

**Incidental and intentional acquisition of multiword expressions  
from audio-visual input: The effects of typographically enhanced  
captions and repetition**

by

Elvenna Majuddin

A thesis submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Doctor of Philosophy

Victoria University of Wellington

2020



## **Abstract**

This research project aims to extend the line of inquiry on pedagogical interventions intended to help second language (L2) learners make better progress of their mastery of multiword expressions (MWEs). Existing studies on these interventions revealed a propensity towards exclusivity in terms of input modality, item type and learning condition. Firstly, there are far more MWE studies in the context of unimodal input, e.g., written input. It is only recently that the potential of audio-visual input (i.e., L2 viewing) has been explored for MWE learning. Secondly, previous studies have by and large focused on certain types of MWEs, such as collocations. While there is merit in focusing on a certain type of item, such studies do not represent the materials that L2 learners are often exposed to. Further, authentic videos entail diverse MWE types, providing a stronger reason to include more than one type of target item. Thirdly, many MWE interventions are investigated exclusively under one of the learning conditions, i.e., intentional or incidental learning conditions. Hulstijn's (2001) criterion is adopted to distinguish these two learning conditions, in that the presence of test announcement characterises the intentional learning condition. Due to this tendency towards a dichotomy of learning conditions, many factors known to facilitate MWE learning have been investigated under one of the learning conditions only.

Two such factors are repetition and typographic enhancement. While repetition is well established as beneficial for MWE acquisition, evidence for this is mainly furnished by studies on incidental learning through written input. Therefore, the aim of this research project is to assess how repetition, operationalised as repeated viewing, influences MWE acquisition under both learning conditions. Similarly, although typographic enhancement has been shown to draw learners' attention and promote MWE uptake, this positive evidence is mostly observed in incidental learning studies. As such, whether typographically-enhanced MWEs are indeed learned better than unenhanced MWEs under intentional learning conditions is still under-researched. Importantly, whether typographic enhancement in captioned viewing leads to superior learning compared to normal captions is unknown. This is one of the aims of the research project, in which different caption conditions are created to explore their effectiveness in facilitating MWE learning. Of further interest is whether MWE learning under different caption conditions would modulate the effect of repetition. This is motivated by the assumption that typographic enhancement might eliminate the need for repetition.

To answer the research questions, two studies differentiated by the presence of test announcement were carried out. For both studies, ESL learners watched a video containing

target MWEs under one of six conditions, which differed in terms of caption condition (no captions, normal captions or enhanced captions) and the number of viewing times (once or twice). MWE learning was assessed through tests that tap into form and meaning knowledge at the level of recall and recognition. Though not part of the research questions, the effects of caption condition and repetition on content comprehension were also assessed. The findings of both studies revealed trends that are consistent with literature on MWE learning and vocabulary learning in general. Firstly, both types of captions promoted better form recall knowledge compared to uncaptioned viewing. This was found to be true under both incidental and intentional learning conditions. Secondly, typographically enhanced captions led to better form recall compared to normal captions, but only under the intentional learning conditions. Under the incidental learning conditions, the effects of L2 viewing with typographically enhanced captions on form recall appeared to be similar to viewing with normal captions. The findings also suggest that the presence of typographically enhanced captions reduced the number of viewings needed to make incidental gains in form recall knowledge. In addition, while repeated viewing under all caption conditions led to better knowledge of form under the incidental learning conditions, the effect of repetition was not found under the intentional learning conditions. This aligns well with the supposition that fewer repetitions are needed for intentional learning. Thirdly, neither repetition nor caption condition had an effect on the acquisition of MWE meanings under both learning conditions. Finally, vocabulary knowledge played a significant role in the amount of MWE learning that takes place, especially so when learners were not forewarned of MWE tests. Taken as a whole, the findings of this research project support the use of captions for L2 viewing as a way to foster MWE acquisition, at least at the level of form acquisition. The use of typographically enhanced captions, however, may have adverse effects on content comprehension. As such, the findings of this research project have meaningful implications concerning when typographically enhanced captions and repeated viewing should be used to optimise MWE learning through L2 viewing.

## **Acknowledgements**

This research project is not only a product of my mental and physical labour, it is also a testament to the strong support system that I have, without which I would have never been able to see this through to completion. First and foremost, I would like to thank my supervisors Dr. Anna Siyanova and Professor Frank Boers, for their patience, guidance and wisdom. Anya never failed to offer me insightful advice and words of encouragement in every meeting, and her confidence in me made me a better researcher. Frank's physical absence went almost unnoticed, as he was with me at every step of the journey, providing astute comments and recommendations to improve the quality of my work. I am eternally grateful to Frank for staying on as my supervisor, even after leaving Victoria University of Wellington (VUW).

I would like to express my gratitude to Lisa Woods as she was the reason I am able to perform the statistical analyses required for my data. Thank you for always making time to see me and for providing me with patient and indispensable guidance. I am also grateful to VUW for the Victoria Doctoral Scholarship and the research grants, which gave me the opportunity to present my research at overseas conferences. My thanks also go to Kate O'Malley, who provided proof-reading services in accordance with the Editorial Advice Policy of VUW. I am also thankful to my examiners, Dr. Irina Elgort, Associate Professor David Hirsh and Associate Professor Elke Peters, for their comments and recommendations to further improve the quality of my thesis.

I would also like to thank Ms. Chu Went Tyng and the rest of the academic and administration staff of Tunku Abdul Rahman University College (TARUC Sabah branch), for allowing me to return for data collection, and for providing me with all the help I needed to run my experiments smoothly. My thanks also go to the TARUC students who participated in my studies and made their schedules more flexible to fit in time for my experiments.

I am also grateful to my friends and LALS PhD colleagues who spared time to help improve my test materials and rate the many test papers. A special thanks go to the members of the LALS vocabulary group for being my sounding board. Thank you also to my fellow PhD friends, especially Mengzhu and Pansa, for the warmth of our friendship and for always motivating me to keep to my goals and timeline. To my close friends, Rishun, Erlind, Lany, Wini, Yi Yin, Jordan and Matthew, you have helped me in more ways than you know, and, for that, I am forever indebted.

Finally, my deepest gratitude goes to my family as I drew my greatest strength from them. I thank my mother, Latihim, for her unwavering support and her unconditional love. I thank my sister, Adai, whose roles include being my confidante and the family caretaker so that I could go on this journey. I thank the Shotiwuth clan for their faith in me.

Most of all, I would like to thank the two most important men in my life. I thank my late father, Majuddin Kambatu, who spurred me on to continue my academic pursuits. I lost my father in my second year of PhD study, but I did not lose the fighting spirit he instilled in me. To my husband, Kouji Shotiwuth, thank you for the huge sacrifice you made in order for my dream to materialise. It takes a lot of courage to uproot one's life and start anew in another country, but you never once doubted our decision. Thank you for your unfailing love, for all the meals that have nourished me, and for being there in the darkest days. My PhD journey could come to completion because you were by my side, patiently and selflessly cheering me on.

It takes a village to complete a PhD, so once again my heartfelt thanks go to everyone who is a part of my village of support. Thank you for believing in me.

***Dedication***

*To my late father, Majuddin Kambatu, who ignited my love and passion for lifelong learning,*

*and*

*To my mother, Latihim Madusin, who endured my absence with great patience and belief that  
it would all be worth it.*

*I owe my success to both of you.*





## TABLE OF CONTENTS

Abstract.....	i
Acknowledgements.....	iii
List of tables.....	xi
List of figures.....	xiv
List of plots.....	xiv
<b>Introduction.....</b>	<b>1</b>
<b>Chapter One: Literature review.....</b>	<b>5</b>
1.1 Incidental and intentional learning of MWEs.....	5
1.2 Vocabulary learning and multimodal input.....	7
1.3 Repetition and vocabulary learning.....	14
1.4 Typographic enhancement and vocabulary learning.....	24
1.5 Summary.....	35
<b>Chapter Two: Methodology.....</b>	<b>37</b>
2.1 Research questions.....	37
2.2 Participants.....	38
2.2.1 Participants in the actual study.....	38
2.2.2 Participants in the pilot study.....	40
2.3 Materials.....	41
2.3.1 Video 1: Fresh off the Boat (Season 1, Episode 2).....	41
2.3.2 The vocabulary in Video 1.....	42
2.3.3 Video 2: Raising Hope (Season 1, Episode 7).....	43
2.3.4 The vocabulary in Video 2.....	44
2.4 Target MWEs.....	45
2.4.1 Initial steps.....	45
2.4.2 Norming of target items.....	46
2.4.3 Norming tasks.....	47
2.4.4 Target items.....	50
2.5 Testing instruments.....	53
2.5.1 Form recall pretest.....	53
2.5.2 Word deletion procedure.....	53
2.5.3 Immediate form recall posttest (Gap-fill transcript-based test).....	54
2.5.4 Delayed form recall posttest.....	55

2.5.5 Delayed form recognition posttest.....	55
2.5.6 Delayed meaning recall posttest.....	56
2.5.7 Delayed meaning recognition posttest.....	56
2.5.8 Listening comprehension test.....	57
2.6 Procedure.....	59
2.6.1 Study 1: The incidental learning conditions.....	59
2.6.2 Study 2: The intentional learning conditions.....	62
2.7 Overall schedule.....	63
2.8 Scoring criteria and scoring procedure of the MWE tests.....	64
2.8.1 Scoring of the gap-fill tests (form recall tests).....	64
2.8.2 Scoring of the multiple-choice tests (form recognition and meaning recognition tests).....	67
2.8.3 Scoring of the meaning recall tests .....	68
2.8.4 Scoring criteria and scoring procedure of the listening comprehension tests.....	69
<b>Chapter Three: Study 1 (Incidental learning) – Analysis and results.....</b>	<b>71</b>
3.1 Research questions and organization of the chapter.....	71
3.2 Explanations of statistical tools and terms.....	71
3.3 Model development and comparison.....	73
3.4 Form recall - Results.....	75
3.4.1 Analysis of learning gains between the form recall pretest and immediate posttest.....	76
3.4.2 Analysis of learning gains between the form recall pretest and delayed posttest.....	80
3.4.3 Form recall item-level analysis.....	84
3.5 Form recognition – Results.....	90
3.6 Meaning recall – Results.....	93
3.7 Meaning recognition – Results.....	96
3.8 Comprehension test – Results.....	99
3.9 Summary of findings.....	102
<b>Chapter Four: Study 2 (Intentional learning) – Analysis and results.....</b>	<b>105</b>
4.1 Research questions and organization of the chapter.....	105

4.2 Explanations of statistical tools and terms.....	105
4.3 Form recall - Results.....	107
4.3.1 Analysis of learning gains between the form recall pretest and immediate posttest.....	107
4.3.2 Analysis of learning gains between the form recall pretest and delayed posttest.....	110
4.3.3 Analysis of learning gains between the form recall pretest and immediate posttest.....	112
4.4 Form recognition - Results.....	117
4.5 Meaning recall - Results.....	119
4.6 Meaning recognition - Results.....	122
4.7 Comprehension test - Results.....	125
4.8 Summary of findings.....	127
<b>Chapter Five: Discussion and implications.....</b>	<b>129</b>
5.1 Form recall under the incidental and intentional learning conditions.....	132
5.1.1 Factors that predict form recall under the incidental learning condition.....	132
5.1.2 Factors that predict form recall under the intentional learning condition.....	137
5.1.3 Summary of the factors the predict form recall under both learning conditions.....	140
5.2 Form recognition under the incidental and intentional learning conditions.....	142
5.2.1 Factors that predict form recognition under the incidental learning condition.....	142
5.2.2 Factors that predict form recognition under the intentional learning condition.....	146
5.2.3 Summary of the factors the predict form recognition under both learning conditions.....	148
5.3 Meaning recall under the incidental and intentional learning conditions.....	149
5.3.1 Factors that predict meaning recall under both learning conditions.....	149

5.3.2 Summary of factors that predict meaning recall under both learning conditions.....	153
5.4 Meaning recognition under the incidental and intentional learning conditions.....	154
5.4.1 Factors that predict meaning recognition under both learning conditions.....	154
5.4.2 Summary of the factors that predict meaning recognition under both learning conditions.....	158
5.5 Factors that influence listening comprehension.....	158
5.6 Summary of all findings.....	161
5.7 Pedagogical implications.....	164
5.8 Limitations and directions for future research.....	166
<b>Conclusion</b> .....	171
References.....	175
Appendices.....	190

## LIST OF TABLES

Table 1. Number of participants in Study 1 and the mean VST score (out of 140) for each condition.....	39
Table 2. Number of participants in Study 2 and the mean VST score (out of 140) for each condition.....	39
Table 3. The mean VLT score for each condition (for both studies).....	40
Table 4. Lexical coverage of Video 1.....	42
Table 5. Lexical coverage of Video 2.....	44
Table 6. Target MWEs in Video 1.....	51
Table 7. Target MWEs in Video 2.....	52
Table 8. Schedule for both Study 1 and Study 2.....	63
Table 9. Scoring criteria for the form recall tests.....	65
Table 10. Scoring criteria for the delayed meaning recall posttest.....	68
Table 11. Summary of statistical models used for each test.....	73
Table 12. Added covariates included in the model for all tests (for Study 1).....	74
Table 13. Descriptive statistics for the form recall pretest, immediate posttest and delayed posttest based on both scoring systems (for Study 1) .....	76
Table 14. Absolute short-term gains based on both scoring systems (for Study 1)...	77
Table 15. Output of the best-fit model for predicting short-term gains based on the lenient scoring system (Study 1) .....	79
Table 16. Output of the best-fit model for predicting short-term gains based on the strict scoring system (for Study 1) .....	80
Table 17. Absolute long-term gains based on both scoring systems (for Study 1)....	81
Table 18. Output of the best-fit model for predicting long-term gains based on the lenient scoring system (for Study 1) .....	82
Table 19. Output of the best-fit model for predicting long-term gains based on the strict scoring system (for Study 1) .....	84
Table 20. Output of the best-fit model (with core variables only) for the form recall item-level analysis (for Study 1) .....	85
Table 21. Estimate of the VST slope at each test time.....	87
Table 22. Summary of added covariates included in the item-level analysis (for Study 1) .....	88
Table 23. Descriptive statistics of the form recognition test (for Study 1) .....	90

Table 24. Output of the best fit model (with added covariates) for the form recognition test (for Study 1) .....	92
Table 25. Descriptive statistics for the meaning recall test (for Study 1).....	93
Table 26. Output of the best fit model (with core variables only) for the meaning recall test (for Study 1) .....	94
Table 27. Output of the best fit model (with added covariates) for the meaning recall test (for Study 1) .....	96
Table 28. Descriptive statistics for the meaning recognition test (for Study 1).....	96
Table 29. The output of the best-fit model (with core variables only) for the meaning recognition test (for Study 1) .....	97
Table 30. The output of the best-fit model (with added covariates) for the meaning recognition test (for Study 1) .....	99
Table 31. Descriptive statistics for the comprehension test (for Video 1).....	100
Table 32. The output of the best-fit model for the comprehension test (for Video 1)	101
Table 33. Added covariates included in the model for all tests (for Study 2).....	106
Table 34. Descriptive statistics for the form recall pretest, immediate posttest and delayed posttest based on both scoring systems (Study 2) .....	107
Table 35. Absolute short-term gains based on both scoring systems (for Study 2)...	108
Table 36. Output of the best-fit model for predicting short-term gains based on the lenient scoring system (for Study 2) .....	109
Table 37. Output of the best-fit model for predicting short-term gains based on the strict scoring system (for Study 2) .....	109
Table 38. Absolute long-term gains based on both scoring systems (for Study 2)....	110
Table 39. Output of the best-fit model for predicting long-term gains based on the lenient scoring system (for Study 2) .....	111
Table 40. Output of the best-fit model for predicting long-term gains based on the strict scoring system (for Study 2) .....	111
Table 41. Output of the best-fit model (with core variables only) for the form recall item-level analysis (for Study 2) .....	113
Table 42. Estimate of the VST slope for each caption condition.....	115
Table 43. Summary of added covariates included in the item-level analysis (for Study 2) .....	116
Table 44. Descriptive statistics of the form recognition test (for Study 2) .....	117

Table 45. Output of the best-fit model (with core variables only) for the form recognition test (for Study 2) .....	118
Table 46. Descriptive statistics for the meaning recall test (for Study 2).....	119
Table 47. Output of the best fit model (with core variables only) for the meaning recall test (for Study 2) .....	120
Table 48. Descriptive statistics for the meaning recognition test (for Study 2).....	122
Table 49. The output of the best-fit model (with core variables only) for the meaning recognition test (for Study 2) .....	123
Table 50. The output of the best-fit model (with added covariates) for the meaning recognition test (for Study 2) .....	124
Table 51. Descriptive statistics for the comprehension test (for Video 2).....	125
Table 52. The output of the best-fit model for the comprehension test (for Video 2)	125
Table 53. Overview of findings: Significant predictors and interactions found for each MWE test under both studies.....	130

## LIST OF FIGURES

Figure 1. Figurativeness rating task (taken from Google Form) .....	48
Figure 2. Compositionality rating task (taken from Google Form) .....	49
Figure 3. Ambiguity rating task (taken from Google Form) .....	50
Figure 4. An example of a still from Video 1 (Fresh off the Boat) with normal captions.....	61
Figure 5. An example of a still from Video 1 (Fresh off the Boat) with enhanced captions....	61
Figure 6. Procedure of Study 1 and Study 2.....	62

## LIST OF PLOTS

Plot 1. Predicted short-term gains under the lenient scoring system (Study 1).....	78
Plot 2. Predicted short-term gains under the strict scoring system (Study 1).....	80
Plot 3. Predicted long-term gains under the lenient scoring system (Study 1).....	82
Plot 4. Predicted long-term gains under the strict scoring system (Study 1).....	83
Plot 5. Predicted probability of getting an item correct in the form recognition test (Study 1)..	91
Plot 6. The effects of pretest score and VST score on the predicted probabilities of getting a score of 0, 0.5 and 1 on a target item in the meaning recall test (Study 1).....	95
Plot 7. Predicted probability of getting an item correct in the meaning recognition test (Study 1).....	98
Plot 8. Predicted probability of getting an item correct in the comprehension test (Study 1).....	101
Plot 9. The effects of pretest score and VST score on the predicted probabilities of getting a score of 0, 0.5 and 1 on a target item in the meaning recall test (Study 2).....	121
Plot 10. Predicted probability of getting an item correct in the meaning recognition test (Study 2).....	123
Plot 11. Predicted probability of getting an item correct in the comprehension test (Study 2)..	126



## INTRODUCTION

Over the past two decades, research on multiword expressions (henceforth, MWEs) has flourished, providing insights into the challenges they pose to second language (L2) learners and the potential pedagogic interventions that could help them overcome these challenges. A few reasons have been put forward as explanations for L2 learners' slow rate of MWE acquisition. The first relates to the sheer number of MWEs that make up a large proportion of language, with some estimates indicating that up to 50% of language we encounter on a daily basis is formulaic (e.g., Erman & Warren, 2000; Hill, 2000). Despite their ubiquity, however, only a small number of high-frequency formulas (e.g., Shin & Nation, 2008) appear repeatedly in a short span of natural discourse. Encountering the same MWEs in an input text is therefore highly unlikely (e.g., Boers & Lindstromberg, 2009). The second reason why MWEs are difficult to acquire is that some of them consist of familiar words and are semantically transparent, as a result of which they may go unnoticed. Other MWEs, however, are semantically non-transparent even though their constituent words look familiar, and in such cases (e.g., idioms and phrasal verbs) the difficulty lies with comprehension.

