10 times across Teaching Sessions 3 through 12. The Final Attitude Survey was completed in Teaching Session 13 after the participants had viewed all 10 episodes.

## **5.5. Materials**

Two surveys were created to survey the participants' attitudes towards learning through television programs. The Episode Attitude Survey was created to assess attitudes immediately following viewing episodes of television while the Final Attitude Survey was designed to survey the participants' attitudes after having studied English through watching television for 10 weeks.

### **5.5.1. Item design**

A combination of rating scale items and forced-choice items were used on the Episode Attitude Survey (EAS) and the Final Attitude Survey (FAS).

#### **5.5.1.1. Rating scale items**

The rating scale items on the EAS and FAS consisted of a question and a corresponding rating scale. The response scales were Likert-type scales which were selected because of their popularity and effectiveness in gathering opinions and attitudes with regards to language-related issues (J. D. Brown, 2001; Dörnyei, 2003). The questions and rating scale format were the product of results obtained from the pilot study. Ten-point bipolar scales were used in the pilot study but changes were made to the scales for Study 4 to improve reliability and validity. The rating scales in the pilot study had verbally and numerically delineated endpoints (1 and 10) and numeric labels for the response categories (2 to 9) in between. Originally, a scale with a broad range was chosen because the wider range encourages more precision in rating and respondents facing such a scale are more likely to think of it as an equal interval rating scale than one with fewer categories (Krosnick, 1999). The results of the pilot, however, showed that respondents tended to use only the middle categories and with little consistency.

Based on the findings from the pilot study, the more common (Hatch & Lazaraton, 1991) and more highly recommended (Schaeffer & Presser, 2003) 9-point scale was chosen. The rationale was that the 9-point scale still had the positive attributes of the 10-point scale but the number of categories lent themselves better to the inclusion of verbal labels throughout the scale. The 9-point scale also offers respondents a middle alternative that increases reliability by reducing the amount of random measurement error, and does not decrease validity (O'Muircheartaigh, Krosnick, & Helic, 2001).

In addition to endpoint labels, three more verbal labels were spaced evenly throughout the categories (at response categories 3, 5 and 7) for a total of five verbal labels. The verbal labels increase the perception that the continuum is divided into equal-sized units and maximize

validity. Using numeric labels with and between the verbal labels leads to assumptions that the categories are equidistant and adds definition to the response categories (Klockars & Yamagishi, 1988).

The pilot study rating scales were a mix of bipolar (verbal endpoints using semantic opposites e.g. *Good to Bad*) and unipolar (verbal endpoints using a single semantic term with varying degrees e.g. *Very enjoyable to Not at all enjoyable*). As respondents have been shown to favour common words and phrases (Mittelstaedt, 1971), wherever possible, the rating scales were standardized as unipolar. Unipolar rating scales make fewer assumptions about the evaluative scale as they do not assume that the poles are indeed opposites. They also have been shown to have a more even distribution of responses between the positive and negative poles (Schaeffer & Presser, 2003). The five verbal labels used in conjunction with the numeric labels consisted of a standard set of adverbs or adverbial phrases of intensity plus adjectives that were appropriate for the question asked in each item. The following adverbial expressions were chosen: *not at all*, *slightly*, *somewhat*, *pretty*, and *very* because they have been shown to represent equidistant categories (Schaeffer & Presser, 2003). While it is acknowledged that reliability increases when verbal labels for all categories are provided (Alwin & Krosnick, 1991), informal trials of nine adverbial phrases showed that the more phrases used, the less people could decide on their relative placement on the continuum.

Not all rating scale items used unipolar rating scales. Questions that had a frequency component necessitated the use of a frequency-based rating scale. In the pilot study, these items also used a 10-point bipolar scale with only numerals as descriptors between the verbal labels *Never* and *Very Often*. For the same reasons described above, the rating scales for frequency items were changed to 9-point scales with five descriptors. The adverbial descriptors used for these items were: *never*, *occasionally*, *fairly often*, *very often* and *always*. These descriptors were chosen because they were shown to have equal distant scaling properties (Pohl, 1981).

The rating scales had the same format throughout. While it is believed to make little psychometric difference, it is important for the format to be clear and consistent with sufficient instruction and examples (Dörnyei, 2003). The numeric labels were underscored by a double-headed arrow to express a continuum. Centered over the appropriate digit were the verbal labels. Each label was in an equal sized box in an effort to portray distinct, equal categories. Participants were presented with examples of the rating scales as part of the Television Viewing Practice they completed prior to viewing the episodes of television. An example of a rating scale item is shown in Figure 5.1.

Figure 5.1 Question 2 of the EAS

*What do you think your comprehension level of this episode was?*

|  |   |                  |   |                  |   |             |   |           |
|--|---|------------------|---|------------------|---|-------------|---|-----------|
| Not at All<br>Good   |   | Slightly<br>Good |   | Somewhat<br>Good |   | Pretty Good |   | Very Good |
| 1  | 2 | 3                | 4 | 5                | 6 | 7           | 8 | 9         |
| ◀──▶ |   |                  |   |                  |   |             |   |           |

#### 5.5.1.2. Forced-choice items

Based on the feedback from the pilot study, a series of forced-choice items were created for the FAS. Areas that participants claimed to have difficulty with were noted and used as the basis for these items. These forced-choice items asked participants to choose between two possible answers: they had difficulty with a particular attribute of the study or they did not have difficulty. The benefits of this type of item are that they are simple and efficient for gathering large amounts of information about specific topics (Krosnick, 1999) and they have been shown to encourage deep processing and fewer satisficing<sup>6</sup> responses (Sudman & Bradburn, 1982).

#### 5.5.1.3. Translations

After all the survey items for each comprehension test were created, they were translated into Japanese. All translations were done by a person fluent in the L1 and L2. This person had also done the translations for the comprehension tests described in Study 1. The translations were done in consultation with the researcher to ensure that the intended meaning of the English items was expressed in the Japanese items. The translated questions and ratings scales were informally piloted with native speakers of Japanese with special attention focused on the translated adverbial expressions and the adverbs of frequency ensuring they communicated the same equidistant properties.

#### 5.5.2. Episode Attitude Survey

The Episode Attitude Survey was designed to monitor the participants' feelings towards various aspects of language learning through L2 television across the 10 episodes in the study. This survey allowed for an analysis of the differences in the episodes as indicated by the participants' responses. Responses came immediately after viewing the episode. This provides a good indication of the participants' beliefs because there is little time between viewing and

<sup>6</sup> Satisficing is a portmanteau of satisfy and suffice.

responding. The EAS consisted of four items that used 9-point unipolar rating scales and accompanying questions. Two of these items asked the participants to respond on the *Not at all Good to Very Good* scale. These items surveyed the participants' perceived comprehension level of the episode, and how well they thought they learned English from viewing the episode. A third item asked about the enjoyment the participants felt from watching the episode and used the *Not at all Enjoyable to Very Enjoyable* rating scale. The final item asked the participants about their perceptions of the usefulness of the episode for learning English and used the *Not at all Useful to Very Useful* continuum. The EAS, in English and Japanese, is shown in its entirety in Appendices D1 and D2.

### 5.5.3. Final Attitude Survey

The Final Attitude Survey was designed to survey the participants' overall feelings towards aspects of viewing television after they had viewed all 10 episodes in the study. The FAS consisted of 12 rating scale and five forced-choice items. Two of the 12 rating scale items were frequency items that asked participants about how often they had previously watched television in order to learn English and how often they planned to in the future. These two items used the 9-point rating scale that contained adverbs of frequency ranging from *Never* to *Always*. Of the 10 remaining rating scale items, three asked the participants to respond on the *Not at all Good to Very Good* continuum. These three items inquired about the participants' beliefs concerning their understanding of the first episode of *Chuck* they viewed, their understanding of the final episode they viewed, and their overall understanding of the episodes they viewed. Three other items asked the participants about how they believed their English listening skills, their vocabulary, and their understanding from the first to the final episode were affected by watching the episodes of *Chuck* all semester. These items had the participants respond on the *Not at all Improved to Very Improved* continuum. Participants responded from *Not at all Useful to Very Useful* on two items that asked how useful they thought watching *Chuck* all semester was for their overall English ability and how useful they believed viewing English-language television was for language learning. One item asked the participants' opinion on whether they believed the time spent watching television was an appropriate use of time. Participants responded on the *Not at all Good Use of Time to Very Good Use of Time* continuum. A single item used the *Not at all Enjoyable to Very Enjoyable* rating scale and asked the participants how they felt about watching television in class.

The five forced-choice items on the FAS asked participants to identify which aspects of watching television throughout the study they found difficult. These aspects included: the

speed of the dialogue, the characters' names, the pronunciation of dialogue, the vocabulary, and the content in the episodes. The FAS, in English and Japanese, is shown in its entirety in Appendices D4 and D5.

## 5.6. Results

### 5.6.1. Episode Attitude Survey

The participants in Study 4 completed the EAS after each comprehension test. Four items surveyed their attitudes towards learning English from the episodes and their comprehension of the episodes. The responses to these items were tabulated for each episode that participants viewed. When a participant missed viewing an episode, their responses were not included in the calculation of the mean responses. This occurred an average of 4.3 times per episode with a minimum of 0 times in the Final Episode and a maximum of 8 times in Episode 2. In the following discussion of the items and the response patterns, the items are grouped together by their common field of investigation. The four Episode Attitude Survey Questions (EASQ) and their accompanying rating scales are shown in Figures 5.2, 5.3, and 5.4.

Figure 5.2 EASQ 1

|               |  |   |                       |   |                       |   |                     |   |                   |
|---------------|--|---|-----------------------|---|-----------------------|---|---------------------|---|-------------------|
|               | Not at All<br>Enjoyable  |   | Slightly<br>Enjoyable |   | Somewhat<br>Enjoyable |   | Pretty<br>Enjoyable |   | Very<br>Enjoyable |
|               | 1  | 2 | 3                     | 4 | 5                     | 6 | 7                   | 8 | 9                 |
| <b>EASQ 1</b> | How did you feel about studying English through watching this program? |   |                       |   |                       |   |                     |   |                   |

Figure 5.3 EASQ 2 and EASQ 3

|               |  |   |                  |   |                  |   |                |   |           |
|---------------|--|---|------------------|---|------------------|---|----------------|---|-----------|
|               | Not at All<br>Good   |   | Slightly<br>Good |   | Somewhat<br>Good |   | Pretty<br>Good |   | Very Good |
|               | 1  | 2 | 3                | 4 | 5                | 6 | 7              | 8 | 9         |
| <b>EASQ 2</b> | What do you think your comprehension level of this episode was?                  |   |                  |   |                  |   |                |   |           |
| <b>EASQ 3</b> | What do you think your level of English learning was from watching this episode? |   |                  |   |                  |   |                |   |           |

Figure 5.4 EASQ 4

|               |  |   |                    |   |                    |   |                  |   |                |
|---------------|--|---|--------------------|---|--------------------|---|------------------|---|----------------|
|               | Not at All<br>Useful   |   | Slightly<br>Useful |   | Somewhat<br>Useful |   | Pretty<br>Useful |   | Very<br>Useful |
|               | 1  | 2 | 3                  | 4 | 5                  | 6 | 7                | 8 | 9              |
| <b>EASQ 4</b> | To what extent do you think watching this episode was useful for studying English? |   |                    |   |                    |   |                  |   |                |

#### 5.6.1.1. EAS items related to learning English through viewing the episodes

Episode Attitude Survey Questions 1, 3 and 4 were used to measure the participants' beliefs about learning English from television for each of the 10 episodes. Each question asked the participants to respond on a different response scale providing insight into different aspects of learning English through viewing television. Through answering these questions, the participants expressed how much they enjoyed learning English through viewing television (EASQ 1), what they believed their level of learning was from the episodes (EASQ 3), and how useful they believed viewing the episodes was for studying English (EASQ 4). The participants responded to EASQ 1 on the *Not at All Enjoyable* to *Very Enjoyable* rating scale, EASQ 3 on the *Not at All Good* to *Very Good* scale, and EASQ 4 on the *Not at all Useful* to *Very Useful* scale.

The mean responses to EASQ 1 (*How did you feel about studying English through watching this program?*) trended higher over the course of the 10 episodes. For the Initial Episode, the mean response was 5.6 and reached 6.8 by the Final Episode. Between Episode 1 and Episode 8, the mean responses ranged from 6.4 (Episode 1) and 6.8 (Episode 8). The mean response of 6.8 for the Final Episode was marginally lower than *Pretty Enjoyable*. The difference between the participants' responses to EASQ 1 following the Initial Episode and following the Final Episode was examined with a Wilcoxon Test. This test, an alternative to a paired-samples *t*-test, is used with non-parametric data such as that provided by the Likert-type rating items used on the EAS. The results of the Wilcoxon Test indicated that the participants' responses towards the enjoyability of studying English through viewing an episode of *Chuck* were significantly higher following the Final Episode than following the Initial Episode,  $Z = 7.72, p < .001$ . The effect size as measured by *r* was .36 a value corresponding to a medium treatment effect. The mean response to EASQ 1 gradually and generally increased across the 10 episodes the participants viewed. This suggests that as the participants became more familiar with the series and the content therein, their enjoyment in viewing the series increased.

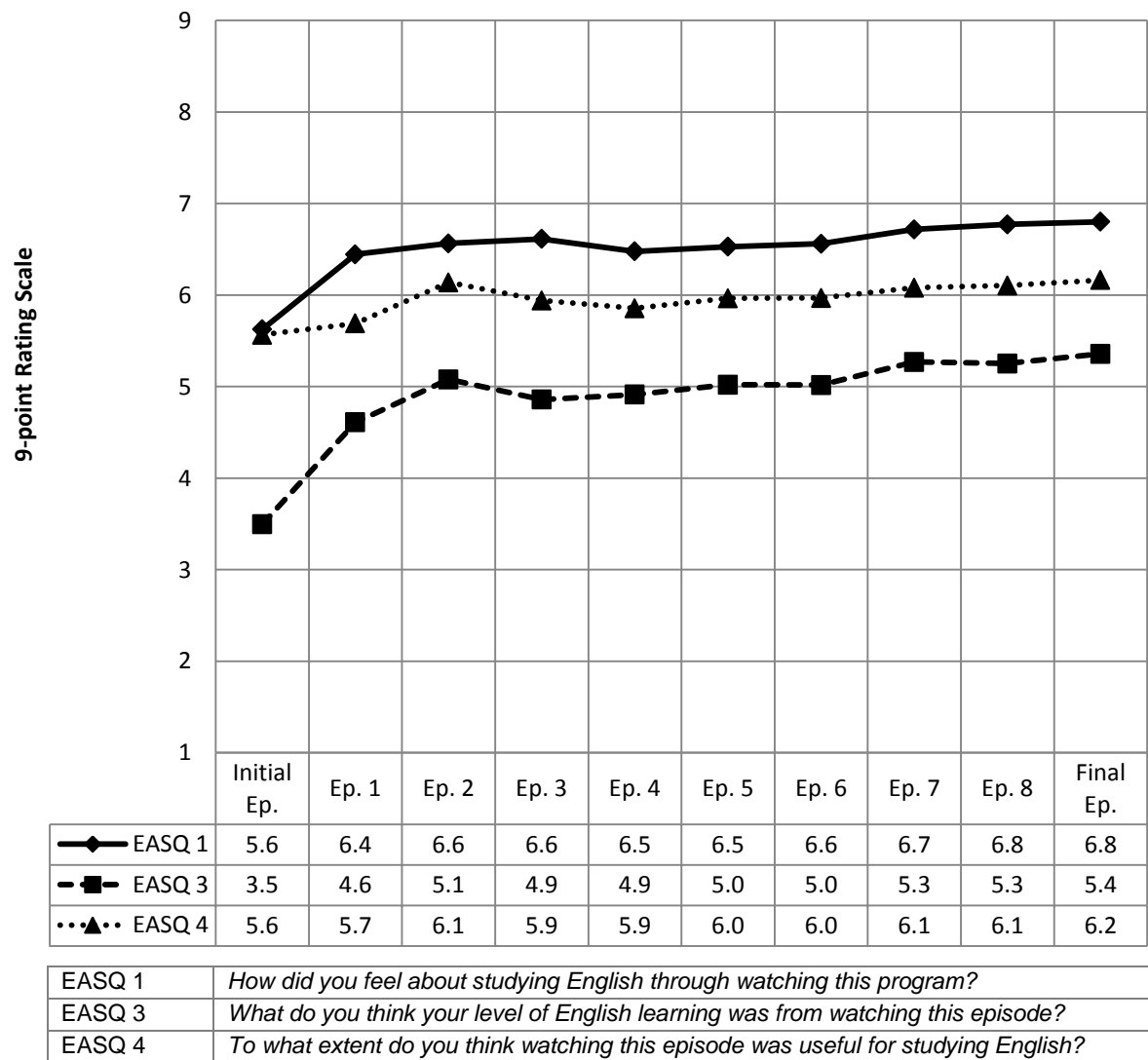
The responses for EASQ 3 (*What do you think your level of English learning was from watching this episode?*) began with a mean score of 3.5 for the Initial Episode and ended with a mean response of 5.4 for the Final Episode. For this item, there was a relatively large increase from the Initial Episode (Episode A1 or B1) to the mean response (4.6) for Episode 1. Following Episode 1, the mean responses ranged from 4.9 (Episode 3 and Episode 4) to 5.3 (Episode 7 and Episode 8). A Wilcoxon Test indicated that the participants' rating of their level of English learning from watching an episode was significantly higher following the

Final Episode than following the Initial Episode,  $Z = 10.86$ ,  $p < .001$ . The effect size, as measured by  $r$ , was .51, a value corresponding to a large treatment effect. This pattern of responses for EASQ 3 indicates that the participants believed that their language learning became better the more episodes they viewed, reaching a level at or slightly higher than *Somewhat Good* (response category 5) for the episodes in the latter half of the study.

For EASQ 4 (*To what extent do you think watching this episode was useful for studying English?*), the increase in mean response scores between the Initial Episode and the Final Episode was 0.6 with the mean response for the Initial Episode at 5.6 and 6.2 for the Final Episode. From Episode 1 to Episode 8 the mean responses ranged from 5.7 (Episode 1) to 6.1 (Episode 2, 7 and 8). A Wilcoxon Test indicated that the participants' rating of the usefulness of an episode of *Chuck* for studying English was significantly higher following the Final Episode than following the Initial Episode,  $Z = 4.51$ ,  $p < .001$ . The effect size, as measured by  $r$ , was .21, a value corresponding to a small treatment effect. Responses to EASQ 4 indicate the participants generally believed that the episodes of television were above *Somewhat Useful* (response category 5) for studying English and their estimation of this increased from the start of the study to the end.

The mean responses to EASQs 1, 3 and 4 across all 10 episodes of *Chuck* are summarized in Figure 5.5. Each question on the survey is represented by a different line with data points indicating the mean response of each episode. Taken together these results indicate the participants believed that the enjoyability, benefits and usefulness of learning English from television increased as they viewed multiple episodes. Response patterns to these items are supported by the results from the Wilcoxon Tests that indicate that mean response scores to these items significantly increased from the Initial Episode to the Final Episode.

Figure 5.5 Results across all 10 episodes for EASQs 1, 3 and 4

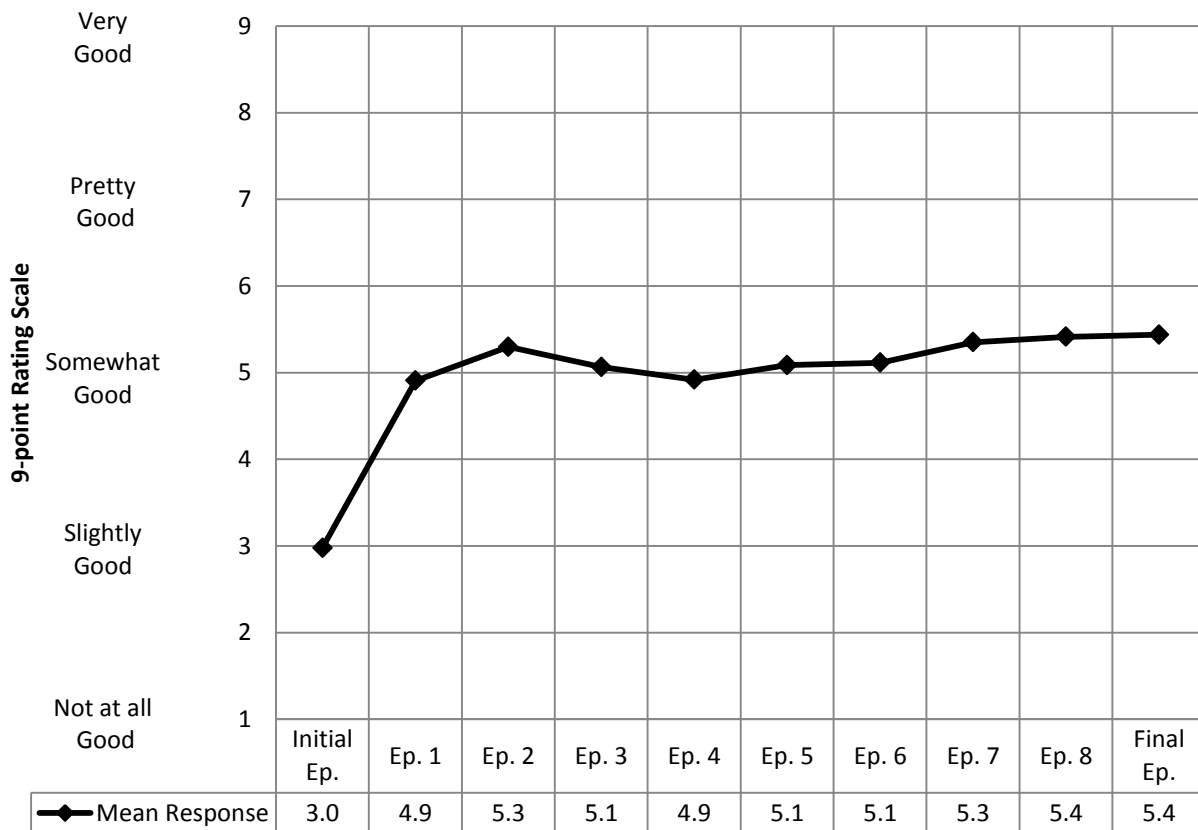


#### 5.6.1.2. EAS item related to comprehension of the episodes

The participants' beliefs about their comprehension of each of the 10 episodes was measured by EASQ 2 (*What do you think your comprehension level of this episode was?*). The mean responses to this item began at 3.0 following the Initial Episode and increased considerably to 4.9 for Episode 1. The mean response score rose to 5.4 by the Final Episode viewed. The mean responses from Episode 1 to 8 ranged from 4.9 (Episode 1 and 4) to 5.4 (Episode 8). Results from a Wilcoxon Test indicated that the participants' evaluation of their comprehension was significantly higher following the Final Episode than following the Initial Episode,  $Z = 11.99$ ,  $p < .001$ . The effect size, as measured by  $r$ , was .57, a value corresponding to a large treatment effect. The mean responses to EASQ 2 across the 10 episodes are summarized in Figure 5.6. It includes data points for each episode and each episode's mean score located in the data table. The responses to EASQ 2 indicate that the

participants' beliefs about their comprehension of the episodes quickly increased as they gained familiarity with the characters and the story, and then the mean responses became more episode-dependent at a level slightly below or over *Somewhat Good* (response category 5).

Figure 5.6 Results across all 10 episodes for EASQ 2 (*What do you think your comprehension level of this episode was?*)



### 5.6.2. Final Attitude Survey

Upon viewing the 10 episodes of *Chuck*, participants completed the FAS. This asked the participants to reflect on the viewing of the episodes and to indicate their beliefs on a variety of topics. The results of the rating scale items are presented in three sections: items related to comprehension, items related to vocabulary learning, and items related to general language learning. These are followed by the results of the forced-choice items concerning perceptions of difficulty in viewing television. On the FAS, the number of participants who failed to respond to items was recorded but no attempt was made to replace their missing data because there was insufficient data from the respondents to use Expectation Maximization (Allison, 2001).

#### 5.6.2.1. FAS items related to comprehension

Four items on the FAS examined the participants' beliefs about their comprehension of the episodes. Final Attitude Survey Question 1 (FASQ 1) asked the participants to consider their level of understanding of the first episode they viewed while FASQ 2 asked them to consider their level of understanding of the tenth episode. FASQ 3 asked the participants to think about their overall understanding of the episodes they viewed throughout the study. FASQs 1, 2 and 3 asked the participants to respond on the *Not at All Good* to *Very Good* response scale. FASQ 7 asked the participants about the degree to which they thought their comprehension of *Chuck* improved from the first to the last episode and had them respond on the *Not at All Improved* to *Very Improved* response scale. Tables 5.2 and 5.3 show the results for FASQs 1, 2, 3, and 7. Each table shows the numerical and verbal labels for each response scale, and the number of participants that responded in each category for each question. The X category is the number of non-respondents for each item. The percentage in each response category and the mean response score are included.

The mean response score for FASQ 1 (*What do you think your understanding of the first episode of Chuck you watched this semester was?*) was 3.3 and the mode was response category 3 with 22.1% of the respondents. The mean response for FASQ 2 (*What do you think your understanding of the last episode of Chuck you watched this semester was?*) was 5.7 and the mode was response category 7 (27.9% of respondents). For FASQ 1, 59.7% of the participants responded in the lowest three response categories compared with 4.9% in the highest three categories. In contrast, for FASQ 2, 40.3% of the participants responded in the highest three response categories compared to 16.9% in the lowest three. For FASQ 7 (*To what degree do you think your understanding of the television program Chuck improved from the first episode watched to the last?*), the mean response score was 6.1 with 27.4% of participants selecting response category 7 as the mode. Approximately 45% of participants indicated that they believed their improvement to be from *Pretty Improved* (response category 7) to *Very Improved* (response category 9). Taken together, these response patterns for FASQs 1, 2 and 7 indicate a considerable shift in the participants' perceived comprehension from the first to the final episode.

The mean response for FASQ 3 (*What do you think your overall understanding of the episodes of Chuck you watched this semester was?*) was 5.3 and the mode was response category 5 (23.5% of the respondents). A majority (69.0%) of the respondents chose category 5 (*Somewhat Good*) or higher indicating that the participants in Study 4 thought that they had a reasonable level of understanding of the television viewed throughout the study. For these

four items concerned with comprehension, the mean responses indicate that the participants believed that their comprehension increased from the first to the last episode and that for a majority of the participants, their overall comprehension of the episodes corresponded to *Somewhat Good* or better.

Table 5.2 Responses to FASQs 1, 2 and 3

|               |   | Not at All<br>Good |             | Slightly<br>Good |            | Somewhat<br>Good |       | Pretty<br>Good |      | Very<br>Good |
|---------------|---|--------------------|-------------|------------------|------------|------------------|-------|----------------|------|--------------|
|               | X   | 1                  | 2           | 3                | 4          | 5                | 6     | 7              | 8    | 9            |
| <b>FASQ 1</b> | What do you think your understanding of the first episode of <i>Chuck</i> you watched this semester was?    |                    |             |                  |            |                  |       |                |      |              |
| Responses     | 6   | 38                 | 47          | 50               | 31         | 33               | 10    | 7              | 0    | 4            |
| %             | 2.7%  | 16.8%              | 20.8%       | 22.1%            | 13.7%      | 14.6%            | 4.4%  | 3.1%           | 0.0% | 1.8%         |
|               |   |                    | <b>Mean</b> | <b>3.3</b>       |            |                  |       |                |      |              |
| <b>FASQ 2</b> | What do you think your understanding of the last episode of <i>Chuck</i> you watched this semester was?     |                    |             |                  |            |                  |       |                |      |              |
| Responses     | 6   | 2                  | 4           | 32               | 13         | 37               | 41    | 63             | 19   | 9            |
| %             | 2.7%  | 0.9%               | 1.8%        | 14.2%            | 5.8%       | 16.4%            | 18.1% | 27.9%          | 8.4% | 4.0%         |
|               |   |                    |             | <b>Mean</b>      | <b>5.7</b> |                  |       |                |      |              |
| <b>FASQ 3</b> | What do you think your overall understanding of the episodes of <i>Chuck</i> you watched this semester was? |                    |             |                  |            |                  |       |                |      |              |
| Responses     | 6   | 3                  | 7           | 35               | 19         | 53               | 48    | 38             | 12   | 5            |
| %             | 2.7%  | 1.3%               | 3.1%        | 15.5%            | 8.4%       | 23.5%            | 21.2% | 16.8%          | 5.3% | 2.2%         |
|               |   |                    |             | <b>Mean</b>      | <b>5.3</b> |                  |       |                |      |              |

Table 5.3 Responses to FASQ 7

|               |  | Not at All<br>Improved |      | Slightly<br>Improved |      | Somewhat<br>Improved |            | Pretty<br>Improved |       | Very<br>Improved |
|---------------|--|------------------------|------|----------------------|------|----------------------|------------|--------------------|-------|------------------|
|               | X  | 1                      | 2    | 3                    | 4    | 5                    | 6          | 7                  | 8     | 9                |
| <b>FASQ 7</b> | To what degree do you think your understanding of the television program <i>Chuck</i> improved from the first episode watched to the last? |                        |      |                      |      |                      |            |                    |       |                  |
| Responses     | 6  | 0                      | 3    | 23                   | 13   | 35                   | 44         | 62                 | 27    | 13               |
| %             | 2.7%   | 0.0%                   | 1.3% | 10.2%                | 5.8% | 15.5%                | 19.5%      | 27.4%              | 11.9% | 5.8%             |
|               |  |                        |      |                      |      | <b>Mean</b>          | <b>6.1</b> |                    |       |                  |

#### 5.6.2.2. FAS items related to vocabulary learning

A single item on the FAS asked the participants about their beliefs about the effects viewing television had on their English vocabulary learning. FASQ 6 had the participants respond on the *Not at All Improved* to *Very Improved* scale. Table 5.4 presents the number of responses in each category and the overall mean. The mean response to FASQ 6 (*To what extent do you think that watching the television program Chuck all semester affected your*

*English vocabulary?*) was 4.5 and the mode was response category 5 (24.3%). There were responses in all nine response categories indicating a wide range of individual beliefs about vocabulary improvement. The vast majority of participants (94.2%) responded that they believed their vocabulary was *Improved* (response categories 2 - 9) indicating that they believed that over the course of the study they experienced improvement in their English vocabulary.

Table 5.4 Responses to FASQ 6

|               |  | Not at All Improved |      | Slightly Improved |       | Somewhat Improved |       | Pretty Improved |      | Very Improved |
|---------------|--|---------------------|------|-------------------|-------|-------------------|-------|-----------------|------|---------------|
|               | X  | 1                   | 2    | 3                 | 4     | 5                 | 6     | 7               | 8    | 9             |
| <b>FASQ 6</b> | To what extent do you think that watching the television program <i>Chuck</i> all semester affected your English vocabulary? |                     |      |                   |       |                   |       |                 |      |               |
| Responses     | 6  | 7                   | 20   | 47                | 31    | 55                | 29    | 22              | 7    | 2             |
| %             | 2.7%   | 3.1%                | 8.8% | 20.8%             | 13.7% | 24.3%             | 12.8% | 9.7%            | 3.1% | 0.9%          |
|               |  |                     |      | Mean              | 4.5   |                   |       |                 |      |               |

### 5.6.2.3. FAS items related to general language learning

Seven items on the FAS asked the participants about their beliefs on language learning from viewing the episodes of *Chuck*. FASQ 4 asked the participants about how they believed viewing television affected their overall English ability. FASQ 5 was concerned with the participants' perceived improvement in English listening skills. FASQs 8 and 9 surveyed the participants about their frequency of studying English through viewing television before and after participation in Study 4. FASQs 10 and 11 surveyed the participants on their feelings towards viewing television in their English-language class. FASQ 10 asked the participants to rate the extent to which viewing television in class was useful. FASQ 11 surveyed the participants about whether viewing television in class was enjoyable. FASQ 12 asked participants about their perceptions of the usefulness of viewing television for language learning. Tables 5.5 to 5.9 show the results for these items.

FASQ 4 and FASQ 12 asked the participants how useful they believed watching English-language television was for improving their English ability and for language learning in general. Both items asked participants to respond on the *Not at All Useful* to *Very Useful* rating scale. For FASQ 4 (*To what extent do you think that watching the television program Chuck all semester had an effect on your overall English ability?*), 72.9% of the participants responded from *Somewhat Useful* (response category 5) to *Very Useful* (response category 9). The mean response was 5.6 and the mode was response category 5 (22.1% of participants).

The responses for FASQ 12 (*To what extent do you think watching English-language television programs is useful for language learning?*) were generally positive with 86.4% of participants indicating a rating of 5 (*Somewhat Useful*) or above. The mode was response category 7 with 27.9% of respondents and the mean response rating was 6.8. The responses to these two items indicate a positive attitude by the participants towards the usefulness of viewing the episodes of *Chuck* for their English ability and the usefulness of watching television in general for language learning.

Table 5.5 Responses to FASQs 4 and 12

|                |   | Not at All<br>Useful |      | Slightly<br>Useful |             | Somewhat<br>Useful |            | Pretty<br>Useful |       | Very<br>Useful |
|----------------|---|----------------------|------|--------------------|-------------|--------------------|------------|------------------|-------|----------------|
|                | X   | 1                    | 2    | 3                  | 4           | 5                  | 6          | 7                | 8     | 9              |
| <b>FASQ 4</b>  | To what extent do you think that watching the television program <i>Chuck</i> all semester affected your overall English ability? |                      |      |                    |             |                    |            |                  |       |                |
| Responses      | 6   | 1                    | 5    | 30                 | 19          | 50                 | 43         | 48               | 8     | 16             |
| %              | 2.7%  | 0.4%                 | 2.2% | 13.3%              | 8.4%        | 22.1%              | 19.0%      | 21.2%            | 3.5%  | 7.1%           |
|                |   |                      |      |                    | <b>Mean</b> | <b>5.6</b>         |            |                  |       |                |
| <b>FASQ 12</b> | To what extent do you think watching English language television programs is useful for language learning?                        |                      |      |                    |             |                    |            |                  |       |                |
| Responses      | 6   | 2                    | 2    | 10                 | 11          | 32                 | 23         | 63               | 23    | 54             |
| %              | 2.7%  | 0.9%                 | 0.9% | 4.4%               | 4.9%        | 14.2%              | 10.2%      | 27.9%            | 10.2% | 23.9%          |
|                |   |                      |      |                    |             | <b>Mean</b>        | <b>6.8</b> |                  |       |                |

A single item asked participants about how they thought viewing the 10 episodes of *Chuck* affected their listening skills. FASQ 5 (*To what extent do you think that watching the television program Chuck all semester affected your English listening skills?*) asked participants to respond on the *Not at All Improved* to *Very Improved* scale. The mean score on FASQ 5 was 5.5 and the mode was response category 5 (23.9%). A large majority of participants (71.3%) responded that they believed their listening was *Somewhat Improved* (response category 5) or higher. There were, however, responses in categories 1 to 9 indicating that, while most participants believed that over the course of the study they experienced improvement in their listening ability, there was a wide range of beliefs about how much improvement there was.

Table 5.6 Responses to FASQ 5

|               |  | Not at All<br>Improved |      | Slightly<br>Improved |      | Somewhat<br>Improved |       | Pretty<br>Improved |      | Very<br>Improved |
|---------------|--|------------------------|------|----------------------|------|----------------------|-------|--------------------|------|------------------|
|               | X  | 1                      | 2    | 3                    | 4    | 5                    | 6     | 7                  | 8    | 9                |
| <b>FASQ 5</b> | To what extent do you think that watching the television program <i>Chuck</i> all semester affected your English listening skills? |                        |      |                      |      |                      |       |                    |      |                  |
| Responses     | 6  | 2                      | 11   | 29                   | 17   | 54                   | 38    | 39                 | 21   | 9                |
| %             | 2.7%   | 0.9%                   | 4.9% | 12.8%                | 7.5% | 23.9%                | 16.8% | 17.3%              | 9.3% | 4.0%             |
| <b>Mean</b>   |  |                        |      |                      |      | <b>5.5</b>           |       |                    |      |                  |

FASQs 8 and 9 surveyed the participants on their frequency of studying English via viewing television outside of Study 4. Participants responded on the *Never to Always* scale for both of these items. The mean rating for FASQ 8 (*To what extent did you try to learn English through watching television programs before you entered this course?*) was 4.8, which corresponds to more frequent than *Occasionally*. The mode was category 3 (26.1%), which corresponds to *Occasionally* on the rating scale. For this item there were responses across the full range of options indicating that the participants had varying degrees of experience learning from television. FASQ 9 asked the participants to consider how often they believed they would try to learn from watching television in the future. The mode for this item was response category 3 (19.6%) and the mean was 4.8. For FASQ 9 (*To what extent do you think you will try to learn English through watching television programs after you finish this course?*), 55.2% of the participants responded in response category 5 (*Fairly Often*) or higher compared with 52.2% for FASQ 8. While these statistics show a minimal increase from FASQ 8 to 9, they do indicate a slightly more positive attitude towards studying from television in the future and the intention of some participants to continue to use the medium to learn English.

Table 5.7 Responses to FASQs 8 and 9

|               |  | Never |      | Occasionally |            | Fairly Often |       | Very Often |      | Always |
|---------------|--|-------|------|--------------|------------|--------------|-------|------------|------|--------|
|               | X  | 1     | 2    | 3            | 4          | 5            | 6     | 7          | 8    | 9      |
| <b>FASQ 8</b> | To what extent did you try to learn English through watching television programs before you entered this course?             |       |      |              |            |              |       |            |      |        |
| Responses     | 6  | 11    | 11   | 59           | 21         | 28           | 35    | 38         | 12   | 5      |
| %             | 2.7%   | 4.9%  | 4.9% | 26.1%        | 9.3%       | 12.4%        | 15.5% | 16.8%      | 5.3% | 2.2%   |
|               |  |       |      | <b>Mean</b>  | <b>4.8</b> |              |       |            |      |        |
| <b>FASQ 9</b> | To what extent do you think you will try to learn English through watching television programs after you finish this course? |       |      |              |            |              |       |            |      |        |
| Responses     | 6  | 6     | 5    | 55           | 28         | 53           | 22    | 37         | 10   | 4      |
| %             | 2.7%   | 2.7%  | 2.2% | 24.3%        | 12.4%      | 23.5%        | 9.7%  | 16.4%      | 4.4% | 1.8%   |
|               |  |       |      | <b>Mean</b>  | <b>4.8</b> |              |       |            |      |        |

FASQs 10 and 11 asked the participants how they felt about watching television in their English language class. The two items used the same question (*How did you feel about watching the television programs in class?*) but different rating scales. For FASQ 10, the mean rating was 7.2 and the mode was response category 9 (32.3%) indicating that participants perceived watching television in class to be a *Very Good Use of Time*. A very large majority (92.9%) responded in the top two thirds of the response scale with 71.2% of the participants responding in the top third. FASQ 11 had the highest mean score (8.0) on the FAS with 55.3% of responses in category 9 (*Very Enjoyable*). The responses by the participants on these two items indicate that the majority of participants believed the watching of episodes of *Chuck* in class to be an enjoyable, worthwhile use of class time.

Table 5.8 Responses to FASQ 10

|                |   | Not at All<br>Good Use<br>of Time |      | Slightly<br>Good Use<br>of Time |      | Somewhat<br>Good Use<br>of Time |            | Pretty<br>Good Use<br>of Time |      | Very<br>Good Use<br>of Time |
|----------------|---|-----------------------------------|------|---------------------------------|------|---------------------------------|------------|-------------------------------|------|-----------------------------|
|                | X   | 1                                 | 2    | 3                               | 4    | 5                               | 6          | 7                             | 8    | 9                           |
| <b>FASQ 10</b> | How did you feel about watching the television programs in class? |                                   |      |                                 |      |                                 |            |                               |      |                             |
| Responses      | 6   | 1                                 | 3    | 4                               | 2    | 24                              | 25         | 67                            | 21   | 73                          |
| %              | 2.7%  | 0.4%                              | 1.3% | 1.8%                            | 0.9% | 10.6%                           | 11.1%      | 29.6%                         | 9.3% | 32.3%                       |
|                |   |                                   |      |                                 |      | <b>Mean</b>                     | <b>7.2</b> |                               |      |                             |

Table 5.9 Responses to FASQ 11

|                |   | Not at All<br>Enjoyable |      | Slightly<br>Enjoyable |      | Somewhat<br>Enjoyable |      | Pretty<br>Enjoyable |             | Very<br>Enjoyable |
|----------------|---|-------------------------|------|-----------------------|------|-----------------------|------|---------------------|-------------|-------------------|
|                | X   | 1                       | 2    | 3                     | 4    | 5                     | 6    | 7                   | 8           | 9                 |
| <b>FASQ 11</b> | How did you feel about watching the television programs in class? |                         |      |                       |      |                       |      |                     |             |                   |
| Responses      | 6   | 0                       | 2    | 0                     | 0    | 13                    | 9    | 50                  | 21          | 125               |
| %              | 2.7%  | 0.0%                    | 0.9% | 0.0%                  | 0.0% | 5.8%                  | 4.0% | 22.1%               | 9.3%        | 55.3%             |
|                |   |                         |      |                       |      |                       |      |                     | <b>Mean</b> | <b>8.0</b>        |

#### 5.6.2.4. FAS items related to perceptions of difficulty

There were five forced-choice items on the Final Attitude Survey. The items asked the participants to indicate whether they had difficulty with certain aspects of viewing episodes of television. The aspects surveyed were: the vocabulary in the episodes, the content of the episodes, pronunciation, characters' names, and the speed of the dialogue. Participants responded with either *Yes* that they had difficulty or *No* that they did not. The number of responses and the relative percentages for the *Yes* or *No* categories for each item are displayed in Table 5.10. Ninety-five percent of the participants indicated that they had difficulty with the vocabulary (FAS Forced-choice Item 1) in the episodes. On the other hand, 49 of the 220 respondents (22.3%) suggested that they had difficulty with the content in the episodes (FAS Forced-choice Item 2). A large majority of the participants indicated that the pronunciation (FAS Forced-choice Item 3) and the speed of the dialogue (FAS Forced-choice Item 5) in the episodes were difficult for them. For pronunciation, 81.8% of the participants indicated difficulty, and for speed of the dialogue it was 88.6%. For the characters' names (FAS Forced-choice Item 4) in the episodes, 45.9% of the participants perceived difficulty while 54.1% did not. Overall, a large majority of participants claimed to have difficulties with the vocabulary, pronunciation, and the speed of the dialogue in the episodes. A large majority also claimed to not have difficulty with the content of the episodes, and the participants were fairly evenly divided between having difficulty and not having difficulty with the characters' names.

Table 5.10 Responses to FAS Forced-choice Items 1 to 5

| Forced-choice Item            | Yes Responses |       | No Responses |       |
|-------------------------------|---------------|-------|--------------|-------|
|                               | Participants  | %     | Participants | %     |
| 1. Vocabulary in the episodes | 209           | 95.0% | 11           | 5.0%  |
| 2. Content of the episodes    | 49            | 22.3% | 171          | 77.7% |
| 3. Pronunciation              | 180           | 81.8% | 40           | 18.2% |
| 4. Characters' names          | 101           | 45.9% | 119          | 54.1% |
| 5. Speed of the Dialogue      | 195           | 88.6% | 25           | 11.4% |

## 5.7. Summary of findings

The main findings of the attitude surveys from Study 4 can be summarized as follows:

1. The mean responses concerning participants beliefs about the enjoyability of, benefits from, and usefulness of language learning through viewing television significantly increased from the first to the tenth episodes viewed. The mean response for the item surveying enjoyability of learning through television was between *Somewhat* and *Pretty Enjoyable* by the final episode viewed. The mean response for the item concerning the level of English learning was between *Somewhat* and *Pretty Good* and the item concerning the usefulness of television for studying English was between *Somewhat* and *Pretty Useful*.
2. Across the 10 episodes of television viewed, the participants' perceptions of their comprehension of the episodes rose dramatically after the Initial Episode then the mean responses became more episode-dependent. The level of perceived comprehension was significantly higher following viewing the Final Episode at a level slightly over *Somewhat Good*.
3. The participants were generally positive about language learning through viewing English-language television. The majority of participants thought that learning from television was at least a *Pretty Good Use of Time* and *Pretty Enjoyable* making these the two highest rated aspects. The majority of the participants thought that viewing television had at least a *Somewhat Useful* effect on their overall English ability, was at least *Pretty Useful* for language learning in general, and had *Somewhat Improved* their listening skills. Participants, however, showed very little difference in the frequency they had previously used television to learn English and the frequency with which they thought they would use it to learn English in the future.

4. Following viewing all the episodes of *Chuck*, the majority of EFL learners believed that their comprehension improved from the first to the last episode viewed and their overall comprehension of the episodes was *Somewhat Good* or better.
5. While there was a wide range of perceptions about their vocabulary learning as indicated by responses in all nine categories stretching from *Not at All Improved* to *Very Improved*, slightly over half (50.8%) of participants believed their vocabulary was *Somewhat Improved* or better.
6. The majority of participants in this study maintained that they had difficulties with the vocabulary, the pronunciation, and the speed of the dialogue in the episodes of *Chuck*. Conversely, a large majority did not have difficulty with the content of the episodes. Participants were fairly evenly divided between having difficulty and not having difficulty with the characters' names.

## **5.8. Discussion**

The design of the present study expanded on previous research on language learners' attitudes towards viewing videos that has generally utilized short videos, single incidences of video viewing, language learning videos, documentaries, or segments of longer narrative videos. The results from this study provide insight into learners' beliefs about learning from viewing multiple full-length episodes of television programs intended for an English-speaking audience.

### **5.8.1. Attitudes towards language learning**

In answer to the first research question, the mean responses for EASQ 1, EASQ 3 and EASQ 4 were all significantly higher following the Final Episode than they were following the Initial Episode. The results indicate language learners' beliefs about the efficacy of viewing television for language learning may improve as they view more episodes. The findings support previous research indicating that language learners enjoy learning through viewing videos (Chung & Huang, 1998; Chung, 1999; Gruba, 2006; Wang, 2012) and that viewing videos are valuable for language learning (Baltova, 1994; Chung, 1999; Sueyoshi & Hardison, 2005). The results from Study 4 go beyond previous research, however, by showing attitudes towards the enjoyability of episodes, perception of learning from episodes, and the usefulness of episodes for studying English can improve through viewing multiple episodes of the same television program.

In answer to research question 2, the results from the seven items on the FAS concerned with general language learning through viewing television indicate positive beliefs towards the overall experience. Mean responses to these items ranged from 4.8 to 8.0. The lowest mean, which was slightly below *Fairly Often*, was for the two items that asked participants about the frequency with which they had studied through television in the past and would continue to do so in the future. The item with the highest mean response, directly between the *Pretty* and *Very Enjoyable* response categories, surveyed participants on how much they enjoyed watching television programs in their English class. Taken as a whole, the responses to these items indicate that EFL learners have a positive attitude towards language learning through viewing television. Participants believed that viewing television had a useful effect on their English ability and their listening ability improved. Participants claimed that they intended to study English through viewing television in the future at the same frequency as they had in the past but it is difficult to say whether this prior learning was part of individual study or in a classroom situation. Participants indicated that viewing television in a language class was a good use of time and an enjoyable experience. Finally, participants believed that watching English-language television was useful for language learning. These findings support previous studies (Baltova, 1994; Chung, 1999; Sueyoshi & Hardison, 2005) that indicated that language learners believed viewing videos was an effective way to learn a language. The findings from Study 4 expand on this earlier research by indicating that learners can have positive attitudes towards language learning through viewing video when watching multiple episodes over an extended period of time and not just when viewing short isolated videos.

### **5.8.2. Attitudes towards comprehension**

In answer to the third research question, the results suggest that EFL learners believed their comprehension of the episodes of television improved as they viewed more episodes. One item on the EAS and four items on the FAS were used to measure changes in participants' beliefs in their comprehension of episodes across the study. Participant responses to FASQ 7 indicate that they believed that their comprehension of the episodes improved from the first to the final episode viewed. This finding was supported by responses to FASQs 1 and 2 where the mean response concerning comprehension of the first episode (3.3) was slightly higher than *Slightly Good* and the mean response for the final episode (5.7) was between *Somewhat* and *Pretty Good*. Results from the FAS reinforce the participants' responses to EASQ 2 recorded following each of the 10 episodes viewed. For this item, participants' ratings of their

comprehension of the episodes improved quickly after the Initial Episode (*Slightly Good*: 3.0) ending up significantly higher following the Final Episode (*Somewhat Good*: 5.4).

Research question 4 was addressed by the participants' responses to FASQ 3 which asked about their beliefs about their overall comprehension of the episodes they viewed. The participants thought that their overall comprehension of the 10 episodes they viewed was slightly higher than *Somewhat Good* (5.3).

These results concerning comprehension are encouraging as they show that learners in an EFL environment may have a generally positive attitude towards their comprehension of television. The results for research questions 3 and 4 support previous studies where learners have indicated positive attitudes towards comprehension of video (Brett, 1997; Chung & Huang, 1998; Sueyoshi & Hardison, 2005; Wang, 2012). The findings here, however, illustrate language learners' perceptions of comprehension of authentic television programs and that these perceptions of comprehension can improve as more episodes are viewed. These findings are important because if learners' attitude towards comprehension is positive across viewing multiple episodes it may translate into a more pleasurable experience for learners where they put more effort into learning through viewing television (Gruba, 2006).

### **5.8.3. Attitudes towards vocabulary learning**

Research question 5 investigated the participants' beliefs about learning vocabulary through viewing episodes of *Chuck*. The responses to FASQ 6 indicate participants believed that their English vocabulary improved through watching the 10 episodes. Participants had a mean response of 4.5 which was over half way between *Slightly* and *Somewhat Improved*. There were responses in all response categories from *Not at All Improved* to *Very Improved* indicating that participants had a range of beliefs regarding the improvement of their vocabulary through viewing television. These results support the findings from Wang's (2012) study where most of the participants believed that viewing television helped them to learn vocabulary. The vocabulary learning in that study, however, was part of a series of explicit vocabulary tasks. In Study 4, participants were commenting on whether they believed they incidentally picked up vocabulary through viewing the episodes. These findings are noteworthy because they indicate that learners tend to believe they incidentally learn vocabulary through watching television.

#### 5.8.4. Perceptions of difficulty

Research question 6 in this study investigated the difficulties that participants had with certain aspects of viewing episodes of television. Difficulties were examined by the five FAS Forced-choice Items. The majority of participants indicated that they had difficulty with three of the five aspects surveyed. Ninety-five percent of the participants indicated that they had difficulty with the vocabulary in the episodes, 88.6% indicated difficulty with the speed of the dialogue, and 81.8% indicated difficulty with the pronunciation in the program. Learners in previous studies have also indicated difficulty with these aspects of viewing video. In Chung and Huang's (1998) study, the majority of participants ranked vocabulary within their top two problems when viewing videos. In studies by Wang (2012) and Chung and Huang (1998), participants indicated that the speed of the dialogue in videos was problematic. Participants in Wang's study also cited the pronunciation by native speakers of English and their tendency to use connected speech as problematic. Identifying sources of difficulty that language learners experience when viewing television is important because these problem areas can be focused on in preteaching activities. Preteaching activities may ease some of the burdens that the learners experience when watching television.

The majority of participants indicated that they did not have difficulty with one aspect of viewing television. Seventy-seven percent of participants in Study 4 claimed not to have difficulty with the content in the episodes of *Chuck*. This finding may appear to contrast the previous research of Vanderplank (1988), whose participants claimed to have difficulty with the content of unfamiliar programs. However, the reason the majority of participants in Study 4 claimed not to have difficulty with the content may have been due to the fact that they viewed multiple episodes of the same program. This may have allowed the participants to build up knowledge of the program incrementally with each episode viewed aiding their understanding of the content. Participants in Vanderplank's study watched a variety of different programs. This finding is important because when learners have difficulties with content they can become demotivated to learn through viewing videos (Gruba, 2006).

For one aspect of viewing television, there was not a large majority of participants indicating difficulty or no difficulty. A small majority (54.1%) of participants believed that the characters in the episodes of *Chuck* they viewed were not a source of difficulty. This is an important finding because as Kobleva (2012) suggests, a lack of knowledge of the proper names in a listening text can lead to comprehension difficulties. Unfortunately, 45.9% of the participants had some difficulty with proper names which may have led to problems with comprehension. More research on the difficulties that learners perceive with proper names in

episodes of television is needed to shed light on how proper names affect comprehension and the consequences of having problems with proper names.

## **5.9. Limitations**

One limitation of Study 4 is the depth at which the survey items were able to obtain information from the participants. Supplemental long form questions and interviews would have added to the richness of the data collected. This is apparent in the results from certain rating scale and forced-choice items where it would have been particularly useful to have more information about the participants' answers. For example, participants, on average, indicated that they thought they improved their English vocabulary through viewing the episodes of *Chuck*. It would be helpful to know which words they thought they learned and how they thought they learned the vocabulary. It also would have been beneficial to ask follow-up questions to the participants that indicated difficulty with the proper names in the episodes. It may have been possible to determine whether it was the regularly occurring characters in *Chuck* that were the problem or the more episode-specific characters that were the source of difficulty. Follow-up questions in these areas could lead to the design of pedagogical interventions based on the information provided by the participants on how they thought their vocabulary learning could be improved or how difficulties with character names could be resolved. While the survey information from Study 4 provides unique information on language learners' attitudes towards viewing multiple episodes of television, supplemental interviews and free response items could have provided more specific information about the beliefs of the participants towards language learning through viewing television.



### ***Study 5: Language learning through viewing television with captions***

#### **6. Introduction**

The previous studies in this thesis have shown the potential for language learning from viewing episodes of television. Study 1 demonstrated that language learners were able to increase their comprehension from the first to the final episode of successive television viewed. Across the 10 episodes, participants had average scores of 62.9% on the comprehension tests and the participants with the highest comprehension scores averaged just over 83%. Study 2 showed that language learners were able to acquire vocabulary from viewing episodes of television at a rate comparable to that from reading. Study 3 showed that the lexical coverage of the episodes had a significant effect on comprehension for the majority of episodes viewed, but no significant effect on vocabulary gain. Study 4 demonstrated that the participants had positive attitudes toward the experience of viewing 10 episodes of television. From the first to the last episode viewed, response scores increased indicating participants believed that their comprehension of the episodes and the enjoyability, benefits and usefulness of learning English from television increased as they viewed multiple episodes of television. Participants also indicated that they believed that viewing the episodes of television improved their English vocabulary.

There were some findings from the earlier studies, however, that were not as positive as might have been anticipated. In Study 1, while there were participants who had reasonably high comprehension scores, there were also those that failed to obtain a substantial level of comprehension. Participants with the lowest comprehension scores across the 10 episodes of television averaged just 31.9% on the comprehension tests. In Study 2, the frequency of occurrence of the target words was found to have a small statistically significant effect on vocabulary acquisition for only one of the vocabulary tests. Frequency of occurrence, however, has been shown to be a strong determiner of uptake in listening (Elley, 1989; Vidal, 2003) and reading studies (Hirai, 1999; Horst et al., 1998). In Study 3, the small effect sizes for the relationship between lexical coverage and comprehension and the lack of significant findings for an effect of lexical coverage on vocabulary acquisition were somewhat unanticipated. Lexical coverage was hypothesized to have a greater effect on comprehension of television and vocabulary acquisition through viewing television (Webb & Rodgers,

2009a). In previous research involving listening (Bonk, 2000; Elley, 1989; Stæhr, 2008; van Zeeland & Schmitt, 2012) and reading (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Liu & Nation, 1985), lexical coverage was found to have significant effects on comprehension and vocabulary acquisition. In the survey results from Study 4, the majority of participants indicated that they had difficulty with the vocabulary, pronunciation, and speed of the dialogue in the episodes of television they viewed. Allowing language learners access to captions while viewing the episodes of *Chuck* may be a possible method of improving on the findings from the previous studies in this thesis.

Captions are a transcription of the spoken text that appears simultaneously at the bottom of the television screen as a video plays (Chung, 1999). Captions are also referred to as teletext subtitles, closed captions, and same language subtitles in different parts of the world (Vanderplank, 2010). In academic works they are referred to as bimodal, unilingual and intralingual subtitles (Danan, 2004). In this study, captions are differentiated from subtitles in that the audio and the on-screen text are in different languages for subtitles while for captions they are in the same language<sup>7</sup>. Originally, captions were intended as a service for the deaf and hearing impaired but they have long been used in language learning situations. The option of displaying captions while viewing a television program is usually standard on commercial DVDs.

There are five reasons why the presence of captions while viewing television may lead to increased comprehension. First, captions allow learners to use their reading skills to enhance their aural comprehension (Garza, 1991). Lower proficiency language learners from a population similar to the participants in this study have been shown to have listening comprehension levels lower than their reading comprehension (Hirai, 1999). Learners may be able to use their more advanced reading ability with the captions to increase their comprehension of episodes of television. Second, research has indicated that the presence of captions does not distract from observation of onscreen details that support comprehension consequently the value of imagery in television programs is not compromised (Danan, 2004). Third, language learners can more easily break down the input from the television into meaningful units by viewing captions (Ellis, 2005). Fourth, captions may also help learners to visualize the aural input, especially if the input is to some extent beyond their comprehension ability (Danan, 2004). The fifth reason captions may increase comprehension is that they

---

<sup>7</sup> There have been studies (Baltova, 1999; Hui, 2007; Markham, Peter, & McCarthy, 2001; Markham & Peter, 2003; Mitterer & McQueen, 2009; Yoshino, Kano, & Akahori, 2000) that have investigated the efficacy of subtitles for language learning but examining the effects of subtitles on learning through viewing television is beyond the scope of this research.

facilitate additional cognitive processes, such as greater depth of spoken-word processing (Bird & Williams, 2002). In these ways, captions might allow for increased comprehension of television for participants with low English proficiency.

Incidental vocabulary learning from viewing television may also be enhanced when captions are present. Language learners have been shown to rely more heavily on captions when they encounter unknown vocabulary, vocabulary with unfamiliar pronunciation, or when the vocabulary is part of dialogue delivered particularly quickly (Winke, Gass, & Sydorenko, *in press*). The connectedness of English speech may make it difficult for lower level language learners to discriminate between individual words. Captions, however, divide the boundaries between words making it easier for learners to recognize the vocabulary in language chunks (Bird & Williams, 2002; Winke et al., *in press*). Captions also do not reflect accents making speech with accents unfamiliar to viewers more accessible (Bird & Williams, 2002). Language learners also use captions and knowledge of the written form of vocabulary to recall meaning more accurately (Danan, 2004; Neuman & Koskinen, 1992) and strengthen the connection between the aural and written form (Bird & Williams, 2002; Garza, 1991). These attributes of captioned video point to the potential for increased incidental vocabulary acquisition when captions accompany English-language television.

The presence of captions could also clarify the relationship between a language learner's lexical coverage of episodes of television and both comprehension and incidental vocabulary acquisition. The presentation of the vocabulary in captions may allow participants to more easily differentiate between the vocabulary that they know and the vocabulary they do not know (Danan, 2004; Neuman & Koskinen, 1992). Recognizing more known vocabulary increases lexical coverage and this in turn may increase comprehension. Recognizing vocabulary that is unknown could also aid in learning vocabulary from television. When participants recognize a word that is unknown to them, they may be more apt to learn the word (Nassaji, 2003) from context and imagery, especially with repeated encounters with the word. Captions may also assist in the recognition of the proper nouns in the episodes. The initial capitalization of proper nouns in captions makes them more identifiable. This ease of recognition is not available when the spoken language is presented aurally. Proper nouns are considered to be known in lexical coverage estimates and easier recognition of proper nouns makes the lexical coverage estimates more valid. In these ways, the captions present when viewing television may ameliorate the relationship between lexical coverage and both comprehension and incidental vocabulary acquisition.

Captions may also affect language learners' attitudes towards learning from television. In general, both low and high proficiency language learners have indicated positive attitudes towards captions (Taylor, 2005) and learners have reported that they experience less anxiety when captions are available (Winke et al., 2010). Language learners have also reported being more motivated to view television in their L2 when it is accompanied by captions (Vanderplank, 1988; Winke et al., in press). The findings from these studies suggest that the language learners in Study 5 may report more positive attitudes towards language learning through viewing television than those reported in Study 4.

Further research investigating whether captions can improve language learning from viewing 10 episodes of a television drama needs to be performed. The present study seeks to do this through the following four experiments:

1. Captions and Comprehension
2. Captions and Incidental Vocabulary Acquisition
3. Captions and Lexical Coverage
4. Captions and Attitudes

The following sections present the methodological details that are common to all four experiments.

### 6.1. Participants

There were 44 male and 29 female participants in their first year of university from three separate classes in this study. All of the participants had studied English for a minimum of seven years. The English proficiency level of the participants can be considered pre-intermediate to intermediate within the context of the university. The classes that provided the participants for Study 5 were all taught by the researcher. Details on the number of participants in each class and their university major are shown in Table 6.1.

Table 6.1 Participants in Study 5 prior to exclusions

| Class | Major          | Year of Study | Gender |    | Number of Participants |
|-------|----------------|---------------|--------|----|------------------------|
|       |                |               | M      | F  |                        |
| 1     | Commerce       | 1             | 12     | 13 | 25                     |
| 2     | Commerce       | 1             | 13     | 12 | 25                     |
| 3     | Human Kinetics | 1             | 19     | 4  | 23                     |
| Total |                |               | 44     | 29 | 73                     |

### 6.1.1. Exclusion of participants

Exclusion procedures differed depending on the analyses that were undertaken. Twenty-two participants were excluded from the analyses of comprehension, lexical coverage and comprehension, and attitudes. Participants were excluded if they were absent from: Teaching Session 1 (VLT and Ethics Approval), Teaching Session 2 (Tough and Sensitive Vocabulary Pre-Tests), Teaching Session 3 (Viewing Episode A1), Teaching Session 12 (Viewing Episode B2), and Teaching Session 13 (Tough and Sensitive Vocabulary Post-Tests and Final Attitude Survey). Participants were also excluded from the study if they were absent from viewing more than one episode from Episode 1 to 8 (Teaching Sessions 4 to 11). It was believed that missing a single episode would not be a serious detriment to comprehension but missing two or more would have a negative effect. When participants were absent from a single teaching session for Episode 1 through Episode 8 they did not complete the comprehension test. Missing comprehension test scores were replaced using the Expectation Maximization Algorithm (Allison, 2001) which is explained in Section 2.8.2. Data for participants who failed to complete an Episode Attitude Survey accompanying a comprehension test or the Final Attitude Survey was treated as missing and not included in the analysis. The results from 51 participants were left for analysis after these exclusions. Table 6.2 shows the starting sizes of the three classes and the number of participants excluded from each.

Table 6.2 Participants excluded from the comprehension, lexical coverage and comprehension, and attitudes analyses for Study 5

| Class | Starting Size | Number of Exclusions | Final Number of Participants |
|-------|---------------|----------------------|------------------------------|
| 1     | 25            | 6                    | 19                           |
| 2     | 25            | 5                    | 20                           |
| 3     | 23            | 11                   | 12                           |
| Total | 73            | 22                   | 51                           |

Thirty-three participants were excluded from the incidental vocabulary acquisition and lexical coverage and vocabulary acquisition portions of this study. Participants were excluded if they were absent from any of the treatment and testing sessions. They were excluded if they missed viewing an episode of *Chuck* because they would not have the opportunity to encounter episode-specific vocabulary and may have reduced encounters with some target words. These exclusions left 40 participants in the analyses of incidental vocabulary learning and lexical coverage and vocabulary acquisition. The number of exclusions is illustrated in

Table 6.3 which shows the starting sizes of the three classes and the number of excluded participants from each.

Table 6.3 Participants excluded from the incidental vocabulary acquisition, and lexical coverage and vocabulary acquisition analyses for Study 5

| Class | Starting Size | Number of Exclusions | Final Number of Participants |
|-------|---------------|----------------------|------------------------------|
| 1     | 25            | 11                   | 14                           |
| 2     | 25            | 8                    | 17                           |
| 3     | 23            | 14                   | 9                            |
| Total | 73            | 33                   | 40                           |

### 6.1.2. Comparison of participants to previous studies

Throughout Study 5, the participants are referred to as being part of the Captions Group. Participants from Studies 1, 2, 3 and 4 are referred to as the No Captions Group. This delineation is used for comparing the results from the present study to the previous studies in this thesis highlighting the effects of captions on aspects of language learning. This classification is also used, where applicable, in the review of previous research.

#### 6.1.2.1. Vocabulary knowledge of the Captions and No Captions groups

Before comparing the results from the participants in Study 5 with the results from the participants in earlier studies, the vocabulary knowledge of the Captions Group and the No Captions Group were compared. The two groups were found to have similar levels of vocabulary knowledge before completing the treatments. This indicates that the groups of learners had a similar level of English language proficiency before participating in these studies. The details of this analysis can be found in Section 6.4.3.5.

#### 6.1.2.2. Incidental vocabulary acquisition Control Group

The 73 participants in the Control Group from Study 2 were also used in Study 5 for the purposes of comparison. Their results on the Tough and Sensitive Vocabulary Tests were compared to the results of the participants in the Captions Group as part of the analysis of the effects of viewing captioned television on incidental vocabulary acquisition in Experiment 2. For a complete description of the Control Group, see Section 3.3.1.

## 6.2. Procedure

The overall schedule and the in-class procedures for Study 5 were identical to those explained in the procedure sections from Studies 1, 2, 3 and 4 (see Sections 2.4, 3.4, 4.5 and 5.4). The experimental procedure for Study 5 was repeated with three different university

classes and took place in the second semester of the Japanese university school year which ran from October 2010 to January 2011. The setting for Study 5 was also identical to that of Studies 1, 2, 3 and 4.

### 6.2.1. Overall schedule

Study 5 took place over thirteen teaching sessions in one university semester. Normally, each teaching session was separated by a week but because of national and school holidays there were instances where the teaching sessions were separated by two or three weeks. The schedule for the study is shown in Figure 6.1.

Figure 6.1 Schedule for Study 5

| Teaching Session | Study 5 – Captioning – Schedule   |
|------------------|---|
| 1                | Human Ethics Committee Ethics Form & Vocabulary Levels Tests (2,000, 3,000, & 5,000)  |
| 2                | <i>Chuck</i> -specific 60-Item Tough Vocabulary Pre-Test, Television Viewing Practice, & <i>Chuck</i> -specific 60-Item Sensitive Vocabulary Pre-Test |
| 3                | Episode A1  |
| 4                | Episode 1   |
| 5                | Episode 2   |
| 6                | Episode 3   |
| 7                | Episode 4   |
| 8                | Episode 5   |
| 9                | Episode 6   |
| 10               | Episode 7   |
| 11               | Episode 8   |
| 12               | Episode B2  |
| 13               | <i>Chuck</i> -specific 60-Item Tough Vocabulary Post-Test, Final Attitude Survey, <i>Chuck</i> -specific 60-Item Sensitive Vocabulary Post-Test       |

### 6.2.2. Viewing order of the episodes

In Study 5, participants viewed Episode A1 first, followed by Episodes 1 to 8, then finally Episode B2. In Studies 1, 2, 3 and 4, there were two viewing groups that first viewed either Episode A1 or Episode B1, and Episode B2 or Episode A2 last. This was because prior to these studies it was unknown if one of these episodes was more difficult than the other which might have been a factor in analyzing first to final episode comprehension gains. Analysis of the comprehension gains in these studies revealed no significant difference between the comprehension scores of the participants who viewed Episode A1 or Episode B1 first. Consequently, it was decided to have all participants in Study 5 view the episodes of *Chuck* in the same order.

### **6.3. Materials**

The materials used in Study 5 were essentially the same as those used in Studies 1, 2, 3 and 4 but included additional materials based on the inclusion of captions while viewing episodes. The materials from Studies 1, 2, 3 and 4 are described in full in Sections 2.7, 3.5, 4.6 and 5.5. Additional survey items on the Episode Attitude Survey and the Final Attitude Survey were included to examine the participants' attitudes towards the presence of captions.

#### **6.3.1. Captions**

In Study 5, captions were shown at the bottom of the screen while the participants viewed the episodes. The captions available on the DVDs of *Chuck* and used in this study are called off-line pop-on captions. In this style of captions, the aim is for a verbatim transcription but limitations of space and time mean the text must sometimes be abridged. In this process, unnecessary words are removed but the spoken text is never rephrased. The captions are presented on a maximum of two lines, at a reading speed of approximately 200 words per minute or less, and shown for a minimum of 1.5 seconds for each 32 characters of text. Pop-on captions occasionally include captions that are not part of the spoken text. Speaker identification captions are present when there are multiple speakers on screen, when the speaker is off-screen, or when the speaker is not obvious. They include the name of the character followed by a colon (e.g. *CHUCK:*) preceding the transcript of the spoken text. There are also descriptive captions which present non-speech information used to create mood, set scene and context, and describe music and sound effects. These captions are enclosed in closed brackets (e.g. (*phone ringing*)) (Canadian Association of Broadcasters, 2008). Captions were displayed for every episode viewed in Study 5.

#### **6.3.2. Episode Attitude Survey**

For Study 5, an additional item was added to the Episode Attitude Survey (EAS) developed in Study 4. This survey accompanied each comprehension test for the 10 episodes. This item, like the original four, was a 9-point unipolar rating scale and question. The question surveyed the extent to which the participants believed the captions accompanying the video were useful for their understanding of the episode. This item was added to examine the participants' attitudes towards the captions that were present during viewing of the episodes of *Chuck*. The EAS for Study 5, in English and Japanese, is shown in its entirety in Appendix D6 and D7.

### **6.3.3. Final Attitude Survey**

The Final Attitude Survey (FAS) for Study 5 also had additional items. The original FAS in Study 4 had 12 questions with 9-point rating scales and five forced-choice questions. The FAS for Study 5 had an additional four rating scale items. These items were added to examine the participants' beliefs about the captions accompanying the episodes. Three items enquired about the perceived usefulness of the captions for understanding the episodes, for vocabulary learning, and for listening ability. The fourth item asked the participants about the amount of attention they paid to the captions while viewing the episodes of television. There was also an additional forced-choice item on this survey which asked participants to choose which method of input, captions or audio, they relied on more to understand the episodes. The FAS for Study 5, in English and Japanese, is shown in its entirety in Appendix D8 and D9.

## **6.4. Study 5: Experiment 1 – Captions and Comprehension**

### **6.4.1. How does the presence of captions affect comprehension?**

Many studies have investigated the effects of captions on comprehension of video (Baltova, 1999; Chung, 1999; Etemadi, 2012; Garza, 1991; Guichon & McLornan, 2008; Guillory, 1998; Huang & Eskey, 1999; Latifi et al., 2011; Markham et al., 2001; Markham & Peter, 2003; Taylor, 2005; Winke et al., 2010; Yoshino, Kano, & Akahori, 2000). Six of these studies, (Chung, 1999; Guillory, 1998; Huang & Eskey, 1999; Latifi et al., 2011; Markham et al., 2001; Markham & Peter, 2003) had similar and comparable methodologies, and the same level of rigor in reporting their results and warrant detailed examination of their findings. A number of the studies (Baltova, 1999; Etemadi, 2012; Garza, 1991; Guichon & McLornan, 2008; Winke et al., 2010; Yoshino et al., 2000) were not generalizable or comparable because of their unique comprehension tests or insufficiently reported results but their findings are worth noting. One study (Taylor, 2005) had findings that contrasted with the majority of research and a detailed investigation of possible explanations for these results is necessary. Throughout the survey of the most relevant studies, the percent difference between the results for the Captions Group and those of the No Captions Group is used to illustrate the degree to which the presence of captions affected comprehension. This provides a means for direct comparison of studies that use different measurement instruments. For the purposes of comparison and summary, the number of participants, the L1 and target language of the participants, the type of comprehension test, the treatment conditions, and the type and length of the input video for each of these studies are described in Figure 6.2.

Figure 6.2 Summary of previous research examining comprehension of video with and without captions

| <b>Study</b>                     | <b>Participants</b>                                 | <b>L1 Language</b>  | <b>Target Language</b> | <b>Comprehension Test</b>                       | <b>Treatment</b>   | <b>Input Video</b>                         | <b>Video Length</b> |
|----------------------------------|---|---------------------|------------------------|---|--|--|---------------------|
| Garza (1991)                     | 140 college students                                | English and Russian | Russian and English    | 10 multiple-choice items per video in L1 and L2 | Groups viewed the videos with captions or no captions  | 5 videos in 5 genres                       | 2 to 4 minutes      |
| Guillory (1998)                  | 202 college students                                | English             | French                 | 7 short-answer items per video in L1            | Groups viewed the videos with captions, keyword captions, or no captions   | 2 segments of an educational video         | Not reported        |
| Baltova (1999)                   | 93 high school students                             | English             | French                 | 10 open-ended short-answer items in L2          | Groups viewed the video with L2 audio and captions twice followed by the L2 audio and no captions once, with L1 audio and L2 subtitles then with L2 audio and captions then finally with L2 audio only, or with L2 audio only with no captions three times | 1 documentary                              | 7.5 minutes         |
| Chung (1999)                     | 183 college students                                | Chinese             | English                | 10 multiple-choice items per video in L1        | Participants viewed a video with captions, no captions, preceded by advance organizers, and with captions and advance organizers   | 4 segments of an educational video         | Not reported        |
| Huang and Eskey (1999)           | 30 adult ESL students                               | Not reported        | English                | 16 multiple-choice items in L2                  | Groups viewed the video twice with captions or no captions   | 1 episode from an educational video series | 21 minutes          |
| Yoshino, Kano and Akahori (2000) | 32 university students and 104 jr. college students | Japanese            | English                | Written recall in L1 and L2                     | Groups viewed each video either with captions, subtitles, no captions, or audio-only (no images)   | 4 segments of music videos                 | 35 to 38 seconds    |

Figure 6.2 Summary of previous research examining comprehension of video with and without captions (continued)

| <b>Study</b>                           | <b>Participants</b>    | <b>L1 Language</b>  | <b>Target Language</b>               | <b>Comprehension Test</b>  | <b>Treatment</b>  | <b>Input Video</b>               | <b>Video Length</b> |
|--|------------------------|---------------------|--------------------------------------|--|---|----------------------------------|---------------------|
| Markham, Peter and McCarthy (2001)     | 169 college students   | English             | Spanish                              | Written summary in L1  | Groups viewed the video with captions, subtitles, or no captions              | 1 documentary                    | 7 minutes           |
| Markham and Peter (2003)               | 213 college students   | English             | Spanish                              | 20 multiple-choice items in L2                                   | Groups viewed the video with captions, subtitles, or no captions              | 1 documentary                    | 7 minutes           |
| Taylor (2005)                          | 85 college students    | English             | Spanish                              | A free recall procedure in L1 and 15 multiple-choice items in L1 | Groups viewed the video with captions or no captions                          | 1 educational video              | 10 minutes          |
| Guichon and McLoran (2008)             | 40 college students    | French              | English                              | Written summary in L2  | Groups watched the video with captions, subtitles, no captions, or audio-only | 1 news report                    | 3 minutes           |
| Winke, Gass and Sydorenko (2010)       | 150 college students   | English and Kannada | Russian, Spanish, Arabic and Chinese | Multiple-choice items in English (number of items not reported)  | Groups viewed the videos twice with captions or no captions                   | 3 documentaries                  | 3 to 5 minutes      |
| Latifi, Mobalegh, and Mohammadi (2011) | 39 college students    | Persian             | English                              | 10 multiple-choice items per segment in L2                       | Groups viewed the videos with captions, subtitles or no captions              | 15 segments of an animated movie | 2 minutes           |
| Etemadi (2012)                         | 44 university students | Persian             | English                              | 10 multiple-choice items per video in L2                         | Groups watched each video with captions or no captions                        | 2 documentaries                  | 20 and 30 minutes   |

The results from six studies indicate that there is a positive effect from captions on comprehension (Chung, 1999; Guillory, 1998; Huang & Eskey, 1999; Latifi et al., 2011; Markham et al., 2001; Markham & Peter, 2003). Guillory (1998) compared the use of full text captions, keyword captions, and no captions on comprehension of two educational videos. Keyword captions consisted of words identified as important to the script. To measure comprehension, participants in each treatment group completed 14 short answer items that focused on the recall of details and inferencing from the information presented in each video. The participants in the Captions Group ( $M=10.1$ ) outperformed the Keyword Captions Group ( $M=9.2$ ) and the No Captions Group ( $M=7.3$ ). The results of the Captions Group and the Keyword Captions Group were significantly higher than those of the No Captions Group but there was no significant difference between the two captions groups. The Captions Group's mean score was 38.1% higher than the No Captions Group's. While the higher comprehension scores for the Captions Group compared to the No Captions Group are of primary concern here, it is also apparent from this study that learners have the capability to manage longer verbatim captions as indicated by the similarity of the comprehension scores for the two captions groups.

Using a video designed for language learning, Huang and Eskey (1999) had half their participants view a captioned version of an educational video twice while the other half viewed an uncaptioned version twice. A comprehension test, presented aurally, followed the second viewing. The Captions Group ( $M=10.87$  out of 16) scored significantly higher than the No Captions Group ( $M=7.67$ ). The participants who viewed the video with captions had 41.7% higher comprehension scores than those who did not have access to captions indicating that the support the captions provide can lead to increased comprehension of videos intended for language learners.

Chung (1999) compared the comprehension of educational videos viewed under four treatment conditions: advance organizers preceding the videos, captions accompanying the videos, both advance organizers and captions accompanying the videos, and video-only. Four groups viewed each video under a different treatment condition in a Latin square design. Comprehension tests were administered in the participants' L1 following each viewing session. The mean score out of 10 for the participants when watching a video preceded by an advance organizer was 6.98, 7.66 when the video had captions, 7.98 when the video had captions and was preceded by an advance organizer, and 6.69 when the video was presented with no captions. The results from the captions treatment were 14.4% higher than the results from the no captions treatment. The captions treatment and the advance organizer with

captions treatment had significantly higher results than those from the advance organizer-only and the no captions treatments. Chung theorizes that captions help language learners to bridge the gap between the development of their reading and listening skills increasing their listening comprehension ability.

In a pair of studies, Markham, Peter and McCarthy (2001) and Markham and Peter (2003) compared the effects of captions, subtitles, and no captions on language learners' comprehension of a documentary. In the 2001 study, comprehension was measured with a written summary which was scored by the number of idea units produced. On this test, the mean scores were 10.97, 8.47 and 12.40 for the Captions Group, the No Captions Group, and the Subtitles Group, respectively. The Captions Group's mean score was 29.5% higher than the No Captions Group. In the 2003 study, comprehension was measured with a multiple-choice listening test. The mean scores (out of 20) were 10.12, 7.81 and 13.33 for the Captions Group, the No Captions Group, and the Subtitles Group, respectively. The Captions Group's mean score was 29.6% higher than the No Captions Group. The Captions Group significantly outperformed the No Captions Group in both studies. The results from these studies indicate the beneficial effect captions can have on comprehension whether measured through multiple-choice items or a written summary.

Latifi, Mobalegh and Mohammadi (2011) studied the effects of subtitles, captions, and no captions on English language learners' comprehension of 15 short sections of an animated movie. The general theme of the movie was explained to the participants before viewing. The participants in each treatment group viewed each film clip twice with key vocabulary and idiomatic expressions explained after the second viewing. The mean score of the 15 comprehension tests was highest for the Subtitles Group (7.17 out of 10), followed by the Captions Group (6.33), then by the No Captions Group (5.25). The Captions Group performed significantly better than No Captions Group but there was no statistical difference between the Subtitles and Captions groups. The Captions Group's mean score was 20.6% higher than the No Captions Group. Even when used in conjunction with decontextualized segments of a film, the presence of captions in this study led to improved comprehension compared with no captions.

Two studies examined the effects of captions on the comprehension of videos but used less generalizable and comparable comprehension tests. In the studies, comprehension was measured by summarization-style activities. Baltova (1999) conducted a study to investigate the comprehension of French language learners viewing a video with different combinations of L2 audio, L1 audio, L1 subtitles, or L2 captions. Comprehension was measured by

counting the number of relevant idea units written in response to eight open-ended questions administered immediately after viewing and again after a two-week delay. Participants viewed the video in either Condition 1 (L1 audio and L2 subtitles, then L2 audio and L2 captions, then finally L2 audio and no captions), Condition 2 (L2 audio and L2 captions twice followed by L2 audio and no captions), or Condition 3 (L2 audio and no captions three times). There was a significant difference between the results from Condition 2 (Immediate  $M = 8.79$  out of 22, Delayed  $M = 6.44$  out of 22) and Condition 3 (Immediate  $M = 4.36$ , Delayed  $M = 3.40$ ) suggesting that captions aided comprehension. However, the extra viewing with the L2 audio and no captions in Condition 2 may have added to comprehension beyond the mere presence of captions. Guichon and McLornan (2008) studied the effects of viewing a short news report with captions, subtitles, no captions, or audio-only on comprehension. Participants in four groups viewed or listened to the video twice, took notes, and then had 20 minutes to write as thorough a summary as possible. On average, the Captions Group referenced 30.2% of the central semantic units, followed by the Subtitles Group (29.7%), the No Captions Group (25.1%), and the Audio-only Group (19.7%). Due to a small sample size, the significance of the results was not calculated. However, taken together the results of these two studies indicate that captions may increase comprehension of videos even when comprehension is measured by summary-based tests.

Other studies have been conducted that examined the effects of captions on the comprehension of videos, but their results have been insufficiently reported such that it is impossible to gauge the degree to which captions affected comprehension. It is still worthwhile to report their results as they help to build a picture of the role that captions play in facilitating comprehension. Garza (1991) compared the comprehension of five discursive videos for both Russian and English language learners. Half the participants viewed the videos twice with captions and half without captions. The comprehension test consisted of content-based items that included informational paraphrases and basic deduction. The results revealed significantly superior performance for the captioned condition in both language learning groups. Yoshino, Kano and Akahori (2000) studied the comprehension of short music videos for four treatment groups: Captions, No Captions, Subtitles, and Audio-only. After viewing a video twice with a short break between, participants were asked to write down every word, phrase, clause, or sentence they could remember from the videos in either their L1 or L2. The participants' were graded by word recall ratio and the accuracy of the recalled information across a variety of criteria. The Captions and Subtitles groups had significantly superior recall and accuracy when compared to the No Captions or Audio-only

groups. Winke, Gass and Sydorenko (2010) studied the effects of captions on comprehension for foreign learners of Spanish. Treatment groups viewed documentaries twice with captions or without captions and completed multiple-choice items concentrating on the main points of the video. The group of language learners who viewed the videos with captions had significantly higher comprehension scores than those who viewed the videos without captions. A Cohen's *d* of 1.1 indicates a large effect size for the treatment. Etemadi (2012) had two groups of participants view each of two documentaries either with captions or without. Comprehension was tested with 10 multiple-choice questions for each video. The Captions Group had significantly higher test scores than the No Captions Group. These studies point to increased comprehension by language learners when captions are present while viewing videos but unfortunately the degree to which captions aided comprehension is not evident.

In contrast with the previously cited research, one study did not find that the presence of captions led to higher comprehension of videos. Taylor (2005) explored the effects of captions on low-level Spanish-language learners' comprehension of a textbook-related video. One group of participants watched the video with captions and a second group watched it without captions. Upon viewing the video, the participants completed a free recall procedure in their L1 and a multiple-choice test. Immediately before viewing the video, the participants were presented with a list of vocabulary as an aid to comprehension. No significant differences were found between the comprehension scores of the Captions Group and the No Captions Group. In fact, on average the No Captions Group had scores 7.6% higher than the Captions Group's on the multiple-choice test (Captions:  $M=7.23$  out of 15, No Captions:  $M=7.78$ ), and 24.9% higher on the recall test (Captions:  $M=1.93$ , No Captions:  $M=2.41$ ). The study appears to indicate that captioning might not be effective for enhancing low-level learners' comprehension. However, numerous low scores on the recall tests (47% of participants with little experience studying Spanish had scores of zero) make the findings less convincing. A lack of motivation rather than a lack of comprehension may be a better explanation for the findings as it seems unreasonable to fail to be able to provide any meaningful answers on a free recall procedure done in the participants' L1.