These broad explanations suggest that in order for L2 learners to master a large repertoire of MWEs, exposure to a vast amount of input is crucial so as to ensure repeated encounters. After all, repetition is one of the conditions that facilitates learning (Webb & Nation, 2017; Uchihara, Webb & Yanagisawa, 2019). Secondly, to ensure that MWEs encountered in input do not escape learners' attention, 'noticing' of these phrases is crucial. As posited by Schmidt (2001), attention or noticing is required for intake. It is, therefore, unsurprising that a substantial amount of studies have investigated whether MWE learning can indeed be enhanced through repetition and noticing-inducing techniques such as typographic enhancement. What is surprising, however, is that these studies have focused almost exclusively on written input, and, to a limited extent, bimodal input.

The limited attention given to the potential use of audio-visual input or L2 viewing for MWE learning is surprising, considering that L2 viewing has been proven to be beneficial for the acquisition of single words (e.g., Feng & Webb, 2019; Neuman & Koskinen, 1992; Peters & Webb, 2018; Rodgers, 2013). Further, studies have shown that English as a Foreign Language (EFL) learners are more exposed to television than written input (e.g., Peters, Noreillie, Heylen, Bulté & Desmet, 2019), possibly owing to its entertainment value, which appeals to all learners including those with low L2 proficiency and low motivation levels (Lin

& Siyanova-Chanturia, 2014). The potential of authentic L2 viewing has also been pointed out through lexical analyses of large samples of movies and TV shows, which revealed that the vocabulary demand for L2 viewing is lower than for reading (Webb & Rodgers, 2009a). Clearly, there are merits in watching TV for vocabulary learning, which has prompted some researchers to advocate extensive TV viewing as a way to bolster incidental L2 vocabulary learning (Webb, 2015). A recent study by Feng and Webb (2019) has also shown that viewing is as beneficial as reading and listening for promoting incidental vocabulary learning. Additionally, the distribution of MWEs in television has been shown to be similar to everyday speech (Lin, 2014). This provides a stronger reason to believe that the potential benefit of L2 viewing for single words might extend to MWE learning. Whether this is true, however, has not been researched extensively.

Indeed, while there is now a shift from written input to L2 viewing for MWE learning, relatively little is still known about the potential of L2 viewing. For instance, although a recent study by Puimège and Peters (2019) showed that incidental MWE learning is possible through L2 viewing, their study shed light on the benefits of uncaptioned viewing only. No comparison was made between captioned and uncaptioned viewing. Studies on L2 viewing on single word acquisition have almost consistently furnished evidence of the superiority of captioned over uncaptioned viewing for vocabulary learning (e.g., Montero Perez, Van Den Noortgate & Desmet, 2013). It would be interesting to assess whether the same applies to MWE learning. This provides the impetus for the research project, where captioned and uncaptioned viewing are compared in order to assess their effectiveness for MWE learning.

Given the aforementioned issues around learners' noticing of phrases as one of the prerequisites for acquisition, it is worth exploring ways to ensure that learners direct their attention to MWEs they encounter through L2 viewing. In the context of MWE acquisition through written input, the use of typographic enhancement, i.e., making items stand out through means such as underlining, **bolding** and *italicizing* to induce noticing and facilitate MWE uptake, has been shown to have positive effects. For instance, reading studies have demonstrated that enhanced MWEs attract longer fixation times (Choi, 2018), and are remembered better than unenhanced items (Boers, Demecheleer, He, Deconinck, Stengers, & Eyckmans, 2017; Sonbul & Schmitt, 2013; Szudarski & Carter, 2016). While these reading studies have offered evidence that typographic enhancements facilitate the noticing and learning of enhanced MWEs, we cannot simply assume that typographic enhancement will have the same positive effect in the case of captions. This is because of the real-time nature of

viewing, and because there is more to look at and take in (such as moving images) besides the captions alone. In other words, compared to written input, learners have less time to fixate on anything, including typographically-enhanced items, when viewing an input video. As such, the facilitative role of typographic enhancement for MWE learning in the context of audio-visual material is still in question. The results of previous the reading studies suggest that typographically-enhanced captions may lead to better MWE knowledge compared to normal captions. Whether this is true, however, is what this research project intends to shed light on. In essence, besides viewing with no captions and normal captions, a condition under which the learners view an input video with typographically-enhanced captions is also created to determine whether it is more effective for MWE learning compared to the other two caption conditions.

Another purported benefit of using typographic enhancement is that it reduces the number of encounters needed for similar MWE gains to be made (Szudarski & Carter, 2016). As mentioned earlier, repeated encounters of unknown MWEs facilitate learning. Evidence of this has been found in MWE studies on written input (e.g., Pellicer-Sánchez, 2017) and bimodal input (e.g., Webb, Newton & Chang, 2013). The texts used in these studies, however, were modified to allow for the inclusion of the necessary number of target items required for multiple encounters. As such, the ecological validity of such texts may be in question, since it is rather unlikely that L2 learners would come across such ‘heavily seeded’ texts. Further, such modification would be harder to create in an authentic video. However, while an authentic video is ecologically valid, it is impossible to control an item’s frequency of occurrence. One way to counter this would be to operationalise repetition as repeated viewing. This way of operationalisation is also supported by Durrant and Schmitt’s (2010) reading study, which found that (at least in the initial stage of learning new MWEs) verbatim repetition is better for MWE learning compared to varied repetition (i.e., encountering the same MWE in different contexts). As the effects of repetition on MWE learning have only been investigated in the context of written input, this research project aims to assess whether repeated viewing, as compared to one viewing, leads to superior MWE gains. Further, comparisons of one and two viewings under all three caption conditions (i.e., uncaptioned, normal captions and typographically-enhanced captions) could also shed light on whether typographic enhancement really reduces the number of repetitions needed to make comparative MWE gains.

It is said that gains made under intentional learning are often much larger than in incidental learning (Webb, 2002; Laufer, 2003). However, few empirical MWE studies can

support this claim, as the bulk of MWE studies have the propensity to focus exclusively on one type of learning, either incidental or intentional. As such, the effectiveness of most interventions in MWE studies has been predominantly investigated under one learning condition. For instance, it is generally accepted that fewer repetitions are needed for intentional learning compared to incidental learning. However, whether this is the case for MWE learning is unknown as no existing MWE studies compared the outcome of repetition under both learning conditions. Similarly, typographic enhancement has been treated as one of the ways to induce noticing in instances where learners are engaged with the content of an input text. How typographically-enhanced captions alter learners' engagement with salient MWEs under intentional learning conditions remains to be explored. As such, this research project also investigates the ways in which learning conditions modulate the effects of caption condition and repetition. To this end, two studies were carried out. Taking a leaf from Hulstijn's (2001) distinction, these two studies are differentiated at the level of test announcement, in that in Study 1 the learners receive no forewarning of MWE tests, while in Study 2 the learners are explicitly informed of the tests.

In summary, the research project aims to not only explore the potential of L2 viewing for MWE learning, but also shed light on how repetition and typographic enhancement could further boost MWE acquisition. Additionally, the effects of repetition and typographic enhancement are also evaluated in conjunction with the two learning conditions, incidental and intentional. The exploration of these questions in this thesis begins with a survey of relevant literature in Chapter One. This is followed by Chapter Two, in which the methodology undertaken to answer these research questions is detailed. In Chapter Three, the results of the MWE tests in Study 1 (i.e., the incidental learning conditions) are presented. Next, Chapter Four presents the results of the MWE tests in Study 2 (i.e., the intentional learning conditions). Chapter Five discusses the findings of both studies. This chapter posits reasons for the similarities and differences found between the two studies, and discusses the limitations and pedagogical implications. Finally, the thesis is concluded by bringing together the most important findings of the two studies.

## **CHAPTER ONE:**

### **LITERATURE REVIEW**

This chapter presents a review of literature on MWEs and studies on the pedagogical interventions meant to enhance the learning of MWEs. The chapter will first discuss the definitions adopted in this research. This will then be followed by a review of the studies concerning vocabulary learning through multimodal input. Following this, the chapter will discuss studies investigating the effects of repetition on MWE learning. The chapter then turns to typographic enhancement studies. The chapter concludes by summarising the gaps in the literature.

#### **1.1 Incidental and intentional learning of MWEs**

Besides multiword expressions, a plethora of terms has been proposed for lexical units above the word level (Wray, 2002). Similarly, different definitions of this phenomenon exist. For the purpose of the current research, MWEs are defined as phrases that are considered conventional by a proficient language user. They vary along the continua of frequency, length, fixedness, abstractness and figurativeness/literality (e.g., Siyanova-Chanturia & Van Lancker Sidtis, 2018).

A large proportion of language is made up of MWEs. An oft-cited finding by Erman and Warren (2000) suggests that MWEs make up 50% percent of discourse. This suggests that in order for L2 learners to possess MWE knowledge on par with native speakers of English, they need a large repertoire of MWEs. Indeed, MWE competence has often been associated with native-like proficiency, including the dimensions of fluency, accuracy and lexical richness (e.g., Boers, Eyckmans, Kappel, Stengers & Demecheleer, 2006; Stengers, Boers, Housen & Eyckmans, 2010; 2011). However, acquiring the breadth and depth of MWE knowledge to rival those of native speakers proves to be difficult, even for highly-advanced L2 learners. Literature has shown that compared to native speakers, highly advanced L2 learners may rely on a narrow range of MWEs (e.g., Durrant & Schmitt, 2010). Further, longitudinal studies (e.g., Li & Schmitt, 2010; Qi & Ding, 2011) have shown that L2 learners' development of MWE knowledge is slow and uneven. In short, it is well established that L2 mastery of MWE is important but difficult to achieve. Perhaps guided by this knowledge, the last decade has seen a proliferation of studies on pedagogic interventions that are intended to boost L2 learners' mastery of MWEs.

Despite the growing number of studies on these pedagogic interventions, there is an observable trend: the bulk of these experimental studies tend to focus on certain types of MWEs or parts of speech, such as adjective + noun collocations (e.g., Durrant & Schmitt, 2010) and verb + noun collocations (e.g., Webb et al., 2013). Confining the choices of target items to certain types, however, necessitated the modification of authentic input in order to include sufficient numbers of items. On the one hand, there is merit in focusing on one type of MWE, as MWEs are not equal in the way that they are processed (e.g., Columbus, 2010), their memorability (e.g., Peters, 2016), or in the challenges they pose to L2 learners (e.g., Boers, 2019). On the other hand, if learners are expected to engage with authentic materials, it means that they will encounter more than one type of MWE. It is then more ecologically valid to keep the materials intact and focus on more than one type of MWE. Further, research has suggested that the risk of inter-item interference is increased when learners are presented with MWEs that are syntactically similar (Boers, Demecheleer, Coxhead & Webb, 2014). In summary, there is a need for more studies that look at how the acquisition of different types of MWEs can be facilitated through authentic input. This is the line of inquiry that the present research intends to expand.

In addition, there is also a need for studies involving the use of authentic multi-modal input. This is because the majority of existing studies on pedagogic interventions meant to boost L2 learners' MWE acquisition focus on unimodal input, i.e., reading, and, to a much lesser extent, reading-while-listening. Studies on single-word acquisition have consistently furnished evidence that L2 vocabulary can be learned through audio-visual input, i.e., L2 viewing (see sub-section 1.1). This provides strong grounds to suspect that L2 viewing may boost MWE acquisition as well. However, there are a limited number of studies investigating this proposition.

Studies on pedagogic interventions for the learning of MWEs have also shown a propensity towards a dichotomy of learning conditions. In other words, many of the interventions intended to facilitate MWE learning have been investigated exclusively under one of the learning conditions, i.e., the intentional or incidental learning conditions. Nation and Webb (2011) define intentional learning as learning that occurs when a learner "deliberately decides to learn a particular word or set of words and focuses on this learning" (p. 307). In essence, a task that aims at committing lexical information to memory is considered intentional learning. In contrast, incidental learning occurs when the primary goal of a task is something

else (Nation & Webb, 2011), such as engaging with meaningful content. As such, vocabulary learning becomes a by-product of the message-focused activity.

Hulstijn (2001), however, argues that the theoretical distinctions between intentional and incidental learning are hard to maintain in practice, as the difficulty lies in determining whether a student's learning is truly deliberate or not. According to him, in operational terms, "incidental and intentional learning can simply be distinguished in terms of the use of prelearning instructions" (p. 268). In other words, the critical operational feature that distinguishes between the two types of learning is learners' expectation of a test. In the current study, Hulstijn's operational terms are adopted. The same operational terms were also used by Nation and Webb (2011) and Pellicer-Sánchez and Boers (2018).

As mentioned earlier, while there has been increasing research on the learning of MWEs, few studies have investigated the effects of an intervention under both incidental *and* intentional learning of MWEs. In the realm of vocabulary learning, the learning gains made by L2 learners under intentional learning conditions are often said to be much larger compared to incidental learning (Webb, 2002; Laufer, 2003). Whether the same could be said about MWEs is yet to be determined owing to lack of studies that actually compare the outcome of MWE learning under both learning conditions. Additionally, as the subsequent sections will highlight, factors that have been known to facilitate MWE uptake such as typographic enhancement and repetition have mostly been investigated under incidental learning conditions. Whether these factors further accelerate MWE learning under intentional learning conditions is not known.

Taken together, existing studies on pedagogical interventions for MWE acquisition have revealed a tendency towards exclusivity in terms of input modality, item type and learning condition. There is limited research that compares how interventions such as typographic enhancement and repetition can boost the acquisition of diverse types of MWEs under both incidental and intentional learning conditions. Importantly, whether these interventions are effective in the context of authentic multimodal input (e.g., watching TV in L2) remains to be investigated. In the following sections, these research gaps will be highlighted through a review of relevant existing studies, starting with those focusing on the potential efficacy of multimodal input for MWE learning.

## **1.2 Vocabulary learning and multimodal input**

Though the majority of research on vocabulary learning focuses on reading as a source of input, there has been increasing research on vocabulary acquisition through multimodal or

audio-visual input. Audio-visual input serves as a valuable source of L2 vocabulary learning for a few reasons. Firstly, surveys have revealed that people spend more time watching television than reading. In her survey that investigates Flemish EFL learners' exposure to English language media, Peters (2018) found that more than 40% of the learners reported regular watching of English TV with and without subtitles. In addition, these learners, aged 16 and 19, reported limited exposure to written input such as books and magazines. Additionally, Peters (2018) found a positive relationship between the exposure to non-subtitled TV programmes and the learners' vocabulary knowledge.

Other surveys have revealed that watching TV is the preferred leisure activity for all the 18 countries surveyed (OECD, 2009), with daily household viewing reported to range from 2.43 hours in Sweden to 8.18 hours in the United States (OECD, 2007). Additionally, in a survey on media consumption by Roy Morgan Research (2015), data collected from 11 countries across the Asia-Pacific region showed that people spend an average of 8.2 hours to 29.5 hours a week watching television. The leisure factor attached to watching TV has made it more motivating for L2 learners to learn from this medium (Gieve & Clark, 2005; Sueyoshi & Hardison, 2005). Previous studies have also shown that watching TV is one of the main ways EFL learners receive out-of-school exposure to English (e.g., Peters et al., 2019). Additionally, a recent study by Feng and Webb (2019) demonstrated that viewing is as effective as reading and listening for L2 incidental learning of single words. Further, where MWEs are concerned, research on prosody has prompted researchers to emphasize the need for L2 learners to be exposed to spoken input as MWEs are found to be prosodically salient (Lin, 2018), a characteristic that may promote their uptake.