Taken as a whole, the results from Chung (1999), Guillory (1998), Huang and Eskey (1999), Latifi, Mobalegh, and Mohammadi (2011), Markham and Peter (2003), and Markham, Peter and McCarthy (2001) indicate that providing language learners with captions can be an effective method for increasing comprehension of video. Learners with a variety of target languages and a range of proficiency levels had substantial gains in comprehension compared to learners who were not provided with captions. In this research, the scores on

comparable comprehensions tests were on average 27.95% higher for participants who viewed videos with captions compared to participants who viewed the videos without captions. Comprehension scores were from 14.7% to 41.7% higher for participants viewing videos with captions. Two features of these studies, however, are striking: the types of videos and the length of the videos. Three of these studies used a documentary, two used videos intended for language learners, and one used segments of an animated movie as the input video. The amount of viewing time (where stated) in these studies ranged from 7 to 30 minutes. The relative brevity of the input videos and the types of videos viewed indicate that further research is needed to investigate the effects of captions on comprehension of television programs.

#### **6.4.2. Research questions**

Study 5: Experiment 1 was designed to answer the following research questions:

1. Does comprehension of episodes of English-language television viewed with captions change from the first episode to the tenth episode viewed?
2. Does comprehension of episodes of English-language television viewed with captions change across successive episodes viewed?
3. Does comprehension of English-language television viewed with captions improve with greater vocabulary knowledge?
4. Is comprehension of television with captions superior to comprehension of television without captions?

#### **6.4.3. Results**

##### **6.4.3.1. First episode to final episode comprehension gains**

To examine how comprehension of *Chuck* viewed with captions changed from the first to the final episode viewed, the comprehension test scores for Episode A1 and Episode B2 were analyzed. As shown in Table 6.4, the mean raw score for the first episode viewed was 45.67 out of a possible 76 (60.1%) and on the final episode viewed it was 44.75 out of a possible 71 (63.0%). Table 6.5 shows the mean scores for these episodes in terms of ability measures known as CHIPs scores which are the raw test scores analyzed by the Rasch Model to obtain interval data. For a full explanation of CHIPs see Section 2.8.3.

Table 6.4 Mean scores of raw data on the comprehension tests for Episode A1 (out of 76) and Episode B2 (out of 71)

| Comprehension Test | Mean  | Mean % | <i>N</i> | <i>SD</i> | Median | Minimum | Maximum |
|--------------------|-------|--------|----------|-----------|--------|---------|---------|
| Episode A1         | 45.67 | 60.1%  | 51       | 5.369     | 46     | 29      | 55      |
| Episode B2         | 44.75 | 63.0%  | 51       | 5.741     | 45     | 30      | 56      |

Table 6.5 Means of CHIPs scores on the comprehension tests for Episode A1 and Episode B2

| Comprehension Test | Mean | <i>N</i> | <i>SD</i> | Median | Minimum | Maximum |
|--------------------|------|----------|-----------|--------|---------|---------|
| Episode A1         | 52.3 | 51       | 1.682     | 52.3   | 47.2    | 55.3    |
| Episode B2         | 53.5 | 51       | 2.248     | 53.5   | 47.9    | 58.3    |

A paired-samples *t*-test was conducted to compare comprehension scores (in CHIPs) of the first and tenth episode viewed. There was a significant difference in the comprehension scores for Episode A1 ( $M=52.3$ ,  $SD=1.682$ ) and Episode B2 ( $M=53.5$ ,  $SD=2.248$ );  $t(50)=3.815$ ,  $p < .001$ . The effect size as measured by  $d$  was 0.61, a value corresponding to a medium treatment effect. These results suggest the viewing of eight successive episodes between the first and the final episodes led to a medium, reliable effect on comprehension scores for the participants in Study 5 who viewed the episodes of *Chuck* with English-language captions.

#### 6.4.3.2. Comprehension across all 10 episodes

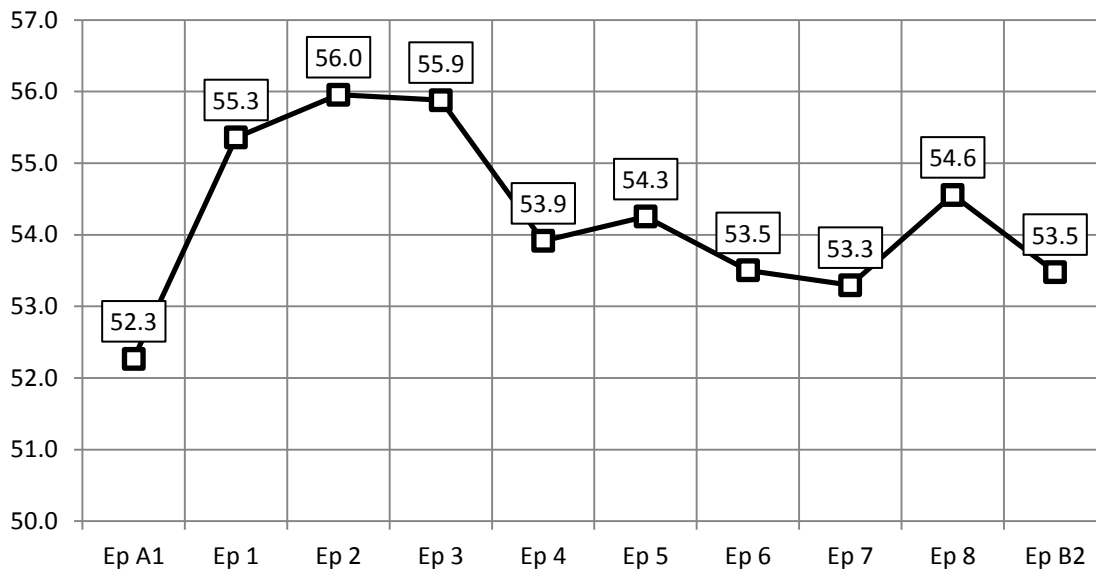
To examine how comprehension of episodes of *Chuck* viewed with captions changed over the duration of the study, the comprehension test scores of the 10 episodes were analyzed. Table 6.6 shows the results, in CHIPs and raw scores, and the descriptive statistics for each comprehension test. While the differences between comprehension tests in terms of mean CHIPs scores may seem minimal, examination of the raw score percentages shows that there was considerable difference between the results of some comprehension tests.

Table 6.6 Comprehension test scores, expressed in CHIPs and raw scores, for all episodes

|                 | Episode |       |       |       |       |       |       |       |       |       |
|-----------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|                 | A1      | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | B2    |
| CHIPs Mean      | 52.3    | 55.3  | 56.0  | 55.9  | 53.9  | 54.3  | 53.5  | 53.3  | 54.6  | 53.5  |
| CHIPs <i>SD</i> | 1.68    | 3.17  | 2.83  | 3.73  | 2.77  | 3.42  | 3.29  | 2.99  | 2.88  | 2.25  |
| CHIPs Median    | 52.3    | 55.5  | 56.0  | 56.2  | 54.3  | 54.8  | 53.4  | 53.1  | 54.5  | 53.5  |
| CHIPs Min.      | 47.2    | 46.4  | 48.7  | 46.8  | 47.2  | 45.4  | 45.3  | 46.1  | 47.1  | 47.9  |
| CHIPs Max.      | 55.3    | 61.9  | 62.1  | 64.7  | 59.6  | 62.2  | 60.0  | 59.2  | 62.2  | 58.3  |
| Raw Mean        | 45.7    | 49.9  | 55.5  | 54.7  | 46.5  | 48.0  | 47.7  | 48.2  | 52.4  | 44.7  |
| Raw %           | 60.1%   | 71.3% | 73.0% | 72.0% | 65.5% | 66.7% | 63.7% | 62.6% | 67.2% | 63.0% |
| Raw <i>SD</i>   | 5.37    | 7.99  | 6.45  | 9.49  | 7.66  | 8.57  | 9.00  | 8.21  | 8.27  | 5.74  |
| Raw Median      | 46      | 51    | 56    | 56    | 48    | 50    | 48    | 48    | 53    | 45    |
| Raw Min.        | 29      | 25    | 36    | 28    | 27    | 24    | 24    | 27    | 29    | 30    |
| Raw Max.        | 55      | 63    | 67    | 71    | 60    | 64    | 63    | 63    | 70    | 56    |

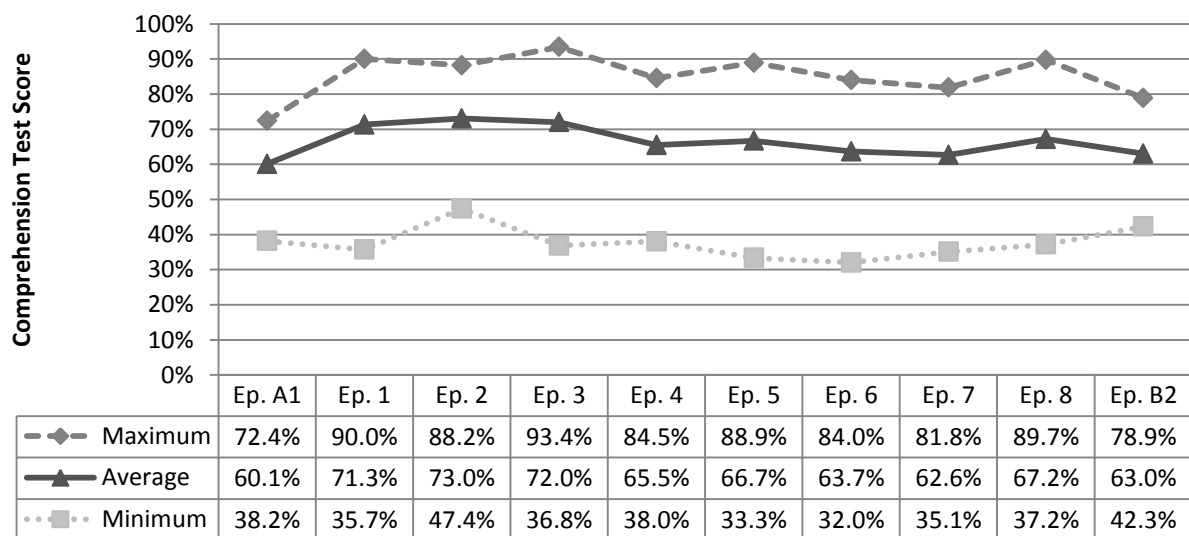
The CHIPs scores show that comprehension varied by episode but after Episode A1 the general trend was an increase in comprehension. Figure 6.3 plots the mean scores for the comprehension tests measured in CHIPs across the 10 episodes. From the first episode the participants viewed to the final episode viewed there is a mean increase of 1.2 CHIPs. After an initial score of 52.3 CHIPs for Episode A1, there is a large rise in comprehension to 55.3 CHIPs for Episode 1 and then to the peak mean score of 56.0 CHIPs for Episode 2. Mean comprehension test scores from Episode 3 to Episode 8 ranged from 53.3 CHIPs (Episode 7) to 55.9 CHIPs (Episode 3). The comprehension scores over the 10 episodes of *Chuck* viewed with captions suggest that participants improved comprehension with successive viewings but there was a considerable difference between comprehension test results for individual episodes.

Figure 6.3 Mean CHiPs comprehension scores across all 10 episodes



There was a good deal of variation in the participants' comprehension test results. This is visible in Figure 6.4 that plots the maximum, average and minimum comprehension test score for each of the 10 episodes. The average comprehension score is consistently over 60% and over 70% for three episodes. The minimum comprehension test score ranged from 32.0% (Episode 6) to 47.4% (Episode 2) while the maximum comprehension test score ranged from 72.4% (Episode A1) to 93.4% (Episode 3). This variation between participants' comprehension test scores indicates that there were members of the sample that were able to achieve a considerable level of comprehension while others were not. The figure also shows that within these proficiency levels there was a general trend of increasing comprehension.

Figure 6.4 Maximum, average and minimum comprehension scores expressed as percentages across all 10 episodes



#### 6.4.3.3. Comparison of comprehension of episodes viewed with captions and without captions

To examine how comprehension of episodes of *Chuck* viewed with captions compared to comprehension without captions, the comprehension test scores from the 10 episodes from Study 5 were compared to those from Study 1. The participants in Study 1 viewed the same episodes and completed the same comprehension tests for each but viewed them without captions. The 51 participants in Study 5 viewed Episode A1 first and Episode B2 last as did 208 participants from Study 1. Another 113 participants from Study 1 viewed Episode B1 as the Initial Episode and Episode A2 as the Final Episode. The results for the two groups of participants in Study 1 were analyzed together, the rationale for which is discussed in Section 2.9.1. For the Captions Group, the results from Episode A1 are denoted as the Initial Episode while the combined results from Episodes A1 and B1 are for the No Captions Group. Similarly the results from Episode B2 are designated as the Final Episode for the Captions Group, and the combined results from Episodes B2 and A2 are the Final Episode for the No Captions Group. Table 6.7 presents the results in CHIPs and descriptive statistics for each comprehension test for the Captions and No Captions groups.

Table 6.7 Mean CHIPs scores on the comprehension tests for all episodes in Study 5 ( $N=51$ ) and Study 1 ( $N=321$ )

| Episode         | Captions Group (Study 5) |           |      |      | No Captions Group (Study 1) |           |      |      |
|-----------------|--------------------------|-----------|------|------|-----------------------------|-----------|------|------|
|                 | CHIPs Score              | <i>SD</i> | Min. | Max. | CHIPs Score                 | <i>SD</i> | Min. | Max. |
| Initial Episode | 52.3                     | 1.68      | 47.2 | 55.3 | 51.0                        | 2.11      | 44.9 | 55.8 |
| Episode 1       | 55.3                     | 3.17      | 46.4 | 61.9 | 55.0                        | 2.82      | 45.9 | 62.3 |
| Episode 2       | 56.0                     | 2.83      | 48.7 | 62.1 | 55.6                        | 2.84      | 47.4 | 62.5 |
| Episode 3       | 55.9                     | 3.73      | 46.8 | 62.8 | 55.3                        | 2.65      | 45.0 | 66.2 |
| Episode 4       | 53.9                     | 2.77      | 47.2 | 59.6 | 52.5                        | 2.57      | 42.6 | 58.9 |
| Episode 5       | 54.3                     | 3.42      | 45.4 | 61.5 | 53.7                        | 2.65      | 44.8 | 61.9 |
| Episode 6       | 53.5                     | 3.29      | 45.3 | 60.0 | 53.0                        | 2.65      | 46.3 | 60.1 |
| Episode 7       | 53.3                     | 2.99      | 46.1 | 59.2 | 52.5                        | 2.50      | 45.3 | 60.2 |
| Episode 8       | 54.6                     | 2.88      | 47.1 | 62.2 | 54.3                        | 2.70      | 45.3 | 61.4 |
| Final Episode   | 53.5                     | 2.25      | 47.9 | 58.3 | 53.2                        | 2.62      | 43.4 | 59.9 |

To compare how comprehension of *Chuck* changed from the first to the tenth episode with and without captions, the comprehension test scores for the Initial Episode and Final Episode were analyzed. The mean CHIPs score for the Initial Episode for the Captions Group was 52.3 and for the Final Episode it was 53.5. For the No Captions Group, the mean score in CHIPs was 51.0 on the Initial Episode and 53.2 on the Final Episode. The group that viewed the

episodes with captions had a mean gain of 1.2 CHIPs and the group that viewed the episodes without captions had a mean gain of 2.2 CHIPs. The descriptive statistics for the two treatment groups for these episodes are presented in Table 6.8.

Table 6.8 Mean gains in CHIPs from the Initial Episode to the Final Episode in Study 5 ( $N=51$ ) and Study 1 ( $N=321$ )

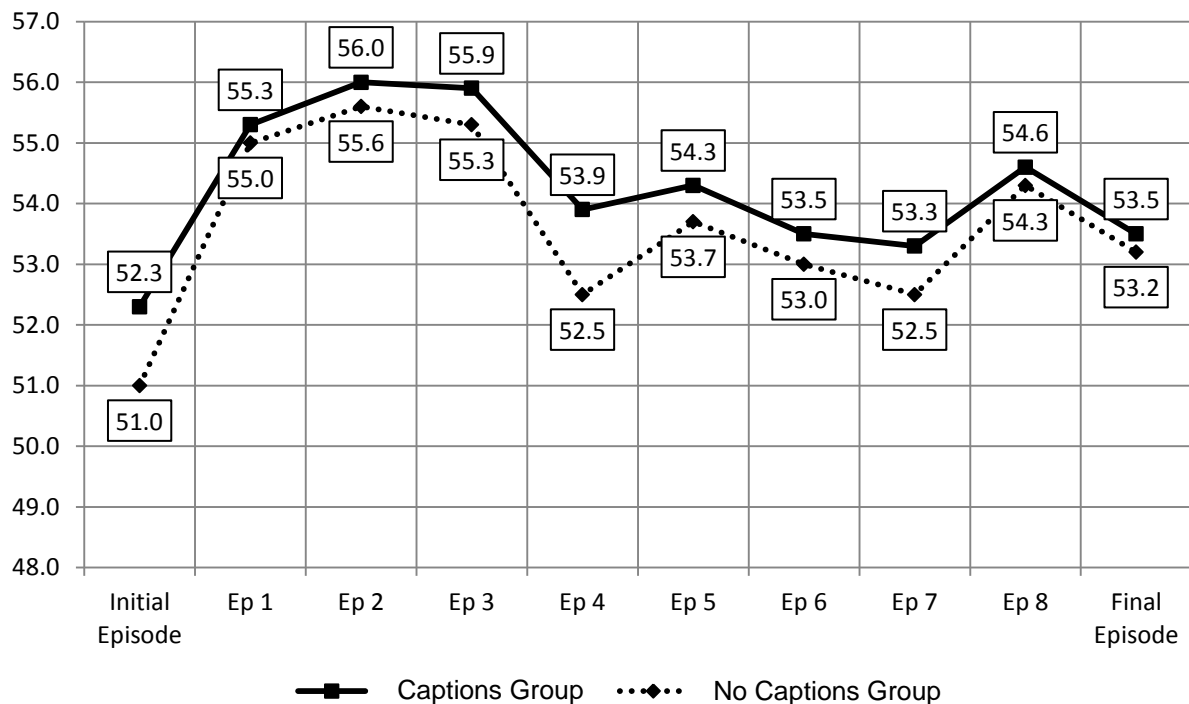
| Episode         | Captions Group (Study 5) |           |      |      | No Captions Group (Study 1) |           |      |      |
|-----------------|--------------------------|-----------|------|------|-----------------------------|-----------|------|------|
|                 | CHIPs Score              | <i>SD</i> | Min. | Max. | CHIPs Score                 | <i>SD</i> | Min. | Max. |
| Initial Episode | 52.3                     | 1.68      | 47.2 | 55.3 | 51.0                        | 2.11      | 44.9 | 55.8 |
| Final Episode   | 53.5                     | 2.25      | 47.9 | 58.3 | 53.2                        | 2.62      | 43.4 | 59.9 |
| Mean Gain       | 1.2                      | 2.27      | -5.7 | 5.8  | 2.2                         | 2.44      | -6.1 | 9.3  |

For both the groups, a significant comprehension gain from the Initial to the Final Episode has previously been established. The details of these analyses can be found in Sections 2.9.1. and 6.4.3.1. To investigate whether there was a difference between the groups in terms of comprehension increase, gains between the Initial and Final episodes viewed for the groups were compared with an independent-samples *t*-test. Prior to performing *t*-tests on the data, an informal analysis of the distribution of the gain scores using a histogram and normal Q-Q plot revealed no serious threats to the assumption of normality. There was a significant difference in comprehension gain between the participants who viewed the episodes with captions ( $M=1.212$ ,  $SD=2.268$ ) and those participants who viewed the episodes without captions ( $M=2.250$ ,  $SD=2.439$ );  $t(370) = 2.850$ ,  $p < .01$ . The effect size, as measured by  $d$ , was 0.44, a value corresponding to a small treatment effect. These results indicate that the participants who viewed the 10 episodes of *Chuck* without captions had significantly greater gains in comprehension from the Initial Episode to the Final Episode than the participants who viewed these episodes with captions.

For all 10 episodes of *Chuck*, the CHIPs scores for both groups show that comprehension varied by episode but after the Initial Episode the general trend was an increase in comprehension. The comprehension scores were also consistently higher for the participants in the Captions Group. The difference in CHIPs between the Captions Group and the No Captions Group for the Initial Episode was 1.3 but by the Final Episode the difference was only 0.3. The average difference between the two groups across the 10 episodes was 0.65 CHIPs. The episode with greatest difference was Episode 4 (1.4 CHIPs) and the episodes with least difference (0.3 CHIPs) were Episode 1, Episode 8, and the Final Episode. Results on the comprehension tests followed a similar pattern with episodes that produced lower or higher comprehension scores for the Captions Group also producing lower or higher comprehension

scores for the No Captions Group. For both groups of participants, the episode with the lowest comprehension score was the Initial Episode and the episode with the highest comprehension score was Episode 2. Figure 6.5 plots the mean scores on the comprehension tests measured in CHIPs across the 10 episodes for the Captions and No Captions groups. The comprehension scores over the episodes viewed suggest that participants who had access to captions, on average, had higher levels of comprehension.

Figure 6.5 Mean CHIPs comprehension scores across all 10 episodes for the Captions Group and the No Captions Group



To determine whether there were any statistically significant differences in mean comprehension scores between the Captions and No Captions groups, independent-samples *t*-tests were conducted comparing the comprehension scores for the two treatment groups for each of the 10 episodes. Prior to performing *t*-tests on the data, an informal analysis of the distribution of these scores using a histogram and normal Q-Q plot revealed no serious threats to the assumption of normality. The *t*-tests revealed that there was a significant difference between those who watched with captions and those without captions for three of the 10 episodes of *Chuck*: Initial Episode [ $t(77.466) = 4.983, p < .001$ ], Episode 4 [ $t(370) = 3.447, p = .001$ ], and Episode 7 [ $t(370) = 2.191, p < .05$ ]. The effect size as measured by *d* was 0.69 for the Initial Episode which equates to a medium effect size, 0.51 for Episode 4 (medium effect size), and 0.31 for Episode 7 (small effect size). These results indicate that for certain

episodes of television viewing with captions can result in significantly higher comprehension scores but the degree to which they affect comprehension depends on the episode.

#### 6.4.3.4. Vocabulary knowledge and comprehension of television with captions

To investigate whether comprehension of English-language television viewed with captions improves with greater vocabulary knowledge a Pearson product-moment correlation coefficient was computed to assess the relationship between the amount of vocabulary knowledge participants had and their comprehension of episodes of *Chuck*. Vocabulary knowledge was quantified by combining participants' results on the three levels of the Vocabulary Levels Test and analyzing them under the Rasch Model. This produces a vocabulary ability measure for each participant. The analysis revealed that there were small significant correlations between the two variables for only four of the 10 episodes: Episode 1 ( $r = .295, n = 40, p < .05$ ), Episode 3 ( $r = .261, n = 40, p < .05$ ), Episode 5 ( $r = .278, n = 40, p < .05$ ), and Episode 8 ( $r = .249, n = 40, p < .05$ ). The correlations for each of the 10 episodes and vocabulary knowledge can be seen in Table 6.9. These results indicate that for certain episodes of television viewed with captions, vocabulary knowledge can be a small but significant factor.

Table 6.9 Pearson correlation results for the comparisons between vocabulary knowledge<sup>†</sup> and comprehension scores for episodes of *Chuck*

| Ep. A1 | Ep. 1 | Ep. 2 | Ep. 3 | Ep. 4 | Ep. 5 | Ep. 6 | Ep. 7 | Ep. 8 | Ep. B2 |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| .217   | .295* | .212  | .261* | .127  | .278* | .206  | .229  | .249* | .047   |

Note. <sup>†</sup> vocabulary knowledge score is a product of a Rasch analysis of the combined VLT 2,000, 3,000, and 5,000 results, \*  $r$  is significant at  $p < .05$

#### 6.4.3.5. Comparison of vocabulary knowledge and comprehension of television viewed with and without captions

To compare how greater vocabulary knowledge affected comprehension of English-language television viewed with and without captions, the correlations between the amount of vocabulary knowledge participants had and their comprehension of episodes of *Chuck* were compared. To investigate whether the two groups had similar levels of vocabulary knowledge before completing the treatments, an independent-samples t-test was conducted to compare the vocabulary knowledge scores (in CHIPs) from the Captions Group ( $M = 50.91, SD = 4.06$ ) and No Captions Group ( $M = 51.46, SD = 3.87$ ). There was no significant difference in the vocabulary knowledge scores;  $t(370) = .939, p = .370$ . For the Captions Group, there were small significant correlations between the two variables for four of the 10 episodes while for

the No Captions Group there were small to medium significant correlations for all 10 of the episodes. The correlations for each of the 10 episodes for both treatment groups are presented in Table 6.10. The results of the computation of the Pearson product-moment correlation coefficient for the Captions Group and No Captions Group are detailed in Sections 6.4.3.4. and 2.9.3., respectively. This comparison indicates that while vocabulary knowledge is related to increased comprehension of certain episodes of television viewed with captions, vocabulary knowledge is more consistently correlated with comprehension when episodes of television are viewed without captions. These findings suggest that vocabulary knowledge may be more important for comprehension when language learners encounter the spoken dialogue of television only aurally, and the added support of captions appears to reduce the effect of increased vocabulary knowledge on comprehension. However, the relatively small population size ( $N = 51$ ) for the Captions Group is likely a factor in the fewer episodes with a statistically significant correlation for this treatment (Nakagawa & Cuthill, 2007). Replication of this study with more participants may improve understanding of the relationship between vocabulary knowledge and the presence of captions when viewing episodes of television.

Table 6.10 Pearson correlation results comparing vocabulary knowledge<sup>†</sup> and comprehension scores for episodes of *Chuck* for the Captions and No Captions groups

| Group       | Initial Episode | Episode 1 | Episode 2 | Episode 3 | Episode 4     |
|-------------|-----------------|-----------|-----------|-----------|---------------|
| Captions    | .217            | .295*     | .212      | .261*     | .127          |
| No Captions | .206***         | .248***   | .180***   | .229***   | .171**        |
|             | Episode 5       | Episode 6 | Episode 7 | Episode 8 | Final Episode |
| Captions    | .278*           | .206      | .229      | .249*     | .047          |
| No Captions | .305***         | .261***   | .235***   | .222***   | .206***       |

Note. <sup>†</sup> vocabulary knowledge score is a product of a Rasch analysis of the combined VLT 2,000, 3,000, and 5,000 results, \*  $r$  is significant at  $p < .05$ , \*\*  $r$  is significant at  $p < .01$ , \*\*\*  $r$  is significant at  $p \leq .001$

#### 6.4.4. Summary of findings

The main findings of Study 5: Experiment 1 can be summarized as follows:

1. There were significant increases in comprehension scores from the first to the last episode viewed with captions which on average equated to a 2.9% or 1.2 CHIPs score increase in comprehension.

2. The comprehension scores of television viewed with captions trended higher over the 10 episodes with large differences between the comprehension scores of participants on individual episodes. After the first episode viewed, average comprehension scores ranged, on the nine subsequent episodes, from 62.5% or 53.3 CHIPs to 72.9% or 56.0 CHIPs.
3. Greater vocabulary knowledge was found to have a small significant correlation with higher comprehension scores for four out of 10 episodes viewed with captions.
4. For both the Captions Group and No Captions Group there was a significant increase in comprehension test results from the Initial Episode to the Final Episode but there were significantly greater gains in comprehension scores for the No Captions Group. The mean CHIPs scores for the Captions Group for all 10 episodes were higher than for the No Captions Group. However, only three of the episodes were found to be significantly higher. For the Captions Group, there were small significant correlations between comprehension and vocabulary knowledge for four of the 10 episodes while for the No Captions Group there were small to medium significant correlations for all 10 of the episodes.

## **6.5. Study 5: Experiment 2 - Captions and Incidental Vocabulary Acquisition**

### **6.5.1. Introduction**

Numerous studies have examined vocabulary learning from viewing videos with captions (Baltova, 1999; Etemadi, 2012; Huang & Eskey, 1999; Hui, 2007; Markham et al., 2001; Markham, 1999; Mitterer & McQueen, 2009; Neuman & Koskinen, 1992; Sydorenko, 2010; Vanderplank, 1990, 1988; Winke et al., 2010; Yuksel & Tanriverdi, 2009). The majority of the studies compared vocabulary acquisition from captioned and uncaptioned videos. Two of these studies were qualitative explorations of language learners' use of captions to learn vocabulary through watching videos (Vanderplank, 1988, 1990). A number of studies tested language learners' gains in knowledge of the form component of target words through viewing captioned videos (Hui, 2007; Markham, 1999; Mitterer & McQueen, 2009; Sydorenko, 2010), while the majority of the studies tested for gains in knowledge of the meaning component of target words (Baltova, 1999; Etemadi, 2012; Huang & Eskey, 1999; Hui, 2007; Markham et al., 2001; Neuman & Koskinen, 1992; Sydorenko, 2010; Winke et al., 2010). One study tested gains in the use component of vocabulary knowledge in addition to measuring gains in form and meaning (Neuman & Koskinen, 1992), and another study used a rating scale to measure the development of word knowledge pre- and post-viewing of a captioned video (Yuksel & Tanriverdi, 2009). These studies are summarized in Figure 6.6 which provides the number of participants, the L1 and target language of the participants, the type of vocabulary test, a summary of the treatment conditions, and the type and length of the input video in each study.

Figure 6.6 Summary of previous research comparing vocabulary acquisition through video with and without captions

| Study                       | Participants                    | L1 Language   | Target Language | Vocabulary Test   | Treatment   | Input Video                                   | Video Length                      |
|-----------------------------|---------------------------------|---|-----------------|---|---|---|-----------------------------------|
| Vanderplank (1988)          | 23 university students          | Various   | English         | Observation of viewing sessions and results of associated L2 tasks                                    | Participants viewed the programs with L2 captions and were asked to consider the frequency they used captions, note unfamiliar words and phrases, and provide specific feedback for particular programs | A variety of television programs from the BBC | Approx. 1 hour of video per week  |
| Vanderplank (1990)          | 25 university students          | French, Spanish, Italian, German, Moroccan, Arabic and Portuguese | English         | Observation of viewing sessions and results of associated L2 tasks                                    | Participants viewed the programs with captions and were asked to note unfamiliar vocabulary and complete oral and written tasks   | A variety of television programs from the BBC | Approx. 4 hours of video per week |
| Neuman, and Koskinen (1992) | 129 junior high school students | Khmer, Lao, Vietnamese and Spanish                                | English         | 90 target words measured by recognition tests, multiple-choice tests and inclusion in retelling tasks | Groups viewed a program with or without captions, read and listened to a script of the program, or worked from a related textbook only  | 9 segments of a children's science program    | 5 to 8 minutes each               |

Figure 6.6 Summary of previous research comparing vocabulary acquisition through video with and without captions (continued)

| Study                              | Participants            | L1 Language       | Target Language | Vocabulary Test   | Treatment  | Input Video                          | Video Length      |
|------------------------------------|-------------------------|-------------------|-----------------|---|--|--------------------------------------|-------------------|
| Baltova (1999)                     | 93 high school students | English           | French          | 30-item cloze test in L2  | Groups viewed the video with L2 audio and captions twice followed once by the L2 audio and no captions, with L1 audio and L2 subtitles then with L2 audio and captions then finally with L2 audio only, or three times with L2 audio only with no captions | 1 documentary                        | 7.5 minutes       |
| Huang and Eskey (1999)             | 30 university students  | Various languages | English         | 8-item multiple-choice test in L2   | Groups viewed the videos with captions or no captions  | 1 episode from an educational series | 21 minutes        |
| Markham (1999)                     | 118 students            | 15 languages      | English         | Two 50-item multiple-choice tests in L2 delivered aurally                                   | Groups viewed the videos with captions or no captions  | 2 excerpts from educational programs | 12 and 13 minutes |
| Markham, Peter and McCarthy (2001) | 169 university students | English           | Spanish         | 10-item multiple-choice test in L2  | Groups viewed the video with L2 captions, L1 subtitles, or video only  | 1 documentary                        | 7 minutes         |
| Hui (2007)                         | 182 university students | Chinese           | English         | 10 target words tested by listening recognition test, spelling test, and meaning test in L2 | High- and low-level groups viewed the video with either subtitles, captions or video only  | 1 documentary                        | 16 minutes        |

Figure 6.6 Summary of previous research comparing vocabulary acquisition through video with and without captions (continued)

| Study                             | Participants            | L1 Language | Target Language | Vocabulary Test  | Treatment  | Input Video                                     | Video Length             |
|-----------------------------------|-------------------------|-------------|-----------------|--|--|---|--------------------------|
| Mitterer and McQueen (2009)       | 120 university students | Dutch       | English         | Participants listened to and repeated excerpts from the videos   | Groups viewed one of two videos with either subtitles, captions or no captions         | An episode of a sitcom and excerpts from a film | 25 minutes each          |
| Yuksel and Tanriverdi (2009)      | 104 university students | Turkish     | English         | Pre- and post-tests of 10 target words using the VKS   | Groups viewed the videos with captions or no captions                                  | A segment of a situation comedy                 | 9 minutes and 14 seconds |
| Sydorenko (2010)                  | 26 university students  | English     | Russian         | 28 target words' written and aural forms were tested via a recognition test, translation test, and word knowledge test | Groups saw the videos with audio and captions, with audio only, and with captions only | 3 segments from a comedy series                 | 2 to 3 minutes each      |
| Winke, Gass, and Sydorenko (2010) | 17 university students  | English     | Spanish         | Test of prior knowledge of target words and translation test presented in written and aural form                       | Groups viewed the videos twice with or without captions                                | 3 documentaries                                 | 3 to 5 minutes each      |
| Etemadi (2012)                    | 44 university students  | Persian     | English         | 10-item multiple-choice test for each video in L2  | Groups viewed the videos with captions or no captions                                  | 2 documentaries                                 | 30 and 20 minutes        |

### **6.5.2. Incidental vocabulary learning from captioned video**

Before quantitatively examining how vocabulary acquisition from videos viewed with captions compares with video viewed without captions, it is useful to examine how captions have been shown to help language learners learn vocabulary. The feedback from the participants in a pair of studies by Vanderplank (1988, 1990) illustrates the ways that captions may influence vocabulary learning. These studies are unique for the amount of time participants spent viewing captioned television. In the first study, Vanderplank showed captioned programs to ESL students over the course of a nine-week period and in the second study participants watched captioned television for three months. Both studies were both informal in nature and do not present quantifiable results for vocabulary acquisition. They do, however, indicate a positive and cumulative effect on vocabulary knowledge from viewing captioned television regularly over long periods. Participants were encouraged to give feedback on specific programs and to reflect on the viewing experience with regards to the degree they depended on the captions, noticed unfamiliar words or phrases, and observed discrepancies between the spoken and captioned text. In the second study, participants were also assigned oral and written tasks before or after viewing.

Vanderplank's results indicate that participants believed their vocabulary knowledge improved through viewing captioned television programs. A key finding was that participants were able to note and ask about words and expressions they had never seen before which are considered important steps in the process of learning vocabulary (Gass, 1999). Improvement in vocabulary knowledge (such as the ability to identify place names, knowledge of the orthographic form, and the ability to discern phonological forms in unfamiliar accents) was believed to be a direct result of having the captions available. Participants were also able to compare their own lexical knowledge with that presented in the programs and were capable of extracting language from the programs, adapting it, and making use of it for their own purposes. One caveat from the second study (Vanderplank, 1990) was that the participants had difficulty recalling the precise language used in some of the programs and summaries of the programs contained very general language and descriptions rather than the specific terminology contained in the television shows. This finding could have implications if target vocabulary is tested for productive knowledge or if vocabulary knowledge is not measured using instruments that do not allow for measurement of partial knowledge (Nation & Webb, 2011). The overall findings from both of Vanderplank's (1988, 1990) studies indicate that language learners who regularly viewed captioned television learned new vocabulary and observations of the language produced in related tasks supported this. There is, however, no

comparison to viewing uncaptioned programs through which it may have been possible to see the degree to which captions made a difference in the acquisition of vocabulary. The following section examines studies that make a direct comparison between captioned and uncaptioned video allowing for a better understanding of the degree to which captions may aid vocabulary acquisition.

#### **6.5.2.1. Acquisition of the meaning component of vocabulary knowledge from captioned video**

Studies that attempt to measure gains in vocabulary knowledge through viewing videos with and without captions do not usually attempt to measure all components of word knowledge but rather focus on different components of word knowledge with specifically designed tests. When comparing the research on vocabulary gains from viewing captioned and noncaptioned videos it is practical to organize the studies by the components of word knowledge being tested. Nation's (2001) vocabulary knowledge framework classifies word knowledge most broadly into three components: form, meaning, and use. While Nation breaks down these components into finer delineations it is beyond the scope of this survey of relevant research to classify the studies in such a detailed manner. By grouping the results of previous research by the component of vocabulary knowledge tested, vocabulary gains can be more easily compared and contrasted. While the studies cited here often have a variety of treatment conditions, the focus of this survey of literature is between each study's Captions Group and No Captions Group in order to examine how the presence of captions might affect incidental vocabulary acquisition.

In studies examining viewing videos with captions and incidental vocabulary acquisition, the most common form of vocabulary knowledge test focused on measuring an increase in knowledge of form-meaning connection of the target words. Results from the majority of these studies indicate that the presence of captions when viewing videos leads to increased vocabulary learning when compared to videos without captions. In one such study, Neuman and Koskinen (1992) measured knowledge of form-meaning connection through a multiple-choice test that followed viewing segments of an educational program. This study examined the difference between vocabulary acquisition from viewing videos with captions, videos without captions, and learning from reading and listening to a script of the program. Knowledge of the target words was measured by a pre-test that had the participants indicate whether they knew the meaning of the words. Results from this test were used to adjust the results on the post-test for prior knowledge. On the post-test, the results for the Captions

Group ( $M=56.56$  out of 90) were significantly higher than those of the No Captions Group ( $M=52.34$ ). Results for both the Captions and No Captions Groups were significantly higher than the Reading and Listening Group ( $M=40.59$ ). The results of this study suggest that vocabulary learning may be improved through the inclusion of captions.

Baltova (1999) examined how learning and retention of vocabulary was affected by viewing a documentary presented with different combinations of audio and captions: Condition 1 - video viewed with L1 (English) audio and L2 (French) subtitles, then viewed with L2 audio and L2 captions, then finally viewed with L2 audio and no captions, Condition 2 - video twice viewed with L2 audio and L2 captions followed by viewing with L2 audio and no captions, and Condition 3 - video viewed three times with L2 audio and no captions. Knowledge of form-meaning connection of 30 French target words from the video was measured by a cloze test immediately after the final viewing and again two weeks later. For both the immediate test (Condition 1:  $M=14.55$ , Condition 2:  $M=21.85$ , Condition 3:  $M=11.66$ ) and the delayed test (Condition 1:  $M=11.65$ , Condition 2:  $n=19.35$ , Condition 3:  $n=12.16$ ), the participants who viewed the video under Condition 2 (two viewings with captions and once without) had significantly higher scores than those participants with other viewing conditions. While the design of this study makes it difficult to ascertain whether the presence of captions had the greatest effect on vocabulary learning or whether the combinations of the captioned and noncaptioned viewing had a more significant effect, the study does point to the efficacy of video with captions over video without captions for learning form-meaning connection of vocabulary.

In a study using a video designed for language learners, Huang and Eskey (1999) examined the effects of captioned video on vocabulary and phrase acquisition. Two groups of participants, considered to be intermediate proficiency by a standardized placement exam, viewed an educational video twice, and then completed a ten-minute comprehension and vocabulary test delivered aurally. The Captions Group had a mean score of 5.07 out of 8 items compared with 3.47 for the No Captions Group. There was a significant difference between the groups. The relatively few target words that might be overtly presented in a learner-centered video makes comparison to longer television programs challenging. The study, however, does indicate that there is greater potential to learn form-meaning connection of vocabulary, using aural tests, when a video is presented with captions.

Markham, Peter and McCarthy (2001) studied the effects of captions, no captions, or subtitles on knowledge of vocabulary taken from a science-themed video. Three groups of participants viewed the video once and answered multiple-choice questions focused on form-

meaning connection of words appearing in the video. The Subtitles Group ( $M=8.24$  out of 10) significantly outperformed the Captions Group ( $M=5.67$ ) which in turn significantly outperformed the No Captions Group ( $M=4.67$ ) on the vocabulary test.

Hui (2007) studied the effects of captions, no captions, and subtitles on incidental vocabulary acquisition from viewing a documentary. Three groups of high and low proficiency participants viewed the video twice then were tested on 10 target words determined by a recognition-based pre-test. Knowledge of the target words was measured in three ways: recognition of spoken phonological form, production of orthographic form (results of form-related tests are detailed in Section 6.5.2.2), and form-meaning connection. On the form-meaning connection test, the Subtitles Group scored the highest followed by the Captions Group and the No Captions Group for both high (Subtitles Group:  $M=6.20$ , Captions Group:  $M=5.77$ , No Captions Group:  $M=4.07$ ) and low (Subtitles Group:  $M=4.65$ , Captions Group:  $M=4.03$ , No Captions Group:  $M=3.30$ ) proficiency levels. There was a significant difference between the Captions Group and the No Captions Group for participants at both levels of proficiency.

Sydorenko (2010) investigated language learners viewing videos under one of three conditions: captions, no captions, or captions but no audio. In addition to measuring receptive knowledge of the spoken and written form of the target vocabulary (results detailed in Section 6.5.2.2), gains in knowledge of form-meaning connection were measured through the percentage of target words correctly translated (L2 to L1). Half the target words tested were presented aurally and half were presented in their written form. Analysis of the results indicated that the Captions Group (Written:  $M=36\%$ , Aural:  $M=35\%$ ) was not significantly higher than the No Audio Group (Written  $M=28\%$ , Aural:  $M=24\%$ ), but significantly higher than the No Captions Group (Written:  $M=25\%$ , Aural:  $M=18\%$ ). The results from this study indicate the positive effects of captions on acquisition of form-meaning connection and that knowledge of word meaning is gained even when words are presented in videos only in the written form.

As part of a larger study examining the effects of the viewing order of captions and no captions treatments on vocabulary learning and comprehension, Winke, Gass, and Sydorenko (2010) examined the effects of captions on vocabulary learning. Two groups of language learners first viewed three short videos twice with or without captions and then completed post-tests. Language learners first completed an L2 to L1 translation test with half the target words presented in written form and half presented aurally. The test was counterbalanced for mode of presentation. Next, participants completed a rating scale-based test of prior

knowledge that was used to ascertain the participants' knowledge of target words prior to viewing. Vocabulary identified as unknown before viewing formed the basis of each participant's list of target words on the translation test. The results revealed that the Captions Group (Written:  $M=17\%$ ; Aural:  $M=14\%$ ) had a significantly higher percentage of correct items both aurally and in the written form compared with the No Captions Group (Written:  $M=4\%$ ; Aural:  $M=4\%$ ). This study indicates that the presence of captions can lead to gains in knowledge of form-meaning connection regardless of whether the words are tested aurally or written.

One study did not find a positive effect for the presence of captions on the acquisition of form-meaning connection through viewing videos. For each of two videos, Etemadi (2012) identified 10 target words that were deemed to be advanced and beyond the level of the participants. Knowledge of the vocabulary was measured by multiple-choice cloze items administered post-viewing. Two groups of participants viewed both videos either with or without captions in a crossover design. The raw scores from the vocabulary tests were not reported but the author indicated that no significant difference was found between watching the video with or without captions. There are, however, some possible reasons why the presence of captions did not lead to superior vocabulary learning in this study. One possible explanation was that knowledge of the target words differed prior to viewing. While the researcher believed that the words were beyond the ability of the participants, they could possibly have had knowledge of some of the target words. Ten target words is also relatively few for videos 20 to 30 minutes in length and there is no mention of the frequency of occurrence of the words in the videos. If the target words only occurred once then lack of encounters might also be a reason for poor acquisition. Another possible reason for the lack of a significant difference between the viewing conditions may be the design of the test items themselves. Knowledge of the target words was measured by having participants choose the word that best completed a sentence. The sentences functioning as stems for each item were example sentences taken from a dictionary so an abundance of contextual support was included. This may mean that participants used this contextual information to answer the items rather than using knowledge of meaning obtained from the videos. The limitations of the study are merely hypotheses as to why captions did not facilitate vocabulary acquisition. However, what the limitations do indicate is a lack of rigor in research design which makes it hard to accept the findings with any great confidence.

#### **6.5.2.2. Acquisition of other components of vocabulary knowledge through viewing captioned video**

There have been a number of studies that have investigated the effects of captions accompanying video on the acquisition of the form component of word knowledge. For two of these studies (Hui, 2007; Sydorenko, 2010), the tests were part of a series of tests measuring multiple components of vocabulary knowledge. More detailed explanations of the methodologies and findings from these studies can be found in Section 6.5.2.1 which presents previous research on the meaning component of vocabulary knowledge. Hui (2007) studied the effects of captions, no captions, or subtitles on recognition of the spoken form and production of the orthographic form for 10 target words. On the aural word recognition test the participants in the Captions Group scored the highest followed by the Subtitles Group and the No Captions Group for participants in both the high (Captions Group:  $M=7.93$ , Subtitles Group:  $M=7.07$ , No Captions Group:  $M=5.17$ ) and low (Captions Group:  $M=6.16$ , Subtitles Group:  $M=5.52$ , No Captions Group:  $M=4.37$ ) proficiency levels. On the orthographic test, the Captions Group scored higher than the Subtitles Group and the No Captions Group for both the high (Captions Group:  $M=6.83$ , Subtitles Group:  $M=5.47$ , No Captions Group:  $M=4.37$ ) and low (Captions Group:  $M=5.03$ , Subtitles Group:  $M=4.03$ , No Captions Group:  $M=3.47$ ) proficiency levels. The Captions Group produced significantly higher scores on both tests at both proficiency levels. This study demonstrates the possible gains in both written and spoken form through watching a video with captions.

Sydorenko (2010) measured receptive knowledge of the spoken and the written forms of target vocabulary after participants viewed short videos. Knowledge of the aural form of half the target words and the written form of half the target words was measured by a recognition test that also included nonwords as a control. The Captions Group (Recognition: Written  $M=73\%$ ; Recognition: Aural  $M=67\%$ ) scored significantly higher on the written items, while the No Captions Group (Recognition: Written  $M=63\%$ ; Recognition: Aural  $M=69\%$ ) scored significantly higher on the items tested aurally. When the written and aural items were combined to measure overall learning of the target words there was no significant difference between the groups. While the results of this study are somewhat limited by the small number of participants in each treatment group (Captions Group:  $n=8$ , No Captions Group:  $n=9$ ) and would be greatly aided by more participants at a range of proficiency levels, overall, the results of this study indicate that captions can lead to increased recognition of written form.

For two studies, vocabulary tests focused exclusively on the form component of vocabulary knowledge (Markham, 1999; Mitterer & McQueen, 2009). Markham (1999)

examined acquisition of 100 target words found in two videos. The major difference between the two videos was that Video 1 had a high correlation between the audio and the imagery while the imagery for Video 2 provided little support for the text. The hypothesis was that there would be greater learning of the words from the video with the more closely related video and audio. The target words had a frequency of occurrence in the videos of 2 to 5 times with an average of 3 occurrences per video. Vocabulary knowledge was measured aurally in a multiple-choice cloze-type test with sentences taken directly from the videos. The phonological similarity of the options in each multiple-choice item meant that knowledge of form was measured but not exclusively as some knowledge of the target words' meanings was necessary to complete the sentences. Two groups of participants viewed each video either with or without captions in a crossover design. The availability of captions led to a significantly higher number of word forms recognized (Video 1 with captions  $M=72.1\%$ , Video 2 with captions  $M=66.5\%$ ; Video 1 without captions  $M=64.1\%$ , Video 2 without captions  $M=55.5\%$ ). The findings suggest that the presence of captions in videos helps to improve L2 aural word recognition and this improvement occurs regardless of whether the images and the spoken text have a strong relationship. This might be particularly important for narrative videos like television dramas where, unlike documentaries or educational programs, the goal of the program is to entertain rather than inform and there may be less intentional relationship between images and the spoken text.

Mitterer and McQueen (2009) compared the effects of captions and subtitles on language learners' adaptation to an unfamiliar regional accent. The hypothesis was that subtitles and captions indicate which words are being spoken, and this has the potential to allow viewers to retune speech-sound categories based on their knowledge about how foreign words ought to sound. Participants watched video material with either strongly accented Australian English or strongly accented Scottish English. For each video, three groups watched the videos with one of L2 captions, L1 subtitles, or no captions. The participants exposed to Scottish English served as a control group for the Australian English treatment and vice versa. The participants were tested on 160 audio-only excerpts from both the Australian and Scottish materials regardless of whether they had been exposed to the language or not. The tests were scored according to the number of content and function words that were repeated correctly for each excerpt. For the Australian accented items, the Captions Group was 9% better on encountered excerpts and 14% better on unencountered excerpts than the Control Group, and 8% and 6% better than the Subtitles Group. For the Scottish accented items, the Captions Group was 6% better on encountered excerpts and 8% better on unencountered excerpts than the Control

Group, and 3% better and equal to the Subtitles Group. Overall, captions were shown to have a positive effect on adaptation to unfamiliar accented speech for both the encountered and unencountered items when compared to subtitles or no captions. The improved adaptation associated with viewing with captions suggests that the listeners benefited from recognizing the written forms of the words they were hearing and used this knowledge to retune their phonetic perceptions. This increase in phonological knowledge through captions could possibly lead to an increase in other components of vocabulary knowledge as more words are recognized in successive encounters.

In addition to measuring the meaning component of word knowledge, Neuman and Koskinen (1992) also measured the form and use components in a study examining vocabulary acquisition from viewing three units of educational television with captions, without captions, and reading and listening to the script. Knowledge of the written form of the target words was measured with a word recognition test. The use component of vocabulary knowledge was measured by a retelling task that counted the participants' frequency of use of the target words, and a sentence anomaly test where participants indicated whether or not a sentence used a target word correctly. On the recognition test, the Captions Group (Unit 1:  $M=22.15$ , Unit 2: 21.23, Unit 3: 22.46 out of 30), while scoring higher for all three units, had significantly greater scores than the No Captions Group (Unit 1:  $M=20.17$ , Unit 2: 17.97, Unit 3: 20.17) on only the Unit 2 vocabulary. The Captions Group (Unit 1:  $M=6.16$ , Unit 2: 4.34, Unit 3: 2.75) had a higher frequency of target word usage in the retelling tasks across all three units and was significantly higher than the No Captions Group (Unit 1:  $M=5.19$ , Unit 2: 2.78, Unit 3: 1.70) on the first two units. The results of the sentence anomaly test were similar with the Captions Group (Unit 1:  $M=20.85$ , Unit 2: 19.24, Unit 3: 21.23 out of 27) scoring higher across all three units but only significantly higher than the No Captions Group (Unit 1:  $M=20.28$ , Unit 2: 17.50, Unit 3: 20.38) on the vocabulary from Unit 2. The Captions Group had significantly higher scores than the Reading & Listening Group on all vocabulary tests and tasks except for the retelling task for Unit 3. The results of this study indicate that when videos are accompanied by captions there is better incidental vocabulary learning across a variety of different components of vocabulary knowledge but the significance of the results are dependent on the target words and videos being tested.

Finally, there was a single study that used Wesche and Paribakht's (1996) Vocabulary Knowledge Scale (VKS) to measure development of vocabulary knowledge. This 5-point rating scale focuses not on one component of word knowledge but treats knowledge of form, meaning and use as a developmental continuum. Yuksel and Tanriverdi (2009) investigated

incidental vocabulary learning from watching a portion of a television program with and without captions. Participants rated the target words pre- and post-viewing using the VKS. A *t*-test comparing the pre-test averages for both treatment groups (Captions Group:  $M=1.86$ ,  $SD=0.26$ ; No Captions Group:  $M=1.74$ ,  $SD=0.37$ ) demonstrated that there were no significant difference between the groups. These results indicate that the majority of participants believed they had either never seen the majority of the target words before or recognized the forms but did not know the meanings of the target words. The post-test averages (Captions Group:  $M=2.77$ ,  $SD=0.41$ ; No Captions Group:  $M=2.57$ ,  $SD=0.33$ ) indicate that the majority of participants believed that for most of the target words they either had seen the word before but did not know what it meant, or they thought they recognized the word and thought they knew what it meant. Analysis showed that both groups had significant gains from pre- to post-test but that there was no statistical difference between the gains of the two groups. The researchers hypothesized that the development in knowledge of the target words stemmed from encountering the words in context and that viewing the video clip helped them develop their vocabulary knowledge regardless of the presence or absence of captions. However, because the participants had been primed for the target words in the pre-test, it makes it difficult to assess to what extent the video clip was the only impetus for the development of vocabulary knowledge. The study indicates the potential for language learners to notice unknown vocabulary in authentic television programs. However, greater exposure to the language may be necessary to improve on the small gains in vocabulary knowledge reported in this study. This could conceivably be accomplished through viewing full and successive episodes of the same television series.

Overall, the results of Sydorenko (2010), Winke, Gass, & Sydorenko (2010), Hui (2007), Markham, Peter, & McCarthy (2001), Huang & Eskey (1999), Baltova (1999), and Neuman & Koskinen (1992) indicate that having captions available to language learners when viewing videos can lead to acquisition of form-meaning connection of vocabulary occurring in videos. Learners at a range of proficiency levels across an array of target languages experienced significant gains over learners who did not have access to captions when viewing videos. Other studies (Hui, 2007; Markham, 1999; Mitterer & McQueen, 2009; Sydorenko, 2010) also demonstrated gains in other components of vocabulary knowledge including form and use. The variety of tests and the different number of items examined in these studies makes direct comparison of the findings impractical. Two comparable features, however, are the types of videos used and the amount of exposure that language learners had to captioned videos. The following types of videos were shown to participants: segments of a comedy

series, documentaries, educational videos, an episode of a comedy series, and segments of a movie. Aside from the episode of a comedy series these videos are not a representative sample of what a language learner might choose to learn English from over a prolonged period of time. Watching a single episode of a television series may also not be very typical viewing behavior. The amount of viewing ranged from approximately seven minutes to 25 minutes. Considering that a standard American comedy series is 22 minutes and a drama is 44 minutes, typical viewing would likely involve greater exposure to language than shown in these studies. This highlights the need for further research on vocabulary learning through viewing greater amounts of captioned videos and through viewing videos that are more characteristic of common television viewing habits.

### **6.5.3. Research questions**

Study 5: Experiment 2 was designed to answer the following research questions:

7. Does viewing English-language television with captions lead to increased word knowledge for Japanese EFL learners?
8. Do English language learners with greater vocabulary knowledge incidentally learn more vocabulary through television viewed with captions than learners with less vocabulary knowledge?
9. Are unknown words that occur more frequently in captioned television programs more likely to be learned?
10. Are unknown words that occur across a greater range of episodes of a captioned television program more likely to be learned?
11. Is incidental vocabulary acquisition greater from viewing television with captions than without captions?

### **6.5.4. Results**

#### **6.5.4.1. Analysis of the results from the Tough and Sensitive Tests**

The effect of viewing the 10 episodes on the vocabulary knowledge of the 40 participants from Study 5 was examined by measuring knowledge of the *Chuck*-specific target words at two sensitivities before and after viewing. The mean number of target words known on the Tough Test was 33.2 while on the Sensitive version it was 38.6. For a target word to be considered known by a participant, it had to be answered correctly on both the pre- and post-test. The mean number of words known indicates that the participants knew between 53.3% and 64.3% of the target words before viewing episodes of *Chuck* depending on the sensitivity

of the test. On the Tough Test, the maximum number of targets words known was 44 words and the minimum number known was 18. On the Sensitive Test, the maximum was 50 and the minimum was 22. Therefore, there was a lot of variation in the amount of vocabulary learning possible between individuals.

Vocabulary gain was determined by subtracting the number of target words identified as known from the number of targets words correct on the post-test. The mean vocabulary gains on the Tough and Sensitive Tests were 6.03 and 5.93 words respectively. There was considerable variance in gain as the minimum gains, maximum gains, and standard deviations in Table 6.11 indicate.

Table 6.11 Vocabulary gains for Tough and Sensitive Tests for Study 5 participants ( $N=40$ )

| Vocabulary Test | Mean Real Gain | <i>SD</i> | Median | Minimum | Maximum |
|-----------------|----------------|-----------|--------|---------|---------|
| Tough           | 6.03           | 2.741     | 6      | 0       | 14      |
| Sensitive       | 5.93           | 3.214     | 6      | 1       | 14      |

In addition to calculating the raw gains on the two vocabulary tests it was also necessary to determine the relative vocabulary gains of the participants in Study 5. Because participants who knew more of the target words had less room for improvement than participants who knew fewer target words, absolute word gains may not give a complete picture of vocabulary learning. For this reason, relative vocabulary gain was used to take into consideration the varying opportunities for gain. For the participants who viewed the episodes of *Chuck* with captions, the mean, median, minimum and maximum relative gains on the Tough and Sensitive Tests are shown in Table 6.12. The mean relative gains were 23.55% and 28.84% for the Tough and Sensitive tests, respectively. There were large differences between the minimum and maximum relative gains for both tests.

Table 6.12 Relative vocabulary gains for Tough and Sensitive Tests for Study 5 participants ( $N=40$ )

| Vocabulary Test | Mean Relative Gain | Median | Minimum | Maximum |
|-----------------|--------------------|--------|---------|---------|
| Tough           | 23.55%             | 23.67% | 0%      | 53.85%  |
| Sensitive       | 28.84%             | 28.17% | 6.67%   | 78.57%  |

#### **6.5.4.2. The effects of viewing captioned television on incidental vocabulary acquisition**

To determine whether the raw vocabulary gains on the Tough and Sensitive Tests were significant, a series of *t*-tests were carried out. Prior to performing *t*-tests on the data, an informal analysis of the distribution of the gain scores on the Tough and Sensitive Tests using

a histogram and normal Q-Q plot revealed no serious threats to the assumption of normality. Paired samples *t*-tests showed that the difference between mean gains for the Tough,  $t(39) = 13.903, p < .001$ , and Sensitive,  $t(39) = 11.661, p < .001$ , tests were significant. There were large treatment effects for both the Tough ( $d=0.81$ ) and the Sensitive Test ( $d= 0.81$ ). A paired sample *t*-test showed that there was no significant difference between the mean gains on the Tough and Sensitive Tests,  $t(39) = .191, p < .850$ , for the participants. These results indicate that the participants in Study 5 made significant, large gains in the number of target words they knew through viewing episodes of English-language television.

The results of independent samples *t*-tests showed no significant difference between the Captions Group and the Control Group (described in Section 3.3.1) on either the Tough,  $t(111) = 1.611, p < .151$ , or Sensitive,  $t(111) = 1.104, p < .272$ , tests. The results for the Control Group on the Tough and Sensitive Tests are presented in Table 6.13. These results indicate that while the participants in Study 5 made significant gains in the number of target words they learned through viewing episodes of English-language television with captions, they were not significantly different from the 73 participants in the Control Group who did not view any episodes of *Chuck* and studied English in a more conventional setting for the same period of time.

Table 6.13 Vocabulary gains for Tough and Sensitive Tests for Control Group participants ( $N=73$ )

| Vocabulary Test | Mean Real Gain | <i>SD</i> | Median | Minimum | Maximum |
|-----------------|----------------|-----------|--------|---------|---------|
| Tough           | 5.32           | 1.914     | 5      | 0       | 9       |
| Sensitive       | 5.37           | 2.118     | 5      | 0       | 9       |

#### 6.5.4.3. The effects of vocabulary knowledge on relative vocabulary gain through viewing English-language television with captions

A Pearson product-moment correlation coefficient was used to assess the relationship between the relative vocabulary gains that resulted from viewing television with captions and the vocabulary knowledge of the participants. Vocabulary knowledge was calculated by combining the results of the three Vocabulary Levels Tests at the 2,000-, 3,000-, and 5,000-word levels and analyzing the results using the Rasch Model to obtain an interval measure of vocabulary knowledge. There was a significant correlation between the relative vocabulary gain on the Tough Test,  $r = .307, N=40, p < .05$ , and the Sensitive Test,  $r = .270, N=40, p < .05$ , and vocabulary knowledge. The correlation for the Tough Test corresponds with a medium effect size and the correlation for the Sensitive Test corresponds with a small effect

size (Hopkins, 2010). These results indicate that participants with more vocabulary knowledge were better able to increase their relative vocabulary learning through watching television with captions.

#### **6.5.4.4. The effects of frequency and range of occurrence on incidental vocabulary acquisition**

To investigate whether words that occur more frequently and words that occurred across a greater range of captioned television episodes were more likely to be learned by the participants in Study 5, gains in the number of participants who learned the target words were analyzed. The mean raw gains for items on the Tough and Sensitive Tests were 4.02 and 3.95 people respectively. That is, across the 60 items on the vocabulary tests, an average of approximately four of the 40 participants learned each item. There was, however, considerable variation in the number of participants who learned items. On both the Tough and Sensitive Test, the minimum increase in participants gaining knowledge of an item was 0. On the Tough Test, the maximum increase was 12 participants and on the Sensitive Test it was 19 participants. Table 6.14 presents a summary of these results. The results for each item, including the number of participants who knew each target word, learned a target word, did not learn the target word, and the relative gain, on both the Tough and Sensitive tests are presented in Appendix E3.

Table 6.14 Vocabulary gains for Tough and Sensitive Tests for items ( $N=60$ )

| Vocabulary Test | Mean Real Gain | <i>SD</i> | Median | Minimum | Maximum |
|-----------------|----------------|-----------|--------|---------|---------|
| Tough           | 4.02           | 3.111     | 4      | 0       | 12      |
| Sensitive       | 3.95           | 3.591     | 3.5    | 0       | 19      |

To take into account the differing numbers of participants who knew items and therefore differing amounts of possible gain, the relative gain of the items was calculated. The mean relative gains for items were 32.40% and 30.85% on the Tough and the Sensitive tests, respectively. There were large differences between the minimum and maximum relative gains for both tests. The descriptive statistics for this analysis are presented in Table 6.15. The results for each item, including the number of participants who knew each target word, learned a target word, did not learn the target word, and the relative gain, on both the Tough and Sensitive tests are presented in Appendix E3.

Table 6.15 Relative gains for Tough and Sensitive Tests for items ( $N=60$ )

| Vocabulary Test | Mean Relative Gain | Median | Minimum | Maximum |
|-----------------|--------------------|--------|---------|---------|
| Tough           | 32.40%             | 28.71% | 0%      | 100%    |
| Sensitive       | 30.85%             | 25.83% | 0%      | 100%    |

The 60 target words that the Tough and Sensitive Test items were based on occurred from 5 to 54 times throughout the 10 episodes of *Chuck*. The target words had an average frequency of occurrence of 10.2 times across these episodes. The range of occurrence for the target words was from one episode to 10 episodes with an average range of 3.7 episodes. A Pearson product-moment correlation coefficient was used to assess the relationship between the relative gain for target words and the frequency and range of occurrence of the target words across the 10 episodes of *Chuck*. There was a medium size significant correlation between the relative gain for items on the Tough Test and frequency of target word occurrence,  $r = .392$ ,  $N=60$ ,  $p < .01$ . The relative gains for items on the Sensitive Test had a small non-significant negative correlation,  $r = -.021$ ,  $N=60$ ,  $p = .873$ , with the frequency of target word occurrence. There was a small non-significant positive correlation,  $r = .194$ ,  $N=60$ ,  $p = .137$ , between the relative item gains for the Tough Test and the range of occurrence of the target words, and a small significant negative correlation for the Sensitive Test,  $r = -.274$ ,  $N=60$ ,  $p < .05$ . It is also useful to consider range and frequency together and analyze a target word's relative frequency of occurrence across the 10 episodes rather than range or frequency alone. Relative frequency is the overall frequency of occurrence of a target word in the episodes divided by the number of episodes (range) that the target word occurred in. There was a medium size significant correlation between the relative frequency of target word occurrence and the relative gain for items on both the Tough Test ( $r = .344$ ,  $N=60$ ,  $p < .01$ ) and the Sensitive Test ( $r = .426$ ,  $N=60$ ,  $p < .01$ ).

#### 6.5.4.5. Comparison of incidental vocabulary acquisition through viewing English-language television with and without captions

To compare incidental vocabulary acquisition from viewing television with and without captions, the vocabulary gains for the Captions and No Captions Groups were compared. For the Captions Group, the mean number of target words known on the Tough Test was 33.2 (out of a possible 60) while on the Sensitive Test it was 38.6. The No Captions Group knew fewer target words with a mean of 32.1 known on the Tough Test and 37.2 on the Sensitive Test. On the Tough Test, the mean vocabulary gain for the Captions Group was 6.03 words and for the No Captions Group the mean gain was 6.36 words. The mean gain for the Captions Group on

the Sensitive Test was 5.93 words and for the No Captions Group the mean was 6.78 words. The mean vocabulary gains for each treatment group on both vocabulary tests along with the standard deviation, median gain, minimum gain, and maximum gain are presented in Table 6.16. To examine the effects of captions on incidental vocabulary acquisition, independent samples *t*-tests were computed on the vocabulary gains from the Captions Group and No Captions Group. Prior to performing the *t*-tests, an informal analysis of the distribution of the gain scores on the Tough and Sensitive Tests for both treatment groups using a histogram and normal Q-Q plot revealed no serious threats to the assumption of normality. Independent samples *t*-tests showed no significant difference between mean vocabulary gains for the Captions and No Captions Groups on the Tough, ( $t(225) = 0.685, p < .494$ ), and Sensitive, ( $t(225) = 1.409, p < .159$ ) Tests. These results indicate that the vocabulary gains made by the participants in the Captions Group were not significantly different from those made by the participants in the No Captions Group.

Table 6.16 Vocabulary gains for Tough and Sensitive Tests for participants in the Captions Group ( $N=40$ ) and the No Captions Group ( $N=187$ )

| Vocabulary Test | Group       | Mean Real Gain | <i>SD</i> | Median | Minimum | Maximum |
|-----------------|-------------|----------------|-----------|--------|---------|---------|
| Tough           | Captions    | 6.03           | 2.74      | 6      | 0       | 14      |
|                 | No Captions | 6.36           | 2.86      | 6      | 1       | 20      |
| Sensitive       | Captions    | 5.93           | 3.21      | 6      | 1       | 14      |
|                 | No Captions | 6.78           | 3.52      | 6      | 1       | 23      |

Participants who knew more of the target words had less room for improvement than participants who knew less target words. Relative vocabulary gain takes into consideration the varying opportunities for gain that absolute gains do not. The relative gain scores from the Captions Group were compared to the No Captions Group. On the Tough Test, the mean relative vocabulary gain for the Captions Group was 23.55% and for the No Captions Group the mean relative gain was 23.03%. The mean for the Captions Group on the Sensitive Test was 28.84% and for the No Captions Group the mean gain was 29.61%. The results for the relative gains of the participants in the Captions and No Captions Group on the Tough and Sensitive Tests are shown in Table 6.17. Independent samples *t*-tests showed no significant difference between relative vocabulary gains for the Captions and No Captions Groups on the Tough, ( $t(225) = 0.274, p < .724$ ), and Sensitive, ( $t(225) = 0.313, p < .753$ ) Tests. The results of this analysis indicate that the presence of captions did not have an effect on the relative

vocabulary gains of participants in the Captions Group compared to the participants in the No Captions Group.

Table 6.17 Relative vocabulary gains for Tough and Sensitive Tests for participants in the Captions Group ( $N=40$ ) and the No Captions Group ( $N=187$ )

| Vocabulary Test | Group       | Mean Relative Gain | Median | Minimum | Maximum |
|-----------------|-------------|--------------------|--------|---------|---------|
| Tough           | Captions    | 23.55%             | 23.67% | 0%      | 53.85%  |
|                 | No Captions | 23.03%             | 22.73% | 2.94%   | 46.67%  |
| Sensitive       | Captions    | 28.84%             | 28.17% | 6.67%   | 78.57%  |
|                 | No Captions | 29.61%             | 28.57% | 4.35%   | 70.00%  |

A Pearson product-moment correlation coefficient was computed to assess the relationship between the relative vocabulary gains and vocabulary knowledge for participants who viewed the 10 episodes of *Chuck* with and without captions. This was done to investigate whether participants who had more vocabulary knowledge made greater target vocabulary gains than participants with less vocabulary knowledge. To investigate whether the two groups had similar levels of vocabulary knowledge before completing the treatments, an independent-samples t-test was conducted to compare the vocabulary knowledge scores (in CHIPs) from the Captions Group ( $M = 51.25$ ,  $SD = 4.08$ ) and No Captions Group ( $M = 50.12$ ,  $SD = 3.44$ ). There was no significant difference in the vocabulary knowledge scores;  $t(225) = 1.825$ ,  $p = .069$ ). For the Captions Group, there was a significant correlation between vocabulary knowledge and relative vocabulary gain on both the Tough ( $r = .307$ ,  $N=40$ ,  $p < .05$ ) and the Sensitive ( $r = .270$ ,  $N=40$ ,  $p < .05$ ) test. For the No Captions Group, correlations were not significant for the both the Tough ( $r = .021$ ,  $N=187$ ,  $p = .773$ ) and the Sensitive ( $r = .080$ ,  $N=187$ ,  $p = .277$ ) test. This comparison indicates those participants with more vocabulary knowledge were better able to learn vocabulary from television when captions were available but this was not the case when captions were not presented.

#### 6.5.5. Summary of findings

The main findings of Study 5: Experiment 2 can be summarized as follows:

1. Incidental vocabulary learning from viewing 10 episodes of English-language television with captions was on average approximately six words regardless of the sensitivity of the test. These gains were not found to be significantly different from the vocabulary gains of the Control Group.
2. Greater vocabulary knowledge was found to have a significant correlation with greater vocabulary gains made through watching television with captions.

3. The frequency with which the target words occurred in the 10 episodes of *Chuck* and the relative gains for these target words was found to have a medium-size correlation for the Tough version of the vocabulary test. There was no significant correlation between frequency and vocabulary gains on the Sensitive Test.
4. The range of episodes in which a target word occurred had a small significant negative effect on the relative gain of the target words on the Sensitive Test.
5. There were comparable word gains on both the Tough Test (Captions Group: 6.03 words, No Captions Group: 6.36 words) and the Sensitive Test (Captions Group: 5.93 words, No Captions Group: 6.78 words) with no significant difference between vocabulary gains (raw and relative) for the two treatment groups. The Captions Group had a significant correlation between vocabulary knowledge and relative vocabulary gain on both the Tough and Sensitive Test while no significant correlation was found for the No Captions Group for either test.

## **6.6. Study 5: Experiment 3 – Captions and Lexical Coverage**

### **6.6.1. Introduction**

There has been no prior research examining the relationship between lexical coverage and comprehension and the relationship between lexical coverage and incidental vocabulary acquisition for captioned television. Estimates of the lexical coverage necessary for comprehension and for vocabulary acquisition to occur are based on research investigating the coverage necessary for reading (Carver, 1994; Hirsh & Nation, 1992; Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Laufer & Sim, 1985; Laufer, 1989; Nation, 2006; Schmitt et al., 2011) and listening (Bonk, 2000; Nation, 2006; Stæhr, 2009; van Zeeland & Schmitt, 2012) comprehension. The one study examining television and lexical coverage was the corpus-driven research by Webb and Rodgers (2009a) that found that knowledge of 3,000 word families plus proper nouns and marginal words resulted in 95% lexical coverage of television programs. They hypothesized that this coverage level might be sufficient for adequate comprehension and for incidental learning to occur. Webb and Rodgers did not theorize, however, on how the presence of captions might affect comprehension and incidental vocabulary acquisition or whether the presence of captions may reduce the coverage level at which adequate comprehension and incidental vocabulary learning might occur. Captions may allow language learners to more fully utilize the lexical coverage they have of a television program by allowing them to make use of their ability to recognize vocabulary in its written form. This could be even more significant as language learners have been shown to have reading comprehension abilities higher than their listening comprehension abilities (Hirai, 1999). Recognizing a higher percentage of the vocabulary in a television program through viewing captions could lead to greater comprehension and greater probability of vocabulary acquisition. For a detailed description of all studies related to lexical coverage and television see Sections 4.1 and 4.2 in Study 3.

### **6.6.2. Research questions**

Study 5: Experiment 3 was designed to answer the following research questions:

1. Does comprehension of English-language television viewed with captions improve with increased lexical coverage?
2. Does lexical coverage affect comprehension when English-language television is viewed with and without captions?

3. Do English-language learners with greater lexical coverage incidentally learn more vocabulary through viewing television with captions than those learners with less lexical coverage?
4. Does lexical coverage affect incidental vocabulary learning when English-language television is viewed with and without captions?

### 6.6.3. Results

#### 6.6.3.1. Lexical coverage of the episodes of *Chuck*

To determine whether aspects of viewing television with captions: comprehension and vocabulary acquisition, improved with increased lexical coverage it was first necessary to determine the participants' lexical coverage for each of the 10 episodes. Lexical coverage is the percentage of vocabulary in a text a language learner at a particular vocabulary level should know. Results of the 2,000, 3,000 and 5,000 levels of the Vocabulary Levels Test (VLT) were used to estimate the participants' lexical coverage of the episodes. This was done in the same manner described in Section 4.7.1 of Study 3. Table 6.18 presents the lexical coverage of each episode at the 2,000-, 3,000-, and 5,000-word levels. These estimates of lexical coverage include proper nouns and marginal words.

Table 6.18 The lexical coverage (measured in percent) of the 10 episodes at the 2,000-, 3,000- and 5,000-word family levels (including proper nouns and marginal words)

| Word Level | Ep. A1 | Ep. 1 | Ep. 2 | Ep. 3 | Ep. 4 | Ep. 5 | Ep. 6 | Ep. 7 | Ep. 8 | Ep. B2 |
|------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| 2,000      | 94.09  | 94.28 | 94.93 | 94.33 | 94.69 | 94.72 | 93.52 | 93.71 | 93.95 | 94.58  |
| 3,000      | 95.60  | 95.70 | 96.87 | 95.94 | 96.05 | 96.10 | 95.61 | 95.47 | 96.08 | 96.12  |
| 5,000      | 97.58  | 97.17 | 97.95 | 97.47 | 97.75 | 97.58 | 97.33 | 97.29 | 97.62 | 97.57  |

The 51 participants from Experiment 1 of Study 5 (Captions and Comprehension) were included in the analysis of lexical coverage and comprehension (see Section 6.1 for a description of these participants). Their mean score on the 2,000-word level was 20.0 out of 30 with 16 participants achieving the mastery score of 24 out of 30. On the 3,000-word level, their mean score was 14.5 out of 30 with two participants achieving the mastery score of 24. No participants achieved mastery of the 5,000-word level although their mean score was only slightly less than the 3,000-word level (14.3 out of 30). One participant achieved a score of 24 on the 5,000-word level test but did not obtain a mastery score on the 3,000-word level which is a prerequisite for mastery of a subsequent level. Table 6.19 displays the results for

participants in the analysis of lexical coverage and comprehension on the three levels of the VLT including the mean scores and the number of participants achieving mastery of each level.

Table 6.19 Scores on VLT for the analysis of lexical coverage and comprehension

| Levels Test | N  | Mean | <i>SD</i> | Median | Minimum | Maximum | Mastery <sup>†</sup> |
|-------------|----|------|-----------|--------|---------|---------|----------------------|
| VLT 2000    | 51 | 20.0 | 6.0       | 21     | 5       | 29      | 16                   |
| VLT 3000    | 51 | 14.5 | 7.2       | 14     | 3       | 26      | 2                    |
| VLT 5000    | 51 | 14.3 | 6.8       | 14     | 5       | 24      | 0                    |

Note. <sup>†</sup>Mastery of VLT set at 24 out of 30

The results of the 40 participants from Experiment 2 (Captions and Incidental Vocabulary Acquisition) were used for the analysis of lexical coverage and incidental vocabulary learning (see Section 6.1 for a description of these participants). For these participants, their mean score on the 2,000-word level was 20.2 out of 30 with 13 participants achieving the mastery score of 24 out of 30. On the 3,000-word level, their mean score was 14.3 out of 30 with two participants achieving the mastery score. No participants achieved mastery of the 5,000-word level although their mean score was the same as the 3,000-word level test at 14.3 out of 30. Table 6.20 displays the results for the participants on the three levels of the VLT including the mean scores and the number of participants achieving mastery of each level.

Table 6.20 Scores on VLT for the analysis of lexical coverage and incidental vocabulary learning

| Levels Test | N  | Mean | <i>SD</i> | Median | Minimum | Maximum | Mastery <sup>†</sup> |
|-------------|----|------|-----------|--------|---------|---------|----------------------|
| VLT 2000    | 40 | 20.2 | 6.1       | 22     | 5       | 29      | 13                   |
| VLT 3000    | 40 | 14.3 | 5.3       | 15     | 3       | 26      | 2                    |
| VLT 5000    | 40 | 14.3 | 4.1       | 15     | 7       | 20      | 0                    |

Note. <sup>†</sup>Mastery of VLT set at 24 out of 30

The participants who achieved mastery of the VLT at the 2,000-word level had lexical coverage ranging from 93.52% (Episode 6) to 94.93% (Episode 2) for the episodes of *Chuck*. Their average coverage level across the episodes was 94.28%. The participants with mastery of the VLT at the 3,000-word level had an average lexical coverage of 95.95% for the 10 episodes viewed. Episode 7, with 95.47% coverage, was the lowest for these participants, while Episode 2 was the highest at 96.87% coverage. It is not possible to estimate the coverage of the participants who did not obtain a mastery score of any level of the VLT but their coverage would be less than those with mastery of the 2,000-word level.

### 6.6.3.2. Lexical coverage and comprehension of television with captions

The participants and their scores on the comprehension tests for the 10 episodes of *Chuck* viewed with captions were grouped by their performance on the VLT. Table 6.21 shows the mean comprehension test scores for those participants who failed to master the VLT at the 2,000-word level. Table 6.22 shows the results of the 16 participants who achieved mastery of the 2,000-word level. Table 6.23 shows the comprehension scores of the two participants who achieved mastery at the 3,000-word level. The mean comprehension scores for these episodes are expressed in terms of ability measures known as CHIPs scores which are raw test scores analyzed by the Rasch Model to obtain interval data in a range from 1 to 100. For a full explanation of CHIPs see Section 2.8.3. The mean CHIPs score for participants not reaching mastery of the 2,000-word level across all 10 episodes viewed was 53.6. The mean CHIPs score of participants that reached the 2,000-word level was 55.5, while the mean for the group reaching the 3,000-word level was 55.2. The highest mean CHIPs score for any single episode was by the group of participants reaching mastery of the 3,000-word level and was 58.3 (Episode 1). The lowest mean CHIPs score for any single episode was by the group of participants not reaching mastery of the 2,000-word level and was 51.9 (Episode A1).

Table 6.21 Mean comprehension scores for participants not reaching mastery of any VLT ( $n=33$ )

| Mean Score | Ep. A1 | Ep. 1 | Ep. 2 | Ep. 3 | Ep. 4 | Ep. 5 | Ep. 6 | Ep. 7 | Ep. 8 | Ep. B2 |
|------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| CHIP       | 51.9   | 54.2  | 55.3  | 55.0  | 53.1  | 53.5  | 52.9  | 52.7  | 53.8  | 53.4   |
| Raw        | 44.5   | 47.0  | 53.9  | 52.5  | 44.4  | 46.4  | 46.1  | 46.5  | 50.3  | 44.5   |
| Raw %      | 58.5%  | 67.2% | 70.9% | 69.0% | 62.6% | 64.4% | 61.5% | 60.4% | 64.5% | 62.7%  |

Table 6.22 Mean comprehension scores for participants reaching mastery of the VLT at the 2,000-word level ( $n=16$ )

| Mean Score | Ep. A1 | Ep. 1 | Ep. 2 | Ep. 3 | Ep. 4 | Ep. 5 | Ep. 6 | Ep. 7 | Ep. 8 | Ep. B2 |
|------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| CHIP       | 53.1   | 57.4  | 57.1  | 57.5  | 55.6  | 55.7  | 54.6  | 54.8  | 56.0  | 53.4   |
| Raw        | 48.3   | 54.8  | 58.0  | 58.6  | 51.1  | 51.2  | 50.5  | 52.1  | 56.5  | 44.5   |
| Raw %      | 63.5%  | 78.3% | 76.3% | 77.1% | 72.0% | 71.1% | 67.3% | 67.6% | 72.4% | 62.7%  |

Table 6.23 Mean comprehension scores for participants reaching mastery of the VLT at the 3,000-word level ( $n=2$ )

| Mean Score | Ep. A1 | Ep. 1 | Ep. 2 | Ep. 3 | Ep. 4 | Ep. 5 | Ep. 6 | Ep. 7 | Ep. 8 | Ep. B2 |
|------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| CHIP       | 51.9   | 58.3  | 57.4  | 57.4  | 52.6  | 54.8  | 53.8  | 53.4  | 56.4  | 56.0   |
| Raw        | 44.5   | 57.0  | 59.0  | 59.5  | 42.5  | 50.0  | 49.0  | 48.5  | 58.0  | 51.0   |
| Raw %      | 58.6%  | 81.4% | 77.6% | 78.3% | 59.9% | 69.4% | 65.3% | 63.0% | 74.4% | 71.8%  |

Differences among the mean comprehension scores for each episode by participants with different lexical coverage levels were assessed using a series of one-way ANOVAs. A Levene test of homogeneity of variance conducted prior to the ANOVAs did not indicate the assumption of homogeneity of variance was significantly violated for any episode, ( $p > .05$ ). There was a significant effect of lexical coverage, as represented by mastery of the 2,000- and 3,000-word levels, on comprehension for three of the 10 episodes: Episode 1 [ $F(2,48) = 8.03$ ,  $p = .001$ ,  $\eta^2 = .25$ ], Episode 4 [ $F(2,48) = 5.30$ ,  $p < .01$ ,  $\eta^2 = .18$ ], and Episode 8 [ $F(2,48) = 3.65$ ,  $p < .05$ ,  $\eta^2 = .13$ ]. Post hoc analysis (LSD) revealed that for these three episodes the comprehension scores from the group of participants with mastery of the 2,000-word level was significantly higher than the comprehension scores from the group of participants without mastery of any level of the VLT. Thus, coverage had a significant but small to moderate effect (as represented by  $\eta^2$ ) on mean comprehension scores for three episodes of *Chuck*.

Across the three episodes where significant differences were identified, the 2,000-word level plus proper nouns and marginal words averaged 94.3% lexical coverage. Coverage from the 2,000-word level for these episodes ranges from 93.95% in Episode 8 to 94.69% in Episode 2. The average comprehension score for these three episodes for the participants with mastery of the 2,000-word level was 74.2% while the average comprehension score for participants without mastery of the 2,000-word level was 64.8%. For these three episodes, participants with mastery of 2,000-word level had a range of raw comprehension scores from 72.0% in Episode 4 to 78.3% in Episode 1. The participants without mastery of the 2,000-word level had a range of raw comprehension scores from 62.6% in Episode 4 to 70.9% in Episode 2. These analyses indicate that for English-language television viewed with captions, increased lexical coverage can lead to increased comprehension but comprehension can also be somewhat episode dependent. For episodes where differences in vocabulary level and coverage were not found to be significantly linked to comprehension other factors such as the content of the episode, the relation of the episode to previous episodes viewed, or composition of the comprehension tests may be interfering with the comprehension and coverage relationship.

#### **6.6.3.3. Comparison of lexical coverage and comprehension of television viewed with and without captions**

To investigate how lexical coverage affected comprehension of English-language television viewed with and without captions, the relationship between lexical coverage and comprehension for the Captions Group and No Captions Group was compared. Lexical

coverage (including proper nouns and marginal words) for both groups was calculated by using the participants' results on the VLT at the 2,000-, 3,000-, and 5,000-word levels. On the 2,000-word level, the mean score for the Captions Group was 20.0 out of 30 with 16 participants achieving the mastery score of 24 out of 30, while for the No Captions Group the mean score was 20.7 with 95 participants reaching the mastery level. Two participants in the Captions Group ( $M=14.5$ ) reached the mastery level of the 3,000-word level compared with 14 participants reaching mastery level in the No Captions Group ( $M=14.8$ ). No participants met the mastery level requirements of the 5,000-word level from the Captions Group ( $M=14.3$ ) while three participants from the No Captions Group ( $M=13.5$ ) did so. Table 6.24 displays the results of the three levels of the VLT for the Captions and No Captions groups including the mean scores and the number of participants achieving mastery at each level.