Corpus studies have also shed light on the potential benefit of audio-visual materials as a source of vocabulary learning. Rodgers and Webb (2011) analysed the scripts of 288 television programmes to investigate the difference in vocabulary reoccurrences between related and unrelated TV programmes, and found that there is potential for vocabulary learning through both types of programmes. In addition, the study found that low frequency words (4000 to 14,000 level) word families were encountered 10 or more times in the related TV programmes. This echoes their earlier finding (Webb & Rodgers, 2009a), where they discovered that low frequency words occur repeatedly within a small amount of viewing time, making it possible for incidental vocabulary learning to take place with regular viewing. Clearly, there are merits in watching TV for vocabulary learning.



Whether the presence of captions or intralanguage subtitles further enhances vocabulary acquisition through audio-visual material is another strand of research that has been receiving increasing attention. Initially created for the hard-of-hearing, captions have benefitted many other types of viewers, L2 learners included. Price's (1983) study was the first empirical study that investigated the effects of captioning. Her study found that even with one viewing, captioned television brought about positive gains, and learners were found to acquire more 'cultural script'. Many other benefits of captioning have since been reported. Apart from exposing learners to authentic input, captioned video decreases learners' anxiety and increases their motivation (Vanderplank, 1988). Further, Garza (1991) posits that the addition of textual modality renders segments more accessible and comprehensible. Also, captions help learners visualise what they hear, which may aid form-meaning mapping (Garza, 1991; Winke, Gass & Sydorenko, 2010). A psycholinguistic study by Bird and Williams (2002) further revealed that in a multimodal situation, captions are heavily relied on and they promote a greater depth of spoken-word processing. Some investigations on how captions help people learn languages are grounded in multimedia learning (e.g., Gass, Winke, Isbell & Ahn, 2019; Winke et al., 2010). Mayer's (2009) theory of multimedia learning, which is based on Paivio's (1986, 2006) dual coding theory, claims that learners attempt to make connections between words and pictures, so learning is deeper and longer lasting when both are present as opposed to when only words *or* pictures are used.

Many empirical studies have since shown favourable effects of captioned over uncaptioned viewing on the learning of single words. In one of the first studies of these kinds, Neuman and Koskinen (1992) investigated the effects of captioned television on young learners' incidental learning of English words. The television segment used in the experimental conditions was a science production for a children's television show. The participants ( $N = 129$ ), who were of various L1s and proficiency levels, were assigned to four different conditions: (a) captioned video, (b) no captions, (c) read and listen to the text and (d) textbook only. Throughout the nine-week study, unit-based tests on word recognition were carried out. Additionally, participants took a word meaning test at the end of the nine weeks. The results revealed that the learning gains for both viewing groups were higher than for the reading-while-listening condition. Additionally, significant differences were found between the captioned and uncaptioned conditions in the meaning test, and in one of the unit-based recognition tests. This study lends evidence to the positive role of captioning for vocabulary learning.

Indeed, the superiority of captioned viewing over uncaptioned viewing for L2 learning of novel single words was also confirmed in a meta-analysis by Montero Perez et al. (2013). In their meta-analysis of existing studies at the time, comparisons on the effectiveness of both types of viewing on video comprehension (15 studies) and the learning of new vocabulary (10 studies) were made. The results revealed that for both purposes captioned viewing is more beneficial than uncaptioned viewing. However, it is important to note that the meta-analysis included studies that employ tests that vary in terms of format and modality. Further, the majority of the studies that were included made use of a written test format, i.e., learners were presented with the written format of the target form. This is different from how the input was presented in the uncaptioned conditions. As noted by Mohd Jelani and Boers (2018), this congruency of input-modality—test-modality may have given an advantage to those participants who had seen the written forms in the captions. As such, there is reason to suspect that the reported benefit of captioning over uncaptioned viewing may be partly due to the use of the written test format, which does not mirror the modality of the input under uncaptioned viewing.

The findings of Sydorenko's (2010) study, which made use of aural test prompts, lend support to this supposition. Using mixed methods, Syrodenko looked at the vocabulary gains, the amount of attention learners gave to each input modality, and the strategies they used to acquire vocabulary from the multimodal input. Twenty-six Russian beginner learners of various L1s watched three Russian comedy clips under one of three conditions: with audio and captions, with audio only, or with captions only. They then answered three comprehension questions, followed by a word recognition test, a translation test and a word knowledge test that assessed learners' knowledge of meaning. In the recognition test, non-words were included. Half of the non-words and target words were presented in writing, while the other half were presented aurally. The participants were required to indicate if they had seen or heard the word during the treatment stage. The translation task was carried out only with the target words, also with half of them presented on paper and the other half aurally. To determine if the target words were unknown to the participants prior to the treatment, a written word knowledge test was administered. The results revealed an interaction between the recognition of form scores and input modality, as the caption groups scored higher in the written recognition test, but were outperformed by the video-with-audio-only condition in the aural recognition test. No such interaction, however, emerged in the translation task. In addition, the combination of the three modalities was found to be more effective compared to the uncaptioned condition on both

the written and aural sections of the meaning test. In summary, this study provides evidence of input-modality—test-modality congruency. It is, however, important to note that any gains in this study may not be considered as incidental as participants were informed beforehand that they would be tested on both text comprehension and language.

A study that did look at incidental vocabulary learning and found evidence for the role of test modality was conducted recently by Mohd Jelani and Boers (2018). Using a similar research design, i.e., written and aural test prompts, they sought to investigate whether aural test prompts would yield favourable vocabulary learning gains for captioned video. Sixty-six Malaysian EFL learners watched a ten-minute TED talk with and without captions. Apart from the two experimental groups, one group of a higher proficiency level was also used to determine words that were unlikely to be known by the experimental groups. Fifteen target words were chosen, and another 15 words were added as foils to be used for the posttests. After watching the input video twice under their respective conditions, two tests were administered. In the word recognition test, the participants were required to indicate if they had encountered the words in the test. Seven target words and foils were presented on an answer sheet, while the remainder were presented aurally. For the latter, the participants were asked to write down the words they heard before indicating whether they remembered encountering them. In the word meaning test, the participants were required to provide the meaning of the words that they remembered encountering by providing a synonym in English or an L1 translation. While both groups indicated that they remembered more words presented in the written form, it was found that the captioned group remembered the target words more than the uncaptioned condition. This suggests that captioned video has an advantage over uncaptioned video. However, the captioned group also responded “yes” to more of the foils. Therefore, when all 30 items were examined closer, it was discovered that the captioned group no longer outperformed the uncaptioned group in the word recognition test. As for the word meaning test, the participants who saw the captioned video performed significantly better than the uncaptioned condition. This superiority, however, was found to be attributed solely to their performance in the test with written prompts as both conditions were found to show no difference in the aural prompt test. Further, no correlation was found between the captioned condition’s scores on the written and aural parts of the task. On the one hand, the findings suggest that the role of test modality should be taken into account so as to avoid input-modality—test-modality incongruence. On the other, an effect of this incongruence has only been found in two studies. Others have found no evidence of a mediating role of test modality.

One such study was conducted by Winke et al. (2010). Adopting a similar design to Sydorenko (2010), Winke et al. investigated the effectiveness of captioned videos in enhancing learners' comprehension of content and learning of vocabulary. The participants were 150 native speakers of English studying one of four L2s, i.e., Spanish, Russian, Arabic and Chinese. Each group of participants watched three short clips ranging from approximately three to five minutes. The videos, which were documentaries about animals, were originally in English and were dubbed into the four L2s. Each language group was further divided into two groups. The Spanish learners saw the clips twice with no captions or with captions. A vocabulary test and a comprehension test were administered after each clip had been watched twice. Analogous to Sydorenko's (2010) research design, the vocabulary test was conducted using two different test prompts. Half of the target words were presented in writing, and the other aurally. The participants had to provide English translations to the target words. They then took a test of prior knowledge of vocabulary, where they were required to indicate their familiarity with the key words on a scale. The results revealed no evidence of modality congruency effect. This is because watching captioned videos led to significantly higher scores compared to viewing non-captioned videos in both vocabulary tests with written *and* aural prompts. The results of this study, however, should be interpreted cautiously as there were fewer than ten participants under each condition.

Markham's (1999) study provides more robust statistical evidence for the lack of modality congruence effects. With the objective of comparing listening word recognition between caption and no caption groups, Markham had 118 advanced L2 learners watch two 12 to 13-minute long educational programmes either with or without captions. Two 50-item multiple-choice listening tests based on the videos were then administered. In both tests, the stems came directly from the video, and the stems along with multiple-choice options were presented aurally. Participants had to distinguish the word that appeared in the video from the distractors. The results revealed that captions assisted participants in recognising words in the video that also appeared in the aurally-presented word recognition test. Fewer words were recognised by the participants who saw the video without captions. While orthographic representation of the word was found to assist the uptake of the phonological form of the word, it is not known whether the 50 words chosen to be tested in the word recognition test were familiar to the participants. Further, in both this study as well as Winke et al.'s (2010), the learners were tested after viewing each clip. Therefore, there is a possibility that learners may have been more (or differently) engaged in the subsequent clips after having been tested for

the first one. As posited by Montero Perez, Peters and Desmet (2015), learners' engagement with input may be altered when there is test anticipation. Thus, it is important to establish whether the presence or absence of test announcement modulates the effects of captioning. This is one of the aims of the present study.

In summary, the inconsistent results of these captioned versus uncaptioned studies make it hard to conclude that the positive effects of captions have been overestimated due to input-mode—test-mode congruency. In addition, these studies differ in many respects such as learners' proficiency levels, the aspects of word knowledge being investigated, and the format of the test prompts. As there is no conclusive evidence on the role of test modality, this research project does not include the role of test modality as one of its foci. Overall, there are limited studies that have found evidence of input-mode—test-mode congruency, and current consensus points to benefits of captioning in facilitating the learning of single words, regardless of whether this learning is assessed by means of aural or written test prompts.

Surprisingly, whether the positive benefit of captioning applies to the learning of MWEs has not yet been investigated. In fact, thus far, only one study has looked at how MWEs could be incidentally learned through L2 viewing. The study was conducted recently by Puimège and Peters (2019), who also sought to assess whether any item- and learner-related factors may influence the learning of single words and MWEs. The participants were 20 Flemish learners who watched an uncaptioned 30-minute reality TV show containing 15 target single words and 20 target MWEs. Three vocabulary tests tapping on to different aspects of vocabulary knowledge, i.e., form recall, form recognition and meaning recall were administered two weeks before and two weeks after participants watched the video. The results of the form-recognition test, however, were not further analysed as participants appeared to score higher in the pretest compared to the posttest, suggesting that a lot of guessing may have taken place. At the level of meaning recall, evidence of single word and MWE learning was not found. At the level of form recall, however, some learning effects were found for both single words and MWEs. Four factors influenced the form recall of single words, i.e., corpus frequency, item length, concreteness and learners' vocabulary size. As for the MWEs, five factors were found to predict recall. Firstly, similar to the learning of single words, the bigger the vocabulary size of the learners, the bigger their learning gains tended to be. Secondly, shorter MWEs stood a better chance of being recalled correctly. Mutual information score (a statistical measure of the strength of word partnerships) was also a significant predictor, but with stronger word partnerships actually recalled less well, because these MWEs involved low-

frequency words. Further, more frequent MWEs tended to be better recalled compared to less frequent MWEs. The most significant predictor, however, was collocate-node relationship, with the adjective-noun combinations learned best.

That uncaptioned viewing leads to incidental MWE learning at the level of form recall suggests that there is potential for bigger learning gains through captioned viewing. This is because MWEs are produced relatively fast, and function words such as articles and prepositions tend to be phonologically reduced in MWEs (Bybee, 2002). Consequently, real time viewing without captions may pose a challenge for learners as they are unable to pick up the precise lexical composition of the MWEs. As such, the presence of captions may help learners to discern the spoken words (Vanderplank, 2010), segment speech into distinct words, and notice the phonologically-reduced words. Whether captioned viewing is indeed beneficial for MWE learning, however, has yet to be investigated.

In summary, although the superiority of captioned viewing has been established in literature concerning the uptake of single words, not much is known about its potentially positive effects on the uptake of MWEs. This is the gap that this research project seeks to close. To this end, the videos in this research project were manipulated to create varying types of caption conditions to investigate the extent to which MWEs are picked up by L2 learners. Further, the effects of two other manipulations, namely repetition and typographic enhancements, are also investigated as these two factors are often found to influence vocabulary learning. The subsequent sections present reviews of studies on repetition and typographic enhancement.

### **1.3 Repetition and vocabulary learning**

Repetition, or frequency of occurrence, has been proven time and time again to aid the learning of unknown words (all else being equal). A recent meta-analysis by Uchihara et al. (2019), which included 26 previous studies on repetition, serves as further proof of its facilitative role, as a medium-strength relationship was found between number of exposures to a word and learners' subsequent knowledge of these words. Though most of the repetition studies concern unimodal input, i.e., reading, there is robust evidence that provided there are enough encounters, learning gains can occur (Elgort, Brysbaert, Stevens & Van Assche, 2018; Elgort & Warren, 2014; Horst, Cobb & Meara, 1998; Pellicer-Sánchez & Schmitt, 2010; Rott, 1999; Webb, 2007). There is yet to be a consensus on the number of repetitions needed for single word learning to take place—and such a consensus is probably unrealistic given the multitude of variables involved. Some studies have reported that as many as 10 encounters are

needed (e.g., Pigada & Schmitt, 2006; Pellicer-Sánchez & Schmitt, 2010; Webb, 2007), while others have found that six encounters may suffice for successful acquisition (e.g., Rott, 1999).

While there is a substantial number of studies on the effects of repetition on the learning of single words, research investigating the effects of repetition on MWE learning is far more limited. Further, these studies have largely focused on the effects of repetition on a particular type of MWE, notably collocations. The results of such studies have been inconclusive. While some studies found no compelling evidence for the role of repetition, others found that more repetitions lead to better collocational knowledge. Peters' (2014) study is an example of the latter. Peters investigated how form recall of both single words and collocations is affected by repetition and the time of the posttest administration. To this end, two learning conditions, which differed in terms of the time the posttest was administered, were created. Both groups were exposed to 12 single words and 12 collocations through a word list that contained their definitions, followed by eight non-communicative focus-on-forms activities, such as fill-in-the-gaps and matching exercises. To investigate the effects of frequency, the target items appeared in differing frequency throughout the activities: occurring once, three times, or five times. In the pretest, the Flemish EFL participants ( $N = 35$ ) were provided with a list of Dutch words and collocations for which they had to supply the correct English translations. As for the posttest, the participants were provided with the Dutch translation and the English definition or synonym, to guide them to fill in the correct English word or collocation. The posttest was taken two times by both groups, but they differed in terms of the duration between each test. In group 1, the first posttest was taken immediately after the treatment, with the second posttest taken two weeks after that. As for group 2, the first posttest was taken one week after treatment, followed by the second posttest, two weeks later. The overall results revealed that regardless of the time lapse between the treatment and the first posttest, the effect of repetition is significant on participants' recall of both single words and collocations. Further, the more often a target item occurred during the vocabulary activities, the higher the learning gains. Specifically, for both groups, there was a significant difference between 1 and 5 occurrences irrespective of the times of the first posttest. Secondly, the effects of frequency were found to be durable in the second posttest. Although the positive effects of repetition on collocational knowledge is demonstrated in this study, it is important to note that the treatment involved activities that explicitly targeted the lexical items. In other words, though the posttests were unannounced, learners' attention was drawn to the lexical items due to the focus-on-form

activities. The effect of repetition may be less pronounced, or non-existent, when learners' attention is not explicitly drawn to the MWEs.

Such was the finding of Pellicer-Sánchez's (2017) study. Contrary to Peters' (2014) study, the target MWEs in Pellicer-Sánchez's study were encountered incidentally, as the treatment only involved the reading of a text containing the target MWEs and did not include any activities that explicitly drew learners' attention to the target MWEs. The target MWEs were six adjective-pseudonoun collocations, and the objective of the study was to find out whether learners would acquire knowledge of the form and meaning of new nouns only, or whether collocational knowledge would also be developed. The pseudowords replaced the nouns in existing adjective-noun collocations, while the adjectives were chosen by creating a list of the most frequent collocates for the nouns. Some examples of the target collocations include '*old hoster*' and '*small berrow*'. To assess the effects of repetition, two versions of the reading text were created. Version A contained the six adjective-pseudowords repeated eight times, while in Version B the pseudowords occurred four times with the adjective collocates, and four times without the adjectives. Forty-one L2 learners of diverse L1s read either Version A or Version B of the text. One week later, a combination of paper-and-pen and interview tests were carried out. The first test was a form recognition test for the pseudowords in multiple-choice format. In the second test, participants' ability to recall the meaning of the pseudowords was assessed through an interview. This was followed by a meaning recognition test, which required the participants to choose the meaning of a pseudoword in a multiple-choice format. Next, a collocation recall test was conducted via an interview. This was done by asking the participants to state collocates that frequently appeared with the pseudowords in the story. In the last test, collocation recognition was assessed by asking participants to choose the correct collocates from the options provided. Results revealed that for both groups, recall was harder than recognition. Participants were also found to acquire both the form-meaning link of the pseudowords, and the form of collocations. Although the eight-encounters group had higher scores in the collocation recall posttest compared to those who only had four encounters, the difference fell short of statistical significance. As acknowledged by the author, the results have to be interpreted in light of the use of the pseudowords. Unlike familiar words which do not attract much attention (Godfroid, Boers & Housen, 2013), novel pseudowords may have drawn the learners' attention, which may have rendered the effect of repetition less strong.

A study that did use real collocations and found a significant effect of repetition was carried out by Webb et al. (2013). It is also the only study thus far that investigates the effects



of repetition in a bimodal context. In the treatment stage, participants read a short story while listening to the audio-recording of these stories under one of the four treatment conditions. The treatment conditions varied according to the number of encounters with the target collocations: 1 encounter, 5 encounters, 10 encounters and 15 encounters. Prior to the treatment stage, the participants, who were Taiwanese university students ( $N = 161$ ), took a pretest in the format of multiple-choice. The 18 target collocations were verb-noun collocations that were (to varying degrees) semantically opaque (i.e., the meaning of the word string cannot be derived from knowledge of the component words) with a low degree of overlap in translation equivalency. In the pretest, the node word (i.e., the base word) was given, and the participants were required to circle the correct collocate from five options. Four posttests meant to tap into different aspects of collocational knowledge were administered after the treatment. The first posttest measured productive knowledge of form, in which participants had to write the collocates for the node words presented. Following that, a receptive knowledge of form test was conducted, which was identical to the pretest. The participants then took the third posttest, in which L1 meanings of the target collocations were provided, and the participants had to recall and supply the corresponding target collocations. In the fourth posttest the participants were required to write the meaning of the target collocations in L1. The results revealed that repetition had a positive effect on the learning of the collocations as more encounters led to more collocational knowledge. Specifically, it was revealed that incidental leaning of the form of collocations at receptive level may only occur after five or more encounters, while more than 15 encounters may be needed for effective learning of collocations at productive level.

While repetition was shown to have a significant effect in Webb et al.'s (2013) study, it should be pointed out that the input text was modified so as to include the necessary multiple instances of the target items. As noted by Pellicer-Sánchez and Boers (2018), this requires “resourcefulness on the part of the material designers” (p. 158). Such texts are also not ecologically valid, and teachers are unlikely to spend time artificially seeding an input material. This is especially so for authentic audio-visual materials, which are difficult to modify for the purpose of including multiple instances of the same MWEs. Therefore, one way to operationalise repetition in the context of L2 viewing is by exposing learners to the same input video twice, i.e., repeated viewings. Learners will then re-encounter the same MWEs in the same context as opposed to different contexts.

Durrant and Schmitt's (2010) study provides empirical support for the superiority of verbatim repetition over varied repetition for MWE learning when learners meet an MWE the

first couple of times. In their study, 20 low-frequency adjective-noun collocations were used as target items. For each of the target items, two sentences were created. There were also 40 matched control sentences. Two lists were then created, with each list containing the following: ten sentences containing a target word-pair, ten sentences containing only the noun part of the collocations, and 20 filler sentences containing other nouns. Eighty-four non-native learners of various L1s were assigned to one of the three conditions: single exposure, verbatim repetition and varied repetition. In the single exposure condition, the participants saw the 40 sentences from the first list once only. The sentences were presented once on a computer screen, and the participants were required to read each sentence aloud into a microphone. In the verbatim condition, the participants were also exposed to the first list of sentences, but the sentences were presented twice. The participants read each sentence aloud each time, with the second round of the process done at a faster rate. In the last condition, the varied repetition condition, the participants were exposed to both sets of the lists, with each target item appearing in two different sentences. Again, the participants read aloud the first list, before moving on to the second list in which the target items appeared in different sentences. A posttest was then administered. The participants were first presented with the adjective part of a target adjective-noun pair. This was immediately followed by the first two letters of the noun from the same pair. Participants had to say the noun if they knew the answer. The results revealed that participants were able to remember the nouns that were seen together with their adjectives more than the ones that appeared alone. It was also revealed that both the verbatim repetition and varied repetition conditions yielded superior levels of recall compared to the single exposure. Importantly, verbatim repetition led to higher gains than varied repetition. The researchers posited that reading different sentences including the collocations might have created extra cognitive burden and reduced the attention learners gave to the collocation proper. Thus, this finding provides support for operationalising repetition in the context of L2 viewing as repeated viewing, so as to reduce the learners' cognitive burden.

Although repetition was clearly shown to lead to superior collocational knowledge in Durrant and Schmitt (2010), it has to be conceded that the sentence-level task is different from the reading of 'texts'. As it is a lab-based study, it is also uncertain how engaged or aware the learners were of the purpose of the treatment. As emphasized earlier, learners' engagement with input differs according to their expectation, so the fact that the repetition conditions performed better could be partly be ascribed to the nature of task and the learners' engagement.

Further, the learners were tested immediately after being exposed to the treatment, so the durability of the effects of repetition is unknown.

There are more studies involving the effects of repetition on MWE, but these will be reviewed later as they involve other variables such as typographic enhancement. What is clear so far is that the handful of reading studies concerning the effects of repetition on MWEs have furnished evidence for the positive role of repetition. Several theories within SLA have been put forward to account for the effects of frequency. The first is usage-based theories. According to usage-based theories, “frequency is a driving force for language acquisition” (Wulff, 2018, p. 22). Token frequency and type frequency, however, perform different roles. The former refers to the number of times a particular construction is encountered in the input, while the latter refers to the amount of “distinct realizations” (Wulff, 2018, p. 22). Usage-based researchers argue that type frequency enables a strong entrenchment of a particular construction due to the varied opportunities afforded to parse the construction in question (Wulff, 2018; Bybee & Hopper, 2001). Related to this issue is another theory that can account for the effects of frequency, namely, the instance-based word learning framework (e.g. Bolger, Balass, Landen & Perfetti, 2008). According to this framework, each time a word is encountered, a contextualized episode occurs, and “memory traces from prior contexts may resonate or reactivate as fragments of decontextualized knowledge” (Bolger et al., 2008, p.127). Similar to what is proposed in usage-based theory, instance-based learning approach posits that in the context of learning word meanings, repetition of a single context will lead to stronger but fewer memory traces. Conversely, contexts that are more varied will provide more traces for the resonance of reactivation process. While these theories seem to suggest that varied repetition is superior to verbatim repetition, Bolger et al. (2018) argue that whether one type of repetition is better than the other depends on the “features of the new encounter in relation to the prior contexts” (p.127). As suggested in Durrant and Schmitt (2010), in the early stages of MWE learning, verbatim repetition seems to work better than varied repetition for the learning of MWE forms. As mentioned above, it is possible that experiencing the same MWE in a similar context may create less resonating memory traces than a random new context, as it may result in more cognitive burden. Additionally, it is challenging to establish conditions that guarantee varied exposures, or exposures to distinct realizations of the same MWE, especially when authentic materials are used in a non-longitudinal study (as is the case of the present study). Put differently, more than one encounter with a target MWE may not always be possible in a short input clip, be it in the same or different contexts. Indeed, there

have been no studies on the effects of repetition on MWE learning in the context of authentic L2 viewing. In fact, even when it comes to the effects of repetition on single word learning, this has only been investigated in three viewing studies. These studies yielded more consistent results in regard to the positive effects of repetition.

Rodgers' (2013) longitudinal study is one example of such studies. In one of the five experiments he conducted, 40 first-year Japanese students took a "Tough" and a "Sensitive" pre-test before watching 10 episodes of a television series, with or without captions, in the span of 12 weeks. The participants then completed the two vocabulary tests again after watching all the episodes. The Tough test was in the format of a multiple-choice in which the distractors shared aspects of form and meaning with the correct answer. The Sensitive test, on the other hand, had distractors that were different from the correct answer in terms of semantics and part of speech. The target items were 60 single words that appeared throughout the episodes with varying frequency of occurrence, between 5 to 54 occurrences. The results of the vocabulary tests for the caption group were then compared to the no caption group. For both the caption and no caption groups, a medium-sized correlation between the number of repetitions and word learning was found, but only for the Tough test. No significant correlation was found between frequency and vocabulary gains on the Sensitive test. It is important to note, however, that in Rodger's study repetition was operationalised as the recurrence of the same target item within and/or across the episodes. In the present study, the manipulation of frequency of occurrence is operationalised as repeated viewings of the same video.

Another viewing study which found an effect of frequency was conducted by Peters, Heynen and Puimège (2016). This study, however, compared the differential effects of subtitles (L1 captions) and captions on vocabulary learning. Two experiments were set up, in which learners were assigned to either the caption or subtitle condition. The two experiments differed in the following ways: learners' educational setting, number of viewings, type of audio-visual input, types of test and test administration. The participants in the first experiment were 28 L1 Dutch EFL learners from a general secondary school. Their vocabulary knowledge was assessed through a vocabulary size test comprising of four parts, which corresponded to four frequency bands. The vocabulary size test was carried out along with the pretests one week before the treatment stage. The pretests contained 50 test items and were administered in aural form only. The participants took a form recognition test first, in which they indicated whether they had seen or heard the given word before. This was immediately followed by a meaning recall test, where participants were required to provide the meaning of the word. The

participants then watched a 13-minute documentary clip twice, before they were tested on the target words through a posttest which was in the same format as the pretest. The results from Experiment 1 revealed that students in the caption condition recognised more target words than those in the subtitle condition. An interaction effect was also found between vocabulary size and types of subtitles. Specifically, the larger the vocabulary size of the participants in the caption condition, the higher their odds of a correct response. As for the meaning recall test, the type of subtitle did not influence learners' scores. Rather, the frequency of occurrence of an item was found to positively correlate with learners' correct responses. Similar to the form recognition test, a positive relationship was found between learners' vocabulary size and the odds of a correct response.

In Experiment 2, 18 participants from a vocational school watched a 20-minute episode of *The Simpsons* once, after taking the vocabulary size test and the pretest. The pretest differed from the first experiment in that other aspects of word knowledge were also assessed. The first test was a written form recall test where learners were asked to provide English definitions for Dutch words. This was followed by a written form recognition test, where learners had to tick words they had heard or seen. In the written meaning recognition test the learners had to choose the correct meaning of the English target words by ticking the correct translation from the multiple-choice given. The tests were repeated again as the posttest. The results revealed that the caption condition did better than the subtitle condition in the form recall test. Additionally, frequency of occurrence and learners' vocabulary size were also found to have a positive interaction with learning gains. Taken as a whole, both experiments showed that the items' frequency of occurrence within the videos positively influences the learning gains. However, as a significant interaction was also found between repetition and the learners' vocabulary size, there is a reason to suspect that the positive effect attested was related to learners' prior vocabulary knowledge.

The effects of prior vocabulary knowledge and repetition on vocabulary acquisition through L2 viewing were also recently investigated by Peters and Webb (2018). As well as repetition and prior vocabulary knowledge, the study also looked at two other factors, namely cognateness and word relevance. It is the first study to use a full-length programme, a one-hour long documentary, as the audio-visual input. Three different aspects of word knowledge were tested: form recognition, meaning recall and meaning recognition. To avoid test effects, the first two tests were taken by the participants in Experiment 1, while the meaning recognition knowledge was tested with a different experimental group in Experiment 2. In both

experiments there was a control group who were not exposed to the input material, but completed the same pretest, immediate posttest and delayed posttest as the experimental groups. All participants were EFL learners with Dutch as their L1. In Experiment 1, the participants first took a vocabulary knowledge test which was a frequency-based multiple-choice meaning recognition test. They then took a test which assessed form recognition and meaning recall. First, the participants had to tick whether they could recognise the target words that were presented both in written and aural forms. The participants were then asked to supply the meaning of the target words. The same test format was used as posttest and delayed posttest, which took place immediately after treatment and one week after treatment, respectively. The results from Experiment 1 indicated that there was no difference in the form recognition test between the experimental and control group. However, it was acknowledged that the results might have been confounded by the test format, as in the immediate posttest the participants were asked if they had heard the target words before, which they had in the pretest. Thus, the form recognition test was determined not to yield valid data to be further analysed. As for the immediate meaning recall test, the experimental group reported a significantly larger relative learning gain compared to the control group. Unfortunately, data for the delayed test hinted that deliberate learning may have happened between the immediate and delayed posttests, so the latter learning gains were not reported. As far as the other variables were concerned, a positive relationship was found between the word learning and prior vocabulary knowledge, frequency of occurrence and cognateness. In Experiment 2, meaning recognition was tested using a multiple-choice format, where participants needed to choose the meaning of a target item from four options, of which two were distractors, one was the correct definition of the target item and another was an “I don’t know” option. The same test format was used for the pretest, immediate posttest and delayed posttest. The procedures and input material were the same as Experiment 1. The results revealed that the relative learning gains of the experimental group were significantly larger than the control group. Similar to Experiment 1, data from the delayed posttest suggest potential test effects or deliberate learning on the participants’ part, and so these test data were not discussed further. As regards the effects of the variables, a similar pattern emerged: a positive correlation was found between word learning and frequency of occurrence, vocabulary knowledge and cognateness.

To summarise, the results of the viewing studies provide compelling evidence for the positive role of repetition in facilitating the uptake of single words. It is then conceivable that the same benefit may apply for the learning of MWEs. This possibility, however, has yet to be explored. Additionally, the reviewed viewing studies so far have operationalised repetition or

frequency of occurrence as repeated encounters with the same MWEs within the same input material or across different episodes in different contexts. As Durrant and Schmitt's (2010) study has shown, verbatim repetition might be more effective for MWE learning compared to varied repetition. In any case, it is unlikely that an MWE will appear several times within one input video. In essence, whether repetition operationalised as repeated viewing leads to better MWE gains remains to be investigated. Furthermore, of the viewing studies on the effects of repetition reviewed so far, only Rodgers (2013) compared the gains between caption and no caption groups. Interestingly, he found that repetition did not create any significant difference between the relative vocabulary gains for the caption and no caption groups. It would be interesting to find out if similar results may apply for MWE learning. As emphasized in the previous section, there are strong grounds to believe that the use of captions may facilitate MWE learning. Coupled with repeated viewing, the gains may be more substantial.

Having said that, the gains reported in the viewing studies are quite marginal, even with repetition. In Peters and Webb (2018), for example, viewing an uncaptioned video led to the meaning recall of 3.86 out of 48 unknown words and the meaning recognition of 3.73 out of 33 unknown words. Viewing captioned video did not lead to more substantial gains either. In Peters et al. (2016) for example, captioned viewing led to very few correct responses in the form recall test, as only 26 out of the 121 responses received were correct. Similarly, low gains were reported in Rodgers (2013), which found that seven-hour long (i.e., ten episodes of the same series) captioned viewing resulted in mean gains of 6.03 words. On the one hand, it could be argued that these relatively small gains may be partly due to the fact that learning took place under the incidental learning condition, providing a reason for captioned viewing to also be investigated in the context of intentional learning. On the other hand, this may point to the fact that captioning alone may not be enough to capture learners' attention to target phrases. In the literature on vocabulary acquisition, noticing has been established as one of the conditions that positively affects learning (Webb & Nation, 2017). To make noticing more likely, vocabulary studies have employed the use of typographic enhancement such as the bolding and underlining of target words. These studies, however, have largely been confined to reading-only input, as there have been very few studies on the effects of typographically-enhanced captions. It is therefore worthwhile to investigate whether MWE learning through repeated viewing may be further enhanced with the aid of typographic enhancement. This proposition seems quite plausible, as the existing studies on the effects of typographic enhancement have almost consistently furnished evidence of its positive role for vocabulary learning. These are the studies that will be addressed next.

#### **1.4 Typographic enhancement and vocabulary learning**

Typographic enhancement is one of the attention-drawing techniques that increases the chances that unknown words will be noticed and subsequently picked up. As posited by Schmidt (1990) in his Noticing Hypothesis, “people learn about the things they attend to and do not learn much about the things they do not attend to” (p. 30). The fact that typographic enhancement can bring about better learning is also supported by literature on input enhancement (Sharwood-Smith, 1993). Set in the realm of grammar acquisition, Sharwood-Smith’s input enhancement theory argues that grammar forms of interest should be salient as they would otherwise go unnoticed by learners. Many studies have since investigated the potential effectiveness of typographic enhancement for grammar acquisition. Findings from these studies, however, have been inconclusive, as some researchers have found positive effects (Alanen, 1995; Lee, 2007; Park, Choi & Lee, 2012), while others found no compelling evidence for the role of typographic enhancement (Izumi, 2002; Leow, 1997; Winke, 2013; Wong, 2003).

Many studies in the realm of vocabulary acquisition have also compared the effects of enhanced and unenhanced textual input to assess their usefulness on learners’ uptake of lexical items. The majority of such studies have focused on the acquisition of single words, and typically investigated in conjunction with input flooding (e.g., Rott, 2007), with the aid of the meaning of the enhanced words in the text itself (e.g., Kim, 2006) or added in glosses (e.g., Ko, 2012; Hulstijn, Hollander and Greidanus, 1996; Watanabe, 1997; De Ridder, 2002). Studies concerning the effects of typographic enhancement on the acquisition of MWE are relatively few, and, similar to the reviews in the previous section, these are studies concerning collocations.

One of the few studies that investigated the effects of typographic enhancement on MWEs and found positive effects was recently conducted by Boers et al. (2017). In their study, intake was operationalised as the ability to identify MWEs, which participants had encountered during the treatment, in the posttest. Apart from investigating whether typographic enhancement of MWEs promoted their intake, another focus of the study was whether typographic enhancement would also foster the intake of other, unenhanced MWEs. To this end, two experiments were conducted. The participants were Dutch L1 university students. There were 38 participants in the first trial, and 43 participants in the second trial. In both



experiments, participants read two texts under one of the three conditions: with 16 target MWEs underlined, with only half underlined, or with none underlined (control). The target MWEs included verb-noun collocations, adjective-noun collocations, nominal compounds and prepositional phrases. Following the reading of the first text, the participants were assessed on their recognition of the MWEs that appeared in the text by presenting them together with a synonymous phrase (e.g., *junk food* vs. *unhealthy food*). The participants had to determine which one of the two options appeared in the text. A second text was then presented to the participants after the MWE recognition task. It was expected that based on their experience on the first test, participants who only had half of the MWEs enhanced would extend their attention to the other unenhanced MWEs. The reading text was again followed by a test on MWE recognition. The second trial was conducted with a different group of participants using the same materials and procedure, but with the order of the texts reversed. The results of trial 1 revealed that the participants who saw half of the MWEs enhanced were able to recognise them better than both the control condition and the all-enhanced group, who saw twice as many enhanced MWEs. While it did not approach statistical significance, it was also found that the half-enhanced condition performed worse for the unenhanced MWEs compared to the baseline condition. Similar results were found in trial 2. The all-enhanced condition significantly outperformed their counterparts. Again, the half-enhanced condition did better in the recognition of the MWEs enhanced in their condition compared to the control group. However, out of three conditions, the half-enhanced condition recorded the poorest recognition of the unenhanced MWEs. In sum, it was found that while typographical enhancement led to the noticing of enhanced items, it might be at the expense of unenhanced MWEs.