Table 6.24 Comparison of scores on VLT for Captions and No Captions groups

| Levels Test | Group       | <i>n</i> | Mean | <i>SD</i> | Median | Min. | Max. | Mastery <sup>†</sup> |
|-------------|-------------|----------|------|-----------|--------|------|------|----------------------|
| VLT 2000    | Captions    | 51       | 20.0 | 6.0       | 21     | 5    | 29   | 16                   |
|             | No Captions | 321      | 20.7 | 5.4       | 21     | 6    | 30   | 95                   |
| VLT 3000    | Captions    | 51       | 14.5 | 7.2       | 14     | 3    | 26   | 2                    |
|             | No Captions | 321      | 14.8 | 5.0       | 15     | 4    | 29   | 14                   |
| VLT 5000    | Captions    | 51       | 14.3 | 6.8       | 14     | 5    | 24   | 0                    |
|             | No Captions | 321      | 13.5 | 4.6       | 14     | 2    | 26   | 3                    |

Note. <sup>†</sup> Mastery of VLT set at 24 out of 30.

For both the Captions Group and the No Captions Group, the participants and their CHIPs scores on the comprehension tests for the 10 episodes of *Chuck* were grouped by their performance on the VLT. Table 6.25 shows the comprehension test scores for participants who failed to master the VLT at the 2,000-word level, participants who achieved mastery of the 2,000-word level, and participants who achieved mastery at the 3,000-word level for both groups. Table 6.25 also shows the comprehension scores of the participants who achieved mastery at the 5,000-word level in the No Captions Group. The mean CHIPs scores across all 10 episodes were 53.6 for participants not reaching mastery at the 2,000-word level from the Captions Group and 53.3 for those in the No Captions Group. The mean CHIPs scores of participants that reached the 2,000-word level were 55.5 for the Captions Group and 54.1 for the No Captions Group. The mean was 55.2 for the Captions Group participants reaching the 3,000-word level and 54.2 for the participants in the No Captions Group. Table 6.25 also shows the difference in mean CHIPs scores between the Captions and No Captions groups for each episode. The mean difference in comprehension scores across the episodes was 0.2 for those not reaching mastery at the 2,000-word level. For those participants with mastery of the

2,000- and 3,000-word level, the mean difference between the treatment groups across the episodes was 1.5 and 1.0 CHIPs respectively.

Table 6.25 Mean comprehension scores in CHIPs for participants in the Captions and No Captions groups who failed to master the VLT at the 2,000-word level, and participants that mastered the 2,000-, 3,000- and 5,000-word levels

| VLT            | Group       | Initial Ep. | Ep. 1 | Ep. 2 | Ep. 3 | Ep. 4 | Ep. 5 | Ep. 6 | Ep. 7 | Ep. 8 | Final Ep. |
|----------------|-------------|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-----------|
| No VLT Mastery | Captions    | 51.9        | 54.2  | 55.3  | 55.0  | 53.1  | 53.5  | 52.9  | 52.7  | 53.8  | 53.4      |
|                | No Captions | 50.8        | 54.6  | 55.4  | 55.0  | 52.4  | 53.3  | 52.7  | 52.3  | 54.0  | 52.9      |
|                | Difference  | 1.1         | -0.4  | -0.1  | 0     | 0.7   | 0.2   | 0.2   | 0.4   | -0.2  | 0.5       |
| 2000           | Captions    | 53.1        | 57.4  | 57.1  | 57.5  | 55.6  | 55.7  | 54.6  | 54.8  | 56.0  | 53.4      |
|                | No Captions | 51.2        | 55.8  | 55.9  | 55.8  | 52.7  | 54.3  | 53.6  | 52.8  | 54.8  | 53.6      |
|                | Difference  | 1.9         | 1.6   | 1.2   | 1.7   | 2.9   | 1.4   | 1     | 2     | 1.2   | -0.2      |
| 3000           | Captions    | 51.9        | 58.3  | 57.4  | 57.4  | 52.6  | 54.8  | 53.8  | 53.4  | 56.4  | 56.0      |
|                | No Captions | 51.1        | 55.6  | 55.8  | 55.8  | 52.8  | 54.9  | 53.9  | 53.6  | 54.5  | 53.8      |
|                | Difference  | 0.8         | 2.7   | 1.6   | 1.6   | -0.2  | -0.1  | -0.1  | -0.2  | 1.9   | 2.2       |
| 5000           | Captions    | --          | --    | --    | --    | --    | --    | --    | --    | --    | --        |
|                | No Captions | 52.9        | 56.6  | 59.3  | 56.4  | 53.9  | 57.9  | 55.3  | 53.7  | 57.0  | 56.7      |
|                | Difference  | --          | --    | --    | --    | --    | --    | --    | --    | --    | --        |

Differences among the mean comprehension scores for each episode by participants with different lexical coverage levels were assessed using one-way ANOVAs. For the Captions Group, there was a significant effect of lexical coverage, as represented by mastery of the 2,000- and 3,000-word levels VLTs, on comprehension for three of the 10 episodes: Episode 1 [ $F(2,48) = 8.03, p = .001, \eta^2 = .25$ ], Episode 4 [ $F(2,48) = 5.30, p < .01, \eta^2 = .18$ ], and Episode 8 [ $F(2,48) = 3.65, p < .05, \eta^2 = .13$ ]. For the No Captions Group, there was a significant effect of lexical coverage, as represented by mastery of 2,000-, 3,000- and 5,000-word levels, on comprehension for six of the 10 episodes: Episode 1 [ $F(3,317) = 4.52, p = .01, \eta^2 = .04$ ], Episode 3 [ $F(3,317) = 2.78, p < .05, \eta^2 = .003$ ], Episode 5 [ $F(3,317) = 6.842, p < .001, \eta^2 = .006$ ], Episode 6 [ $F(3,317) = 3.68, p < .05, \eta^2 = .003$ ], Episode 8 [ $F(3,317) = 2.66, p < .05, \eta^2 = .02$ ], and Final Episode [ $F(3,317) = 3.81, p < .05, \eta^2 = 0.03$ ]. This comparison indicates that lexical coverage, as represented by mastery of the VLT at different levels, had a significant but small effect (as indicated by  $\eta^2$ ) on mean comprehension scores for more episodes of *Chuck* when participants did not have access to captions. Statistical analyses between the treatment groups were not performed because of the small number of participants at the different lexical coverage points indicated by mastery of a level.

Comprehension of episodes of *Chuck* improved with increased lexical coverage for some but not all episodes regardless of whether the participants had access to captions or not. The Captions Group had higher comprehension scores than the No Captions Group for most episodes regardless of the amount of lexical coverage but the difference in comprehension scores between the treatment groups was not consistent within most individual episodes. That is, when mean comprehension scores were higher for the Captions Group at one level of lexical coverage, mean comprehension scores may not have been higher at the other levels of lexical coverage. For example, for Episode 5 participants in the Captions Group with mastery of the 2,000-word level (94.72% lexical coverage) had comprehension scores 1.4 CHIPs higher than those in the No Captions Group. However, the participants in the Captions Group with no mastery of any word level (less than 94.72% lexical coverage) had only a slightly higher (0.2 CHIPs) mean score than those in the No Captions Group. The participants with mastery of the 3,000-word level (96.10% lexical coverage) in the Captions Group had a slightly lower (-0.1 CHIPs) mean comprehension score than those in the No Captions Group. In the No Captions Group, there were more episodes with a significant difference in comprehension between groups of participants at different levels of coverage indicating that lexical coverage was more important when captions were not available to the participants. This may be due in part to the smaller number of participants at each level in the Captions Group.

#### **6.6.3.4. Lexical coverage and incidental vocabulary acquisition through viewing television with captions**

The participants and their vocabulary gain scores on the Tough and Sensitive Tests were grouped by their performance on the Vocabulary Levels Test. Table 6.26 shows the mean real gain scores for the 25 participants who failed to master the VLT at the 2,000-word level and Table 6.27 shows their mean relative gain scores. Table 6.28 shows the mean real gain scores of the 13 participants who achieved mastery of the 2,000-word level and Table 6.29 shows their mean relative gain scores. Table 6.30 shows the mean real gain scores of the two participants who achieved mastery at the 3,000-word level and Table 6.31 shows their mean relative gain scores.

The mean real gain of the participants not reaching mastery of the 2,000-word level across all 10 episodes viewed was 6.48 words on the Tough Test and 6.52 words on the Sensitive Test. This group had mean relative gains scores of 22.50% and 27.41% respectively on the two vocabulary tests. The mean real gain of participants who had mastery of the 2,000-word

level was 5.38 words (25.74% relative gain) on the Tough Test and 4.54 words (28.53% relative gain) on the Sensitive Test. The mean real gain of the two participants who reached mastery of the 3,000-word level was 4.50 words (22.46% relative gain) on the Tough Test and 7.50 words (48.68% relative gain) on the Sensitive Test.

Table 6.26 Vocabulary gains for participants not reaching mastery of any VLT ( $n=25$ )

| Vocabulary Test | Mean Real Gain | <i>SD</i> | Median Real Gain | Minimum Real Gain | Maximum Real Gain |
|-----------------|----------------|-----------|------------------|-------------------|-------------------|
| Tough           | 6.48           | 2.50      | 7                | 2                 | 10                |
| Sensitive       | 6.52           | 3.38      | 6                | 1                 | 14                |

Table 6.27 Relative vocabulary gains for participants not reaching mastery of any VLT ( $n=25$ )

| Vocabulary Test | Mean Relative Gain | Median Relative Gain | Minimum Relative Gain | Maximum Relative Gain |
|-----------------|--------------------|----------------------|-----------------------|-----------------------|
| Tough           | 22.50%             | 21.88%               | 5.56%                 | 43.48%                |
| Sensitive       | 27.41%             | 26.92%               | 9.09%                 | 78.57%                |

Table 6.28 Vocabulary gains for participants reaching mastery of the VLT at the 2,000-word level ( $n=13$ )

| Vocabulary Test | Mean Real Gain | <i>SD</i> | Median Real Gain | Minimum Real Gain | Maximum Real Gain |
|-----------------|----------------|-----------|------------------|-------------------|-------------------|
| Tough           | 5.38           | 3.20      | 5                | 0                 | 14                |
| Sensitive       | 4.54           | 2.67      | 4                | 1                 | 10                |

Table 6.29 Relative vocabulary gains for participants reaching mastery of the VLT at the 2,000-word level ( $n=13$ )

| Vocabulary Test | Mean Relative Gain | Median Relative Gain | Minimum Relative Gain | Maximum Relative Gain |
|-----------------|--------------------|----------------------|-----------------------|-----------------------|
| Tough           | 25.74%             | 26.09%               | 0%                    | 53.85%                |
| Sensitive       | 28.53%             | 30.00%               | 6.67%                 | 58.82%                |

Table 6.30 Vocabulary gains for participants reaching mastery of the VLT at the 3,000-word level ( $n=2$ )

| Vocabulary Test | Mean Real Gain | <i>SD</i> | Median Real Gain | Minimum Real Gain | Maximum Real Gain |
|-----------------|----------------|-----------|------------------|-------------------|-------------------|
| Tough           | 4.50           | 2.12      | 4.5              | 3                 | 6                 |
| Sensitive       | 7.50           | 2.12      | 7.5              | 6                 | 9                 |

Table 6.31 Relative vocabulary gains for participants reaching mastery of the VLT at the 3,000-word level ( $n=2$ )

| Vocabulary Test | Mean Relative Gain | Median Relative Gain | Minimum Relative Gain | Maximum Relative Gain |
|-----------------|--------------------|----------------------|-----------------------|-----------------------|
| Tough           | 22.46%             | 22.46%               | 17.65%                | 27.27%                |
| Sensitive       | 48.68%             | 48.68%               | 47.37%                | 50.00%                |

To investigate whether there was a difference between participants with different vocabulary levels and their relative vocabulary gains on the Tough and Sensitive Tests two one-way ANOVAs were run. The first two ANOVAs compared the relative vocabulary gains on the two vocabulary tests for those participants who did not have mastery of the 2,000-word level, those who had mastery of the 2,000-word level, and those who had mastery of the 3,000-word level. Neither ANOVA found a significant difference between the three groups: Tough Test [ $F(2,37) = 0.386, p = .682$ ], and Sensitive Test [ $F(2,37) = 1.800, p = .180$ ]. These results indicate that lexical coverage, as represented by mastery of the VLT at different levels, did not have a significant effect on relative gain of target words.

Webb and Rodgers (2009a) hypothesized that language learners with knowledge of the 3,000 most frequent word families as well as proper nouns and marginal words, or at least 95% lexical coverage, had the potential for considerable incidental vocabulary learning through watching episodes of television. For the 10 episodes of *Chuck* in this study, mastery of the 2,000-, 3,000- or 5,000-word levels meant an average lexical coverage across the 10 episodes of 94.28% at the 2,000-word level, 95.95% at the 3,000-word level, and 97.53% at the 5,000-word level. The results from Study 5 indicate that there is incidental vocabulary learning at lower coverage levels but whether there is increased incidental learning with increased lexical coverage is unclear. The results from the few participants with mastery of the 3,000-word level in this study do not support Webb and Rodgers' hypothesis.

#### **6.6.3.5. Comparison of lexical coverage and incidental vocabulary acquisition from television viewed with and without captions**

To investigate the relationship between lexical coverage and relative vocabulary gains from watching television with and without captions, one-way ANOVAs were calculated. The ANOVAs compared the Captions and No Captions groups at the different lexical coverage points indicated by mastery of a level. For both the Tough and Sensitive tests, ANOVAs did not find any significant difference between the groups of participants (Captions Group: Tough Test [ $F(2,37) = 0.386, p = .682$ ], Sensitive Test [ $F(2,37) = 1.800, p = .180$ ]; No Captions

Group: Tough Test [ $F(3,183) = 2.584, p = .055$ ], Sensitive Test [ $F(3,183) = 1.553, p = .202$ ]). This comparison indicates that regardless of whether television was viewed with or without captions there was no significant relationship between relative vocabulary gain and mastery of the VLT at the 2,000-, 3,000- or 5,000-word levels. Table 6.32 shows the mean real and relative gain scores for the participants from the Captions and No Captions groups organized by mastery of the different levels of the VLT. The table reveals that there is little difference between lexical coverage points as indicated by mastery of a level and vocabulary gain. No further statistical analyses between the treatment groups were carried out because the earlier within group analyses of the effects of lexical coverage on relative vocabulary gain revealed no significant difference between the different coverage points.

Table 6.32 Comparison of vocabulary gains by mastery of levels of the VLT for the Captions Group and No Captions Group

| Mastery Level  | Group       | <i>n</i> | Tough Test |               | Sensitive Test |               |
|----------------|-------------|----------|------------|---------------|----------------|---------------|
|                |             |          | Real Mean  | Relative Mean | Real Mean      | Relative Mean |
| No VLT Mastery | Captions    | 25       | 6.48       | 22.50%        | 6.52           | 27.41%        |
|                | No Captions | 145      | 6.70       | 22.96%        | 6.99           | 29.16%        |
| VLT 2,000      | Captions    | 13       | 5.38       | 25.74%        | 4.54           | 28.53%        |
|                | No Captions | 39       | 5.15       | 22.40%        | 6.21           | 31.68%        |
| VLT 3,000      | Captions    | 2        | 4.50       | 22.46%        | 7.50           | 48.68%        |
|                | No Captions | 2        | 5.00       | 28.78%        | 2.00           | 15.71%        |
| VLT 5,000      | Captions    | -        | -          | -             | -              | -             |
|                | No Captions | 1        | 7.00       | 46.67%        | 7.00           | 41.18%        |

#### 6.6.4. Summary of findings

The main findings of Study 5: Experiment 3 can be summarized as follows:

1. Comprehension of English-language television viewed with captions was shown to improve with increased lexical coverage for some but not all episodes of *Chuck*. Participants with an average lexical coverage of 94.3% (as indicated by mastery of the 2,000-word level) were found to have significantly higher comprehension than those participants with an average lexical coverage less than 94.3% (as indicated by failure to master the 2,000-word level) for three of the 10 episodes of *Chuck*. For these three episodes the participants with at least 94.3% lexical coverage had a mean comprehension score of 74.2% while the participants with less lexical coverage averaged 64.8%.

2. There were more episodes with a significant difference in comprehension scores between participants with different levels of lexical coverage for the No Captions Group (six episodes) than the Captions Group (three episodes). This indicates that lexical coverage had more effect on comprehension when captions were not available to the participants.
3. Incidental vocabulary acquisition from English-language television viewed with captions was not shown to improve with increased lexical coverage.
4. For both the Captions and No Captions groups increased lexical coverage did not correspond with increased incidental vocabulary acquisition.

## **6.7. Study 5: Experiment 4 – Captions and Attitudes**

### **6.7.1. Learners' attitudes towards learning through viewing captioned videos**

Many studies have surveyed learners' attitudes towards captions accompanying videos used for language learning (Baltova, 1999; Borrás & Lafayette, 1994; Chai & Erlam, 2008; Chung, 1999; Guillory, 1998; Huang & Eskey, 1999; Kothari, Pandey, & Chudgar, 2004; Stewart & Pertusa, 2004; Sydorenko, 2010; Taylor, 2005; Vanderplank, 1988, 1990; Winke et al., in press). These studies investigated learners' beliefs about how the presence of captions when viewing videos affected their comprehension (Baltova, 1999; Chung, 1999; Guillory, 1998; Huang & Eskey, 1999; Taylor, 2005; Vanderplank, 1988, 1990; Winke et al., in press), their vocabulary learning (Baltova, 1999; Huang & Eskey, 1999; Sydorenko, 2010), their affective attitude (Borrás & Lafayette, 1994; Chai & Erlam, 2008; Guillory, 1998; Huang & Eskey, 1999; Stewart & Pertusa, 2004; Taylor, 2005), and their general language learning experience (Baltova, 1999; Stewart & Pertusa, 2004; Sydorenko, 2010; Winke et al., 2010). Difficulties of language learning through viewing captioned videos have also been identified (Baltova, 1999; Chai & Erlam, 2008; Chung, 1999; Guillory, 1998; Sydorenko, 2010; Taylor, 2005; Vanderplank, 1988, 1990).

Language learners have indicated that they believe the presence of captions when viewing videos led to increased comprehension. Whether surveyed through open-ended items (Chung, 1999; Winke et al., 2010) or response scales (Baltova, 1999; Guillory, 1998; Huang & Eskey, 1999), learners across a range of studies have suggested that they had better comprehension when viewing television with captions than without captions. In studies by Vanderplank (1988, 1990), when language learners viewed captioned videos they claimed to develop strategies to utilize captions which they thought led to better comprehension. When learners in Winke, Gass, and Sydorenko's (in press) study faced unfamiliar content they relied more heavily on the captions to construct meaning from the videos. In studies by Guillory (1998) and Taylor (2005) learners indicated that they thought they would have understood very little of the videos without access to captions.

In addition to a positive effect on comprehension, language learners also indicated that captions had a positive effect on their vocabulary learning (Huang & Eskey, 1999; Sydorenko, 2010) and that the reason that they learned more vocabulary was because they could both hear and see the words (Stewart & Pertusa, 2004). Winke, Gass and Sydorenko (2010) found that learners believed that vocabulary learning was improved because the captions allowed for better analysis of the language in the videos. As might be expected, learners in Baltova's (1999) study indicated that captions were particularly helpful for learning the written form of

vocabulary. Overall, language learners have been shown to believe that captions have a positive effect on their vocabulary learning.

The presence of captions has also been shown to promote a positive attitude towards the use of videos for learning English (Guillory, 1998; Taylor, 2005). Studies have found that learners are more motivated to study English when using captioned video, that participants preferred learning English from captioned video in comparison with other more traditional forms of language study, and that learners generally enjoyed learning through viewing captioned video (Chai & Erlam, 2008; Huang & Eskey, 1999). Even when watching captioned videos was a relatively new experience, Stewart and Pertusa (2004) reported that learners perceived little or no negativity towards utilizing the captions. One reason for the positive attitude towards captions may be that their presence relieves some of the anxiety that may surface if learners are unable to aurally follow a video (Borrás & Lafayette, 1994).

Research involving captioned videos has also indicated that language learners believe captions contribute to language learning in general terms. Winke, Gass, and Sydorenko (2010) found that lower proficiency language learners believed they needed multiple input modes to visualize and listen to the dialogue. Sydorenko (2010) reported that learners believed that the presence of captions could reduce the learning burden of videos that might otherwise be considered too difficult. Baltova (1999) reported that when learners experienced videos both with and without captions, they had more confidence in their learning from the treatment that included captions. In Stewart and Pertusa's (2004) study, learners saw the benefits of incorporating captioned videos into their individual study programs in the future.

While in general, language learners' attitudes towards the presence of captions have been positive, learners have also indicated some difficulties with viewing videos with captions. One reported problem was that the captions, as a third channel of input put a strain on attention (Sydorenko, 2010) and learners believed that at times they only read the captions and scanned the images thereby tuning out the audio (Chai & Erlam, 2008; Sydorenko, 2010). Learners, however, who saw the captions as a distraction from the audio track were generally in the minority (Chung, 1999; Guillory, 1998). This was supported by the participants in Taylor's (2005) study who claimed to try and listen to the audio and not rely solely on the captions. In studies by Vanderplank (1988, 1990), where captioned videos were viewed extensively, participants initially reported that the captions were distracting and difficult to use but caption-use became less problematic as they viewed more television. Vanderplank found that this was particularly pronounced for learners that had limited experience viewing television with L1 subtitles as they were slower to adapt to and build strategies for using the

L2 captions. Vanderplank also reported that learners believed they paid more attention to the captions because they thought reading was easier than listening and more helpful for learning. Another commonly cited problem with captions relates to the speed of the dialogue in the videos as learners in some studies have expressed that they would have liked more time to read the captions (Baltova, 1999; Guillory, 1998; Sydorenko, 2010). Winke, Gass and Sydorenko (in press) reported that learners' use of captions depends on their vocabulary knowledge, the speed of the audio in relation to the listener's proficiency level, and the amount of action on the screen. Sydorenko (2010) found that learners believed that the presence of captions may result in some negative effects on vocabulary learning. Learners believed that captions caused them to focus only on known words and made it difficult to learn the aural form of certain words.

The previous research on captioned videos has indicated a number of reactions to the presence of captions. In general, language learners believed that their comprehension of video and vocabulary learning from viewing video increased because of the presence of captions. Learners also indicated that they enjoyed viewing videos more when captions were present and that learning from the videos was improved by the presence of captions. Language learners also reported certain difficulties associated with viewing videos with captions including concentrating too intently on the captions and ignoring the audio and visual input from the video. The research presented here indicating language learners' attitudes towards viewing videos with captions, like the research presented in the comprehension and vocabulary acquisition experiments in Study 5, has involved videos with short running times (Baltova, 1999; Chai & Erlam, 2008; Chung, 1999; Sydorenko, 2010; Taylor, 2005; Winke et al., in press), videos viewed in isolation (Chai & Erlam, 2008; Huang & Eskey, 1999; Taylor, 2005), types of videos more conducive to language learning (educational, documentaries or news clips) (Baltova, 1999; Chung, 1999; Guillory, 1998; Huang & Eskey, 1999; Taylor, 2005; Winke et al., in press), and segments of longer narrative videos (Chai & Erlam, 2008). Further research needs to investigate whether findings from these studies are consistent with language learners' attitudes towards viewing full-length episodes of television programs with captions as this may indicate the potential efficacy of captioned television for use in language learning.

### **6.7.2. Research questions**

Study 5: Experiment 4 was designed to answer the following research questions:

1. Do EFL learners' beliefs about learning through English-language television viewed with captions change through viewing multiple episodes of a program?
2. Do EFL learners' beliefs about their comprehension of English-language television viewed with captions change through viewing multiple episodes of a program?
3. Do EFL learners' beliefs about the usefulness of captions for comprehension of English-language television change through viewing multiple episodes of a program?
4. What are EFL learners' beliefs about language learning through viewing English-language television viewed with captions?
5. What are EFL learners' beliefs about their comprehension of English-language television viewed with captions?
6. What are EFL learners' beliefs about vocabulary learning through viewing English-language television viewed with captions?
7. To what extent do EFL learners believe that captions present while viewing English-language television affect their comprehension, vocabulary learning, and listening ability?
8. To what extent do EFL learners believe that they rely on captions when viewing English-language television?
9. What do EFL learners believe are sources of difficulty when they view English-language television viewed with captions?
10. How do EFL learners' attitudes towards learning from television viewed with captions compare with attitudes towards learning from television viewed without captions?

### **6.7.3. Results**

#### **6.7.3.1. Episode Attitude Survey**

The participants in Study 5 completed the Episode Attitude Survey (EAS) after each comprehension test. Items surveyed the participants' attitudes towards learning English from the episodes, their comprehension of the episodes, and the presence of captions. In Study 4 there were four items on this survey. In Study 5 there was an additional item concerning captioning. When a participant missed viewing an episode, their responses were not included

in the calculation of the mean responses. Thirteen individuals missed a single episode across Episodes 1 to 8. The four Episode Attitude Survey Questions (EASQ) from Study 4 are shown in Section 5.6.1 and EASQ 5, unique to Study 5, and its accompanying rating scale is shown in Figure 6.7.

Figure 6.7 EASQ 5

|               | Not at All<br>Useful  |   | Slightly<br>Useful |   | Somewhat<br>Useful |   | Pretty<br>Useful |   | Very<br>Useful |
|---------------|---|---|--------------------|---|--------------------|---|------------------|---|----------------|
|               | 1   | 2 | 3                  | 4 | 5                  | 6 | 7                | 8 | 9              |
| <b>EASQ 5</b> | To what extent do you think the English captions were useful for understanding this episode of <i>Chuck</i> ? |   |                    |   |                    |   |                  |   |                |

### ***EAS items related to learning English through viewing the episodes***

Episode Attitude Survey Questions 1, 3 and 4 were used to measure the participants' beliefs about learning English from captioned television for each of the 10 episodes. Each question asked the participants to respond on a different response scale providing insight into different aspects of learning English through viewing captioned television. Through answering these questions, the participants expressed how much they enjoyed learning English through viewing television (EASQ 1), what they believed their level of learning was from the episodes (EASQ 3), and how useful they believed viewing the episodes was for studying English (EASQ 4). The participants responded to EASQ 1 on the *Not at All Enjoyable* to *Very Enjoyable* rating scale, EASQ 3 on the *Not at All Good* to *Very Good* scale, and EASQ 4 on the *Not at all Useful* to *Very Useful* scale.

The mean responses to EASQ 1 (*How did you feel about studying English through watching this program?*) trended higher over the course of the 10 episodes. Between Episode 1 and Episode 8 the mean responses ranged between 6.4 (Episode 1) and 7.2 (Episode 8). The final mean response of 7.2 for the Final Episode was marginally higher than *Pretty Enjoyable* (response category 7). The difference between the participants' responses to EASQ 1 following the Initial Episode and following the Final Episode was examined with a Wilcoxon Test. The results of the Wilcoxon Test indicated that the participants' responses towards the enjoyability of studying English through viewing an episode of *Chuck* were significantly higher following the Final Episode than following the Initial Episode,  $Z = 3.77, p < .001$ . The effect size, as measured by  $r$ , was .37, a value corresponding to a medium treatment effect. The mean response for this item generally increased across the 10 episodes the participants

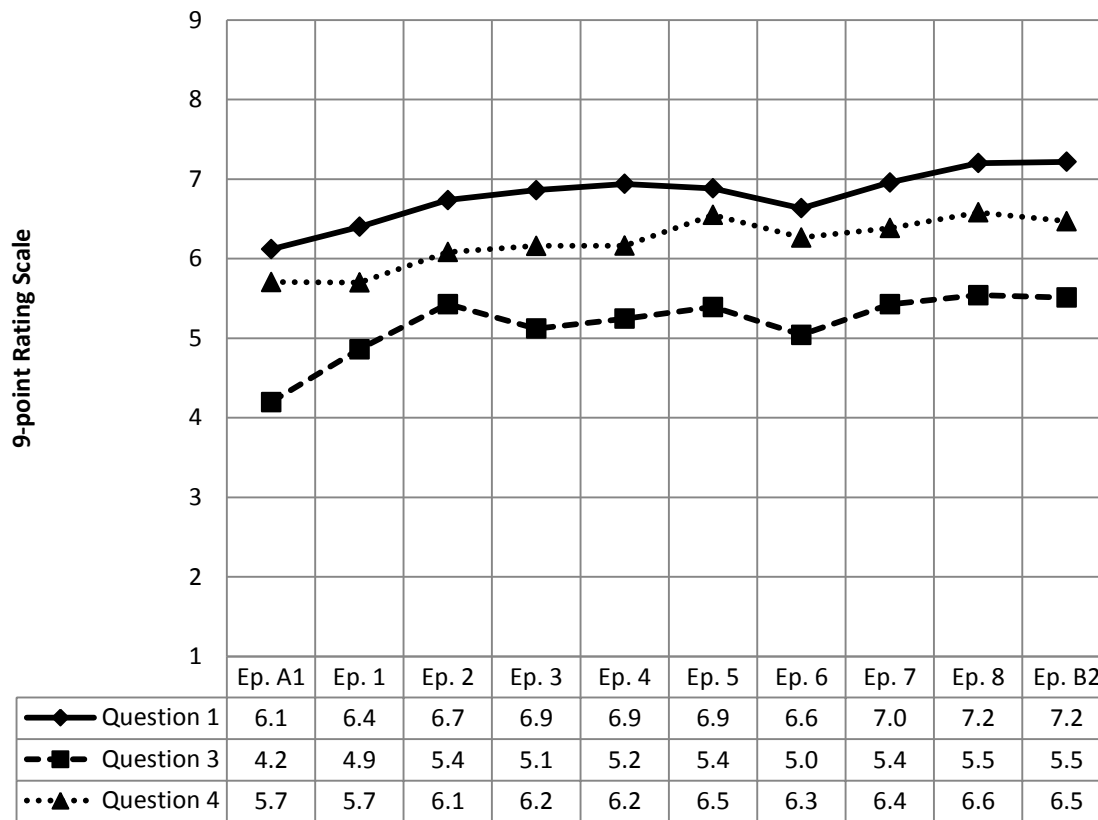
viewed. This suggests that as the participants became more familiar with the series and the content therein, their enjoyment of viewing the series increased.

The responses for EASQ 3 (*What do you think your level of English learning was from watching this episode?*) began with a mean score of 4.2 for the Initial Episode and ended with a mean response of 5.5 for the Final Episode. For this item, there was a relatively large increase from the mean response for the Initial Episode to the mean response (4.9) for Episode 1. Following Episode 1, the mean responses ranged from 5.0 (Episode 6) to 5.5 (Episode 8). A Wilcoxon Test indicated that the participants' rating of their level of English learning from watching an episode was significantly higher following the Final Episode than following the Initial Episode,  $Z = 4.23$ ,  $p < .001$ . The effect size, as measured by  $r$ , was .42, a value corresponding to a medium treatment effect. This pattern of responses for EASQ 3 indicates that the participants believed that their language learning became better as they viewed more episodes reaching a level incrementally higher than *Somewhat Good* (response category 5) in the ninth and tenth episode viewed with captions.

For EASQ 4 (*To what extent do you think watching this episode was useful for studying English?*) the increase in mean response scores between the Initial Episode and the Final Episode was 0.8. The mean responses for the Initial and Final episodes were 5.7 and 6.5, respectively. From Episode 1 to Episode 8, the mean responses ranged from 5.7 (Episode 1) to 6.6 (Episode 8). A Wilcoxon Test indicated that the participants' rating of the usefulness of an episode of *Chuck* for studying English was significantly higher following the Final Episode than following the Initial Episode,  $Z = 3.21$ ,  $p < .001$ . The effect size, as measured by  $r$ , was .31, a value corresponding to a medium treatment effect. These responses indicate that the participants generally believed that the captioned episodes of television were above *Somewhat Useful* (response category 5) for studying English and their estimation of this increased from the start of the study to the end.

The mean responses to EASQs 1, 3 and 4 across all 10 captioned episodes are summarized in Figure 6.8. Each question on the survey is represented by a different line with data points for each episode. Taken together, these results indicate the participants believed that the enjoyability, benefits and usefulness of learning English from captioned television increased as they viewed multiple episodes. Response patterns to these items are supported by the results from the Wilcoxon Tests indicating that mean response scores significantly increased from the Initial to the Final Episode.

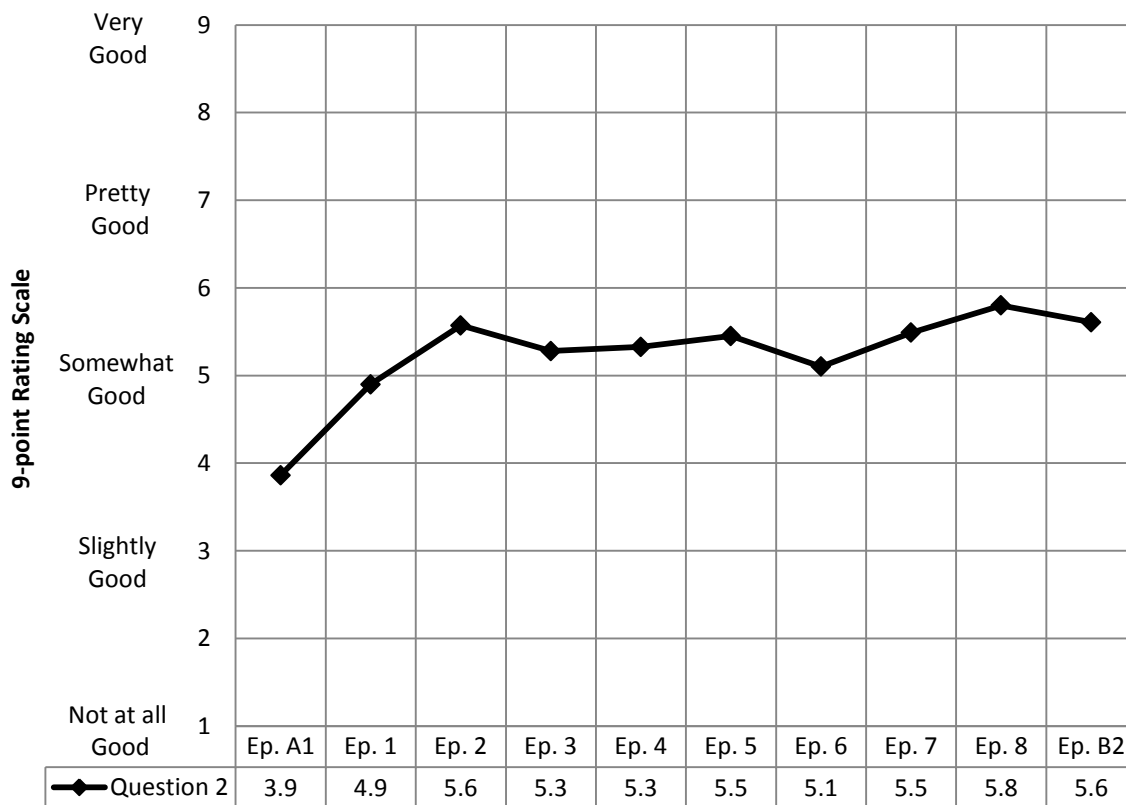
Figure 6.8 Results across all 10 episodes for EASQs 1, 3 and 4



### *EAS item related to comprehension of the episodes*

EASQ 2 (*What do you think your comprehension level of this episode was?*) measured the participants' beliefs about their comprehension of each of the 10 episodes viewed with captions. The mean responses for EASQ 2 began at 3.9 following the Initial Episode. The mean response rose to 5.6 by the Final Episode. The mean responses from Episode 1 to 8 were all higher than the Initial Episode and the scores ranged between 4.9 (Episode 1) and 5.8 (Episode 8). Results from a Wilcoxon Test indicated that the participants' evaluation of their comprehension of an episode of *Chuck* was significantly higher following the Final Episode than following the Initial Episode,  $Z = 4.61, p < .001$ . The effect size, as measured by  $r$ , was .46, a value corresponding to a medium treatment effect. The mean responses to EASQ 2 across the 10 episodes are shown in Figure 6.9. It includes data points for each episode and each episode's mean score is located in the data table. The responses to EASQ 2 indicate that the participants' beliefs about their comprehension of the episodes quickly increased perhaps because they gained familiarity with the characters and the story. From Episode 3 the mean responses became more episode-dependent at a level slightly over *Somewhat Good* (response category 5).

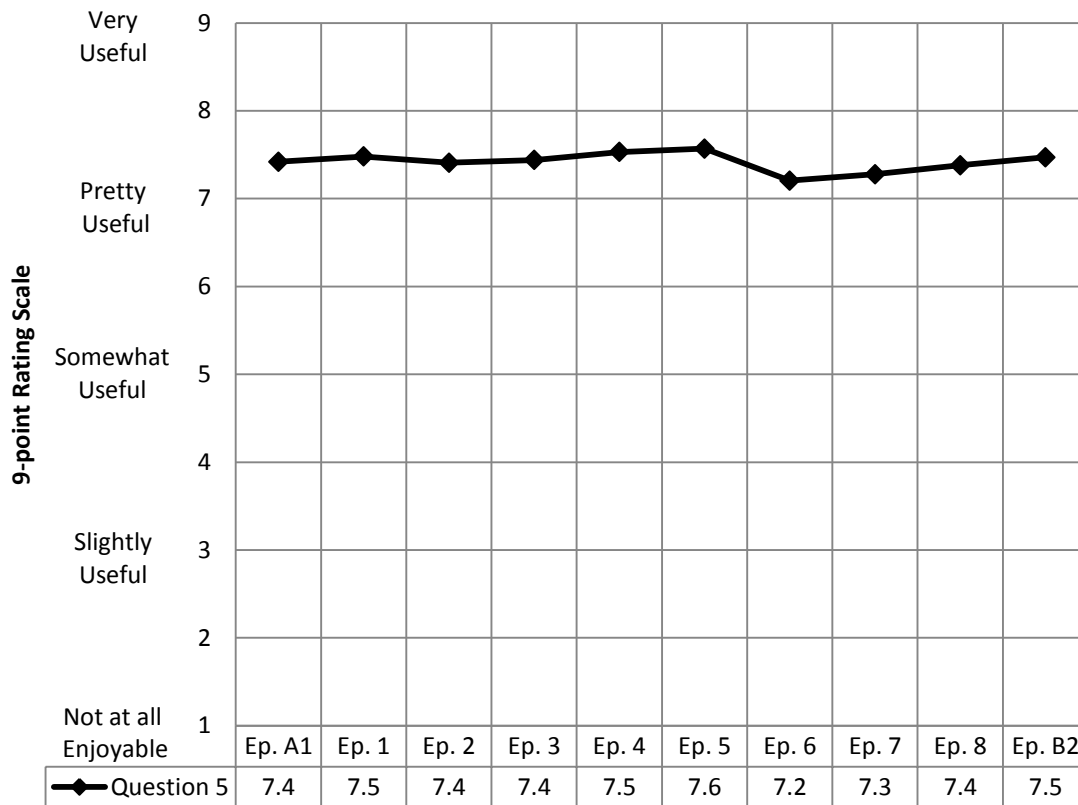
Figure 6.9 Results across all 10 episodes for EASQ 2



#### *EAS item related to the captions present while viewing the episodes*

The caption-specific item EASQ 5 (*To what extent do you think the English captions were useful for understanding this episode of Chuck?*) evaluated the participants' beliefs about the usefulness of the English-language captions for comprehension of the 10 episodes. This item had the least amount of increase in mean response scores across the episodes but was consistently the item that received the highest mean response. The Initial Episode had a mean response of 7.4 while the Final Episode's mean score was only 0.1 higher at 7.5. The highest mean score was in Episode 5 (7.6) and the lowest was in Episode 6 (7.2). A Wilcoxon Test indicated that the participants' rating of the usefulness of the captions for comprehension of an episode was not significantly higher following the Final Episode than following the Initial Episode,  $Z = .087, p < .465$ . This pattern of responses demonstrated that the majority of participants found the captions between *Pretty Useful* (response category 7) and *Very Useful* (response category 9) for comprehension throughout the 10 episodes. The mean responses to EASQ 5 across the episodes are summarized in Figure 6.10. The limited variation in mean score for each episode suggests that the participants continued to use and rely on the captions at the same level regardless of increased comprehension or familiarity with the series.

Figure 6.10 Results across all 10 episodes for EASQ 5

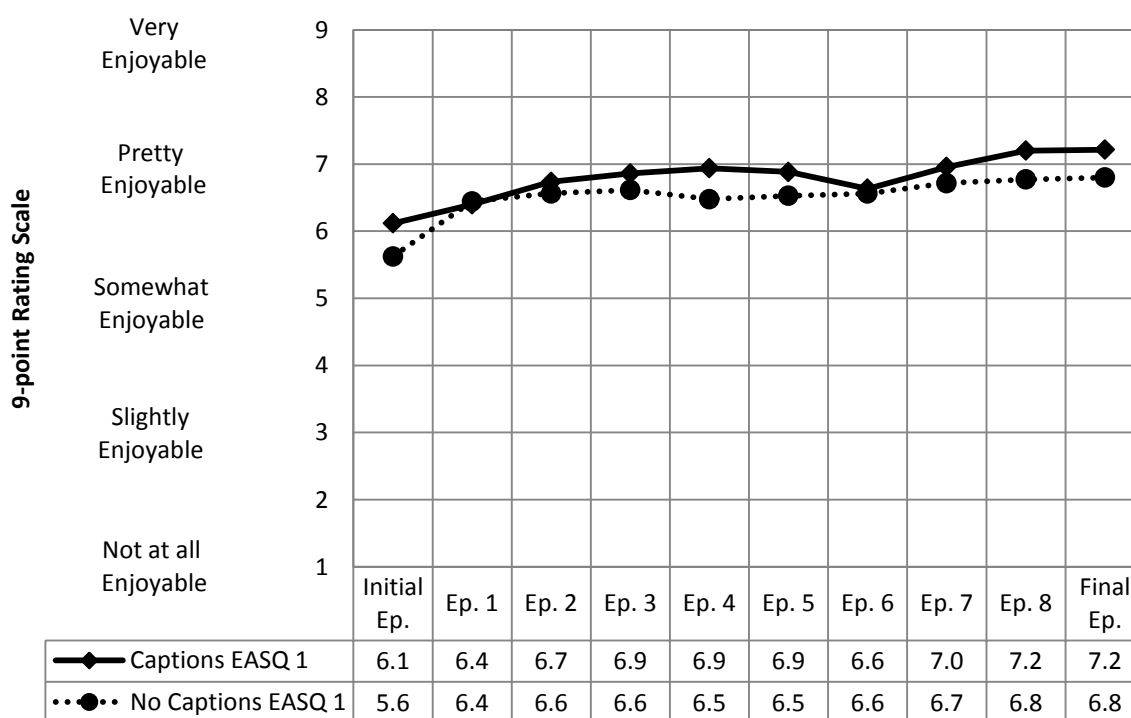


### *Comparison of the results of the EAS for episodes viewed with and without captions*

Participants' responses on the EAS from Study 5 were compared to the responses on the EAS from participants in Study 4 to examine how the attitudes of those that viewed the 10 episodes of television with captions compared with those that did not have access to captions. Participants in Study 4 completed four of the five items: EASQ 1, EASQ 2, EASQ 3 and EASQ 4. The participants in Study 4 (No Captions Group) and participants in Study 5 (Captions Group) viewed the same episodes but some of the participants viewed them in a different order. The participants in the Captions Group viewed Episode A1 first and Episode B2 last. Half the participants in the No Captions Group did so as well but half the participants viewed Episode B1 as the Initial Episode and Episode A2 as the Final Episode. The responses to the EASQs following Episode A1 and Episode B1 were compared with a series of Wilcoxon Rank-Sum Tests. No significant differences were found between responses for the viewing groups: EASQ 1: ( $Z = .585, p < .559$ ), EASQ 2: ( $Z = 1.466, p < .143$ ), EASQ 3: ( $Z = .860, p < .390$ ), and EASQ 4: ( $Z = .599, p < .549$ ). Because the responses to the EASQs were not significantly different regardless of whether Episode A1 or Episode B1 was viewed first, it is assumed that the different viewing order did not have an effect on responses to the EAS.

Figure 6.11 plots the mean responses for EASQ 1 (*How did you feel about studying English through watching this program?*) for all 10 episodes for the Captions Group and the No Captions Group. The responses to EASQ 1 for the two groups were at similar levels across the 10 episodes of *Chuck*. The mean responses for the Captions Group, however, were higher than or equal to those of the No Captions Group for all episodes. The mean responses for both groups started between *Somewhat Enjoyable* (response category 5) and *Pretty Enjoyable* (response category 7). The Captions Group's mean responses rose to the *Pretty Enjoyable* (response category 7) level or beyond for the last three episodes while the responses for the No Captions Group were slightly below that. The episode where the difference between mean responses for the two groups was the largest was the Initial Episode (0.5). The mean difference between the mean scores for each treatment across all episodes was 0.28 with a range of differences from 0 to 0.5.

Figure 6.11 Results across all 10 episodes for EASQ 1 for the Captions and No Captions groups



The mean responses for EASQ 2 (*What do you think your comprehension level of this episode was?*) for the Captions Group and the No Captions Group are shown in Figure 6.12. The mean response for the Initial Episode was 0.9 higher for the Captions Group but the difference between the two treatments in the Final Episode was only 0.2. The difference between the treatment groups across the intervening episodes ranged from 0 to 0.4 with the

Captions Group's mean response equal to or higher than the No Captions Group's in all comparisons. The average difference between the groups across all 10 episodes was 0.3. The No Captions Group's mean response to EASQ 2 for the Initial Episode corresponded to *Slightly Good* (response category 3) and was above *Somewhat Good* (response category 5) by the Final Episode. The Captions Group had a similar increase to over *Somewhat Good*.

Figure 6.12 Results across all 10 episodes for EASQ 2 for the Captions and No Captions groups

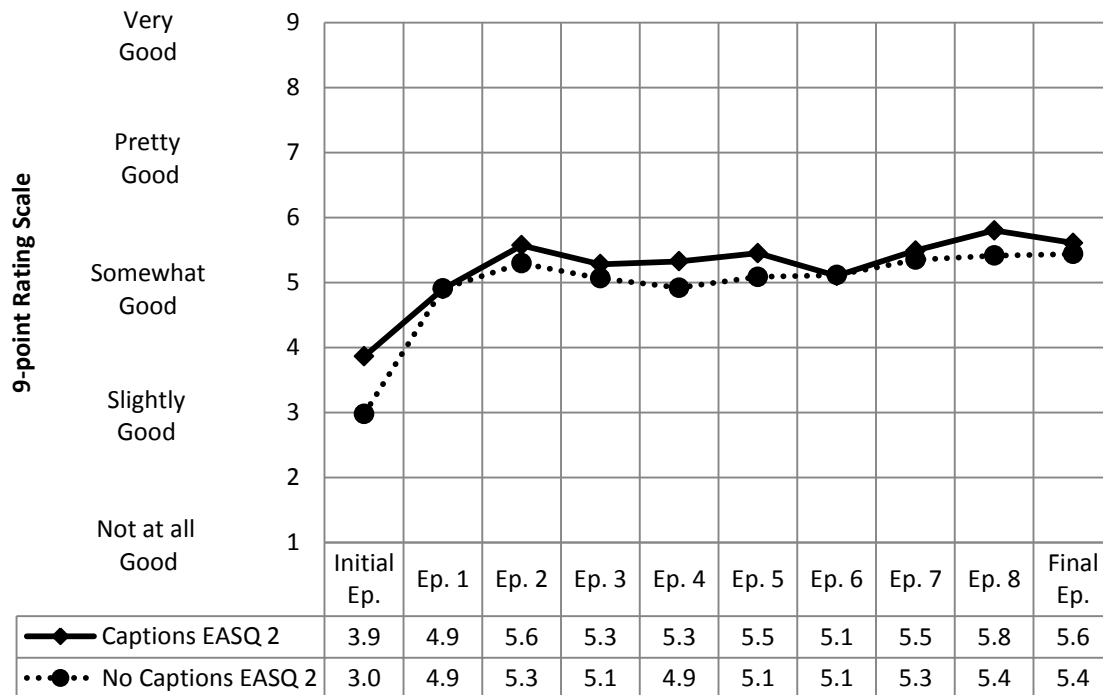
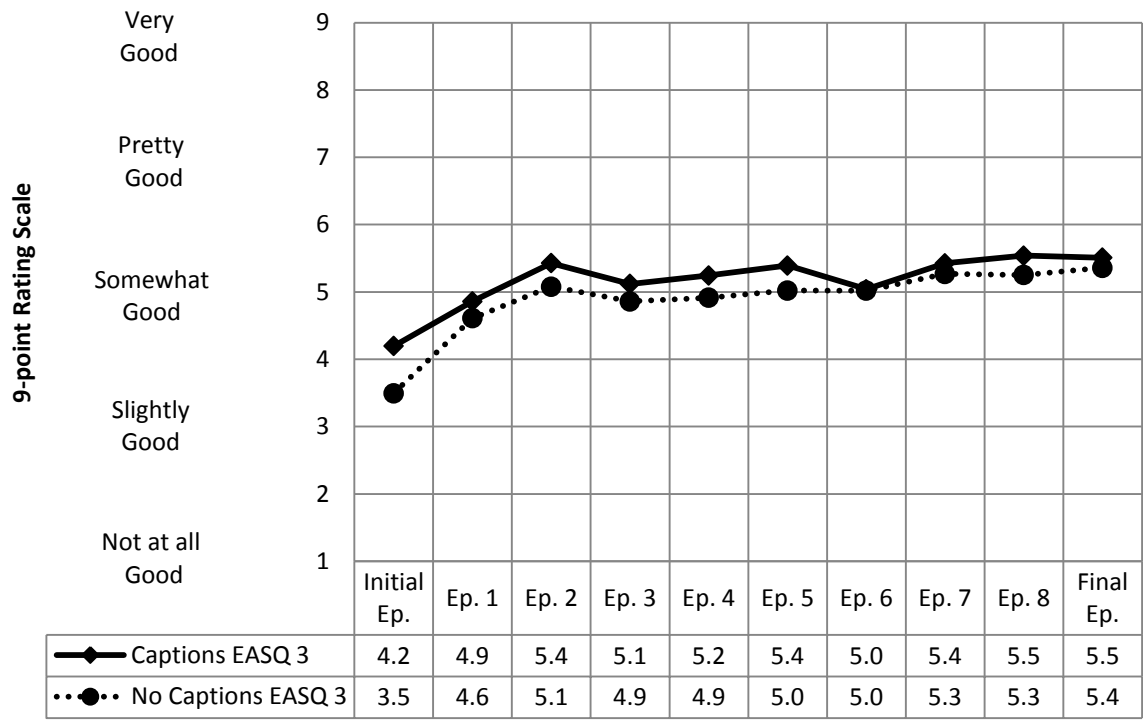


Figure 6.13 plots the mean responses for EASQ 3 (*What do you think your level of English learning was from watching this episode?*) for all 10 episodes viewed by the Captions and No Captions groups. The Captions Group's mean response was 0.7 higher following the Initial Episode, but by the Final Episode the Captions Group's mean response was only 0.1 higher. The mean difference between the groups across the intervening episodes ranged from 0 to 0.4. The mean responses for the Captions Group were higher than or equal to those of the No Captions Group for all episodes. The average difference between the groups across the episodes was 0.26. The mean response for the Captions Group after the Initial Episode was 4.2 which is at a level between *Slightly Good* (response category 3) and *Somewhat Good* (response category 5). The mean score increased (1.3) to above *Somewhat Good* (response category 5) following the Final Episode. The No Captions Group had a more pronounced increase (1.9) with the mean response increasing from just above *Slightly Good* (response

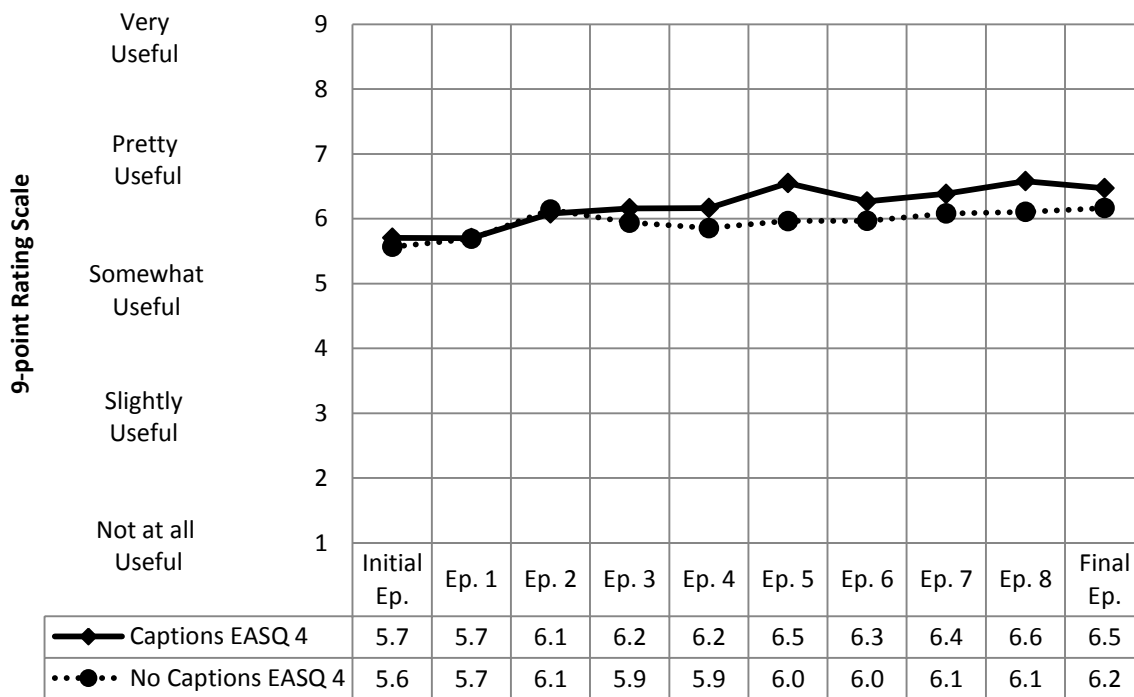
category 5) for the Initial Episode to over *Somewhat Good* (response category 5) for the Final Episode.

Figure 6.13 Results across all 10 episodes for EASQ 3 for the Captions and No Captions groups



A comparison of the mean responses to EASQ 4 (*To what extent do you think watching this episode was useful for studying English?*) for the Captions Group and the No Captions Group across the 10 episodes viewed is presented in Figure 6.14. For both groups, the range of responses was between *Somewhat Useful* (response category 5) and *Pretty Useful* (response category 7) for all episodes viewed. The Captions Group had higher or equal mean responses across all episodes with a mean difference of 0.26. There was a range of differences across the episodes from 0 to 0.5.

Figure 6.14 Results across all 10 episodes for EASQ 4 for the Captions and No Captions groups



To compare how response levels on EASQs 1 through 4 for the Captions and No Captions groups changed between the first episode to the tenth episode viewed, the differences between response scores following the Initial Episode and following the Final Episode for the two treatment groups were compared with a series of Wilcoxon Rank-Sum Tests. For EASQ 1, no significant difference ( $Z = .13, p < .447$ ) between the two treatment groups was found. The difference in response scores between the Initial Episode and the Final Episode for EASQ 2 was significantly higher ( $Z = 1.87, p < .05$ ) for the No Captions Group than for the Captions Group. The effect size, as measured by  $r$ , was .11, which corresponds to a small treatment effect. The difference in response scores between the Initial Episode and the Final Episode for EASQ 3 was significantly higher ( $Z = 1.79, p < .05$ ) for the No Captions Group than for the Captions Group. The effect size, as measured by  $r$ , was .11, which corresponds to a small treatment effect. For EASQ 4, no significant difference ( $Z = 1.11, p < .133$ ) was found between the two treatment groups.

The results of these analyses indicate that there was no difference in the increase in positive attitude towards the enjoyability of viewing the episodes (EASQ 1) and the usefulness of the episodes for studying English (EASQ 2) from the first to the tenth episode regardless of whether captions were present. However, the significantly greater increase in responses from the Initial to the Final Episode viewed for EASQ 2 and EASQ3 indicates that

the participants without captions had more growth in their perceptions of comprehension (EASQ 2) and language learning from episodes (EASQ 4). Much of this growth is a result of the considerably lower response scores following the Initial Episode on these two items for the No Captions Group.

#### **6.7.3.2. Final Attitude Survey**

After viewing all 10 episodes of *Chuck* over the course of Study 5, the participants completed the Final Attitude Survey (FAS). The FAS asked the participants to reflect on the viewing of the episodes and to indicate their beliefs regarding a variety of topics. Twelve of the items on this survey were the same as those that made up the FAS in Study 4. This survey also included items specific to Study 5 that concerned the captions that accompanied the episodes. In the following discussion of the items and the response patterns, the items are grouped together by common interrogative field. The number of participants who failed to respond to items was recorded but no attempt was made to replace their missing data because there was insufficient data from the respondents to use Expectation Maximization (Allison, 2001).

#### ***FAS items related to comprehension***

Four items on the FAS examined the participants' beliefs about their comprehension of the captioned episodes. Final Attitude Survey Question 1 (FASQ 1) asked the participants to consider their level of understanding of the first episode they viewed, while FASQ 2 asked them to consider their level of understanding of the tenth episode. FASQ 3 asked the participants to think about their overall understanding of the episodes they viewed throughout the study. FASQs 1, 2 and 3 asked the participants to respond on the *Not at All Good* to *Very Good* response scale. FASQ 7 asked the participants about the degree to which they thought their comprehension of *Chuck* improved from the first to the last episode and had them respond on the *Not at All Improved* to *Very Improved* response scale. Tables 6.33 and 6.34 show the results for FASQs 1, 2, 3, and 7. Each table shows the numerical and verbal labels for each response scale, and the number of participants that responded in each category for each question. The X category is number of non-respondents for each item. The percentage of responses in each response category and the mean response score are included.

The mean response score for FASQ 1 (*What do you think your understanding of the first episode of Chuck you watched this semester was?*) was 3.6 and the responses were bimodal with 19.6% of participants choosing response categories 2 and 5. The mean response for

FASQ 2 (*What do you think your understanding of the last episode of Chuck you watched this semester was?*) was 6.1 and the mode was response category 7 (27.5% of respondents). In FASQ 1, 50.9% of the participants responded in the lowest three response categories compared to 7.9% in the highest three categories. In contrast, for FASQ 2, 49.1% of the participants responded in the highest three response categories compared to 7.9% in the lowest three. For FASQ 7 (*To what degree do you think your understanding of the television program Chuck improved from the first episode watched to the last?*), the mean response score was 6.3 with a mode of 7 (35.3%). A majority of participants (56.8%) indicated that they believed their improvement to be from *Pretty Improved* (response category 7) to *Very Improved* (response category 9). Taken together, the results from FASQs 1, 2 and 7 indicate a considerable shift in the participants' perceived comprehension from the first to the final episode. The mean response for FASQ 3 (*What do you think your overall understanding of the episodes of Chuck you watched this semester was?*) was 5.6 and the mode was response category 7 (31.4% of respondents). A large majority (78.4%) of the respondents chose category 5 (*Somewhat Good*) or higher indicating that the participants in Study 5 thought that they had a reasonable level of understanding of the television viewed. The mean responses to these four items concerned with comprehension indicate that the participants believed that their comprehension increased from the first to the last episode and that for over 70% of the participants their overall comprehension of the episodes corresponded to *Somewhat Good* or better.

Table 6.33 Responses to FASQs 1, 2 and 3

|               |   | Not at All<br>Good |             | Slightly<br>Good |       | Somewhat<br>Good |            | Pretty<br>Good |       | Very<br>Good |
|---------------|---|--------------------|-------------|------------------|-------|------------------|------------|----------------|-------|--------------|
|               | X   | 1                  | 2           | 3                | 4     | 5                | 6          | 7              | 8     | 9            |
| <b>FASQ 1</b> | What do you think your understanding of the first episode of <i>Chuck</i> you watched this semester was?    |                    |             |                  |       |                  |            |                |       |              |
| Responses     | 0   | 7                  | 10          | 9                | 7     | 10               | 4          | 3              | 1     | 0            |
| %             | 0.0%  | 13.7%              | 19.6%       | 17.6%            | 13.7% | 19.6%            | 7.8%       | 5.9%           | 2.0%  | 0.0%         |
|               |   |                    | <b>Mean</b> | <b>3.6</b>       |       |                  |            |                |       |              |
| <b>FASQ 2</b> | What do you think your understanding of the last episode of <i>Chuck</i> you watched this semester was?     |                    |             |                  |       |                  |            |                |       |              |
| Responses     | 0   | 0                  | 1           | 3                | 4     | 11               | 7          | 14             | 10    | 1            |
| %             | 0.0%  | 0.0%               | 2.0%        | 5.9%             | 7.8%  | 21.6%            | 13.7%      | 27.5%          | 19.6% | 2.0%         |
|               |   |                    |             |                  |       | <b>Mean</b>      | <b>6.1</b> |                |       |              |
| <b>FASQ 3</b> | What do you think your overall understanding of the episodes of <i>Chuck</i> you watched this semester was? |                    |             |                  |       |                  |            |                |       |              |
| Responses     | 0   | 1                  | 0           | 7                | 3     | 12               | 9          | 16             | 3     | 0            |
| %             | 0.0%  | 2.0%               | 0.0%        | 13.7%            | 5.9%  | 23.5%            | 17.6%      | 31.4%          | 5.9%  | 0.0%         |
|               |   |                    |             |                  |       | <b>Mean</b>      | <b>5.6</b> |                |       |              |

Table 6.34 Responses to FASQ 7

|               |  | Not at All<br>Improved |      | Slightly<br>Improved |      | Somewhat<br>Improved |            | Pretty<br>Improved |       | Very<br>Improved |
|---------------|--|------------------------|------|----------------------|------|----------------------|------------|--------------------|-------|------------------|
|               | X  | 1                      | 2    | 3                    | 4    | 5                    | 6          | 7                  | 8     | 9                |
| <b>FASQ 7</b> | To what degree do you think your understanding of the television program <i>Chuck</i> improved from the first episode watched to the last? |                        |      |                      |      |                      |            |                    |       |                  |
| Responses     | 0  | 0                      | 1    | 2                    | 5    | 7                    | 7          | 18                 | 7     | 4                |
| %             | 0.0%   | 0.0%                   | 2.0% | 3.9%                 | 9.8% | 13.7%                | 13.7%      | 35.3%              | 13.7% | 7.8%             |
|               |  |                        |      |                      |      | <b>Mean</b>          | <b>6.3</b> |                    |       |                  |

### ***FAS items related to vocabulary learning***

A single item on the FAS asked the participants about their beliefs about the effects viewing television with captions had on their English vocabulary learning. FASQ 6 had the participants respond on the *Not at All Improved* to *Very Improved* scale. Table 6.35 presents the number of responses in each category and the mean. The mean response to FASQ 6 (*To what extent do you think that watching the television program Chuck all semester affected your English vocabulary?*) was 5.2 and the mode was response category 5 (27.5%). There were responses in all nine response categories indicating a wide range of individual beliefs about the participants' vocabulary improvement. The vast majority of participants (96.1%) responded that they believed their vocabulary was *Improved* (response categories 2 - 9)

indicating that they believed that over the course of Study 5 they experienced improvement in their English vocabulary.

Table 6.35 Responses to FASQ 6

|               |  | Not at All<br>Improved |      | Slightly<br>Improved |             | Somewhat<br>Improved |       | Pretty<br>Improved |      | Very<br>Improved |
|---------------|--|------------------------|------|----------------------|-------------|----------------------|-------|--------------------|------|------------------|
|               | X  | 1                      | 2    | 3                    | 4           | 5                    | 6     | 7                  | 8    | 9                |
| <b>FASQ 6</b> | To what extent do you think that watching the television program <i>Chuck</i> all semester affected your English vocabulary? |                        |      |                      |             |                      |       |                    |      |                  |
| Responses     | 0  | 2                      | 1    | 5                    | 7           | 14                   | 11    | 9                  | 1    | 1                |
| %             | 0.0%   | 3.9%                   | 2.0% | 9.8%                 | 13.7%       | 27.5%                | 21.6% | 17.6%              | 2.0% | 2.0%             |
|               |  |                        |      |                      | <b>Mean</b> | <b>5.2</b>           |       |                    |      |                  |

### *FAS items related to general language learning*

Seven items on the FAS asked the participants about their beliefs on language learning from viewing captioned episodes of *Chuck*. FASQ 4 asked the participants about how they believed viewing television affected their overall English ability. FASQ 5 was concerned with the participants' perceived improvement in English listening skills. FASQs 8 and 9 surveyed the participants about their frequency of studying English through the viewing of television before and after participation in Study 5. FASQs 10 and 11 surveyed the participants on their feelings towards viewing television in their English language class. FASQ 10 asked the participants to rate the extent to which viewing television in class was useful. FASQ 11 surveyed the participants about whether viewing television in class was enjoyable. FASQ 12 asked the participants about their perceptions of the usefulness of viewing television for language learning. Tables 6.36 to 6.40 show the results for these items.

FASQ 4 and FASQ 12 asked the participants how useful they believed watching English-language television was for improving their English ability and for language learning in general. Both items asked participants to respond on the *Not at All Useful* to *Very Useful* rating scale. For FASQ 4 (*To what extent do you think that watching the television program Chuck all semester had an effect on your overall English ability?*), 84.3% of the participants responded from *Somewhat Useful* (response category 5) to *Very Useful* (response category 9). The mean response was 5.9 and the mode was response category 7 (35.3% of participants). The responses for FASQ 12 (*To what extent do you think watching English-language television programs is useful for language learning?*) were generally positive with 96.1% of participants indicating response category 5 (*Somewhat Useful*) or above. The mode of the

responses was category 7 (33%) and the mean response score was 6.9. The responses to these two items indicate a positive attitude by the participants towards the usefulness of viewing the episodes of *Chuck* for their English ability and the usefulness of watching television in general for language learning.

Table 6.36 Responses to FASQs 4 and 12

|                |   | Not at All<br>Useful |      | Slightly<br>Useful |             | Somewhat<br>Useful |            | Pretty<br>Useful |       | Very<br>Useful |
|----------------|---|----------------------|------|--------------------|-------------|--------------------|------------|------------------|-------|----------------|
|                | X   | 1                    | 2    | 3                  | 4           | 5                  | 6          | 7                | 8     | 9              |
| <b>FASQ 4</b>  | To what extent do you think that watching the television program <i>Chuck</i> all semester affected your overall English ability? |                      |      |                    |             |                    |            |                  |       |                |
| Responses      | 0   | 1                    | 0    | 1                  | 6           | 10                 | 12         | 18               | 1     | 2              |
| %              | 0.0%  | 2.0%                 | 0.0% | 2.0%               | 11.8%       | 19.6%              | 23.5%      | 35.3%            | 2.0%  | 3.9%           |
|                |   |                      |      |                    | <b>Mean</b> | <b>5.9</b>         |            |                  |       |                |
| <b>FASQ 12</b> | To what extent do you think watching English language television programs is useful for language learning?                        |                      |      |                    |             |                    |            |                  |       |                |
| Responses      | 0   | 1                    | 0    | 0                  | 1           | 9                  | 6          | 17               | 7     | 10             |
| %              | 0.0%  | 2.0%                 | 0.0% | 0.0%               | 2.0%        | 17.6%              | 11.8%      | 33.3%            | 13.7% | 19.6%          |
|                |   |                      |      |                    |             | <b>Mean</b>        | <b>6.9</b> |                  |       |                |

A single item asked participants about how they thought viewing the 10 episodes of *Chuck* affected their listening skills. FASQ 5 (*To what extent do you think that watching the television program Chuck all semester affected your English listening skills?*) asked participants to respond on the *Not at All Improved* to *Very Improved* scale. The mean score on FASQ 5 was 5.8 and the results were bimodal with 12 of the 51 (23.5%) participants responding in response categories 6 and 7. A large majority of participants (80.3%) responded that they believed their listening was *Somewhat Improved* (response category 5) or higher. There were, however, responses in categories 2 to 9 indicating that, while most participants believed that over the course of this study they experienced improvement in their listening ability, there was wide range of beliefs about how much improvement there was.

Table 6.37 Responses to FASQ 5

|               |  | Not at All<br>Improved |      | Slightly<br>Improved |             | Somewhat<br>Improved |       | Pretty<br>Improved |       | Very<br>Improved |
|---------------|--|------------------------|------|----------------------|-------------|----------------------|-------|--------------------|-------|------------------|
|               | X  | 1                      | 2    | 3                    | 4           | 5                    | 6     | 7                  | 8     | 9                |
| <b>FASQ 5</b> | To what extent do you think that watching the television program <i>Chuck</i> all semester affected your English listening skills? |                        |      |                      |             |                      |       |                    |       |                  |
| Responses     | 0  | 0                      | 2    | 6                    | 2           | 9                    | 12    | 12                 | 7     | 1                |
| %             | 0.0%   | 0.0%                   | 3.9% | 11.8%                | 3.9%        | 17.6%                | 23.5% | 23.5%              | 13.7% | 2.0%             |
|               |  |                        |      |                      | <b>Mean</b> | <b>5.8</b>           |       |                    |       |                  |

FASQs 8 and 9 surveyed the participants on their frequency of studying English by viewing television outside of Study 5. Participants responded on the *Never* to *Always* scale for both these items. The mean rating for FASQ 8 (*To what extent did you try to learn English through watching television programs before you entered this course?*) was 5.0, which corresponds to *Fairly Often*, and the mode was 3 (27.5%), which corresponds to *Occasionally* on the rating scale. For this item there were responses across the full range of options indicating that the participants had varying degrees of experience learning from television. FASQ 9 asked the participants to consider how often they believe they would try to learn from watching television in the future. The mode for this item was response category 6 (19.6%) and the mean was 5.3. In FASQ 9 (*To what extent do you think you will try to learn English through watching television programs after you finish this course?*), 62.6% of the participants responded in response category 5 (*Fairly Often*) or higher compared to 54.9% in FASQ 8. While these statistics do not show very large differences between FASQ 8 and 9, they do point to a more positive attitude towards studying with television in the future and the intention of some participants to continue to use the medium to learn English.

Table 6.38 Responses to FASQs 8 and 9

|               |  | Never |      | Occasionally |             | Fairly Often |       | Very Often |      | Always |
|---------------|--|-------|------|--------------|-------------|--------------|-------|------------|------|--------|
|               | X  | 1     | 2    | 3            | 4           | 5            | 6     | 7          | 8    | 9      |
| <b>FASQ 8</b> | To what extent did you try to learn English through watching television programs before you entered this course?             |       |      |              |             |              |       |            |      |        |
| Responses     | 1  | 2     | 2    | 14           | 4           | 5            | 8     | 9          | 3    | 3      |
| %             | 2.0%   | 3.9%  | 3.9% | 27.5%        | 7.8%        | 9.8%         | 15.7% | 17.6%      | 5.9% | 5.9%   |
|               |  |       |      |              | <b>Mean</b> | <b>5.0</b>   |       |            |      |        |
| <b>FASQ 9</b> | To what extent do you think you will try to learn English through watching television programs after you finish this course? |       |      |              |             |              |       |            |      |        |
| Responses     | 0  | 1     | 1    | 9            | 8           | 7            | 10    | 9          | 4    | 2      |
| %             | 0.0%   | 2.0%  | 2.0% | 17.6%        | 15.7%       | 13.7%        | 19.6% | 17.6%      | 7.8% | 3.9%   |
|               |  |       |      |              | <b>Mean</b> | <b>5.3</b>   |       |            |      |        |

FASQs 10 and 11 asked the participants how they felt about watching television in their English language class. The two items used the same question (*How did you feel about watching the television programs in class?*) but different rating scales. For FASQ 10, the mean rating was 7.3 and the mode was 7 (31.4%) indicating that participants perceived watching television in class to be a *Very Good Use of Time*. All of the participants responded in the top two thirds of the response scale with 78.5% of the participants responding in the top third. FASQ 11 had the highest mean score (8.0) on the FAS with 56.9% of responses in category 9 (*Very Enjoyable*). The responses by the participants on these two items indicate that the majority of participants believed that watching episodes of *Chuck* was an enjoyable, worthwhile use of class time.

Table 6.39 Responses to FASQ 10

|                |   | Not at All Good Use of Time |      | Slightly Good Use of Time |      | Somewhat Good Use of Time |             | Pretty Good Use of Time |       | Very Good Use of Time |
|----------------|---|-----------------------------|------|---------------------------|------|---------------------------|-------------|-------------------------|-------|-----------------------|
|                | X   | 1                           | 2    | 3                         | 4    | 5                         | 6           | 7                       | 8     | 9                     |
| <b>FASQ 10</b> | How did you feel about watching the television programs in class? |                             |      |                           |      |                           |             |                         |       |                       |
| Responses      | 0   | 0                           | 0    | 0                         | 3    | 4                         | 4           | 16                      | 10    | 14                    |
| %              | 0.0%  | 0.0%                        | 0.0% | 0.0%                      | 5.9% | 7.8%                      | 7.8%        | 31.4%                   | 19.6% | 27.5%                 |
|                |   |                             |      |                           |      |                           | <b>Mean</b> | <b>7.3</b>              |       |                       |

Table 6.40 Responses to FASQ 11

|                |   | Not at All<br>Enjoyable |      | Slightly<br>Enjoyable |      | Somewhat<br>Enjoyable |      | Pretty<br>Enjoyable |            | Very<br>Enjoyable |
|----------------|---|-------------------------|------|-----------------------|------|-----------------------|------|---------------------|------------|-------------------|
|                | X   | 1                       | 2    | 3                     | 4    | 5                     | 6    | 7                   | 8          | 9                 |
| <b>FASQ 11</b> | How did you feel about watching the television programs in class? |                         |      |                       |      |                       |      |                     |            |                   |
| Responses      | 0   | 0                       | 0    | 1                     | 0    | 3                     | 3    | 10                  | 5          | 29                |
| %              | 0.0%  | 0.0%                    | 0.0% | 2.0%                  | 0.0% | 5.9%                  | 5.9% | 19.6%               | 9.8%       | 56.9%             |
|                |   |                         |      |                       |      |                       |      | <b>Mean</b>         | <b>8.0</b> |                   |

### *FAS items related to English-language captions*

The FAS for Study 5 had five items that asked the participants about their beliefs about the captions present while viewing the episodes of *Chuck*. Four of the items used 9-point response scales while one of these items was a forced-choice item. Three of the items utilized the *Not at All Useful* to *Very Useful* response scale. FASQ 13 asked the participants about the extent to which the captions were useful for their overall understanding, while FASQ 14 asked about the usefulness of the captions for vocabulary learning. FASQ 15 asked about the effect the presence of the captions had on the participants' listening ability. Results for these three items are presented in Table 6.41.

The mean response for FASQ 13 (*To what extent do you think that the English-language captions displayed as you watched the television program Chuck all semester affected your overall understanding of these episodes?*) was 7.8 and the mode was response category 9 (47.1%). Over 80% of participants responded in the top one third of the response scale (*Pretty Useful* to *Very Useful*). For FASQ 14 (*To what extent do you think that the English-language captions displayed as you watched the television program Chuck all semester affected your overall vocabulary learning?*) the mean response was 7.0 and the mode was response category 7 (39.2%). The mean response for FASQ 15 (*To what extent do you think that the English-language captions displayed as you watched the television program Chuck all semester affected your listening ability?*) was 7.0 and the mode was response category 7 (31.4%). For both FASQ 14 (72.5%) and FASQ 15 (66.7%) a large majority of participants responded in categories 7 to 9. These results suggest that the participants strongly believed that the inclusion of captions while viewing English-language television was useful for their comprehension, vocabulary learning, and listening ability.

Table 6.41 Responses to FASQs 13, 14 and 15

|           |   | Not at All<br>Useful |      | Slightly<br>Useful |      | Somewhat<br>Useful |       | Pretty<br>Useful |       | Very<br>Useful |
|-----------|---|----------------------|------|--------------------|------|--------------------|-------|------------------|-------|----------------|
|           | X   | 1                    | 2    | 3                  | 4    | 5                  | 6     | 7                | 8     | 9              |
| FASQ 13   | To what extent do you think that the English language captions displayed as you watched the television program <i>Chuck</i> all semester affected your overall understanding of these episodes? |                      |      |                    |      |                    |       |                  |       |                |
| Responses | 0   | 1                    | 0    | 1                  | 1    | 1                  | 5     | 6                | 12    | 24             |
| %         | 0.0%  | 2.0%                 | 0.0% | 2.0%               | 2.0% | 2.0%               | 9.8%  | 11.8%            | 23.5% | 47.1%          |
| Mean      |   |                      |      |                    |      |                    | 7.8   |                  |       |                |
| FASQ 14   | To what extent do you think that the English language captions displayed as you watched the television program <i>Chuck</i> all semester affected your overall vocabulary learning?             |                      |      |                    |      |                    |       |                  |       |                |
| Responses | 0   | 1                    | 0    | 0                  | 2    | 5                  | 6     | 20               | 5     | 12             |
| %         | 0.0%  | 2.0%                 | 0.0% | 0.0%               | 3.9% | 9.8%               | 11.8% | 39.2%            | 9.8%  | 23.5%          |
| Mean      |   |                      |      |                    |      |                    | 7.0   |                  |       |                |
| FASQ 15   | To what extent do you think that the English language captions displayed as you watched the television program <i>Chuck</i> all semester affected your listening ability?                       |                      |      |                    |      |                    |       |                  |       |                |
| Responses | 0   | 1                    | 0    | 0                  | 1    | 5                  | 10    | 16               | 5     | 13             |
| %         | 0.0%  | 2.0%                 | 0.0% | 0.0%               | 2.0% | 9.8%               | 19.6% | 31.4%            | 9.8%  | 25.5%          |
| Mean      |   |                      |      |                    |      |                    | 7.0   |                  |       |                |

A single item on the FAS asked participants about the frequency that they paid attention to the captions while viewing episodes of *Chuck*. FASQ 16 (*To what extent did you pay attention to the English captions as you watched episodes of Chuck all semester?*) asked participants to respond on the *Never* to *Always* scale. Table 6.42 shows the number and percentage of respondents in each category as well as the overall mean response. The mean response was 7.6 and the mode was response category 9. Ninety-six percent of participants responded from response category 5 (*Fairly Often*) to response category 9 (*Always*) with 45.1% responding that they *Always* paid attention to the captions. This response pattern indicates that the large majority of participants believed they paid a significant amount of attention to the captions as they viewed the episodes.

Table 6.42 Responses to FASQ 16

|                |  | Never |      | Occasionally |      | Fairly Often |      | Very Often |       | Always |
|----------------|--|-------|------|--------------|------|--------------|------|------------|-------|--------|
|                | X  | 1     | 2    | 3            | 4    | 5            | 6    | 7          | 8     | 9      |
| <b>FASQ 16</b> | To what extent did you pay attention to the English captions as you watched episodes of <i>Chuck</i> all semester? |       |      |              |      |              |      |            |       |        |
| Responses      | 1  | 0     | 1    | 1            | 0    | 5            | 3    | 11         | 6     | 23     |
| %              | 2.0%   | 0.0%  | 2.0% | 2.0%         | 0.0% | 9.8%         | 5.9% | 21.6%      | 11.8% | 45.1%  |
| <b>Mean</b>    |  |       |      |              |      |              |      | <b>7.2</b> |       |        |

The results from FASQ 16 are echoed in the results from Forced-choice Item 1 which surveyed the form of input the participants thought they relied on more: the captions or the audio. The vast majority of participants responded that they relied more on captions (49 out of 51) indicating the significant contribution that captions played in learning from television. Table 6.43 presents the number and percentage of participants who chose each mode of input.

Table 6.43 Responses to Forced-choice Item 1 on the FAS

|          | Participants | %     |
|----------|--------------|-------|
| Captions | 49           | 96.1% |
| Audio    | 2            | 3.9%  |

### ***FAS items related to perceptions of difficulty***

There were five additional forced-choice items on the FAS. The items asked the participants to indicate whether they had difficulty with certain aspects of viewing episodes of television. The aspects surveyed were: the vocabulary in the episodes, the content of the episodes, pronunciation, characters' names, and the speed of the dialogue. Participants responded with either *Yes* that they had difficulty or *No* that they did not. The number of responses and the relative percentages for the *Yes* or *No* categories for each item are displayed in Table 6.44. One hundred percent of the participants in Study 5 indicated that they had difficulty with the vocabulary (FAS Forced-choice Item 2) in the episodes. On the other hand, only six of the 51 participants (11.8%) suggested that they had difficulty with the content in the episodes (FAS Forced-choice Item 3). A large majority of the participants indicated that the pronunciation (FAS Forced-choice Item 4) and the speed of the dialogue (FAS Forced-choice Item 6) in the episodes were difficult for them. For pronunciation, 86.3% of the participants indicated difficulty, and for speed of the dialogue it was 88.2%. For the characters' names (FAS Forced-choice Item 5), 43.1% of the participants perceived difficulty while 56.9% did not. Overall, a large majority of participants in Study 5 claimed to have

difficulties with the vocabulary, pronunciation, and the speed of the dialogue in the episodes. A large majority also claimed not to have difficulty with the content of the episodes, and the participants were fairly evenly divided between having difficulty and not having difficulty with the characters' names.

Table 6.44 Responses to FAS Forced-choice Items 2 to 6

| Forced Choice Item            | Yes Responses |       | No Responses |       |
|-------------------------------|---------------|-------|--------------|-------|
|                               | Participants  | %     | Participants | %     |
| 2. Vocabulary in the episodes | 51            | 100%  | 0            | 0%    |
| 3. Content of the episodes    | 6             | 11.8% | 45           | 88.2% |
| 4. Pronunciation              | 44            | 86.3% | 7            | 13.7% |
| 5. Characters' names          | 22            | 43.1% | 29           | 56.9% |
| 6. Speed of the Dialogue      | 45            | 88.2% | 6            | 11.8% |

***Comparison of the results of the FAS rating scale items for the Captions Group and the No Captions Group***

To examine how participants' attitudes about comprehension of the episodes, vocabulary learning from the episodes, and language learning through viewing television with captions compared with those of the participants who viewed the episodes without captions, responses to items on the Final Attitude Survey from Study 5 (Captions Group) were compared with those from Study 4 (No Captions Group). The two groups had 12 FAS items in common and the responses to these items are presented by common field of investigation: items related to comprehension, items related to vocabulary learning, and items related to general language learning. See Sections 5.6.2 and 6.7.3.2 for a full explanation of each treatment groups' responses to these FAS items.

***Comparison of the responses from the Captions and No Captions groups for the FAS items related to comprehension***

Four items on the FAS examined the participants' beliefs about their comprehension of the episodes of *Chuck*. Tables 6.45 and 6.46 show the results for FASQs 1, 2, 3, and 7 for the Captions Group and the No Captions Group. Each table shows the questions participants were asked to respond to, the numerical and verbal labels for each response scale, the percentage of responses in each category, and the mean response for both groups of participants. FASQ 1 asked the participants to consider their level of understanding of the first episode they viewed while FASQ 2 asked them to consider their level of understanding of the final episode. The mean response on FASQ 1 for the Captions Group was 3.6 and was bimodal with 19.6% of

participants choosing response categories 2 and 5. For the No Captions Group, the mean was 3.3 with a mode of response category 3 (22.1%). While the response patterns of the two groups were similar, participants who viewed the first episode with captions tended to believe their comprehension to be slightly better than those who viewed the episode without captions. This is reflected in the slightly higher group mean and because 35.3% of participants in the Captions Group responded from response category 5 (*Somewhat Good*) to response category 9 (*Very Good*) compared with only 23.9% for the No Captions Group.

The mean response for FASQ 2 concerning the comprehension of the final episode viewed also saw the Captions Group respond with a higher mean (6.1) than the No Captions Group (5.7). Both groups had a mode of response category 7 (Captions Group: 27.5%, No Captions Group: 27.9%). The presence of captions may again have led to stronger beliefs about comprehension in the final episode viewed as indicated by 49.1% of participants in the Captions Group selecting response category 7 (*Pretty Good*) to response category 9 (*Very Good*) compared to 40.3% for the No Captions Group.