The posttest results of Boers et al. (2017) provide indirect evidence that typographic enhancement facilitates noticing. Their findings complement those reported in an earlier study by Bishop (2004), who provided more direct evidence that typographic enhancement attracts attention. Using a computerised task, Bishop sought to determine whether unenhanced MWEs would be clicked less for glosses compared to the more perceptually salient ones. Forty-four participants of various L1s were assigned to read an English passage with typographically-enhanced MWEs or with plain MWEs. All the target forms were provided with glosses when clicked. Prior to treatment, the participants took a vocabulary knowledge scale (VKS) test, which elicits self-reported knowledge of the extent to which a word is known. After reading the treatment passage, participants in both conditions answered 20 true or false items with the reading passage at their disposal. The results indicated that typographically-enhanced MWEs were consulted for glosses more than the unenhanced MWEs. It was also revealed that

participants who read under the enhanced condition scored higher in the comprehension test compared to their counterparts, but the difference was not significant. It is important to interpret the results with caution, as the experiment mainly focused on the clicking of target items. In other words, it is the noticing of target forms that is being assessed here, rather than the learning of new MWEs. A VKS posttest, and a comparison of the pretest-posttest gains would have been useful in establishing whether clicking more, i.e., more consultation with the glosses, would lead to better collocational knowledge. Put differently, although Bishop's (2004) study furnished evidence that typographic enhancement leads to noticing, it is unknown whether it leads to more MWE knowledge.

A study that did show that typographic enhancement promotes the noticing *and* acquisition of MWEs was recently conducted by Choi (2018). Using eye movement technology, two hypotheses were put to test. Firstly, as opposed to unenhanced collocations, enhanced collocations were hypothesised to attract more attention which leads to learning. Secondly, it was predicted that the recall of unenhanced text would be impaired owing to the increased attention focused on the enhanced collocations. Thirty-eight undergraduate Korean students were randomly assigned to read the treatment passage with enhanced text or baseline text. In the enhanced text, the 14 target collocations appeared boldfaced. Prior to the treatment stage, the participants underwent a battery of tests which included a reading test and a collocation pre-test, whereby participants were presented with 14 Korean sentences and an English equivalent with two blanks. With the first letter given, the participants were required to fill in the two blanks. A week later, the participants read the treatment text with an eye-tracking device recording their eye movements. A recall task in cloze format was then administered. A week later, an unannounced collocation post-test in a similar format to the pre-test was carried out. The results revealed that similar reading times between the two conditions were recorded for the known collocations. As for the unknown collocations, the enhanced text condition recorded a statistically-significant longer fixation time compared to the baseline text condition. The enhanced text condition also showed significantly superior performance compared to their counterparts in the collocation posttest. The recall test results further revealed that compared to the baseline text condition, the enhanced text condition recalled 39% more target collocations. However, the baseline text condition was found to outperform the enhanced text condition by 48% in the recall of unenhanced content words. In other words, Choi's study provides evidence that typographic enhancement leads to superior MWE uptake, but that it may come at the expense of learners' uptake of other elements of the text. Choi posited that the

trade-off effect may be explained by the cognitive load theory (Sweller, 1994). In essence, learners have limited cognitive resources and spreading these resources to the encoding of enhanced lexical items while simultaneously taking in the unenhanced text may be challenging. Though not part of the main research questions, the present study also included comprehension tests to assess whether learners' uptake of the content of the input videos is indeed impaired when typographically enhanced captions are present.

The studies reviewed so far focus on how typographic enhancement affects explicit knowledge of aspects of MWEs, by means of tests which measure explicit knowledge. Explicit knowledge is commonly described as conscious and analysable (Ellis, 1993). Implicit knowledge is the type of knowledge that is intuitive and unanalysed (Ellis, 1993, 1999). It has been suggested that knowledge of MWEs may be mostly implicit (Durrant & Schmitt, 2010; Ellis, 2004). This has prompted recent studies to also include the implicit dimension of MWE knowledge in their investigation of the role of typographic enhancement. Sonbul and Schmitt (2013) is an example of such studies. Two different experiments were carried out to assess the effects of different learning conditions on the implicit and explicit knowledge of collocations. The participants for the first experiment were native speakers of English, while the second experiment was conducted with learners of L2 English. In each experiment, the participants saw 15 medical collocations. Ten collocations appeared three times in a passage. Five of the collocations appeared unenhanced (enriched), while the other five were bolded (enhanced). In the third deliberate and learning condition, five collocations were presented without context using PowerPoint presentations, with each collocation shown for 10 seconds. A control group who read the baseline text was also included in each experiment. To assess the effects of each type of exposure on the implicit and explicit knowledge of collocations, two explicit tests and one implicit test were carried out. Using a priming task on E-prime, implicit collocation knowledge was measured by asking participants to determine whether a string of letters which followed the first word of a collocation was a real English word. This was followed by the first explicit test which measured form recall, where participants were required to provide the first word of the collocation with the meaning provided in the margin. In the next explicit test, form recognition was measured by providing five multiple-choice options for the same questions that appeared in the first test. Two weeks later the delayed explicit and implicit knowledge tests were carried out again. The results from the second experiment (i.e., with L2 participants) revealed no gains in implicit knowledge. However, all three types of input conditions led to positive gains in form recall and form recognition. Additionally, the collocations in the

enhanced condition were learned better compared to the enriched condition. This is consistent with the findings of Boers et. al (2017), Bishop (2004) and Choi (2018) in which typographic enhancement was found to yield superior MWE gains. Further, Sonbul and Schmitt found that the enhanced condition led to similar gains as the decontextualised condition. In other words, their study suggests that employing typographic enhancement is as effective as deliberate, decontextualised learning. Another interesting finding was that the decontextualised condition did not lead to higher gains compared to the enriched condition. This suggests that multiple encounters in an incidental learning condition can lead to similar results as a single encounter in intentional learning.

The conceptual replication of Sonbul and Schmitt's study by Toomer and Elgort (2019) also provides clear evidence of the positive effects of typographic enhancement for fostering explicit MWE knowledge. Since implicit knowledge was not found in Sonbul and Schmitt's study, Toomer and Elgort attempted to fill the gap by extending the independent variables. Among the extensions were more learning sessions and a larger number of exposures through multiple texts. Sixty-two upper-intermediate to advanced ESL speakers were exposed to nine occurrences of 15 medical collocations in three learning sessions that lasted for two days. Of the 15 collocations, five were bolded and provided with glosses, five more were bolded without glosses, and the other five appeared without bolding and glosses (i.e., reading-only), creating the three learning conditions. Following the treatment stage, one procedural knowledge test and two declarative knowledge tests were carried out. The first test, a test of procedural knowledge, was a lexical decision task. Reaction times were used to determine participants' procedural knowledge. Faster reaction times for target collocation were assumed to reflect procedural knowledge of target collocations. Subsequently, a cued recall test in the form of sentence gap-filling was administered. The second word of the collocation, along with the meaning, was provided and participants had to provide the first word. In the last test, a form recognition task was administered with multiple-choice options provided. Two weeks later, delayed posttests of the same formats were administered. The results revealed that compared to the collocations encountered in the reading-only condition, those encountered in the bolding condition were significantly more likely to result in correct responses in both the form recall and form recognition tests. This corroborates Sonbul and Schmitt's findings, and lends further support to the role of typographical enhancement as a means of promoting explicit collocational knowledge through repeated occurrences. Toomer and Elgort also found that prior knowledge affected form recall but not form recognition. Additionally, contrary to Sonbul and Schmitt,

Toomer and Elgort found evidence of implicit knowledge for the reading-only condition. In summary, bolding was found to produce explicit knowledge, while implicit knowledge was found to develop in the absence of bolding and glossing. Toomer and Elgort conjectured that this absence of interruption enabled smooth reading with attention allocated only to the forms of collocations. As regards the observed priming effect which was not found in Sonbul and Schmitt's study, Toomer and Elgort posited that a larger number of occurrences, i.e., nine vs. three, may account for this difference. This once again emphasizes the fact that frequency of occurrence or repetition matters in the development of implicit and explicit knowledge of MWEs.

As both repetition and typographic enhancement have been shown to facilitate the uptake of MWEs, it is conceivable that a combination of these factors will lead to even bigger gains. This possibility was investigated by Szudarski and Carter (2016). Forty-one Polish students read six stories containing ten verb-noun collocations and ten adjective-noun collocations. To assess the effects of repetition, which they termed as input flood, half of the target items appeared six times while the other half appeared 12 times. The participants read the stories under one of two conditions, i.e., with the target items typographically-enhanced or not enhanced. There was also a control group that only took part in the pre-test and post-test. Prior to the treatment phase, a Vocabulary Levels Test (VLT) was administered, along with a pre-test on the target collocations. The five pretests were in the same format as the delayed posttests. The first three tests were different variations of form recall tests, which were followed by a meaning recall test and a form recognition test respectively. The results revealed that participants who saw the typographically-enhanced items (i.e., input flood plus input enhancement) significantly outperformed those who read the unenhanced versions (i.e., input flood only), but only in the form recall and form recognition tests. The lack of gains in the meaning recall test was attributed to the fact that learners may only notice the form, without attending to the meaning. The researchers also posited that the lack of semantic gains may be due to the absence of in-depth processing. Unexpectedly, the input-flood only group did not make any gains, suggesting that repetition alone may not be sufficient for the development of MWE knowledge. Repetition in conjunction with typographic enhancement, however, showed inconsistent results. Encountering 12 enhanced collocations resulted in significantly higher scores in form recall compared to encountering six enhanced collocations. For form recognition, however, six enhanced encounters led to significantly higher scores compared to

12 enhanced encounters. This prompted the researchers to suggest that input enhancement might reduce the number of encounters required to make MWE gains.

The studies reviewed so far have consistently pointed to the effectiveness of typographic enhancement in boosting MWE learning, especially when combined with repetition. A common denominator of these empirical studies is the fact that the effects of typographic enhancement were investigated under the incidental learning condition. If learning gains made under intentional learning conditions are more substantial than in intentional learning (Webb, 2002; Laufer, 2003), then it stands to reason that the presence of typographic enhancement coupled with test awareness may further boost MWE learning. However, there have been limited intentional learning studies that have made use of typographic enhancement in their input material.

Although the express purpose was not to investigate the effects of typographic enhancement, Peters (2009) is one of the few studies that looks at the effects of test announcement on the learning of enhanced and unenhanced new vocabulary. Using a between-subject design, Peters investigated the effects of instructions, which differed in specificity, on the learning of novel words and collocations. In group 1, the participants were instructed to focus on new vocabulary in general, while in group 2 the participants were instructed to pay attention to both new vocabulary and collocations. Both groups were explicitly informed that they would be tested on the items that they were required to pay attention to. The treatment passage contained 78 marginal glosses, of which 19 were for the target single words and 21 were for the target collocations. All items were underlined in the treatment passage. Prior to the treatment stage, the 54 advanced EFL learners took a pretest which required them to provide English translations to 40 target items and ten distractors. This was immediately followed by the treatment stage. A posttest which was identical to the pretest was then administered. The results revealed that while both groups showed significant gains, collocations were recalled significantly better than the target words. Importantly, participants who performed the collocation-oriented task and who were aware of the vocabulary test did not recall significantly more target collocations compared to their counterparts. The explanations offered for the lack of differences between the two groups included the fact that collocations were made salient in both groups via the marginal glosses. Although this was not posited by the author, another possible reason might be that the use of typographic enhancement made the collocations stand out as much as the other single words, resulting in both groups engaging with the collocations to the same extent. In other words, the results of this study suggest that the use of typographic

enhancement alone may be as effective as instructing learners to focus on enhanced MWEs. Further, there is reason to suspect that explicit instruction on its own may not result in learners noticing the target phrases.

This was echoed by the findings of Peters (2012), which was a conceptual replication of her previous study. Similar to her earlier study, Peters (2012) sought to investigate the effects of instructional method and typographic salience on the recall of L2 German single words and MWEs. A 3 X 2 factorial design was employed, with the instructional method being the between-subject variable, and typographic salience and type of target items as the within-subject variable. In the experimental condition, the participants were told to focus on and write down single words and MWEs. In the control group, the participants were only instructed to pay attention to and write down new vocabulary, with no explicit mention of collocations. Both groups were informed that a vocabulary test would follow, before reading a text containing 12 target words and 12 formulaic sequences. Half of the target words and MWEs appeared boldfaced and underlined, while the other half were unenhanced. All targets items were provided with L1 translations in the margin. With the text at their disposal, participants were asked to summarise the text after reading it. The 28 participants, whose L1 was predominantly Dutch, then took an immediate posttest, in which they were required to supply German translations to the 24 target items. An identical posttest with the order of items randomised was carried out two weeks later. In line with the findings of her earlier study, Peters found that instructing learners to pay attention to MWEs did not result in superior learning of the target items. Typographic salience, however, had a positive effect as the enhanced items were learned better compared to the unenhanced ones. However, this was only observed in the immediate posttest.

In summary, both of Peters' studies illustrate that typographic enhancement makes test announcement redundant, at least if the learners correctly interpret the enhancement as an effort on the teacher's part to direct their attention to certain language features. A substantial amount of MWEs are made up of comparatively frequent words (Martinez & Murphy, 2011; Nation, 2013), so even with the presence of test announcement, learners may not notice the word strings without typographic enhancement. Even proficient learners have been found to struggle with identifying MWEs in written input (Eyckmans, Boers & Stengers, 2007). Therefore, the use of typographic enhancement increases the chances of MWEs being picked up, especially when learners are aware that they will be tested.

Given the positive benefits of typographic enhancement for the development of MWEs, it is surprising that it has only been investigated for unimodal reading texts, and not yet when the text is part of multimodal input. In other words, there have been limited studies on how MWE learning through L2 viewing can be facilitated through the use of typographically-enhanced captions. To date, only two studies are known to have explored how salient captions could facilitate vocabulary acquisition. The first study was by Montero Perez, Peters, Clarebout and Desmet (2014). Three different caption conditions varying in the amount of text and lexical saliency were set up to investigate their effects on incidental L2 French vocabulary learning and content comprehension. The three experimental groups were: full captioning, keyword captioning and full captioning with highlighted keywords. In addition, there was also a control group in which participants watched the clips without any captions. The 17 French target words included seven nouns, four verbs and six MWEs. Prior to the treatment phase, the L1 Dutch participants took a vocabulary test in multiple-choice VKS format. One month later, three clips ranging from approximately two to four minutes were shown twice to the participants under their respective conditions. After each clip, a comprehension test was administered. After the third comprehension test, four vocabulary tests were then carried out. In the form recognition test, learners were required to indicate whether the words presented to them appeared in the clip. In order to reduce the effect of guessing, a clip association question was added, which had the participants indicate which clip contained the word they claimed to have seen. This was followed by a meaning recall test in the format of VKS, similar to the pretest. A meaning recognition test with multiple-choice options then followed. Contrary to their hypothesis, no difference among the conditions was found in the comprehension test. This finding ties in with Mohd Jelani and Boers's (2018) study, which also found that the no caption condition performed as well as the captioned condition in the comprehension test. With regard to form recognition, the caption conditions significantly outperformed the control group. However, the hypothesis that keyword caption groups would outperform the full caption group on the form recognition and clip association test was not supported. Put differently, saliency was not found to be as crucial as the availability of captions in developing form recognition knowledge. The keyword caption conditions, however, did significantly outperform the no captions group in the meaning recognition test. As for the meaning recall test, the caption conditions did not show significant difference in scores when compared to the control group, perhaps due to the low scores in all the treatment conditions. Additionally, a positive correlation was found between participants' vocabulary size and their text comprehension test scores as well as their vocabulary test scores. In summary, typographic enhancement led to superior knowledge in



meaning recognition, but not form recognition. That form recognition was not better for the keyword caption groups led the authors to posit that saliency did not enhance noticing because forewarning the students of a comprehension test may have directed learners' attention to the target words already. Moreover, as noted by the authors, the target words were crucial for comprehension, and so the learners may have considered them important even when they were unenhanced.

Indeed, as emphasized before, the conditions under which learning takes place matters. The effects of typographic enhancement have been investigated almost exclusively under the incidental learning condition. This is usually done through a test announcement on comprehension, so as to ensure that learners pick up the new words incidentally. While Choi has shown that typographically-enhanced written text draws more attention than unenhanced text, the same is not known for audio-visual input. As shown in Montero Perez et al. (2014), learners may also focus on unenhanced captions even when a vocabulary test is not announced. A better way to assess learners' engagement and fixation with captions is through an eye-tracking study. Montero Perez et al.'s (2015) study is one that employed eye-tracking to investigate the effects of types of captioning as well as test announcement. In their study, the L2 French participants watched the input clips once, either with captions or with keyword captioning. Under each caption condition, half of the participants were informed of the vocabulary tests, while the other half were unaware of the tests. Six weeks prior to the treatment phase, the participants took a vocabulary size test in a format similar to their earlier study (i.e., Montero Perez et al., 2014). An eye-tracking device recorded participants' eye-movements while they watched the clips that contained the French target items which included six verbs, seven nouns and five MWEs. The participants then completed a comprehension test, a form recognition test, and a meaning recall test. All the tests were in a similar format to the previous study. The results revealed that type of captioning had a significant effect only on form recognition as the keyword caption groups were found to outperform the full caption groups. This runs counter to the finding in their earlier study where no significant difference on the word recognition test was found between the full caption and keyword caption groups. The researchers attributed the difference to the amount of viewing as participants in this study only viewed the clips once. Consistent with the results of the authors' earlier study, type of captioning did not have an effect on meaning recall. As regards the effect of test announcement, no significant difference was found between the incidental and intentional groups in the form recognition test. However, participants under the intentional conditions showed superior

knowledge in meaning recall compared to the incidental learning groups. Additionally, participants under the intentional learning conditions recorded longer fixation times compared to their counterparts. Further, under the intentional conditions, the keyword caption group fixated longer on target words compared to the full captions group.

Although visual salience was credited for learners' prolonged focus, it has to be pointed out that visual salience in the study was realised through keyword captioning. Thus, it is unknown whether salience brought about by typographic enhancement may lead to same level of engagement and the same extent of learning. What is clear from the study, however, is the fact that learners alter their engagement with the target items when they are aware of impending vocabulary tests. Therefore, it would be interesting to compare how typographically-enhanced captions would affect vocabulary learning under both learning conditions. The exploration of the role of typographic enhancement under both incidental and intentional learning may also shed light on the *type* of noticing given to the target items under the respective learning conditions. The weaker version of Schmidt's (1994) Noticing Hypothesis states that more noticing or awareness leads to more learning. There are, however, different levels of awareness. Noticing, for example, happens at a low level of awareness (Schmidt, 1995, 2001), otherwise known as noticing at the level of attention (Godfroid et al., 2010). Noticing could also happen at the level of awareness (Godfroid et al., 2010), where there is realisation that a particular feature is being paid attention to (Godfroid & Schmidtke, 2013). Tomlin and Villa (1994) posit that attention can be separated into three different components, i.e., alertness, orientation and detection, and that 'acquisition requires detection, but such detection does not require awareness' (p. 199). In other words, Tomlin and Villa suggest that learning can take place without awareness. They further argue that input enhancement may represent one way of increasing the chances of detection. Taken together, these theories suggest that the effectiveness of typographic enhancement on lexical acquisition may be influenced by different learning conditions. The effect of typographic enhancement under the intentional learning, for instance, may be more even more pronounced as compared to under incidental learning. This is because when learners anticipate that the enhanced MWEs could be targets in a forewarned test, the learners may be more aware of what they are paying attention to.

In short, investigating the roles of captioned viewing and typographic enhancement under both learning conditions is worthwhile not only because there has been limited research in this area, but also because of existing frameworks that suggest that these variables may have differential effects under the different learning conditions.

## 1.5 Summary

The previous sections have indicated a need for further research concerning the effects of several variables on the acquisition of MWEs. Specifically, while repetition and typographic enhancement have been known to be favourable for the development of MWE knowledge, they are often investigated under incidental learning conditions. This begs the following question: does exposure to enhanced MWEs under intentional learning conditions lead to superior knowledge compared to the same exposure under incidental learning conditions? Further, when compared to incidental learning conditions, will repetition under intentional learning conditions lead to bigger gains, or will it appear to be redundant? Additionally, does the combination of repetition and typographic enhancement accelerate learning? These questions are pertinent to the ways in which these interventions could be used so as to maximise MWE learning without massive amounts of time being invested. This is especially relevant since there is only a finite amount of classroom time in most ESL contexts. For example, if exposure to typographically-enhanced items *once* leads to similar incidental gains as exposure to unenhanced target items *twice*, then there is good reason to advocate for the use of typographic enhancement in the interest of preserving precious classroom time.

Importantly, little is known about how MWE learning could be facilitated through L2 viewing. The viewing studies on single word acquisition have shown that audio-visual input does fuel lexical development, so the same positive benefit may extend to the learning of MWEs as well. Additionally, whether captioned viewing is further reinforced through the use of typographic enhancement and repetition remains to be investigated. This then brings us to the same questions pertaining to the effectiveness of these interventions under the two learning conditions. In essence, how L2 viewing could be optimised in order to yield maximum MWE gains under both types of learning conditions needs to be investigated. This is the aim of this research project. The next chapter outlines the specific research questions as well as the methods used to answer the research questions.



## **CHAPTER TWO: METHODOLOGY**

This chapter starts off with the specific research questions that the research project aims to address. This is then followed by the methodology used to answer the research questions, starting with details on the research participants. Section 2.3 then describes the materials used in the study, such as the videos and their vocabulary load. Section 2.4 provides information on the target MWEs and the process of narrowing down the set of target MWEs. Section 2.5 describes the test instruments. This is then followed by Section 2.6, which details the procedure of the study. This chapter concludes with a section on scoring criteria and the scoring procedure.

### **2.1 Research questions**

The previous chapter has highlighted gaps pertaining to research concerning pedagogical interventions meant to boost MWE learning. In particular, little is known about whether L2 viewing of authentic materials promotes MWE knowledge. Further, the role of typographic enhancement and repetition in the context of L2 viewing has yet to be proven as facilitative for MWE development. Moreover, how these factors affect MWE learning under intentional and incidental conditions merits investigation. Guided by these gaps, the research project seeks to answer these specific questions:

1. Is there an effect of caption condition (i.e., no captions, normal captions, enhanced captions) on the incidental and intentional learning of MWEs?
2. Is there an effect of repeated viewing on the incidental and intentional learning of MWEs?
3. Does repeated viewing influence incidental and intentional uptake of MWEs to the same degree under all three caption conditions?

Two studies were undertaken to answer these research questions. Study 1 concerns incidental learning conditions, while Study 2 focuses on intentional learning conditions. The possibility of learners expecting a test was pre-empted by a test announcement before the treatment phase of Study 2. Repetition was operationalised as repeated viewing in this research project as it was impossible to control for the frequency of occurrence of an item in an authentic video. Further, compared to varied repetition, verbatim repetition has been shown to be more effective for MWE learning (Durrant & Schmitt, 2010), providing a stronger reason for

repetition to be operationalised as repeated viewing. The subsequent sections provide details on the methodology used to seek answers to these questions.

## **2.2 Participants**

### ***2.2.1 Participants in the actual study***

The data for the research project was collected in Malaysia. There were originally 133 Malaysian ESL learners recruited for this research project. The same participants were used in both studies, but some had to be excluded depending on their completion of the treatments and tests. In Study 1 there were 122 participants (75 females and 47 males) with complete data, while in Study 2 there were 126 participants (78 females and 48 males). The participants were between the ages of 17 and 22. They were first-year and second-year Diploma students in various fields such as Hospitality Management and Information Technology Systems. They were predominantly Chinese with either Malay or Mandarin as their L1. A summary of the participants' information with regard to age, gender, and first language is attached in Appendix 1.

All the participants had formally learned English for at least ten years. Their vocabulary size test (VST) (Nation & Beglar, 2007) scores were between 4,000 and 9,000 most frequent word families in English, with a mean of 6,600. The participants came from six intact classes, and these intact classes were randomly assigned to one of six experimental conditions, differing in the numbers of exposure to the video (i.e., one or two viewings) and caption condition (i.e., no captions or uncaptioned, normal captions or enhanced captions). The six conditions will also be referred throughout the thesis as UC1, NC1, EC1, UC2, NC2 and EC2, with the numbers indicating the number of viewings, while the letters UC, NC and EC indicate the uncaptioned, normal captions and enhanced captions, respectively. There was no significant difference between the six groups in their scores on the VST (for both studies: Kruskal-Wallis  $\chi^2(5) = 4.57; p = .47$ ). The number of learners under each condition and their mean VST scores are summarised in Tables 1 and 2 below, for Study 1 and Study 2, respectively. The total VST scores (out of 140) of the participants were used as the indicator of their vocabulary knowledge in the study.

**Table 1. Number of participants in Study 1 and the mean VST score (out of 140) for each condition**

Condition	<i>N</i>	Mean ( <i>SD</i> )
Enhanced Captions X 1	19	68.37 (11.31)
Normal Captions X 1	23	73.04 (18.38)
Uncaptioned X 1	15	74.40 (16.63)
Enhanced Captions X 2	25	65.92 (16.75)
Normal Captions X 2	22	71.09 (11.49)
Uncaptioned X 2	18	66.39 (14.02)

*Note: 1 = one viewing; 2 = two viewings.*

**Table 2. Number of participants in Study 2 and the mean VST score (out of 140) for each condition**

Condition	<i>N</i>	Mean ( <i>SD</i> )
Enhanced Captions X 1	20	68.90 (11.3)
Normal Captions X 1	24	73.70 (18.27)
Uncaptioned X 1	15	74.40 (16.63)
Enhanced Captions X 2	25	65.92 (16.75)
Normal Captions X 2	23	70.17 (12.06)
Uncaptioned X 2	19	66.05 (13.70)

*Note: 1 = one viewing; 2 = two viewings.*

Apart from the VST, Schmitt, Schmitt, and Clapham's (2001) Vocabulary Levels Test (VLT) (Version 2) was also administered. The VLT scores were used to assess learners' lexical coverage of the input videos. In its original version, the VLT has five levels: the 2,000 (2K), 3,000 (3K), 5,000 (5K), 10,000 (10K) levels and the Academic Word List (AWL). However, the AWL was not used in the study as the VLT was used for the purpose of measuring learners' knowledge of general high-frequency vocabulary instead of academic vocabulary. The 10K level was also removed as the learners in the study were unlikely to have mastered this level. A score of 24 out of 30 was used as a cut-off point for showing mastery of a level. This cut-off point was decided upon because it has precedence as an indicator of mastery of a level. For example, this cut-off score was used in Xing and Fulcher's (2007) study of the reliability of

the two versions of the VLT. The same cut-off score was also used in Rodger's (2013) study of English language learning through viewing television. Table 3 presents the results of the three levels of the VLT including the mean scores of each condition. As can be seen from the table, all six conditions had mean scores that would indicate mastery of the 2,000 word level. As for the 3,000 word level, the mean scores of the three groups, i.e., NC1, UC1 and NC2, suggested mastery of the 3K level, while the other three groups approached mastery.

**Table 3. The mean VLT score for each condition (for both studies)**

	VLT 2000		VLT 3000		VLT 5000	
	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>
Enhanced Captions X 1	26.00	3.51	21.89	4.46	16.32	4.62
Normal Captions X 1	26.70	3.24	24.35	5.31	18.13	8.92
Uncaptioned X 1	27.13	2.97	24.00	5.07	18.07	6.63
Enhanced Captions X 2	25.60	3.52	22.00	5.55	15.96	7.38
Normal Captions X 2	26.91	5.02	24.27	5.00	18.05	5.91
Uncaptioned X 2	25.59	4.91	22.11	5.70	15.67	5.93

### ***2.2.2 Participants in the pilot study***

Twenty participants (15 females and 5 males) between the ages of 19 and 23, whose backgrounds were similar to the participants in the actual study, took part in a pilot study. The purpose of the pilot study was twofold. The first purpose was to ascertain whether the input videos were easy to follow and were interesting enough to be watched at least twice. Although the actual study only compared the outcome of viewing once and twice, the pilot study sought to find out whether three viewings were also feasible. For each video, five participants watched the video twice, while the other half had three viewings. The participants watched the video with either captions or enhanced captions. The pilot study did not include uncaptioned viewing because the point of interest was how learners react to the use of captions and enhanced captions. The second purpose was to ascertain whether there were any major issues regarding the test instruments, i.e., the form recall pretest and immediate posttests and the comprehension tests (see Section 2.3 below). The form recall pretest was administered two weeks before participants watched the video. Following the viewing session, the participants took the comprehension test and the immediate form recall posttest. A short interview session was also conducted with some of the pilot participants to gain insight into their experience.



The pilot study revealed that the participants found it hard to sustain their interest for a third viewing. This feedback was then used to inform the design of the study, in that the comparison would only be made between one and two viewings. As regards their responses to the form recall test instrument, the information was used to revise the test items accordingly. Statistical analyses were not carried out to analyse the data as there were too few participants under each condition. The interview sessions, however, corroborated that they had no issues with the format and the clarity of the majority of the test items. The participants also revealed that they were able to follow the video regardless of the caption conditions. Additionally, they reported the video to be interesting (and interesting enough for a second viewing) and comprehensible. The following section describes the input videos in detail.

## **2.3 Materials**

When considering the input materials for this research, a few factors had to be taken into account. While the majority of previous studies on captions used educational clips (e.g., Markham, 1999; Peters et. al, 2016; Winke et. al, 2010; Winke, Gass & Sydorenko, 2013) such as documentaries as the input material, it was felt that an educational video might not sustain the learners' interest long enough for them to watch the video from start to finish, or for a second viewing. Apart from being easy to follow in terms of lexical coverage and speed of dialogue, the videos had to be interesting enough for motivational purposes. Another factor that heavily influenced the choice of input material was the number of MWEs contained in the videos. For a sufficient number of (new) MWEs to be encountered, the videos needed to be of considerable length.

Taking into account the aforementioned factors, two episodes from two different American comedy series were chosen. American sitcoms were chosen as the participants were more accustomed to an American accent as opposed to a British accent. Further, Lin's (2014) study of internet television found that the comedy genre contained a high amount of everyday spoken formulaic sequences. In addition, caption studies which used comedy series (e.g., Peters et al., 2016; Sydorenko, 2010) have reported favourable findings for word learning. This provided empirical grounds for choosing sitcoms as the genre of the input material. The following section provides a synopsis of the videos chosen, followed by their vocabulary load.

### ***2.3.1 Video 1: Fresh off the Boat (Season 1, Episode 2)***

As suggested by the title, this series follows the journey of a husband and wife who originally hail from Taiwan. The series revolves around the parents and their children, who are

second generation immigrants, trying to find a balance between fitting in with their American peers and honouring their traditions and roots. The chosen episode was 20-minutes long. The plot of the episode centres around the mother who handles the family business and deals with her children with values that are characteristic of an Asian ‘tiger mother’—a demanding figure who scrutinises and micro-manages what her loved ones do (or don’t do). This particular episode was chosen as it was felt that the participants would be able to relate to certain scenes in the movie, such as hiding their exam marks or being sent to extra classes after school by a stern Asian mother. Interviews conducted with the actual and pilot participants revealed that the participants did indeed enjoy the video and could follow the storyline.

### ***2.3.2 The vocabulary in Video 1***

The vocabulary in Video 1 was analysed to ensure the suitability of the video in terms of lexical coverage. The transcript of the video was first downloaded and double-checked against the video. Next, hyphenated words, contractions and reduced forms were changed to their full form in line with the spellings used in Nation’s (2004) BNC word lists. The transcript was then run through RANGE (Nation & Heatley, 2002). The results of the analysis are shown in Table 4.

**Table 4. Lexical coverage of Video 1**

<b>Word list (1,000)</b>	<b>Cumulative coverage (without proper nouns and marginal words)</b>	<b>Cumulative coverage (with proper nouns and marginal words)</b>
	<b>%</b>	<b>%</b>
<b>1,000</b>	87.88	91.16
<b>2,000</b>	91.59	94.87
<b>3,000</b>	93.09	96.37
<b>4,000</b>	94.10	97.38
<b>5,000</b>	94.55	97.83
<b>6,000</b>	94.93	98.21
<b>7,000</b>	95.27	98.55
<b>8,000</b>	95.50	98.78
<b>9,000</b>	95.69	98.97
<b>10,000</b>	95.84	99.12
<b>11,000</b>	95.95	99.23

<b>12,000</b>	95.95	99.23
<b>13,000</b>	96.03	99.31
<b>14,000</b>	96.07	99.35
<b>Not in the list</b>	96.72	100.00

As can be seen from Table 4 the results of the analysis showed that, together with proper nouns and marginal words, the most frequent 1,000, 2,000 and 3,000 word families provided 91.16%, 94.87% and 96.37% cumulative coverage of the script's total running words, respectively. In literature, a range of 90% to 99% of lexical coverage has been proposed to be sufficient for adequate comprehension, depending on the discourse (Hu & Nation, 2000; Laufer, 1989; Nation, 2006; Stæhr, 2009; Schmitt, Jiang & Grabe, 2011), and 95% to 98% for learners to stand a good chance of guessing the meaning of as yet unfamiliar words (Liu & Nation, 1985; Nation, 2001; Webb & Rodgers, 2009a, 2009b). For listening comprehension, for example, van Zeeland and Schmitt (2013) found that coverage of 90% to 95% provides adequate listening comprehension to L2 learners. In his investigation of the effects of lexical coverage on EFL learners' comprehension of English language television, Rodgers (2013) found that many learners with less than 94% coverage were able to adequately understand the television episodes. Thus, given the participants' VLT scores (indicating a mastery of the 2,000 level, and approaching the mastery of the 3,000 level, see above), there did not seem to be any issue regarding lexical coverage. In sum, the results showed that with knowledge of proper nouns and marginal words, the participants would be able to understand the video.

### ***2.3.3 Video 2: Raising Hope (Season 1, Episode 7)***

*Raising Hope* follows the story of an American family who finds themselves having to raise a little girl they named Hope. This particular episode, which is 21-minutes long, centres around Thanksgiving, which is celebrated in their little town by having a Radish Festival carnival, akin to a town fair. The main highlight of the fair is the 'We're Thankful for the Thanksgiving Float', a parade that goes around the town with a little pilgrim riding on the float. The theme of the episode revolves around the family trying to help Hope become 'The Little Pilgrim', by selling as much candy as possible. Wars between families ensue, with devious tricks and revenge plans concocted in order to help their little ones win the coveted title.