This response pattern indicating that participants with access to captions believed they had slightly better comprehension in the first and final episodes viewed was supported by the responses to FASQ 7. FASQ 7 asked participants about the degree to which they believed their comprehension improved from the first to the last episode viewed. The mean response for the Captions Group was 6.3 with a mode of response category 7 (35.3%) while the mean for the No Captions Group was 6.1 with a mode of response category 7 (27.4%). Of the participants in the Captions Group, 56.8% responded from response category 7 (*Pretty Improved*) to response category 9 (*Very Improved*) while 45.1% of those in the No Captions Group did.

FASQ 3 asked participants what they believed their overall level of understanding of the episodes viewed was. The mean response for the Captions Group was 5.6 compared to 5.3 for the No Captions Group. The mode for the Captions Group was response category 7 (31.4%) and for the No Captions Group it was response category 5 (23.5%). Once again the presence of captions appears to lead to a slightly stronger belief about the amount of comprehension participants perceived themselves to have.

Table 6.45 Responses to FASQ 1, 2 and 3 for the Captions Group and No Captions Group

|               | Not at All<br>Good  |       | Slightly<br>Good |       | Somewhat<br>Good |       | Pretty<br>Good |       | Very<br>Good |            |
|---------------|---|-------|------------------|-------|------------------|-------|----------------|-------|--------------|------------|
|               | 1   | 2     | 3                | 4     | 5                | 6     | 7              | 8     | 9            |            |
| <b>FASQ 1</b> | What do you think your understanding of the first episode of <i>Chuck</i> you watched this semester was?    |       |                  |       |                  |       |                |       |              | <b>M</b>   |
| Captions      | 13.7%   | 19.6% | 17.6%            | 13.7% | 19.6%            | 7.8%  | 5.9%           | 2.0%  | 0.0%         | <b>3.6</b> |
| No Captions   | 16.8%   | 20.8% | 22.1%            | 13.7% | 14.6%            | 4.4%  | 3.1%           | 0.0%  | 1.8%         | <b>3.3</b> |
| <b>FASQ 2</b> | What do you think your understanding of the last episode of <i>Chuck</i> you watched this semester was?     |       |                  |       |                  |       |                |       |              | <b>M</b>   |
| Captions      | 0.0%  | 2.0%  | 5.9%             | 7.8%  | 21.6%            | 13.7% | 27.5%          | 19.6% | 2.0%         | <b>6.1</b> |
| No Captions   | 0.9%  | 1.8%  | 14.2%            | 5.8%  | 16.4%            | 18.1% | 27.9%          | 8.4%  | 4.0%         | <b>5.7</b> |
| <b>FASQ 3</b> | What do you think your overall understanding of the episodes of <i>Chuck</i> you watched this semester was? |       |                  |       |                  |       |                |       |              | <b>M</b>   |
| Captions      | 2.0%  | 0.0%  | 13.7%            | 5.9%  | 23.5%            | 17.6% | 31.4%          | 5.9%  | 0.0%         | <b>5.6</b> |
| No Captions   | 1.3%  | 3.1%  | 15.5%            | 8.4%  | 23.5%            | 21.2% | 16.8%          | 5.3%  | 2.2%         | <b>5.3</b> |

Table 6.46 Responses to FASQ 7 for the Captions Group and No Captions Group

|               | Not at All<br>Improved   |      | Slightly<br>Improved |      | Somewhat<br>Improved |       | Pretty<br>Improved |       | Very<br>Improved |            |
|---------------|--|------|----------------------|------|----------------------|-------|--------------------|-------|------------------|------------|
|               | 1  | 2    | 3                    | 4    | 5                    | 6     | 7                  | 8     | 9                |            |
| <b>FASQ 7</b> | To what degree do you think your understanding of the television program <i>Chuck</i> improved from the first episode watched to the last? |      |                      |      |                      |       |                    |       |                  | <b>M</b>   |
| Captions      | 0.0%   | 2.0% | 3.9%                 | 9.8% | 13.7%                | 13.7% | 35.3%              | 13.7% | 7.8%             | <b>6.3</b> |
| No Captions   | 0.0%   | 1.3% | 10.2%                | 5.8% | 15.5%                | 19.5% | 27.4%              | 11.9% | 5.8%             | <b>6.1</b> |

***Comparison of the responses from the Captions and No Captions groups for the FAS item related to vocabulary learning***

The single item on the FAS that asked the participants about their beliefs about the effects that viewing 10 episodes of *Chuck* had on their English vocabulary learning had the largest difference in average response between the Captions and No Captions groups on the FAS. Table 6.47 shows the results for FASQ 6 for the Captions Group and the No Captions Group. The mean response on FASQ 6 for the Captions Group was 5.2 and the mode was response category 5 (27.5%) while the mean for the No Captions Group was 4.5 with the same mode of response category 5 (24.3%). For the Captions Group, 66.7% of participants indicated that they believed that their vocabulary was from *Somewhat Improved* (response category 5) to *Pretty Improved* (response category 7). In contrast, 46.8% of participants in the No Captions Group responded in the same range of categories. This pattern of responses indicates that

language learners with access to captions believed that they had more improvement in their vocabulary than those learners who did not have access to captions.

Table 6.47 Responses to FASQ 6 for the Captions Group and No Captions Group

|             | Not at All Improved  |      | Slightly Improved |       | Somewhat Improved |       | Pretty Improved |      | Very Improved |     |
|-------------|--|------|-------------------|-------|-------------------|-------|-----------------|------|---------------|-----|
|             | 1  | 2    | 3                 | 4     | 5                 | 6     | 7               | 8    | 9             |     |
| FASQ 6      | To what extent do you think that watching the television program <i>Chuck</i> all semester affected your English vocabulary? |      |                   |       |                   |       |                 |      |               | M   |
| Captions    | 3.9%   | 2.0% | 9.8%              | 13.7% | 27.5%             | 21.6% | 17.6%           | 2.0% | 2.0%          | 5.2 |
| No Captions | 3.1%   | 8.8% | 20.8%             | 13.7% | 24.3%             | 12.8% | 9.7%            | 3.1% | 0.9%          | 4.5 |

***Comparison of the responses from the Captions and No Captions groups for the FAS items related to language learning***

Seven items on the FAS asked the participants about their beliefs about language learning from the episodes of *Chuck*. Two of these items asked participants to respond on the *Not at All Useful* to *Very Useful* rating scale. Table 6.48 shows the results for FASQs 4 and 12. FASQ 4 asked participants about their beliefs on how watching episodes of *Chuck* affected their overall English ability. The Captions Group had a slightly higher mean of 5.9 compared with 5.6 for the No Captions Group. The mode for the Captions Group was response category 7 (35.3% of participants) and for the No Captions Group it was response category 5 (22.1%). In the Captions Group, 84.3% of participants responded from response category 5 (*Somewhat Useful*) to response category 9 (*Very Useful*) while 72.9% of participants in the No Captions Group did. There was very little difference between the response patterns of the Captions Group and the No Captions Group on FASQ 12 which asked participants about the extent to which they believed watching television is useful for language learning. The mean for the Captions Group was 6.9 and for the No Captions Group it was 6.8. For both treatment groups the mode of the responses was response category 7 (Captions Group: 33.3%, No Captions Group: 27.9%). These response patterns for these items point to a positive belief by the participants in both groups, about the usefulness of watching television for their English ability and for language learning in general.

Table 6.48 Responses to FASQs 4 and 12 for the Captions Group and No Captions Group

|                | Not at All Useful   |      | Slightly Useful |       | Somewhat Useful |       | Pretty Useful |       | Very Useful |            |
|----------------|---|------|-----------------|-------|-----------------|-------|---------------|-------|-------------|------------|
|                | 1   | 2    | 3               | 4     | 5               | 6     | 7             | 8     | 9           |            |
| <b>FASQ 4</b>  | To what extent do you think that watching the television program <i>Chuck</i> all semester affected your overall English ability? |      |                 |       |                 |       |               |       |             | <b>M</b>   |
| Captions       | 2.0%  | 0.0% | 2.0%            | 11.8% | 19.6%           | 23.5% | 35.3%         | 2.0%  | 3.9%        | <b>5.9</b> |
| No Captions    | 0.4%  | 2.2% | 13.3%           | 8.4%  | 22.1%           | 19.0% | 21.2%         | 3.5%  | 7.1%        | <b>5.6</b> |
| <b>FASQ 12</b> | To what extent do you think watching English language television programs is useful for language learning?                        |      |                 |       |                 |       |               |       |             | <b>M</b>   |
| Captions       | 2.0%  | 0.0% | 0.0%            | 2.0%  | 17.6%           | 11.8% | 33.3%         | 13.7% | 19.6%       | <b>6.9</b> |
| No Captions    | 0.9%  | 0.9% | 4.4%            | 4.9%  | 14.2%           | 10.2% | 27.9%         | 10.2% | 23.9%       | <b>6.8</b> |

FASQ 5 asked participants about the extent to which they believed that watching the episodes of *Chuck* affected their listening skills. The mean response for the Captions Group (5.8) was higher than that of the No Captions Group (5.5). The responses for the Captions Group were bimodal with 23.5% of participants choosing response categories 6 and 7. The mode for the No Captions Group was response category 5 (23.9%). Generally, the Captions Group felt their English listening skills improved more than participants in the No Captions Group. A total of 39.2% of the Captions Group responded from category 7 (*Pretty Improved*) to category 9 (*Very Improved*) compared to 30.6% from the No Captions Group. Table 6.49 shows the results for FASQ 5 for the Captions Group and the No Captions Group.

Table 6.49 Responses to FASQ 5 for the Captions Group and No Captions Group

|               | Not at All Improved  |      | Slightly Improved |      | Somewhat Improved |       | Pretty Improved |       | Very Improved |            |
|---------------|--|------|-------------------|------|-------------------|-------|-----------------|-------|---------------|------------|
|               | 1  | 2    | 3                 | 4    | 5                 | 6     | 7               | 8     | 9             |            |
| <b>FASQ 5</b> | To what extent do you think that watching the television program <i>Chuck</i> all semester affected your English listening skills? |      |                   |      |                   |       |                 |       |               | <b>M</b>   |
| Captions      | 0.0%   | 3.9% | 11.8%             | 3.9% | 17.6%             | 23.5% | 23.5%           | 13.7% | 2.0%          | <b>5.8</b> |
| No Captions   | 0.9%   | 4.9% | 12.8%             | 7.5% | 23.9%             | 16.8% | 17.3%           | 9.3%  | 4.0%          | <b>5.5</b> |

FASQs 8 and 9 investigated the participants' frequency of studying English by viewing television prior to taking part in this study and their intentions to do so following the study. The participants responded on the *Never* to *Always* scale for both of these items. On FASQ 8, the mean response for the Captions Group was 5.0 (*Fairly Often*) and the mode was response category 3 (27.5%). For the No Captions Group, the mean rating was 4.8, which corresponds to more frequent than *Occasionally* with the mode also category 3 (26.1%). On FASQ 9, the

mean response for the Captions Group was 5.3 and the mode was response category 6 (19.6%). For the No Captions Group, the mean response was 4.8 and the mode was response category 3 (24.3%). The responses to these questions indicate little potential change in the habits of the Captions Group and little to no change in the No Captions Group. The range of responses across all response categories for both survey items and both treatment groups indicates that participants had a wide range of individual beliefs about their intentions for using television for language learning in the future.

Table 6.50 Responses to FASQs 8 and 9 for the Captions Group and No Captions Group

|               | Never  |      | Occasionally |       | Fairly Often |       | Very Often |      | Always |            |
|---------------|--|------|--------------|-------|--------------|-------|------------|------|--------|------------|
|               | 1  | 2    | 3            | 4     | 5            | 6     | 7          | 8    | 9      |            |
| <b>FASQ 8</b> | To what extent did you try to learn English through watching television programs before you entered this course?             |      |              |       |              |       |            |      |        | <b>M</b>   |
| Captions      | 3.9%   | 3.9% | 27.5%        | 7.8%  | 9.8%         | 15.7% | 17.6%      | 5.9% | 5.9%   | <b>5.0</b> |
| No Captions   | 4.9%   | 4.9% | 26.1%        | 9.3%  | 12.4%        | 15.5% | 16.8%      | 5.3% | 2.2%   | <b>4.8</b> |
| <b>FASQ 9</b> | To what extent do you think you will try to learn English through watching television programs after you finish this course? |      |              |       |              |       |            |      |        | <b>M</b>   |
| Captions      | 2.0%   | 2.0% | 17.6%        | 15.7% | 13.7%        | 19.6% | 17.6%      | 7.8% | 3.9%   | <b>5.3</b> |
| No Captions   | 2.7%   | 2.2% | 24.3%        | 12.4% | 23.5%        | 9.7%  | 16.4%      | 4.4% | 1.8%   | <b>4.8</b> |

FASQs 10 and 11 both asked the participants how they felt about watching television in their English-language class but had the participants respond on different rating scales. For both treatment groups, the response patterns to these items were very similar with the mean responses the highest among the 12 shared FAS items. On FASQ 10, a large majority of participants responded from response category 7 (*Pretty Good Use of Time*) to response category 9 (*Very Good Use of Time*) in both treatment groups (Captions Group: 78.5%, No Captions Group: 71.2%). The mean for the Captions Group was 7.3 and the mode was response category 7 (31.4%). The mean for the No Captions Group was 7.2 with a mode of response category 9 (32.3%). For FASQ 11, an even larger majority of participants in both treatment groups (Captions Group: 86.5%, No Captions Group: 86.7%) responded in the top third of the response scale (*Pretty Enjoyable* to *Very Enjoyable*). The mean and mode for both the Captions Group and the No Captions Group was 8.0 and response category 9 (Captions Group: 56.9%, No Captions Group: 55.3%) respectively. The results from these two items indicate that with or without the presence of captions the majority of participants viewing television in their language class believed it to be an enjoyable, worthwhile use of time.

Table 6.51 Responses to FASQ 10 for the Captions Group and No Captions Group

|                | Not at All<br>Good Use<br>of Time                                 |      | Slightly<br>Good Use<br>of Time |      | Somewhat<br>Good Use<br>of Time |       | Pretty<br>Good Use<br>of Time |       | Very<br>Good Use<br>of Time |            |
|----------------|---|------|---------------------------------|------|---------------------------------|-------|-------------------------------|-------|-----------------------------|------------|
|                | 1   | 2    | 3                               | 4    | 5                               | 6     | 7                             | 8     | 9                           |            |
| <b>FASQ 10</b> | How did you feel about watching the television programs in class? |      |                                 |      |                                 |       |                               |       |                             | <b>M</b>   |
| Captions       | 0.0%  | 0.0% | 0.0%                            | 5.9% | 7.8%                            | 7.8%  | 31.4%                         | 19.6% | 27.5%                       | <b>7.3</b> |
| No Captions    | 0.4%  | 1.3% | 1.8%                            | 0.9% | 10.6%                           | 11.1% | 29.6%                         | 9.3%  | 32.3%                       | <b>7.2</b> |

Table 6.52 Responses to FASQ 11 for the Captions Group and No Captions Group

|                | Not at All<br>Enjoyable   |      | Slightly<br>Enjoyable |      | Somewhat<br>Enjoyable |      | Pretty<br>Enjoyable |      | Very<br>Enjoyable |            |
|----------------|---|------|-----------------------|------|-----------------------|------|---------------------|------|-------------------|------------|
|                | 1   | 2    | 3                     | 4    | 5                     | 6    | 7                   | 8    | 9                 |            |
| <b>FASQ 11</b> | How did you feel about watching the television programs in class? |      |                       |      |                       |      |                     |      |                   | <b>M</b>   |
| Captions       | 0.0%  | 0.0% | 2.0%                  | 0.0% | 5.9%                  | 5.9% | 19.6%               | 9.8% | 56.9%             | <b>8.0</b> |
| No Captions    | 0.0%  | 0.9% | 0.0%                  | 0.0% | 5.8%                  | 4.0% | 22.1%               | 9.3% | 55.3%             | <b>8.0</b> |

***Comparison of the responses from the Captions and No Captions groups for the FAS items related to language learning***

Responses by the Captions Group and No Captions Group on the 12 FASQs were compared with a series of Wilcoxon Rank-Sum Tests. For 11 of 12 items there was no significant difference between the two treatment groups (FASQ 1:  $Z = 1.37$ ,  $p < .870$ , FASQ 2:  $Z = 1.29$ ,  $p < .980$ , FASQ 3:  $Z = 1.36$ ,  $p < .870$ , FASQ 4:  $Z = 1.46$ ,  $p < .720$ , FASQ 5:  $Z = 1.40$ ,  $p < .810$ , FASQ 7:  $Z = 1.01$ ,  $p < .137$ , FASQ 8:  $Z = .67$ ,  $p < .253$ , FASQ 9:  $Z = 1.57$ ,  $p < .058$ , FASQ 10:  $Z = .22$ ,  $p < .415$ , FASQ 11:  $Z = .097$ ,  $p < .462$ , FASQ 12:  $Z = .20$ ,  $p < .422$ ). The response patterns of the two treatment groups were found to be significantly different for FASQ 6. The participants' response scores on the item asking about the effect that viewing the episodes of *Chuck* had on their English vocabulary were significantly higher for the Captions Group than the No Captions Group,  $Z = 2.62$ ,  $p < .01$ . The effect size, as measured by  $r$ , was .16 which corresponds to a small treatment effect.

***Comparison of the responses from the Captions and No Captions groups for the FAS forced-choice items related to perceptions of difficulty viewing television***

The results of the five common forced-choice items on the FAS from the Captions Group and the No Captions Group were compared to investigate if the presence of captions affects the areas where participants perceived difficulties. The items asked the participants to indicate

whether they had difficulty with certain aspects of viewing episodes of television. The number of responses and the percentages for either the *Yes* (perceived difficulty) or *No* (did not perceive difficulty) categories for each item for the Captions Group and the No Captions Group are displayed in Table 6.53. For the FAS Forced-choice Item concerning the vocabulary in the episodes, 95% of the No Captions Group indicated difficulty. The participants in the Captions Group who had access to the written form to help them make a form-meaning link reported 100% agreement that they had difficulty with the vocabulary. These results indicate that a barrier to learning from viewing authentic television programs with or without captions might be the low-frequency vocabulary contained within them.

The FAS Forced-choice Item which asked participants whether they thought they had difficulty with the content of the episodes had the largest difference between the responses of the Captions and No Captions Groups. Over 10% more participants in the No Captions Group indicated difficulty with the content of *Chuck*. This suggests that the availability of captions led participants to feel better supported with regards to content. This finding is supported by the significantly higher mean comprehension scores of the Captions Group for certain episodes (see Section 6.4.3.3). Still, the large majority of participants in both groups claimed not to have had difficulty with the content of the episodes.

The response patterns for the FAS Forced-choice Items concerning the pronunciation and the speed of the dialogues in the episodes were very similar for both groups of participants. For the item concerning pronunciation, the Captions Group had 4.5% more participants claim difficulty while for the item concerning the speed of the dialogue there was only 0.4% difference between the groups. For both items, over 80% of participants indicated they had difficulty in these areas. This indicates that regardless of the presence of captions, the pronunciation and the speed of the dialogue in authentic television programs is a challenge for most language learners to cope with.

With regards to the FAS Forced-choice Item asking the participants about their perceptions of difficulty with the characters' names there was not much difference between the groups with 2.8% more participants in the No Captions Group reporting difficulty. The responses were fairly evenly split in each group with just over 54% of participants in both groups reporting no difficulty with the characters' names.

Table 6.53 Comparison of responses to FAS Forced-choice Items for perceptions of difficulty for the Captions Group and the No Captions Group

| Forced Choice Item         | Treatment   | Yes Responses |       | No Responses |       |
|----------------------------|-------------|---------------|-------|--------------|-------|
|                            |             | Participants  | %     | Participants | %     |
| Vocabulary in the episodes | Captions    | 51            | 100%  | 0            | 0%    |
|                            | No Captions | 209           | 95.0% | 11           | 5.0%  |
| Content of the episodes    | Captions    | 6             | 11.8% | 45           | 88.2% |
|                            | No Captions | 49            | 22.3% | 171          | 77.7% |
| Pronunciation              | Captions    | 44            | 86.3% | 7            | 13.7% |
|                            | No Captions | 180           | 81.8% | 40           | 18.2% |
| Characters' names          | Captions    | 22            | 43.1% | 29           | 56.9% |
|                            | No Captions | 101           | 45.9% | 119          | 54.1% |
| Speed of the dialogue      | Captions    | 45            | 88.2% | 6            | 11.8% |
|                            | No Captions | 195           | 88.6% | 25           | 11.4% |

#### 6.7.4. Summary of findings

The main findings of Study 5: Experiment 4 can be summarized as follows:

1. The mean responses concerning the Captions Group participants' beliefs about the enjoyability of, benefits from, and usefulness of language learning through viewing captioned television significantly increased from the first to the tenth episodes viewed. The mean response for the item surveying enjoyability of learning through television was slightly higher than *Pretty Enjoyable* by the final episode viewed. The mean response for the item concerning the level of English learning was marginally higher than *Somewhat Good* and the item concerning the usefulness of television for studying English was between *Somewhat* and *Pretty Useful*.
2. Across the 10 episodes viewed with captions, the participants' perceptions of their comprehension of the episodes rose immediately after the first episode and then the mean responses became more episode-dependent. The level of perceived comprehension was significantly higher after viewing the Final Episode at a level slightly above *Somewhat Good*.
3. The majority of EFL learners' used and relied on the captions throughout the episodes they viewed and considered them between *Pretty Useful* and *Very Useful* for comprehension.
4. Following viewing of all 10 episodes of captioned television, the majority of EFL learners believed their comprehension improved from the first to the last

episode viewed and that their overall comprehension of the captioned episodes was *Somewhat Good* or better.

5. While there was a wide range of perceptions about their vocabulary learning as indicated by responses in all nine categories stretching from *Not at All Improved* to *Very Improved*, 96.1% of participants believed their vocabulary was *Improved* over the course of the study.
6. The participants were generally positive about language learning through viewing television with captions. The majority of participants thought that learning from captioned television was a *Pretty Good Use of Time* and *Pretty Enjoyable* or beyond making these the two highest rated issues. The majority of participants thought that viewing television had a *Somewhat Useful* effect on their overall English ability, was *Somewhat Useful* for language learning in general, and had *Somewhat Improved* their listening skills. Participants, however, showed very little difference in the frequency they had previously used television to learn English and the frequency with which they thought they would use it to learn English in the future.
7. Participants felt strongly about the usefulness of the captions for comprehension, vocabulary learning, and listening ability and strongly favored the captions over the audio. The majority of participants thought that viewing the captioned television was *Pretty Useful* or higher for comprehension, vocabulary learning, and listening ability. The majority of participants also thought they paid attention to the captions from *Very Often* to *Always* and over 95% of participants claimed to rely more on the captions than the audio.
8. The majority of the participants in this study maintained that they had difficulties with the vocabulary, pronunciation, and the speed of the dialogue in the captioned television episodes. Conversely, a large majority did not have difficulty with the content of the episodes. Participants were fairly evenly divided between having difficulty and not having difficulty with the characters' names.
9. On the Episode Attitude Survey, the Captions Group had mean responses that were higher or equal to those of the No Captions Group for all common items and the response patterns across the episodes were similar. On the Final Attitude Survey, the mean responses for the Captions Group were equal to or higher than those from the No Captions Group on all items. However, the

Captions Group only had significantly higher responses on the item asking about vocabulary learning from viewing television. Overall, the results indicated that there was little difference between the attitudes of participants who viewed television with or without captions, but the participants who had access to captions had slightly more positive attitudes.

## **6.8. Discussion**

The focus of Study 5 is to investigate the effects of captions on the four aspects of language learning investigated in Studies 1 to 4: comprehension, vocabulary acquisition, lexical coverage, and attitudes. As such, this discussion of relevant findings concentrates on contrasting the findings from the Captions and No Captions groups.

### **6.8.1. Captions and comprehension**

#### **6.8.1.1. Comprehension gains from the first to the tenth episode of television**

Both the Captions Group and the No Captions Group made significant gains in comprehension from the first to the tenth episode of *Chuck*. Participants in the No Captions Group had significantly greater gains which reflect the lower comprehension scores they had for the first episode. By the tenth episode viewed, there was no significant difference between the comprehension scores of the treatment groups. This suggests that for the first episode viewed the Captions Group may have been able to make use of the captions to improve comprehension. The No Captions Group did not have this added support. However, by the final episode viewed, it appears that the No Captions Group was able to accumulate knowledge of the series to a level where the added support of the captions did not to make a significant difference in comprehension.

#### **6.8.1.2. Comprehension of the 10 episodes of television**

A comparison of the comprehension results for the Captions Group and the No Captions Group across the 10 episodes viewed reveals some apparent differences between the treatment groups. The mean comprehension scores for all episodes were higher when captions were present. The comprehension results for the Captions Group, however, were only shown to be significantly higher than those of the No Captions Group for three of the 10 episodes. The Initial Episode was one of the significantly higher episodes indicating that when the participants were viewing a television series for the first time, the captions were beneficial for comprehension. The two other episodes with a significant difference between comprehension scores (Episodes 4 and 7) were episodes where the No Captions Group had their lowest comprehension scores after the first episode. This indicates that for certain episodes, regardless of the participants' familiarity with a series, the presence of captions can significantly assist comprehension particularly when comprehension of the episode is comparatively difficult.

The difference in comprehension scores between the Captions and No Captions groups in this study do not fully support the findings of prior research. In the majority of previous studies, language learners with access to captions had significantly higher comprehension scores (Chung, 1999; Guillory, 1998; Huang & Eskey, 1999; Latifi et al., 2011; Markham et al., 2001; Markham & Peter, 2003). In these studies, comprehension scores were on average 27.95% higher when learners viewed videos with captions. In Study 5, for the three episodes with a significant difference between comprehension scores of the treatment groups, the mean difference was 4.5% (Initial Episode: 6.4%, Episode 4: 4.7%, Episode 7: 2.4%). The comparatively smaller differences in comprehension scores between the Captions Group and the No Captions Group may be a result of the type of video used in this study. In the six studies cited, three used educational videos, two used documentaries, and one study used segments of an animated movie. The viewing times (where reported) ranged from 7 minutes to 30 minutes. The study with 30 minutes of viewing time was actually 15 2-minute segments of a movie (Latifi et al., 2011). In Study 5, however, participants viewed episodes of television over 40 minutes in length. The series, *Chuck*, used in this study had a season-long story-arc with each episode furthering that storyline. However, each episode was still designed to be understood and enjoyed as a separate entity even with little or no background to the series. Each episode told a complete story from beginning to end. This may not be the case for short documentaries or educational programs which are designed to convey information to viewers. While a movie is designed to tell a complete and comprehensible story, viewing short decontextualized segments may reduce the ability of the viewer to build up knowledge about the story and the characters. In the absence of complete narratives developed in detail over a relatively long period of an episode, it appears that language learners in previous studies might have relied on captions considerably more for comprehension leading to larger differences in comprehension scores between the Captions and No Captions groups. In Study 5, even if learners did not have access to captions they may have been able to comprehend the story more effectively than the learners in previous research did due to the type of video viewed.

#### **6.8.1.3. The effects of vocabulary knowledge on comprehension of episodes of television**

For the Captions Group, there were small significant correlations between vocabulary knowledge and comprehension for four of the 10 episodes, and for the No Captions Group, there were small to medium significant correlations for all 10 of the episodes. This indicates

that greater vocabulary knowledge may be more important for comprehension of episodes of television when captions are not present. When learners are forced to rely solely on the aural form of vocabulary, when they are viewing episodes of television without captions, vocabulary knowledge plays a more significant role. This suggests that captions allow learners with less vocabulary knowledge to utilize their knowledge of the written forms of the words for comprehension. There are additional factors involved in comprehension, as indicated by the small and medium correlations in both groups, but the results do indicate the importance of vocabulary knowledge for comprehension of episodes of television, especially when viewed without captions.

## **6.8.2. Captions and incidental vocabulary acquisition**

### **6.8.2.1. Incidental vocabulary acquisition from viewing television**

On both the Tough Test (Captions Group: 6.03 words, No Captions Group: 6.36 words) and the Sensitive Test (Captions Group: 5.93 words, No Captions Group: 6.78 words), there were comparable vocabulary gains with no significant difference between the mean gains (raw and relative) of the participants in Study 5 and the participants in Study 2. This finding does not support previous research that has indicated that vocabulary learning through viewing videos with captions is superior to vocabulary learning through viewing videos without captions (Baltova, 1999; Huang & Eskey, 1999; Hui, 2007; Markham, 1999; Neuman & Koskinen, 1992; Sydorenko, 2010; Winke et al., 2010). However, differences in the types and lengths of the videos used in this study and those used in prior research may provide an explanation for the contradictory results. The earlier research comparing vocabulary learning from videos viewed with and without captions has most commonly employed documentaries or educational videos. By their very nature, these video types are designed to educate and convey information to viewers. In documentaries and educational programs relevant vocabulary may occur at a relatively high frequency often with explicit and clear links to the imagery in the videos. The episodes of *Chuck* used in this study were designed to entertain.

The videos used in the earlier research ranged in length from 2 to 21 minutes. A single episode of the program used in this study was approximately 42 minutes and participants viewed 10 episodes of *Chuck*. The vocabulary in the episodes would likely have occurred with greater spacing than the target vocabulary in the previous research. For example in the study by Sydorenko (2010) knowledge of 28 target words was measured following viewing approximately 8 minutes of video. The 60 target words from Study 5 occurred across approximately 7 hours of viewing time. In studies like Sydorenko's with short videos with a

relatively high concentration of unknown vocabulary, captions appear to contribute more to learning the form-meaning connection. However, in this study, vocabulary learning from viewing videos with captions was not statistically different from vocabulary learning through viewing videos without captions indicating that captions do not have as much effect on vocabulary learning through viewing full-length episodes of television dramas. While the vocabulary gains from Study 5 may not have been significantly different from the gains in Study 2, they nevertheless show that incidental vocabulary learning through viewing episodes of television with captions does occur.

#### **6.8.2.2. The effects of vocabulary knowledge on relative vocabulary gain**

The Captions Group had significant correlations between vocabulary knowledge and relative vocabulary gain on both the Tough and Sensitive tests while no significant correlations were found for the No Captions Group for either test. This indicates that the participants with greater vocabulary knowledge in the Captions Group were able to use both the aural and written input in captioned television to learn vocabulary more effectively than those with less vocabulary knowledge. These results support previous research that has shown that vocabulary proficiency influences incidental vocabulary acquisition (Day et al., 1992; Horst et al., 1998; Oetting et al., 1995; Zahar et al., 2001). The reason no significant difference was found for the No Captions Group may be a function of how vocabulary knowledge was measured in this study. Vocabulary knowledge was calculated by combining the results of the 2,000-, 3,000-, and 5,000-word levels of the VLT and analyzing the results using the Rasch Model to obtain an interval measure. This was done for both the Captions Group and the No Captions Group. The VLT measures receptive vocabulary knowledge using written forms. It is not clear how receptive vocabulary knowledge measured using aural forms may affect findings. Further research examining the relationship between knowledge of the aural form and incidental vocabulary learning through viewing episodes of television without captions is needed.

#### **6.8.2.3. Frequency and range of vocabulary occurrence and incidental vocabulary acquisition**

The findings from Study 5 on the relationship between vocabulary gain and frequency of occurrence support the findings from Study 2. For the Captions Group, there was a significant medium size correlation of .39 between relative gain of the items on the Tough Test and frequency of occurrence. For the No Captions Group, the correlation was .30. There was no

significant correlation for frequency of occurrence and gains on the Sensitive Test for either treatment group. The findings from Study 5 provide more support for the role that frequency of occurrence has in incidental learning of L2 vocabulary (Schmitt, 2008; Webb, 2008). The small significant negative correlation between the relative item gains for the Sensitive Test and the range of occurrence of the target words ( $r = -.27$ ) and the medium size significant positive correlation between the relative frequency of target word occurrence and the relative gain for items on both the Tough Test ( $r = .34$ ) and the Sensitive Test ( $r = .43$ ) in Study 5 shed light on the role that frequency of occurrence has in incidental vocabulary learning. These findings indicate that target words are better learned when they occur more frequently in one text rather than across a range of texts.

### **6.8.3. Captions and lexical coverage**

#### **6.8.3.1. The effects of lexical coverage on comprehension of episodes of television**

The comparison of the results from the Captions and No Captions groups from the investigation of the effects of lexical coverage on comprehension of television revealed a number of noteworthy findings. First, there was a difference in the number of episodes where there was a significant difference in comprehension between participants with different levels of lexical coverage. In the Captions Group, there were three episodes of *Chuck* where the participants with mastery of the 2,000-word level of the VLT had significantly higher comprehension scores than participants without mastery of the 2,000-word level. This compares with the six episodes for the No Captions Group. This finding indicates that increased lexical coverage may be more important for comprehension when captions are not present. The reason for this may be that the participants had more knowledge of the written form of vocabulary than the aural form. When captions were present, participants with lower lexical coverage were better able to make use of their vocabulary knowledge leading to similar comprehension scores between the participants with different levels of lexical coverage. This suggests that captions allow learners with less lexical coverage to utilize their knowledge of the written forms of words for comprehension. The test results for the episodes where there was a difference in comprehension between the two levels of lexical coverage are shown in Table 6.54.

Table 6.54 Comprehension scores, lexical coverage and effect sizes of episodes where there was a significant difference between participants with and without mastery of 2,000-word level

| Episode | VLT        | Lexical Coverage      | Captions Group |       |          | No Captions Group |       |          |
|---------|------------|-----------------------|----------------|-------|----------|-------------------|-------|----------|
|         |            |                       | Mean Score     |       | $\eta^2$ | Mean Score        |       | $\eta^2$ |
|         |            |                       | %              | CHIPs |          | %                 | CHIPs |          |
| 1       | Less 2,000 | < 94.28%              | 67.2           | 54.2  | .25      | 68.6              | 54.6  | .04      |
|         | 2,000      | 94.28%                | 78.3           | 57.4  |          | 71.8              | 55.8  |          |
| 3       | Less 2,000 | < 94.33%              | -              | -     | -        | 69.6              | 55.0  | .03      |
|         | 2,000      | 94.33%                | -              | -     |          | 72.0              | 55.8  |          |
| 4       | Less 2,000 | < 94.69%              | 62.6           | 53.1  | .14      | -                 | -     | -        |
|         | 2,000      | 94.69%                | 72.0           | 55.6  |          | -                 | -     |          |
| 5       | Less 2,000 | < 94.72%              | -              | -     | -        | 63.5              | 53.3  | .06      |
|         | 2,000      | 94.72%                | -              | -     |          | 66.1              | 54.3  |          |
| 6       | Less 2,000 | < 93.52%              | -              | -     | -        | 61.7              | 52.7  | .03      |
|         | 2,000      | 93.52%                | -              | -     |          | 64.2              | 53.6  |          |
| 8       | Less 2,000 | < 93.95%              | 64.5           | 53.8  | .13      | 65.5              | 54.0  | .02      |
|         | 2,000      | 93.95%                | 72.4           | 56.0  |          | 67.7              | 54.8  |          |
| Final   | Less 2,000 | < 94.34% <sup>a</sup> | -              | -     | -        | - <sup>b</sup>    | 52.9  | .03      |
|         | 2,000      | 94.34% <sup>a</sup>   | -              | -     |          | - <sup>b</sup>    | 53.6  |          |

Note. <sup>a</sup> average lexical coverage of Episode B2 and Episode A2, <sup>b</sup> participants in No Captions Group watched Episode B2 or Episode A2 as the Final Episode so no percent score given

Another salient finding is the difference in effect sizes (as indicated by  $\eta^2$ ) between the two treatment groups. For the Captions Group, for episodes where there was a significant difference between participants with different levels of lexical coverage as defined by mastery of the 2,000-word level, the effect sizes ranged from .13 to .25. This signifies a medium effect size for one of the episodes and a large effect size for two of the episodes. For the No Captions Group, the effect sizes ranged from .02 to .06 which signifies a small effect size for five episodes and a medium effect size for one episode. The effect sizes for the episodes where there was a significant difference between participants with and without mastery of the 2,000-word level are shown in Table 6.54. This finding suggests that when there is a significant difference between learners with different levels of lexical coverage, the effect of captions leads to larger differences in comprehension scores between the groups. Further research needs to explore how the presence of captions can lead to larger differences in comprehension between learners with differing levels of lexical coverage.

In Study 3, it was proposed that 94% lexical coverage may be the level at which comprehension becomes likely. The results from Study 5 appear to support this. However, the presence of captions appears to reduce the effects of lexical coverage allowing learners with

lower coverage to achieve comprehension scores similar to those of learners with higher lexical coverage. The results from Study 5 also support the finding from Study 3 that the relationship between lexical coverage and comprehension differs from episode to episode of the same television program. The assistance to comprehension that captions provide may also differ from episode to episode. There was a relationship between increased coverage and comprehension for six episodes in Study 3 and three in Study 5 with only two of these episodes in common between the studies.

#### **6.8.3.2. The effects of lexical coverage on incidental vocabulary acquisition**

For both the Captions and No Captions groups increased lexical coverage did not correspond with increased vocabulary acquisition. This is somewhat surprising as it was hypothesized that the availability of captions could allow participants to incidentally acquire more vocabulary because they would have been able to utilize knowledge of the written form of the vocabulary and use knowledge of proximal words to learn unknown words. As in Study 3, participants were able to make vocabulary gains with less than 94% coverage which is lower than the 95% coverage level suggested by Webb and Rodgers (2009a, 2009b). This indicates that, regardless of whether captions are present, learners can make vocabulary gains at a level of lexical coverage lower than what was proposed. Perhaps the reason for this is that images present in television may lessen the potential benefits of increased lexical coverage for incidental vocabulary learning. With no participants in either treatment group reaching the 98% coverage level suggested by Nation (2001) and the small number of coverage points in this study further research on the relationship between lexical coverage and incidental vocabulary learning through viewing episodes of television is needed.

#### **6.8.4. Captions and attitudes**

##### **6.8.4.1. Attitudes towards captions**

The attitudes of the participants in Study 5 towards the presence of captions when language learning through viewing episodes of television are revealed in the responses to the items on the Episode Attitude Survey and the Final Attitude Survey that are specific to this study. Findings from Study 5 support previous research where learners indicated that they believed captions were helpful for language learning (Baltova, 1999; Chung, 1999; Guillory, 1998; Huang & Eskey, 1999; Stewart & Pertusa, 2004; Sydorenko, 2010; Winke et al., 2010). EASQ 5 asked participants about their beliefs regarding the usefulness of the captions for understanding each episode of *Chuck*. Across the 10 episodes viewed, participants indicated

that they believed the captions were at a level slightly above *Pretty Useful* for comprehension. This item had the highest mean response scores on the EAS. There was no significant difference between the mean responses following the first episode and final episode indicating that the value that learners perceived captions to have for comprehension did not diminish as more episodes were viewed and they became more familiar with the series. Five items on the FAS also surveyed the participants' attitudes towards the presence of captions when viewing television. Mean response scores for FASQs 13, 14 and 15 indicated that participants believed that captions were slightly more useful for comprehension than for vocabulary learning or listening ability although the mean responses for all the items were at a level of *Pretty Useful* or higher. On FASQ 16, the majority of participants (78.5%) indicated that they paid attention to the captions from *Very Often* to *Always*. This finding was supported by the results from FAS Forced-choice Item 1 where the large majority of participants (96.1%) indicated that they relied more on the captions than the audio. Overall, the participants in Study 5 felt strongly about the usefulness of the English-language captions for comprehension, vocabulary learning, and listening ability, and paid more attention to the captions than the audio when watching the episodes of *Chuck*.

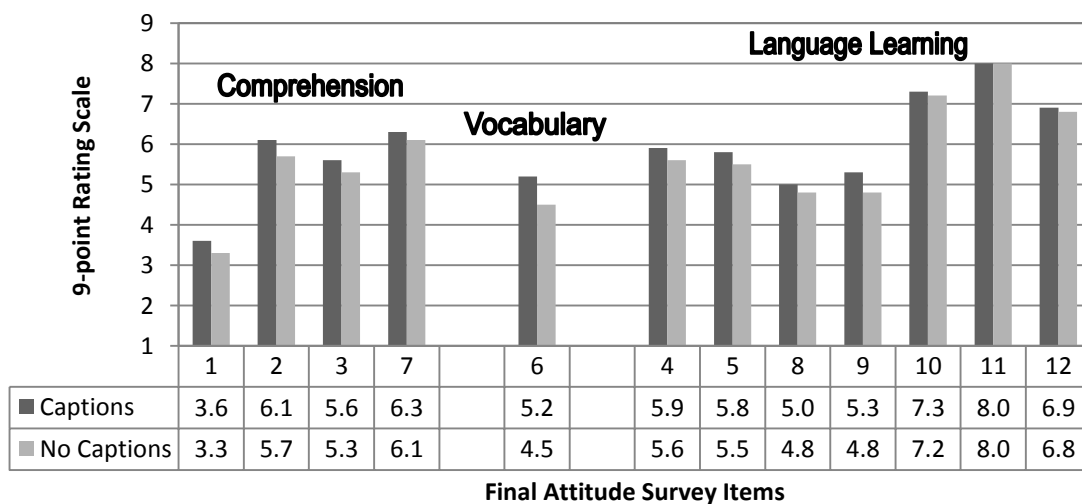
#### **6.8.4.2. Attitudes towards language learning through viewing television**

There are three salient findings that emerge from the comparison of the EAS results from Study 4 and Study 5. First, for the four common EASQs, and across all 10 episodes, the mean responses for the Captions Group were higher or equal to those of the No Captions Group. This indicates that the presence of captions while watching episodes of *Chuck* may have fostered a more positive attitude towards studying English through viewing television in terms of enjoyability of viewing the episodes, comprehension of the episodes, level of learning from the episodes, and usefulness of the episodes for language learning. This supports the findings from previous research that indicated that language learners had more favorable attitudes towards the presence of captions (Baltova, 1999; Chung, 1999; Sydorenko, 2010; Vanderplank, 1988, 1990; Winke et al., 2010). The second finding relates to the initial episode viewed. In this episode, the largest differences between the treatment groups' mean responses for EASQs 1, 2 and 3 (enjoyability of viewing the episode, comprehension of the episode, and level of learning from the episode, respectively) were observed. This indicates that the effects of captions on attitudes might be most observable when learners have not previously seen any episodes or established any background knowledge. The third important finding is the similar pattern of results for the two treatment groups. By examining the trend

lines of the mean responses from the four EASQs across the 10 episodes for both treatment groups (see Figures 6.11, 6.12, 6.13 and 6.14), it is apparent that the response scores generally follow the same pattern. While the mean response scores differ by EASQ, the response patterns between the two treatment groups do not vary in the extreme. This indicates that the participants learning English through watching television had similar experiences which informed their attitudes in similar ways regardless of the presence of captions. While the mean response scores to the four EASQs were shown to trend higher as more episodes were viewed, there are characteristics of individual episodes that affect attitudes towards learning through viewing these episodes regardless of whether or not captions are present.

The mean responses to the 12 FAS items for the Captions and No Captions groups are shown in Figure 6.15 which displays the rating scale items sorted by their common fields of investigation. Across the 12 items, the mean response scores for the Captions Group are higher than or equal to those of the No Captions Group. Analysis, however, indicated that there was no statistical difference between the response scores of participants who viewed the episodes with or without captions for 11 of the 12 FASQs. The Captions Group had significantly higher responses on FASQ 6 which asked participants about their beliefs regarding the effects viewing television had on their English vocabulary. However, results from Experiment 2 in Study 5 indicated no significant difference in incidental vocabulary learning between the participants who viewed the episodes with or without captions.

Figure 6.15 Results for all 12 FASQs for the Captions and No Captions groups



## **6.9. Limitations**

The methodological limitations outlined for the previous four studies in this thesis are also limitations of Study 5. These include: the method of measuring vocabulary knowledge, the selection protocol for the target words on the Tough and Sensitive vocabulary tests, the use of the VLT for estimating lexical coverage, and the lack of open-ended questions and interviews following the attitude survey. Specifically for Study 5, the language proficiency of the participants may have affected the results obtained in the experiments limiting the ability to generalize from the findings. Using the results from the VLT as an indicator of proficiency, it is apparent that there was a paucity of upper-intermediate and advanced language learners. There were very few participants in this study with mastery of the 3,000-word level and no participants with mastery of the 5,000-word level. Replication of the experiments in Study 5 with EFL learners across a wider range of proficiencies may provide a more accurate assessment of the effects of captions on language learning through viewing television.



### **Conclusion**

#### **7. Introduction**

This chapter reviews the results of the five studies that were presented in this thesis and examines how different aspects of language learning were affected by viewing episodes of television. This is followed by brief discussions on the original contributions to the field of learning from video made by this research, the pedagogical utilization of the findings from the studies, and the limitations of this thesis.

#### **7.1. Review of findings**

Study 1 investigated language learners' comprehension of 10 episodes of the television program, *Chuck*. The results showed significant gains in comprehension from the first to the tenth episode viewed with the mean comprehension score increasing approximately 8%. Comprehension across the eight intervening successive episodes was episode dependent with mean scores ranging from 60.2% to 70.8%. Small to moderate significant correlations were found between vocabulary knowledge and comprehension scores for all 10 episodes.

In Study 2, the focus of the research was vocabulary acquisition through viewing episodes of television. Vocabulary gain was measured by two tests measuring form and meaning at differing sensitivities. Results from Study 2 showed that learners were able to gain knowledge of an average of six words which was significantly more than the control group that did not view any of the episodes of television. No significant correlation between vocabulary knowledge and vocabulary gain was found. The frequency of the target words' occurrence in the episodes was found to have a significant medium-size correlation with relative vocabulary gains measured by the Tough Test. However, no significant correlation was found for the gains measured by the Sensitive Test. No significant relationship was found between the target words' range of occurrence across the episodes and their acquisition by the participants. However, a significant correlation was found between the relative frequency of the target words' occurrence and the relative gain of the items on the Tough Test.

Study 3 investigated the effects of lexical coverage on comprehension of television and on vocabulary acquisition through viewing television. Comprehension was shown to improve with increased lexical coverage for some but not all episodes. For 6 of the 10 episodes, participants with 94.21% (average lexical coverage across the six episodes) lexical coverage (mastery of the 2,000-word level plus knowledge of proper nouns and marginal words) were

found to have significantly higher comprehension scores than those participants with lexical coverage less than 94.21%. For one episode, participants with lexical coverage of 96.10% (mastery of the 3,000-word level plus knowledge of proper nouns and marginal words) had significantly higher comprehension scores than those participants with lexical coverage less than 94.72% (no mastery of the 2,000-word level). For two episodes, participants with a mean lexical coverage of 97.58% (mastery of the 5,000-word level plus knowledge of proper nouns and marginal words) had significantly higher comprehension scores than those participants with 94.46% lexical coverage (mastery of the 2,000-word level plus knowledge of proper nouns and marginal words) and participants with less than 94.46% lexical coverage. No significant relationship between incidental vocabulary acquisition and lexical coverage of episodes of television was found.

In Study 4, two surveys examined language learners' attitudes towards learning English through viewing episodes of television. The participants' mean responses on Episode Attitude Survey items that measured attitudes following each episode were all significantly higher following the last episode than following the first episode. These items surveyed participants on their beliefs about their enjoyment of the episode, the usefulness of studying English through viewing the episode, their level of learning from the episode, and their comprehension of the episode. On the Final Attitude Survey that was completed after all the episodes were viewed, the participants were surveyed on their attitudes towards various aspects of language learning through viewing television. Mean responses to these items indicated that the participants thought that learning from television was at least a *Pretty Good Use of Time* and *Pretty Enjoyable*. Participants also indicated that they believed viewing television had at least a *Somewhat Useful* effect on their overall English ability, was at least *Pretty Useful* for language learning in general, and had *Somewhat Improved* their listening skills. Participants also indicated that they thought that their comprehension improved from the first to the last episode viewed and their overall comprehension of the episodes was *Somewhat Good* or better. The participants suggested that they believed their vocabulary was at least *Somewhat Improved* through viewing the episodes. The majority of participants indicated that they had difficulties with the vocabulary, the pronunciation, and the speed of the dialogue in the episodes. Participants were fairly evenly divided about whether or not they had difficulty with the proper names in the series.

The aim of Study 5 was to investigate how the presence of captions affected the aspects of language learning through viewing television that were investigated in Studies 1 to 4. Participants in Study 5 were referred to as the Captions Group and participants from the other

studies that served as comparisons were referred to as the No Captions Group. For comprehension, participants in the Captions Group had higher mean comprehension scores for all 10 episodes of *Chuck* but had significantly higher scores for only three episodes. These episodes were early on in the viewing process or were episodes that might be considered more difficult than other episodes. Gains in comprehension from the first to the last episode were significant for the Captions Group but were significantly less than the gains of the No Captions Group. This was due in part to the significantly higher comprehension scores of the participants with access to captions when viewing the first episode. For the Captions Group, there were small significant correlations between comprehension and vocabulary knowledge for 4 of the 10 episodes while for the No Captions Group there were small to medium significant correlations for all 10 of the episodes.

For vocabulary acquisition, the Captions Group had significant gains through viewing 10 episodes of television for both sensitivities of the vocabulary tests. There was no significant difference, however, between the word gains for the two treatment groups. There were significant small to medium correlations between vocabulary knowledge and relative vocabulary gain on both the vocabulary tests for the Captions Group. However, no significant correlations were found for the No Captions Group.

For the Captions Group, the examination of lexical coverage and comprehension revealed a significant difference in comprehension scores between participants with different levels of lexical coverage for three episodes. In these episodes, participants with a mean lexical coverage of 94.28% (mastery of the 2,000-word level plus knowledge of proper nouns and marginal words) had significantly higher comprehension scores than participants with lexical coverage less than 94.28%. For the No Captions Group, significant differences between the comprehension scores of participants with differing lexical coverage levels was found in six episodes. For both the Captions and No Captions groups, no significant relationship between lexical coverage and vocabulary acquisition was found.

For the investigation of language learners' attitudes towards learning from viewing television, the Captions Group had mean response scores higher than or equal to those of the No Captions Group for the four common items on the Episode Attitude Survey. Comparison of the two treatment groups revealed a significant difference in mean response scores from the first episode to the last episode for two of the four survey items. The No Captions Group had significantly larger gains in mean response scores for the items concerned with the participants' perceived level of comprehension of the episodes and perceived level of learning from the episodes. Following each of the 10 episodes, the Captions Group indicated that they

believed the captions were at least *Pretty Useful* for understanding the episode. On the Final Attitude Survey, the mean responses for the Captions Group were higher than or equal to those from the No Captions Group on all common items but only one of the twelve items had significantly different response scores. The participants' beliefs about the effect that viewing the episodes of *Chuck* had on their English vocabulary were significantly higher for the Captions Group than the No Captions Group.

## **7.2. Original contribution of the research**

The research presented in this thesis offers numerous original contributions to the field of language learning through viewing videos. These include fields of inquiry that have not been previously explored as well as materials and methodologies that have not been employed in prior video-based research.

The unique areas of investigation concerning language learning through video include:

1. Investigation of changes in comprehension from the first to the tenth episode of television viewed
2. Investigation of comprehension of successive episodes of television
3. Examination of the relationship between comprehension of television and vocabulary knowledge
4. Investigation of incidental vocabulary acquisition through viewing multiple full-length episodes of television
5. Analysis of the effects of frequency, range and relative frequency on vocabulary gain through viewing episodes of television
6. Examination of the relationship between lexical coverage and comprehension of video
7. Examination of the relationship between lexical coverage and vocabulary acquisition through viewing video
8. Survey of attitudes towards learning from multiple episodes of television
9. Examination of the effects of captions on language learning through television for the areas of investigation listed in 1 to 8

Many of the materials and methodologies employed in this thesis were also novel to the field of language learning through viewing videos including:

1. Participants viewed 10 episodes of the same television program
2. The total amount of viewing time in the study was over 7 hours
3. The individual videos were over 40 minutes each

4. There was successive viewing of episodes of television
5. Measures of comprehension were based on an explicitly defined comprehension model
6. Comprehension tests were validated with Rasch analysis
7. Participants completed 10 comprehension tests
8. Vocabulary tests at two sensitivities were administered
9. Target words were chosen based on frequency of occurrence
10. Vocabulary test items were created via a rigorous creation procedure
11. Attitudes towards language learning were surveyed after viewing each of 10 episodes

### **7.3. Pedagogical implications**

The findings from this thesis have important implications for language learning through viewing authentic television. The results from the studies indicate the potential benefits to language learners from viewing multiple episodes of television including increasing L2 aural comprehension and vocabulary knowledge. The findings also shed light on learners' attitudes towards viewing television, the minimum lexical coverage learners might have before beginning learning from television, and when it might be advisable to make use of captions. Findings from each of the five studies in this thesis have applicability to utilizing episodes of television for language learning.

In Study 1, participants made gains in comprehension from the first episode to the tenth episode supporting the use of episodes of television programs to increase L2 aural comprehension. This finding indicates that viewing multiple and ideally successive episodes of a television program can lead to comprehension gains. This suggests that it would be useful to inform learners of their potential comprehension gains when they are introduced to the idea of language learning through viewing television. Learners might be more likely to begin learning through television if they understand that their comprehension will improve as they view more episodes. Learners might also be more likely to continue viewing television if they are aware that their comprehension may be lower at first but should improve as they view more episodes. This would serve to counter possible demotivation that could come from low levels of comprehension early in the viewing process. The finding that there was considerable variation in the comprehension scores across the eight successive episodes of *Chuck* also has implications for language learning through viewing television. Characteristics inherent in different episodes of a television program may make some episodes more or less difficult for language learners. At the outset of learning through L2 television, learners would benefit from

knowing that they may understand more of some episodes and less of others. If learners are unaware that they may have comprehension problems with particular episodes, they might be discouraged from watching further episodes when they encounter difficult ones.

The findings from Study 2 indicate that language learners can acquire vocabulary through watching episodes of television supporting Rodgers and Webb's (2011) proposal for utilizing television as an aid to incidental vocabulary learning. In addition to the new form-meaning connections that learners were shown to make, it is not unrealistic to believe that learners also made gains in knowledge of the spoken form of the vocabulary appearing in the episodes as well as strengthening knowledge of words that were already partially known to them. While incidental vocabulary acquisition through viewing television would not be intended to replace explicit vocabulary instruction, it is another potential benefit of language learning through viewing television.

The findings from Study 3 indicate that there is a relationship between lexical coverage and comprehension. Webb and Rodgers (2009a) and Rodgers and Webb (2011) suggested that a vocabulary size of 3,000 word families plus proper nouns and marginal words, or 95% lexical coverage, may be sufficient for adequate comprehension. Findings from this study showed that comprehension is more likely when learners have mastery of the first 2,000 words of English. This level of vocabulary knowledge (plus knowledge of proper nouns and marginal words) provided a mean lexical coverage of over 94% across the 10 episodes viewed. This level of coverage would be a good benchmark for when learners could begin to explore language learning through viewing television. Therefore, it would be advisable to measure learners' vocabulary knowledge before beginning to view episodes of television to ensure that they have reached the minimum level of vocabulary knowledge recommended.

Findings from Study 4 revealed that the participants had generally positive attitudes towards language learning through viewing television. This is important because if learners do not have a positive reaction to this study method it may be difficult to convince them to start or continue viewing episodes of television outside the classroom. Language learners who do not perceive value in a pedagogical method may not be willing to participate in instruction based on that method (Dörnyei & Ushioda, 2011). The mean responses on items that surveyed attitudes towards learning through television trended higher as participants viewed more episodes. This suggests that an approach that involves watching multiple rather than single episodes of television might be most effective. Learners who view multiple episodes of a program may be more apt to continue learning through viewing television as their attitudes towards language learning through television become more positive.

The results from Study 5 concerning the presence of captions when viewing television may also have a bearing on learning through viewing episodes of television. The most salient effect of the availability of captions was increased comprehension of episodes near the beginning of the viewing process and for episodes where comprehension may be more difficult. Consequently, language learners should be encouraged to make use of captions when they first begin viewing a series that is new to them and when they believe they are having comprehension problems with later episodes. Supporting comprehension in these ways would be particularly useful for television series with interrelated episodes where failure to fully comprehend one episode may lead to comprehension problems in subsequent episodes.

As a whole, the results from this thesis may have applicability in an extensive viewing program. Webb and Rodgers (2009a) and Rodgers and Webb (2011) suggested that regular viewing of related episodes of television may result in opportunities to gain vocabulary knowledge and increase comprehension. This has been supported by findings from the studies presented in this thesis. Regular viewing of television would form the basis of extensive viewing. Like in extensive reading programs where language learners choose and read books as a source of L2 input, learners in extensive viewing programs would be encouraged to choose and watch episodes of television that appeal to them. Ideally, learners would have access to complete seasons of television across a wide range of genres to allow for differing tastes and interests. Authentic L2 aural input obtained in this manner may be particularly valuable in EFL environments where sources of L2 listening input are less accessible.

In summary, findings from the studies in this thesis have applicability for establishing language learning programs based on viewing episodes of television as well as indicating the potential benefits to learners participating in these programs. However, research on learning through viewing television is limited to these studies and more research needs to be undertaken.

#### **7.4. Are episodes of television a suitable source of input for language learning?**

In the introduction to this thesis, there was an examination of the five conditions that Nation (2007) maintains must be met in order for input to be considered suitable for inclusion in a language course. The suitability of episodes of television to serve as L2 aural input was considered by comparing the nature of television and past research from language learning through videos to the five conditions. Only the first condition, that the input needs to be processed in large quantities, was shown to be met by what is known about learning from television. For the remaining four conditions, it was hypothesized that episodes of television

could potentially fulfill the requisites but the appropriate previous research to support this was lacking. It was also posited that the presence of captions when viewing episodes of television may create an environment more favorable to satisfying the conditions. The research undertaken in this thesis has advanced what is known about the efficacy of television to meet the requirements necessary for input to be considered suitable for inclusion in a language course.

The second condition for suitable input is that it should be familiar to the language learners (Nation, 2007) so that they can achieve a reasonable level of comprehension. The results from Study 1 of this thesis showed comprehension increased as language learners build familiarity with the television program by viewing multiple and successive episodes. Learners had a mean comprehension score of 63% across the ten episodes they viewed. In this way, episodes of television appear to satisfy the second condition although further research on the level of comprehension that best serves language learning is needed.

The third condition is that learners should be able to gain knowledge of vocabulary using context cues and background knowledge (Nation, 2007). The learners in Study 2 were shown to be able to make gains of an average of 6 words or approximately 25% of the target words that were available to be learned. This finding indicates that learners may have been able to utilize contextual and accumulated knowledge when viewing episodes of television to acquire knowledge of previously unknown vocabulary items thereby satisfying the third condition.

The fourth condition that Nation (2007) proposed is that only a small percentage of the vocabulary in the input texts should be unknown to learners because if a large proportion of the vocabulary were unknown comprehension and vocabulary learning could be impeded. The findings from Study 3 indicated that for the majority of episodes of television, comprehension improved with increased lexical coverage. For vocabulary learning, no significant relationship between incidental vocabulary acquisition and lexical coverage of episodes of television was found. The findings from this study support the supposition that the less unknown vocabulary there is the better comprehension is likely to be. However, the participants' lexical coverage of the episodes was relatively low compared to what was hypothesized from previous research as necessary for reasonable comprehension and vocabulary learning. Further research needs to be undertaken with participants with greater lexical coverage of episodes of television to ascertain what the percentage of unknown words is optimal for learning from television is in terms of both comprehension and vocabulary learning.

The fifth condition that must be met in order for input to be considered suitable for inclusion in language programs is that learners should be interested in the input and want to

understand it (Nation, 2007). The participants in Study 4 indicated that they found learning through viewing episodes of television to be, on average, *Pretty Enjoyable*. Learners' enjoyment of individual episodes also increased from the first to the last episode viewed. Because the learners expressed a level of enjoyment when viewing the episodes of television, it may be possible to say that they made efforts to understand what they were viewing. This is supported by the results from Study 1 showing comprehension increased from the first to the last episode. Through these findings, episodes of television appear to satisfy the fifth condition for the suitability of L2 input.

The effect that captions had on making episodes of television more suitable as a source of input was most salient concerning the second condition. The presence of captions when viewing episodes of television was shown to improve comprehension early in the viewing process and for episodes that may be considered more difficult. In this way, providing learners with captions when beginning to view a television series when the content of the episodes is likely to be most unfamiliar may make television a more suitable source of L2 aural input.

The research undertaken in this thesis advances what is known about the potential of episodes of authentic television to provide input suitable for inclusion in an EFL language program. However, there is a need for further research concerning how best to make use of episodes of authentic television as source of L2 aural input.

## **7.5. Limitations and future research**

Limitations of this research specific to the five individual studies in this thesis are discussed in detail in the chapters for each of the studies. There are, however, five limitations that apply to the thesis as a whole. First, the participants in this research viewed only episodes from a single television program. It is possible that the results of the studies may have been different if another television program had been utilized. For example, it is unknown whether participants would have had similar levels of comprehension or vocabulary gains had they viewed another program with attributes similar to *Chuck*. The second limitation is related to the successive nature of the episodes viewed between the first and the last episode. While it is hypothesized that the successive episodes lead to greater gains in background knowledge it is uncertain whether participants may have made similar gains to comprehension and vocabulary through viewing random episodes from the same series or even random episodes of different programs. The third limitation of this research is the running time of the episodes. Authentic American television programs are, by and large, produced at two lengths: 22 minutes or 42

minutes. Using episodes with 22-minute running times may have led to different results. The reduced length may have held the participants' attention better leading to better comprehension scores and better attitudes towards the viewing process. The fourth limitation is the number of episodes viewed in this research. Participants in these studies viewed 10 episodes from the first season of *Chuck*. It remains unclear what the effect of decreasing or increasing the number of episodes viewed would have on comprehension gain and vocabulary learning. A final limitation is the language of the television program used in the studies. It is unclear whether the findings from viewing English-language television relate to second language learning through viewing television in other languages. While the research undertaken in this thesis has advanced our understanding of how television might be used as a tool for language learning, there are still many aspects of language learning from authentic television that have yet to be researched.

## **7.6. Conclusion**

The purpose of this thesis has been to investigate multiple aspects of language learning through viewing television. The experimental design expanded upon earlier methodologies by employing multiple full-length episodes of television intended for an English speaking audience. The findings provided insight into the following aspects of language learning:

1. Comprehension of the episodes of television
2. Vocabulary acquisition through viewing the episodes
3. Lexical coverage and its effects on comprehension of television and vocabulary acquisition from viewing television
4. Attitudes towards language learning through viewing television
5. Captions and their effects on the aspects of language learning described in 1 to 4

Taken as a whole, this thesis shows the value of using authentic episodes of television for language learning. Since this is one of the first studies that has investigated how comprehension, vocabulary acquisition, lexical coverage, attitudes and captions relate to language learning through viewing multiple episodes of television, there is much that remains to be researched. Hopefully, future research will expand on the findings from this thesis and examine more aspects of language learning through viewing television.

## References

- Alderson, J., Percsich, R., & Szabo, G. (2000). Sequencing as an item type. *Language Testing*, 17(4), 423–447.
- Allison, P. D. (2001). *Missing data*. Thousand Oaks, CA: Sage.
- Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language, Learning & Technology*, 5(1), 202–232.
- Alwin, D. F., & Krosnick, J. A. (1991). The reliability of survey attitude measurement: The influence of question and respondent attributes. *Sociological Methods & Research*, 20(1), 139–181.
- Avila, E., & Sadoski, M. (1996). Exploring new applications of the keyword method to acquire English vocabulary. *Language Learning*, 46(3), 379–395.
- Baltova, I. (1994). The impact of video on the comprehension skills of core French students. *Canadian Modern Language Review*, 50(3), 507–531.
- Baltova, I. (1999). Multisensory language teaching in a multidimensional curriculum: The use of authentic bimodal video in core French. *Canadian Modern Language Review*, 56(1), 31–48.
- Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography*, 64(4), 253–279.
- Bennett-Jones, P. (Producer). (1990). *Mr. Bean* [Television Series]. United Kingdom: ITV.
- Bianchi, F., & Ciabattini, T. (2008). Captions and subtitles in EFL learning: An investigative study in a comprehensive computer environment. In A. Baldry, M. Pavesi, C. Taylor Torsello, & C. Taylor (Eds.), *From didactas to ecolingua* (pp. 69–80). Trieste: Edizioni Università di Trieste.
- 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.
- Bond, T., & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah: Lawrence Erlbaum.
- Bonk, W. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening*, 14(1), 14–31.
- Borrás, I., & Lafayette, R. C. (1994). Effects of multimedia courseware subtitling on the speaking performance of college students of French. *The Modern Language Journal*, 78(1), 61–75.

- Brett, P. (1995). Multimedia for listening comprehension: The design of a multimedia-based resource for developing listening skills. *System*, 23(1), 77–85.
- Brett, P. (1997). A comparative study of the effects of the use of multimedia on listening comprehension. *System*, 25(1), 39–53.
- Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics*, 18(1), 171–191.
- Brown, J. D. (2001). *Using surveys in language programs*. Cambridge: Cambridge University Press.
- Brown, J. I. (1949). The construction of a diagnostic test of listening comprehension. *The Journal of Experimental Educational*, 18(2), 139–146.
- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, 20(2), 136–163.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Reading Behavior*, 26(4), 413–437.
- Chai, J., & Erlam, R. (2008). The effect and the influence of the use of video and captions on second language learning. *New Zealand Studies in Applied Linguistics*, 14(2), 25–44.
- Chang, A. C. S., & Read, J. (2006). The effects of listening support on the listening performance of EFL learners. *TESOL Quarterly*, 40(2), 375–397.
- Chiang, C. S., & Dunkel, P. (1992). The effect of speech modification, prior knowledge, and listening proficiency on EFL lecture learning. *TESOL Quarterly*, 26(2), 345–374.
- Chung, J. M. (1999). The effects of using video texts supported with advance organizers and captions on Chinese college students' listening comprehension: An empirical study. *Foreign Language Annals*, 32(3), 295–308.
- Chung, J. M., & Huang, S. C. (1998). The effects of three aural advance organizers for video viewing in a foreign language classroom. *System*, 26(4), 553–565.
- Conrad, L. (1985). Semantic versus syntactic cues in listening comprehension. *Studies in Second Language Acquisition*, 7(1), 59–69.
- Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta: Translators' Journal*, 49(1), 67–77.
- Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge: Cambridge University Press.

- Day, R. R., Omura, C., & Hiramatsu, M. (1992). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language*, 7(2), 541–551.
- Dörnyei, Z. (2003). Attitudes, orientations, and motivations in language learning: Advances in theory, research, and applications. *Language Learning*, 53(1), 3–32.
- Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation* (2nd ed.). Harlow: Pearson Education.
- Driscoll, R., Atkinson, R., Clifford, A. (Writers), & Birkin, J. (Director). (1995). Tee Off, Mr. Bean [Television series episode]. In P. Bennett-Jones (Producer), *Mr. Bean*. United Kingdom: ITV.
- Dubois, M., & Vial, I. (2000). Multimedia design: The effects of relating multimodal information. *Journal of Computer Assisted Learning*, 16(2), 157–165.
- Dunkel, P., Henning, G., & Chaudron, C. (1993). The assessment of an L2 listening comprehension construct: A tentative model for test specification and development. *The Modern Language Journal*, 77(2), 180–191.
- Duquette, L., & Painchaud, G. (1996). A comparison of vocabulary acquisition in audio and video contexts. *Canadian Modern Language Review*, 53(1), 143–72.
- Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, 24(2), 174–187.
- Ellis, N.C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24(2), 143–188.
- Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27(2), 305–352.
- Etemadi, A. (2012). Effects of bimodal subtitling of English movies on content comprehension and vocabulary recognition. *International Journal of English Linguistics*, 2(1), 239–248.
- Field, J. (1998). Skills and strategies: Towards a new methodology for listening. *ELT Journal*, 52(2), 110–118.
- Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, 40(2), 273–293.
- Freedle, R., & Kostin, I. (1999). Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language Testing*, 16(1), 2–32.
- Frisbie, D. A., & Becker, D. F. (1991). An analysis of textbook advice about true-false tests. *Applied Measurement in Education*, 4(1), 67–83.

- Gardner, D. (2008). Vocabulary recycling in children's authentic reading materials: A corpus-based investigation of narrow reading. *Reading in a Foreign Language*, 20(1).
- 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). Incidental vocabulary learning. *Studies in Second Language Acquisition*, 21(2), 319–333.
- Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28(1), 55–75.
- Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly*, 25(3), 375–406.
- Gravic. (2007). *Remark Office OMR (Version 6)* [Computer software]. Malvern: Gravic.
- Gruba, P. (2004). Understanding digitized second language videotext. *Computer Assisted Language Learning*, 17(1), 51–82.
- Gruba, P. (2006). Playing the videotext: A media literacy perspective on video-mediated L2 listening. *Language Learning & Technology*, 10(2), 77–92.
- Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System*, 36(1), 85–93.
- Guillory, H. G. (1998). The effects of keyword captions to authentic French video on learner comprehension. *Calico Journal*, 15(1-3), 89–108.
- Haladyna, T. M. (1992). The effectiveness of several multiple-choice formats. *Applied Measurement in Education*, 5(1), 73–88.
- Haladyna, T. M., & Downing, S. M. (1993). How many options is enough for a multiple-choice test item? *Educational and Psychological Measurement*, 53(4), 999–1010.
- Haladyna, T. M., Downing, S., & Rodriguez, M. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–334.
- Hansen, C., & Jensen, C. (1994). Evaluating lecture comprehension. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 241–268). Cambridge: Cambridge University Press.
- Hasan, A. S. (2000). Learners' perceptions of listening comprehension problems. *Language, Culture and Curriculum*, 13(2), 137–153.
- Hatch, E. M., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House Publishers.

- Herron, C., Cole, S. P., Corrie, C., & Dubriel, S. (1999). The effectiveness of a video-based curriculum in teaching culture. *The Modern Language Journal*, 83(4), 518–533.
- Herron, C., York, H., Cole, S. P., & Linden, P. (1998). A comparison study of student retention of foreign language video: Declarative versus interrogative advance organizer. *The Modern Language Journal*, 82(2), 237–247.
- Hirai, A. (1999). The relationship between listening and reading rates of Japanese EFL learners. *The Modern Language Journal*, 83(3), 367–384.
- Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 698–696.
- Hopkins, W. G. (2010). Linear models and effect magnitudes for research, clinical and practical applications. *Sportscience*, 14, 49–57.
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review*, 61(3), 355–382.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207–223.
- Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403–430.
- Huang, H. C., & 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(1), 75–96.
- Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, 21(2), 181–193.
- Hui, W. (2007). The effects of captions on Chinese EFL students' incidental vocabulary acquisition. *Media in Foreign Language Instruction*, 30(4), 9–16.
- Hulstijn, J. H. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge: Cambridge University Press.
- In'nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing*, 26(2), 219–244.
- Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, 21(4), 767–787.

- Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal*, 86(4), 546–561.
- Kelly, P. (1991). Lexical ignorance: The main obstacle to listening comprehension with advanced foreign language learners. *International Review of Applied Linguistics in Language Teaching*, 29(2), 135–149.
- Kerr, A. W., Hall, H. K., & Kozub, S. A. (2002). *Doing statistics with SPSS*. London: Sage Publications.
- Kipfer, B. A. (2003). *Roget's Thesaurus, 6th Edition*. New York: Harper.
- Klockars, A. J., & Yamagishi, M. (1988). The influence of labels and positions in rating scales. *Journal of Educational Measurement*, 25(2), 85–96.
- Kobeleva, P. P. (2012). Second language listening and unfamiliar proper names: Comprehension barrier? *RELC Journal*, 43(1), 83–98.
- Koolstra, C. M., & Beentjes, J. W. J. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development*, 47(1), 51–60.
- Kothari, B., Pandey, A., & Chudgar, A. (2004). Reading out of the “idiot box”: Same-language subtitling on television in India. *Information Technologies and International Development*, 2(1), 23–44.
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. New York: Longman.
- Kreiter, C. D., & Frisbie, D. A. (1989). Effectiveness of multiple true-false items. *Applied Measurement in Education*, 2(3), 207–216.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50(1), 537–567.
- Kucera, H., & Francis, W. N. (1967). *A computational analysis of present day American English*. Providence: Brown University Press.
- Kweon, S. O., & Kim, H. R. (2008). Beyond raw frequency: Incidental vocabulary acquisition in extensive reading. *Reading in a Foreign Language*, 20(2).
- Latifi, M., Mobalegh, A., & Mohammadi, E. (2011). Movie subtitles and the improvement of listening comprehension ability: Does it help? *The Journal of Language Teaching and Learning*, 1(2), 18–29.
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Clevedon: Multilingual Matters.

- Laufer, B. (2001). Quantitative evaluation of vocabulary: How it can be done and what it is good for. In Elder, C., McNamara, T., Lumley, T., Grove, L., Iwashita, N., Brown, A., & Hill, K. (Eds.), *Experimenting with uncertainty* (pp. 241–250). Cambridge: Cambridge University Press.
- Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, 59(4), 567–587.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26.
- Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48(3), 365–391.
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30.
- Laufer, B., & Sim, D. D. (1985). Measuring and explaining the reading threshold needed for English for academic purposes texts. *Foreign Language Annals*, 18(5), 405–411.
- Linacre, M. (2010). *Practical Rasch Measurement - Core Topics* [Course notes]. Retrieved from statistics.com website: [www.statistics.com](http://www.statistics.com).
- Liu, N., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELJ Journal*, 16(1), 33–42.
- Long, D. R. (1990). What you don't know can't help you. *Studies in Second Language Acquisition*, 12(1), 65–80.
- Lund, R. J. (1991). A comparison of second language listening and reading comprehension. *The Modern Language Journal*, 75(2), 196–204.
- Maleki, A., & Safaee Rad, M. (2011). The effect of visual and textual accompaniments to verbal stimuli on the listening comprehension test performance of Iranian high and low proficient EFL learners. *Theory and Practice in Language Studies*, 1(1), 28–36.
- Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals*, 32(3), 321–328.
- Markham, P., & Peter, L. A. (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., Peter, L. A., & McCarthy, T. J. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals*, 34(5), 439–445.
- Mayer, R. E., & 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.
- McFadden, K., Barret, K., & Horst, M. (2009). What's in a television word list? A corpus-informed investigation. *Concordia Working Papers in Applied Linguistics*, 2, 78–98.
- Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, & M. Torreblanca-López (Eds.), *Further insights into nonnative vocabulary teaching and learning* (pp. 83–97). Bristol: Multilingual Matters.
- Mittelstaedt, R. A. (1971). Semantic properties of selected evaluative adjectives: Other evidence. *Journal of Marketing Research*, 8(2), 236–237.
- Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE*, 4(11), 1–5.
- Mueller, G. A. (1980). Visual contextual cues and listening comprehension: An experiment. *The Modern Language Journal*, 64(3), 335–340.
- Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20(2), 233–253.
- Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, 82(4), 591–605.
- Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, 37(4), 645–670.
- Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, 5, 12–25.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston: Newbury House.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 3–13). Amsterdam: John Benjamins.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59–82.

- Nation, I. S. P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching*, 1(1), 1–12.
- Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle.
- 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(1), 95–106.
- Nunan, D. (2002). Listening in language learning. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 238–241). Cambridge: Cambridge University Press.
- O’Muirheartaigh, C. A., Krosnick, J. A., & Helic, A. (2001). *Middle alternatives, acquiescence, and the quality of questionnaire data*. Chicago: University of Chicago.
- Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing*, 24(4), 517–537.
- OECD. (2009). *Communications Outlook 2009*. Paris: OECD.
- OECD. (2011). *Communications Outlook 2011*. Paris: OECD.
- Oetting, J. B., Rice, M. L., & Swank, L. K. (1995). Quick incidental learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech and Hearing Research*, 38(2), 434–445.
- Ohtsuka, K., & Brewer, W. F. (1992). Discourse organization in the comprehension of temporal order in narrative texts. *Discourse Processes*, 15(3), 317–336.
- Paivio, A. (1990). *Mental representations: A dual coding approach*. New York: Oxford University Press.
- Paribakht, T. S., & Wesche, M. (1999). Reading and “incidental” L2 vocabulary acquisition. *Studies in Second Language Acquisition*, 21(2), 195–224.
- Park, G. P. (2004). Comparison of L2 listening and reading comprehension by university students learning English in Korea. *Foreign Language Annals*, 37(3), 448–458.
- Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart? *Reading in a Foreign Language*, 22(1), 31–55.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18(1), 1–28.
- Pohl, N. F. (1981). Scale considerations in using vague quantifiers. *The Journal of Experimental Educational*, 49(4), 235–240.
- Rice, M. L., & Woodsmall, L. (1988). Lessons from television: Children’s word learning when viewing. *Child Development*, 59(2), 420–429.

- Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, 17(2), 219–240.
- Richards, J. C. (1990). *The language teaching matrix*. Cambridge: Cambridge University Press.
- Rodgers, M. P. H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, 45(4), 689–717.
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2), 3–13.
- Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, 21(4), 589–619.
- Rubin, J. (1994). A review of second language listening comprehension research. *Modern Language Journal*, 78(2), 199–221.
- Sadighi, F., & Zare, S. (2006). Is listening comprehension influenced by the background knowledge of the learners? A case study of Iranian EFL learners. *The Linguistics Journal*, 1(3), 110–126.
- Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, 6(2), 72–78.
- Schaeffer, N. C., & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology*, 29, 65–88.
- Schmidt-Rinehart, B. C. (1994). The effects of topic familiarity on second language listening comprehension. *The Modern Language Journal*, 78(2), 179–189.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26–43.
- Schmitt, N., & Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 1(1), 1–20.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–88.

- Schwartz, J., Fedak, C., & McG. (Producers). (2007). *Chuck* [Television series]. Los Angeles: NBC.
- Seels, B., Fullerton, K., Berry, L., & Horn, L. J. (2004). Research on learning from television. In *Handbook of research on educational communications and technology* (pp. 249–334). Mahwah: Lawrence Erlbaum.
- Shefelbine, J. L. (1990). Student factors related to variability in learning word meanings from context. *Journal of Literacy Research*, 22(1), 71–97.
- Sherman, J. (1997). The effect of question preview in listening comprehension tests. *Language Testing*, 14(2), 185–213.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 1(2), 147–170.
- Shohamy, E., & Inbar, O. (1991). Validation of listening comprehension tests: The effect of text and question type. *Language Testing*, 8(1), 23.
- Smith, E. V. (2000). Metric development and score reporting in Rasch measurement. *Journal of Applied Measurement*, 1(13), 303–326.
- Song, M. Y. (2008). Do divisible subskills exist in second language (L2) comprehension? A structural equation modeling approach. *Language Testing*, 25(4), 435–464.
- Spaan, M. (2007). Evolution of a test item. *Language Assessment Quarterly*, 4(3), 279–293.
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, 36(2), 139–152.
- Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a Foreign Language. *Studies in Second Language Acquisition*, 31(4), 577–607.
- Stahl, S. A., Jacobson, M. G., Davis, C. E., & Davis, R. L. (1989). Prior knowledge and difficult vocabulary in the comprehension of unfamiliar text. *Reading Research Quarterly*, 24(1), 27–43.
- Stewart, M. A., & Pertusa, I. (2004). Gains to language learners from viewing target language closed-captioned films. *Foreign Language Annals*, 37(3), 438–447.
- Sudman, S., & Bradburn, N. M. (1982). *Asking questions*. San Francisco: Jossey-Bass.
- Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55(4), 661–699.
- 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.

- Thorndike, E. L., & Lorge, I. (1944). *The Teacher's Word Book of 30 000 Words*. New York: Columbia University Teacher's College.
- van Zeeland, H., & Schmitt, N. (2012). Lexical coverage and L1 and L2 listening comprehension: The same or different from reading comprehension. *Applied Linguistics*.
- Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics*, 24(1), 3–25.
- Vandergrift, L. (2007). Recent developments in second language and foreign language listening comprehension research. *Language Teaching*, 40(3), 191–210.
- Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT Journal*, 42(4), 272–281.
- Vanderplank, R. (1990). Paying attention to the words: Practical and theoretical problems in watching television programmes with uni-lingual (CEEFAQ) sub-titles. *System*, 18(2), 221–34.
- Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching*, 43(1), 1–37.
- Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, 24(1), 56–89.
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1), 219–258.
- Wagner, E. (2002). Video listening tests: A pilot study. *Columbia University Working Papers in TESOL & Applied Linguistics*, 2(1), 1–39.
- Wang, Y. C. (2012). Learning L2 vocabulary with American TV drama: From the learner's perspective. *English Language Teaching*, 5(8), 217–225.
- Waring, R., & Nation, I. (2004). Second language reading and incidental vocabulary learning. *Angles on the English Speaking World*, 4, 97–110.
- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader. *Reading in a Foreign language*, 15(2), 130–163.
- Weaver, C. (2005). How entrance examination scores can inform more than just admission decisions. In P. Ross, T. Newfields, Y. Ishida, M. Chapman, & M. Kawate-Mierzejewska (Eds.), *Lifelong Learning: Proceedings of the 4th Annual JALT Pan-SIG Conference* (pp. 114–121). Tokyo: Tokyo Kezai University.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46–65.

- Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, 20(2), 232–245.
- Webb, S. (2011). Selecting television programs for language learning: Investigating television programs from the same genre. *International Journal of English Studies*, 11(1), 117–136.
- Webb, S., & Rodgers, M. P. H. (2009a). Vocabulary demands of television programs. *Language Learning*, 59(2), 335–366.
- Webb, S., & Rodgers, M. P. H. (2009b). The lexical coverage of movies. *Applied Linguistics*, 30(4), 407–427.
- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *Canadian Modern Language Review*, 53(1), 13–40.
- West, M. (1953). *A general service list of English words*. London: Longman.
- Winke, P., Gass, S., & Sydorenko, T. (in press). Factors influencing the reading of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal*.
- Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, 14(1), 65–86.
- Wolff, D. (1987). Some assumptions about second language text comprehension. *Studies in Second Language Acquisition*, 9(3), 307–326.
- Wright, B. D. (1997). A history of social science measurement. *Educational Measurement: Issues and Practice*, 16(4), 33–45.
- Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of Vocabulary Levels Tests. *System*, 35(2), 182–191.
- Yi'an, W. (1998). What do tests of listening comprehension test? - A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing*, 15(1), 21–44.
- Yoshino, S., Kano, N., & Akahori, K. (2000). The effects of English and Japanese captions on the listening comprehension of Japanese EFL students. *Language Laboratory*, 37, 111–130.
- Yuksel, D., & Tanriverdi, B. (2009). Effects of watching captioned movie clip on vocabulary development of EFL learners. *The Turkish Online Journal of Educational Technology*, 8(2), 48–54.
- Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review*, 57(4), 541–572.



## *Notes*

<sup>i</sup> Ostensibly, the results of the Sensitive Test indicate that there may be a ceiling effect for some of the participants in Study 2. In particular, the fact that the mean number of words known on the Sensitive Test was 37.2 out of 60 (62% of the target words) and the fact that the maximum score was 53 out of 60 both point to a potential ceiling effect. However, further examination of the results from the Sensitive Test reveals a number of counterarguments to the possibility of a ceiling effect on this test. First, the single participant who obtained the maximum score of 53 on the Sensitive Test gained knowledge of 2 words through viewing the episodes of television. This indicates that there were still target words that he or she could have gained knowledge of (5 target words remained unlearned for this participant). Second, there were still words available to be learned and measured by the Sensitive Test, even for the most proficient participants. An examination of the results from the learners who knew at least 40 of the target words was undertaken to ascertain if any of these participants were affected by a potential ceiling effect. The average score of these 68 participants was 48.6 and their average gain was 5.1 words. The maximum score for these participants on the Sensitive Post-Test was 55 out of 60. While there may have been less vocabulary acquisition opportunities for these learners, it is believed that this was controlled for by expressing the results of the vocabulary tests as relative gains. In summary, the results indicate there was likely not a true ceiling effect for the Sensitive Test but did approach one for some learners.