While the pilot participants admitted that they enjoyed the humorous video, they also reported finding this video slightly challenging as they were unfamiliar with the concept of

Thanksgiving and selling candy to get on a carnival parade float. Following this feedback, a short explanation accompanied by pictures was prepared for the actual research participants. The explanation involved a short history of Thanksgiving featuring explanations of words such as pilgrims, harvest and radish, along with the concept of a Thanksgiving carnival and parade floats. Care was taken to make sure that the short explanation was the same for all conditions. The feedback on the unfamiliar context was also the reason Video 1 was shown to the actual participants first. Compared to Video 2, Video 1 portrayed a storyline involving Asian values, which the participants could more easily relate to. Therefore, Video 2 was chosen as the input material for Study 2 as this mimics how educational materials or lessons would be introduced in class -- by exposing learners to a concept that is easily grasped first.

### **2.3.4 The vocabulary in Video 2**

Similar to Video 1, the vocabulary in Video 2 was also analysed using RANGE (Nation & Heatley, 2002) after changing the hyphenated words, contractions and reduced forms in accordance with Nation's (2004) BNC word lists. As shown in Table 5, Video 2 is slightly more challenging than Video 1 in terms of lexical coverage. The most frequent 1,000, 2,000 and 3,000 word families provided 85.99%, 91.42% and 94.25% cumulative coverage of the script's total running words, respectively, which is lower compared to the coverage of Video 1. This might seem to pose a problem for the participants, as the VLT results suggest that not all conditions had achieved mastery of the 3,000 word level. However, a recent study on the relationship between vocabulary and viewing comprehension by Durbahn, Rodgers and Peters (2020) suggests that less lexical coverage is needed for adequate viewing comprehension compared to unassisted reading and advanced listening, and that lexical coverage of around 90% is enough to be able to view a TV programme. Therefore, the coverage of the transcript was well within the participants' listening competence, given their profile. Further, participants' high scores in the comprehension test (see Section 3.8) confirm that the video was comprehensible.

**Table 5. Lexical coverage of Video 2**

<b>Word list (1,000)</b>	<b>Cumulative coverage</b>	<b>Cumulative coverage</b>
	<b>(without proper nouns and marginal words)</b>	<b>(with proper nouns and marginal words)</b>
	<b>%</b>	<b>%</b>
<b>1,000</b>	82.72	85.99

<b>2,000</b>	88.15	91.42
<b>3,000</b>	90.98	94.25
<b>4,000</b>	92.30	95.57
<b>5,000</b>	92.87	96.14
<b>6,000</b>	93.89	97.16
<b>7,000</b>	94.19	97.46
<b>8,000</b>	94.64	97.91
<b>9,000</b>	94.75	98.02
<b>10,000</b>	95.20	98.47
<b>11,000</b>	95.28	98.55
<b>12,000</b>	95.28	98.55
<b>13,000</b>	96.00	99.27
<b>14,000</b>	96.08	99.35
<b>Not in the list</b>	96.73	100

---

## 2.4 Target MWEs

### 2.4.1 Initial steps

The majority of studies on MWE learning have the propensity to focus on one ‘type’ of MWE. In order to achieve this, researchers in some reading studies have needed to artificially ‘seed’ the input, so as to ensure there would be enough instances of target MWEs. However, exposing learners to authentic materials entails them encountering many different types of MWEs. It is then more realistic to have MWEs of diverse kinds. Thus, it was decided that the target MWEs would only be restricted in terms of semantics and not structure. Put differently, the target items initially chosen were more figurative than literal, and there was no restriction on the length of the MWEs.

It is acknowledged that not all MWEs are ‘equal’ in the way that they are processed (e.g., Columbus, 2010). However, this inter-item variability applies across all six conditions due to the between-participant design of the experiment. In addition, recent studies have found the risk of inter-item interference is increased when learners are presented with the same type of MWEs (Boers, Demecheleer, Coxhead & Webb, 2014). Thus, including MWEs of diverse kinds is justified as it is more ecologically valid, and it decreases the risk of inter-item interference.

After compiling lists of potential target items contained in both Videos 1 and 2, four teachers of the actual participants were consulted on whether or not they believed that the participants had knowledge of the target items. Guided by this information, the potential target items were narrowed down further. The final set of target items was made up of those that were unknown by 70% of the actual participants in the pretest.

#### ***2.4.2 Norming of target items***

There was a possibility that participants' learning gains may have been influenced by the properties of the target MWEs as they were of diverse kinds. The present study explored the possibility of the effects of item-related variables by including them in the analyses as predictors of secondary interest. The variables included the formal aspects of the target items, such as length, as well as the corpus phrasal frequency and their MI score. Apart from these, three other different properties were investigated, namely compositionality, figurativeness and ambiguity. Traditionally discussed in the context of idioms, these three factors have been shown to play a role in MWE processing and learning. Compositionality refers to how easy it is to get the meaning of an MWE from the meaning of its parts. If the meaning cannot be inferred by adding up the meanings of the constituent parts, then the word string is considered as non-compositional. MWEs that are non-compositional based on such semantic grounds are traditionally labelled as idioms. MWEs vary in their compositionality, and compositionality is associated with transparency of meaning, which is likely to influence the learnability of a given MWE. As such, compositionality is included as one of the secondary predictors of MWE learning. As for figurativeness, results have been contradictory. Some studies suggest that literal MWEs (i.e., those than can be interpreted literally and idiomatically) require less time to read and make judgement (e.g., Mueller & Gibbs, 1987). Other have found that encountering a literal MWE in a sentence involves more processing, as more time is needed to resolve other interpretations of the phrase (e.g., Titone & Connine, 1994). Availability of a possible literal interpretation may decrease the robustness of an idiomatic interpretation. Hence, this may affect how an MWE is learned. Similar arguments have been made for ambiguous MWEs, i.e., MWEs that have two possible senses, literal and figurative. Previous studies suggest that L2 learners' processing of ambiguous MWEs could be slowed down from competing semantic representations (Siyanova-Chanturia, Conklin, & Schmitt, 2011). Some scholars have also called such two-word combinations as duplex collocations (Macis & Schmitt, 2017). As shown by Cieślicka (2006), compared to native speakers, L2 learners are more inclined to activate a literal reading of the content words of an idiom. Therefore, this may mean that upon

encountering a new ambiguous idiom, a learner might erroneously attempt to arrive at the figurative meaning through a literal reading, which will mislead the learners from the correct interpretation of the figurative meaning (e.g., Martinez & Murphy, 2011; Macis, 2018). It is acknowledged that the ambiguity rating task, which was performed by native speakers, might not necessarily reflect the interpretation process of the actual research participants. However, the participants may not necessarily realise that a given word string also has a figurative meaning due to word play in the input videos. Martinez and Murphy (2011), for instance, showed that when the available context does not preclude a literal reading, learners may fail to realise that a word string is an idiom. To illustrate, the target item “*in the bag*” in the present study (from Video 2) appeared in the video in the following context. The main character carries the cash money raised from selling candy in two big paper bags, while exclaiming “*We’ve got this in the bag!*”, alluding to how certain she was of their victory in the candy-selling competition. Therefore, contexts such as these made it necessary for the effect of ambiguity to also be taken into consideration. It has to be reiterated, however, that these three factors (i.e., figurativeness, compositionality and ambiguity) are predictors of secondary interest, so their effects are not the main focus of the study, but are still accounted for after taking into consideration the treatment effects (i.e., caption condition and number of viewings).

#### **2.4.3 Norming tasks**

Figurativeness and compositionality reside on a continuum and cannot be quantified in absolute terms. Similarly, whether or not an MWE is ambiguous is subjective. Therefore, these factors must be quantified by subjective ratings. As such, a norming study that included three rating tasks was created. Thirty native speakers were recruited for the norming study. They were recruited via online surveys and were between the ages of 23 and 44 years old. None of the respondents were in the applied linguistics or linguistics field. There were ten respondents for each of the three rating tasks.

For the figurativeness rating task, the respondents were provided with three examples of expressions that were: (i) not at all figurative, (ii) completely figurative and (iii) somewhat figurative. The respondents were then asked to rate all the target MWEs on a scale of 1 to 7, with 1 being not at all figurative (there is no plausible figurative meaning) and 7 being very figurative (not at all literal). The MWEs included in the rating task consisted of the target items from both Video 1 and Video 2, as well as fillers. An example of the instruction is provided in Figure 1 below.

Figure 1. Figurativeness rating task (taken from Google Form)

## Phrase Rating

For each of the phrases you see please make a judgement on whether they are more literal or figurative. Please rate each phrase on the scale of 1 to 7. A rating of 1 would indicate that a word string is literal, i.e. Not at all Figurative, while a rating of 7 would indicate that a word string is Very Figurative (i.e. not at all literal).

An example of a phrase that may be rated 1 is "blue dress". This is because there is no plausible figurative meaning, and the phrase means "a dress that is blue in colour".

In contrast, a phrase such as "under the weather" may be given a rating of 7, as it is a figurative expression meaning "slightly unwell".

Some phrases may be rated in the middle, where a phrase may be deemed neither 100% literal, nor completely figurative. "A piece of cake", for example, may be given an intermediate judgement as it could either be a literal phrase, or a figurative expression meaning "something easily accomplished".

1. Put someone on the spot

...

	1	2	3	4	5	6	7	
Not at all Figurative	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very Figurative

The rating task for compositionality is similar to figurativeness. The respondents first received an explanation of compositionality, which was defined as "whether or not the meaning of the individual words contribute to the figurative meaning of the whole phrase". Examples of MWEs that are completely decomposable, completely non-decomposable and somewhat decomposable were also included. The respondents were then asked to rate the target MWEs on a scale of 1 to 7, with 1 indicating that there is no link between the individual words and figurative meaning, and 7 indicating that there is a clear link between the individual words and figurative meaning. The MWEs included in the rating task again consisted of the target items from both Video 1 and Video 2, as well as fillers. Figure 2 below provides an example of the instruction.



Figure 2. Compositionality rating task (taken from Google Form)

## Phrase Rating

For each of the phrases you see please make a judgement on whether or not the meaning of the individual words contributes to the figurative meaning of the whole phrase. Please rate each phrase on the scale of 1 to 7. A rating of 1 indicates that there is definitely NO link between the meanings of the individual words and the figurative meaning of the whole phrase, while a rating of 7 indicates that there is a clear link between the meanings of the individual words and the figurative meaning of the whole phrase.

For example, "chew the fat" may be rated as 1, as there is no clear link between the literal meanings of the individual words to the figurative meaning of the whole phrase, which is "to converse about everyday matter".

In contrast, the phrase "play with fire" may be rated closer to 7. This is because there is a clear link between the figurative meaning of the phrase, i.e. "to do something dangerous", and the meaning of the individual words of the phrase.

An example of a word string that would be given an intermediate judgement would be "to save someone's skin". "To save someone's skin" means "to save someone from danger or difficulty". We can work out the figurative meaning here because the meaning of the individual words "save someone" relates to the figurative meaning of the whole phrase.

1. Put someone on the spot (Figurative meaning: To pressure someone to answer a difficult question)

	1	2	3	4	5	6	7	
There is NO link between the individual words and the figurative meaning of the phrase	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	There is a CLEAR link between the individual words and the figurative meaning of the phrase

As for the ambiguity rating task, the respondents were first informed of what ambiguous and unambiguous expressions are. The respondents were required to determine whether the MWEs they saw, which included both the target items and fillers, were ambiguous or unambiguous (see Figure 3 below). Target items that were deemed as unambiguous by more respondents (i.e., at least six out of ten respondents) were coded as 1, while those rated as ambiguous by the majority of the respondents were coded as 2. For target items that received an equal rating for ambiguous and unambiguous, no ambiguity value was attached them (see Table 7 below).

**Figure 3. Ambiguity rating task (taken from Google Form)**

## Phrase Rating

---

Please decide whether or not the phrases you see next are ambiguous. Ambiguous phrases are those that have plausible literal AND figurative meanings. Unambiguous phrases are those that have NO plausible literal meaning.

An example of an ambiguous phrase is "to pull someone's leg". "To pull someone's leg" can be interpreted as a literal action of tugging someone's lower limb, but it also has the figurative meaning "to playfully deceive someone".

In contrast, "be on cloud nine", which means "very happy", is an unambiguous phrase as there is no plausible literal meaning.

Thank you for your help!

### 1. Put someone on the spot

Ambiguous

Unambiguous

The mean ratings for the compositionality, literality and ambiguity for all target items are summarised in Tables 6 and 7 in the next section.

#### **2.4.4 Target items**

As mentioned above, target MWEs that were known by 70% of the participants in the pretest were excluded. There were two such items in Video 1 (i.e., '*on the street*' and '*figure something out*'), while in Video 2, three items (i.e., '*look up to someone*', '*count someone in*' and '*go too far*') were excluded from the analysis. This resulted in a total of 18 target MWEs in Video 1, and 20 target MWEs in Video 2. The target items for both Video 1 and Video 2, along with their lemmatised phrasal frequency, MI score and mean rating for figurativeness and compositionality are detailed in Tables 6 and 7. Note that the values for ambiguity refer to the way it was coded for the analysis (see Chapter 3 and Chapter 4). The lemmatised frequencies and MI scores were obtained from COCA, i.e., the Corpus of Contemporary American English (Davies, 2008) in July 2018.

**Table 6. Target MWEs in Video 1**

<b>Target MWEs</b>	<b>COCA</b>	<b>MI Score</b>	<b>Figurativeness (mean rating)</b>	<b>Compositionality (mean rating)</b>	<b>Ambiguity</b>
on the same page	796	9.93	4.6	4.8	2
turn a profit	484	10.48	4.0	5.0	1
on (someone's) hands	4071	4.52	4.1	3.4	2
tighten up	263	5.43	3.7	4.8	2
kill someone with kindness	19	11.76	7.0	3.4	1
root for	1845	2.48	6.0	2.6	1
(one's) hands are tied	132	5.55	5.1	5.2	2
slippery slope	676	13.87	3.8	4.3	2
put (someone) on the spot	182	5.04	6.2	3.8	2
talk some sense into (someone)	60	5.98	4.1	6.0	1
beg to differ	288	10.98	3.9	4.9	1
whisked away	264	6.38	5.1	4.4	2
chip in	1279	1.54	5.2	2.8	2
bear with (someone)	517	0.35	5.2	4.7	2
go through a rough patch	28	14.51	4.9	5.2	2
let up on (someone)	106	1.3	5.4	3.7	1
look out for (someone)	2174	3.18	4.0	6.0	2
work (something) out with (someone)	25	8.53	2.6	6.0	2

**Table 7. Target MWEs in Video 2**

<b>Target MWEs</b>	<b>COCA</b>	<b>MI Score</b>	<b>Figurativeness (mean rating)</b>	<b>Compositionality (mean rating)</b>	<b>Ambiguity</b>
whip up (something)	1349	7.61	5.0	3.3	2
unsung hero	236	15.32	5.9	4.8	1
make the most of (something)	1352	2.62	5.1	6.5	1
spread the word	754	3.21	4.7	5.8	
the dust has settled	44	7.61	5.0	4.5	2
screw with (someone/something)	156	5.25	5.3	3.3	2
take (something) lying down	14	4.75	5.6	3.6	2
step (something) up	7217	8.82	5.6	4.7	2
have a beef with (someone/something)	58	3.88	6.3	1.8	2
invest in (someone)	12154	5.78	4.3	6.3	1
have a run-in with (someone)	14	3.46	5.4	3.4	
in the bag	830	3.21	4.3	3.3	2
rat out (someone)	121	8.77	6.6	1.9	1
hopped up	210	8.82	4.3	3.5	2
(someone's) day in the sun	96	2.62	5.3	3.7	2
play the ( ) card	98	0.43	5.0	4.0	2
bring out the big guns	10	2.21	5.1	5.0	2
hit the streets	239	3.21	5.7	1.6	2
small potatoes	157	11.35	4.2	2.7	2
caught up in (something)	6656	8.82	4.6	5.4	2

## 2.5 Testing instruments

Several tests were administered to measure MWE learning. Measuring multiple aspects of MWE knowledge provides a more accurate assessment of learning (Nation & Webb, 2011). Further, measuring each aspect at different levels of sensitivity allows us to gauge the extent to which each aspect of knowledge is learned (Nation & Webb, 2011). The MWE tests created therefore measured knowledge of form and meaning at two different sensitivities, recognition and recall. The following details the format of each test administered in both studies.

### 2.5.1 Form recall pretest

In developing the pre-test, it was decided that the target items should remain as obscure as possible so as not to give away the purpose of the study. Form-meaning connection is unlikely to be established when learners are not exposed to the actual forms of the target items. To ensure this, a gap-fill and C-test blend format was used for the pretest. This format has been used in previous studies such as in Schmitt, Dörnyei, Adolphs and Durow (2004). The test was constructed by first taking sentences containing the target items from dictionaries and the corpus. For each MWE, only the content words had missing letters. The function words were kept intact. Contextual clues were present so as to ensure that the MWEs could be guessed. In cases where the phrases taken from the corpus contained only weak contextual clues, the phrases were modified with more contextual clues added. The draft test was first run through RANGE, and words that were not within the first 3,000 words were either substituted by or provided with a synonym within the 3,000 range. The form recall pretest for Study 1 and Study 2 are attached in Appendix 4 and Appendix 11, respectively.

Below is an example of the test item for *'unsung heroes'*:

*Teachers are the uns\_\_\_\_\_ he \_\_\_\_\_ of a great writer's success. They are often not noticed or praised for their hard work.*

In Study 1, five fillers were included in the pretest. In Study 2, there were four fillers. For motivational purposes, the fillers were MWEs that were thought to be familiar to students, such as *'piece of cake'* and *'keep in touch'*.

### 2.5.2 Word deletion procedure

Initially, the conventions of C-test were used whereby the second half of a word was cut off. However, it was found that deleting half the word could lead to guessing. For instance,

the MWE “*lo \_\_\_ up to*” could be easily guessed by a learner, even without full knowledge of the whole target phrase. It was then decided that only the first letter would be given, except for words that start with consonant clusters. If the contextual clues were strong enough, the limited letters would then suffice. The draft pre-test instrument was piloted in person with three American native speakers. For the items whose answers could not be provided by the native speakers, more letters were provided until they could get the whole MWE phrase. The items that were deemed as unnatural by two of the three native speakers were discarded, while the items whose context was unclear were revised according to the feedback received. The revised test was piloted with three native speakers and three non-native Malaysian postgraduate students, all of whom completed the instrument 100% correctly. The final version was piloted with 20 participants (see 2.2.2 for information on the pilot participants). The same test with differently ordered items was also used as a delayed posttest (see 2.5.4 below). As mentioned before, pilot testing revealed no issues, and the scores indicated there was no risk of a ceiling effect.

### ***2.5.3 Immediate form recall posttest (Gap-fill transcript-based test)***

To test learners’ immediate form recall, a gap-fill transcript-based test was used, as an episodic memory test is thought to be more direct. Episodic memories are memories related to “temporally dated episodes or events, and the temporal-spatial relations” (Tulving, 1972, p. 385), i.e., memories where “the words are explicitly associated with the context or circumstance in which they were encountered” (Mohd Jelani & Boers, 2018, p.176). Below is an example of the test item for “*bring out the big gun*” in the immediate form recall posttest:

*It was time for us to (8) br \_\_\_\_\_ out the b \_\_\_\_\_ g \_\_\_\_\_.*

The transcript of the immediate form recall posttest was condensed to include just the main scenes, so as to prevent participants from being demotivated by a 20-page transcript. The main scenes and dialogues that were deemed integral to the video were maintained, and a native speaker who watched the videos verified that the essence or whole plot of the videos was still intact and understandable despite the reduced amount of dialogue. The immediate form recall test for Study 1 and Study 2 are attached in Appendix 5 and Appendix 12, respectively.

For studies that utilise a pretest-posttest design, the format of the tests is recommended to be identical (Nation & Webb, 2011, p. 121), so that learning gains are not confounded by the use of different test formats. However, for the present study, having the immediate posttest in the same format as the pretest may create an extra learning burden. This is because the

contextualised sentences in the pretest, which contained clear clues on the meaning of the MWEs, were different to the ones in the videos. Thus, having learners insert an MWE that is newly acquired into contextualised sentences that are not exactly similar to the video they have just watched could be cognitively demanding. Furthermore, congruent learning tasks and test conditions will lead to better performance (Lotto & de Groot, 1998; Schmitt, 2010). It is acknowledged that an immediate posttest could be used as a successful retrieval event for learners who do well in the immediate posttest. This successful retrieval consolidates knowledge—knowledge that can be subsequently accessed during the delayed posttest. On the other hand, this provides another justification for the different test items and test format of the immediate posttest: what was consolidated in the immediate posttest would need to be transferred to test items that were different in the delayed posttest. Put differently, using different test formats and different test items minimises the risk of test effects, in that the direct transfer of responses on the immediate posttest to the delayed posttest (which is in the same format as the pretest, see 2.5.4 below) could be avoided.

#### ***2.5.4 Delayed form recall posttest***

A more accurate assessment of learning could be gathered by carrying out assessments that tap into multiple aspects of MWE knowledge (Nation & Webb, 2011). With this in mind, four delayed posttests were created. The first posttest was test on form recall. As mentioned before, the delayed form recall posttest was in the same format as the pretest, with the order of the items randomised. The delayed form recall posttest for Study 1 and Study 2 are attached in Appendix 6 and Appendix 13, respectively. The participants were required to fill in the gaps of sentences to create the full MWEs. The items in the delayed posttest were the same as the pretest, with their order randomised. As suggested by Schmitt (2010), whether or not learning has occurred can only be demonstrated by a delayed posttest. Thus, a delayed posttest that is in the same format as the pretest is fitting. Further, a delayed posttest in the same format as the pretest allows for the long-term retention or learning gains to be assessed, without the format being confounded with the time of the test.

#### ***2.5.5 Delayed form recognition posttest***

The second delayed posttest was a receptive knowledge of form test. In this test, multiple-choice options were provided for the gaps. The delayed form recognition posttest for Study 1 and Study 2 are attached in Appendix 7 and Appendix 14, respectively. The options consisted of the correct answer, and three distractors. The distractors were created by taking

authentic responses of the participants in the piloting of the form recall test. Taking authentic responses of pilot participants was thought to make the distractors stronger and more plausible. Care was taken to ensure that the distractors were not only grammatically correct, but also semantically plausible in the context provided.

Below is an example of the test item for the MWE “*the dust has settled*”:

*Let’s just wait until the d \_\_\_\_\_ has s \_\_\_\_\_ before we decide what to do. It’s better to make a decision when the situation has calmed down.*

- |                            |                              |
|----------------------------|------------------------------|
| <i>a) dust has settled</i> | <i>c) daylight has shone</i> |
| <i>b) dusk has started</i> | <i>d) dust has spread</i>    |

### ***2.5.6 Delayed meaning recall posttest***

The third delayed posttest measured meaning recall. In this test, the participants were presented with the MWEs, and were required to provide the meaning of the MWE in their L1 or in English. The delayed meaning recall posttest for Study 1 and Study 2 are attached in Appendix 8 and Appendix 15, respectively. The participants could write as many meanings that they could come up with, according to the context of the video.

### ***2.5.7 Delayed meaning recognition posttest***

The fourth delayed posttest measured meaning recognition, with multiple-choice options provided. Of the four options, one was the correct meaning, while the other three options were distractors. The distractors were created by taking the meanings of MWEs that were semantically related to the target MWE. For instance, for the MWE “*take something lying down*” as illustrated below, the meaning is option (c), but options (a), (b) and (d) are the meanings of the MWEs “*to lie down on the job*”, “*lie in wait*” and “*to lie low*” respectively.

*Take something lying down*

- |   |   |
|---|---|
| <i>(a) To do one’s job poorly</i>                     | <i>(c) To accept something unpleasant without fighting back</i> |
| <i>(b) To remain hidden while preparing to attack</i> | <i>(d) To keep quiet and not be noticed</i>                     |

In instances where a target MWE did not have semantically-related counterparts whose meaning was plausible enough to be a distractor, the literal meaning was also used as one of



the distractors. As in all the other tests, all distractors were kept within the most frequent 3,000 words. The delayed meaning recognition posttest for Study 1 and Study 2 are attached in Appendix 9 and Appendix 16, respectively.

### ***2.5.8 Listening comprehension test***

In order to create the listening comprehension tests, the listening construct had to first be decided upon. The purpose of the listening test is to assess whether the participants are able to remember and understand the content of the videos. The test construct should then be theory-based, rather than syllabus-based (Bachman & Palmer, 1996; Nguyễn, 2017). The comprehension test was created by drawing on established listening constructs as well as taxonomies of listening skills. One of the most commonly cited and widely accepted descriptions of listening involves the idea of top-down and bottom-up processing (e.g., Buck, 2001). Bottom-up processing is when a listener perceives the acoustic input as phonemes first, then moves on to higher syntactic level, while top-down processing involves interacting with their linguistic knowledge to interpret or create meaning for the acoustic input (Buck, 2001). According to Buck (2001), although both processing types happen at the same time, they are not necessarily equal. Rather, the processing happens in different proportions depending on the situation and purpose.

To operationalise the bottom-up and top-down construct, tasks that reflect the ability to perform the two different but complementary processes were identified by drawing on previous research. Skills such as identifying details, facts, supporting ideas and more local points of information constitute the latent ability to perform bottom-up processing (e.g., Aitken, 1978; Richards, 1983; Weir, 1993). The ability to perform top-down processing, on the other hand, is observed through global skills such as listening for gist, controlling idea or topic, making inferences about context and relationships, and making inferences about speakers' attitude and pragmatic meaning (e.g., Buck & Tatsuoka, 1998; Hildyard & Olson, 1978). As these skills are well established in literature, they are used to operationalise the construct of the listening test used in this research.

Two test formats were used, namely multiple-choice and True/False. These formats were chosen as the learners were familiar with them, so there was no need to spend time familiarising the participants with the test format (Frisbie & Becker, 1991; Haladyna, 1992; Kreiter & Frisbie, 1989). Further, these formats are well suited to tap into learners' skills in making inferences, as learners are forced to "choose the most plausible inference" (Rodgers,

2013, p. 31). As argued by Brindley (1998), open-ended items would entail more subjective interpretations that would be hard to score as incorrect. Finally, these test formats also involve no subjective scoring.

To create the items for the comprehension tests, one native speaker and two non-native speakers were invited to watch the videos and identify the idea units. Idea units were identified as distinct events, actions and dialogues spoken. In addition, the viewers were also asked to identify any utterances whose interpretation requires going beyond literal meaning, such as meanings that a character implies or suggests without directly expressing them. These distinct idea units and utterances corresponded to the listening construct and the tasks that operationalise the construct. The videos were watched with breaks in between the integral scenes, which were identified by the researcher in advance. During each break, the potential idea units were discussed with the help of a transcript of the video.

Comprehension questions based on these idea units were then created. The questions were not distributed equally between all constructs and were not forced to be distributed equally. The questions were run through RANGE, and any words that were not within the first 3,000 words were either substituted by, or provided with, a synonym within the 3,000 range. For example, the word “potty” appeared in Video 2 as well as in the comprehension test, but because it is within the 10,000 most frequent words, the word “toilet” was included in parentheses next to the word “potty”.

The potential set of comprehension questions was trialled on three postgraduate students, of whom two were non-native speakers and one was a native speaker. They first watched the videos from start to end, and then proceeded to answer the questions. Analogous to the actual research procedure, a document that showed the characters’ names and faces was projected to help with answering the comprehension questions. A discussion on the difficulty and clarity of the test items took place, and, where necessary, items were then revised according to the feedback received. The revised version was then trialled on one native and one non-native postgraduate student, both of whom completed the test 100% correctly. The comprehension test for Study 1 and Study 2 are attached in Appendix 10 and Appendix 17, respectively.

## 2.6 Procedure

### 2.6.1 Study 1: *The incidental learning conditions*

This study took place in classrooms equipped with a projector, a screen and speakers. In the first meeting the participants were briefed on the research and its purpose before they signed the consent form. The information sheet and consent form are attached in Appendix 2 and 3, respectively. The participants also filled in a background questionnaire that provided insights into the number of years they had studied English, the languages they speak, and their self-perception of their abilities to write, speak, listen and read in English.

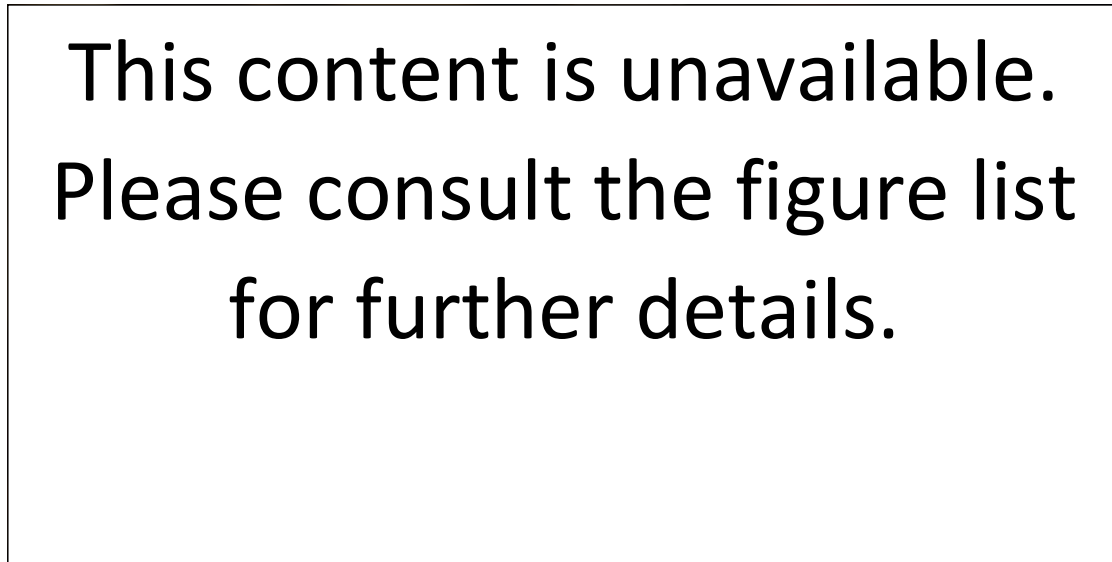
A week later the participants commenced the pre-test stage of the study. Prior to completing the form recall pretest, the instructions were read out aloud, and an example was provided. The participants then had 40 minutes to complete the test. After the pretest, the VLT was administered, which took 45 minutes. The VLT was conducted right after the pretest in order to distract participants from memorising the items in the pretest and to minimise the chances of participants discussing the target items. In Study 2, the pre-test for the target items in the second video was immediately followed by the VST (see Section 2.6.2), which was also done to flush the effects of memory and draw participants' attention away from the target items in the pretest.

Two weeks after the pretest, the participants watched Video 1. Prior to watching the video, the learners were informed that they would be asked comprehension questions post-viewing. There was no mention of MWE tests. The learners then viewed the video under one of the six conditions. As the present study sought to investigate the effects of caption conditions, the participants watched one of the three different versions of the video: without captions, with normal captions or with typographically-enhanced captions. To assess the effects of repetition under each caption condition, some participants watched the video once, while the others watched the video twice. There were then six conditions in total (see Section 2.2.1 for the number of participants in each condition).

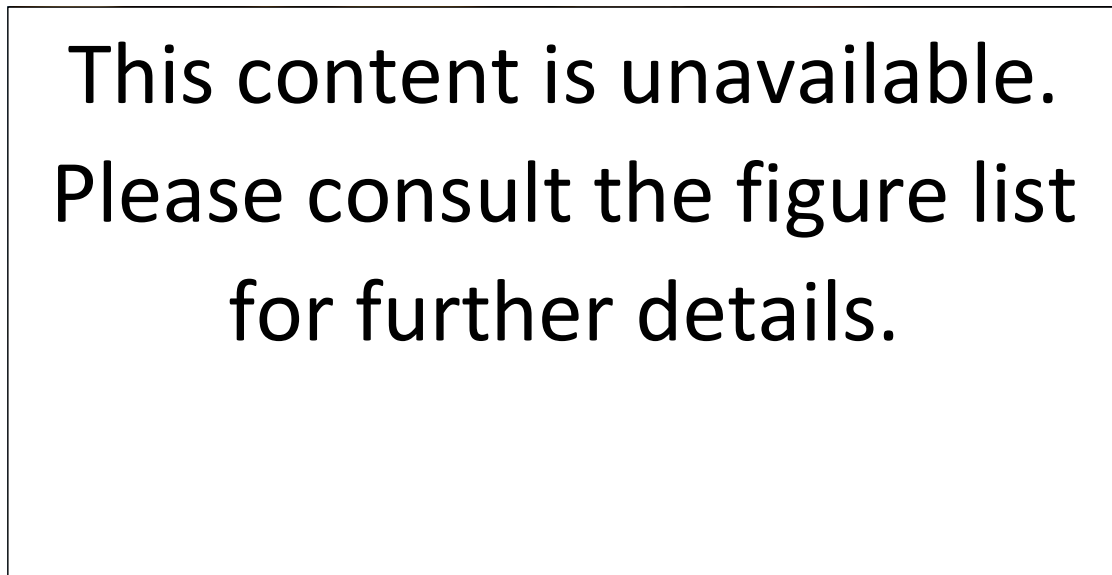
In the normal captions condition, captions were shown at the bottom of the screen without any modification. In instances where the credits appeared at the bottom middle of the screen, the captions were positioned at the bottom left of the screen, so as to avoid overlapping. A maximum of two lines of captions appeared at one time. The MWEs were designed to appear in one line, without 'spilling' into another line. This was done in order to conform to subtitling principles. For instance, Díaz-Cintas and Ramael (2014) proposed that if a relatively short

subtitle could be fit into one line, there is no need to use two lines. Although having a single caption line is ideal, this was not always possible in both input videos. To ensure that caption appearance coincided with speech onset (Vanderplank, 2016), there were instances in the input videos where two lines of captions were needed. However, there was only one instance in Video 2 (Study 2), in which the target item ‘bring out the big guns’ was broken into two lines. Only words uttered by the characters were captioned. Non-speech information such as character names or descriptive captions (e.g., “*baby crying*”) was not captioned (see Figure 4 below). For the enhanced captions condition, all the target items in the videos were bolded and modified using a software called SubtitleEdit. There was no difference in the amount of time the captions were shown for both the normal caption condition and enhanced caption condition. The only difference was the typographic enhancement employed for the latter (see Figure 5 below). In the uncaptioned condition, participants watched both videos without the aid of captions.

**Figure 4. An example of a still from Video 1 (*Fresh off the Boat*) with normal captions**



**Figure 5. An example of a still from Video 1 (*Fresh off the Boat*) with enhanced captions**



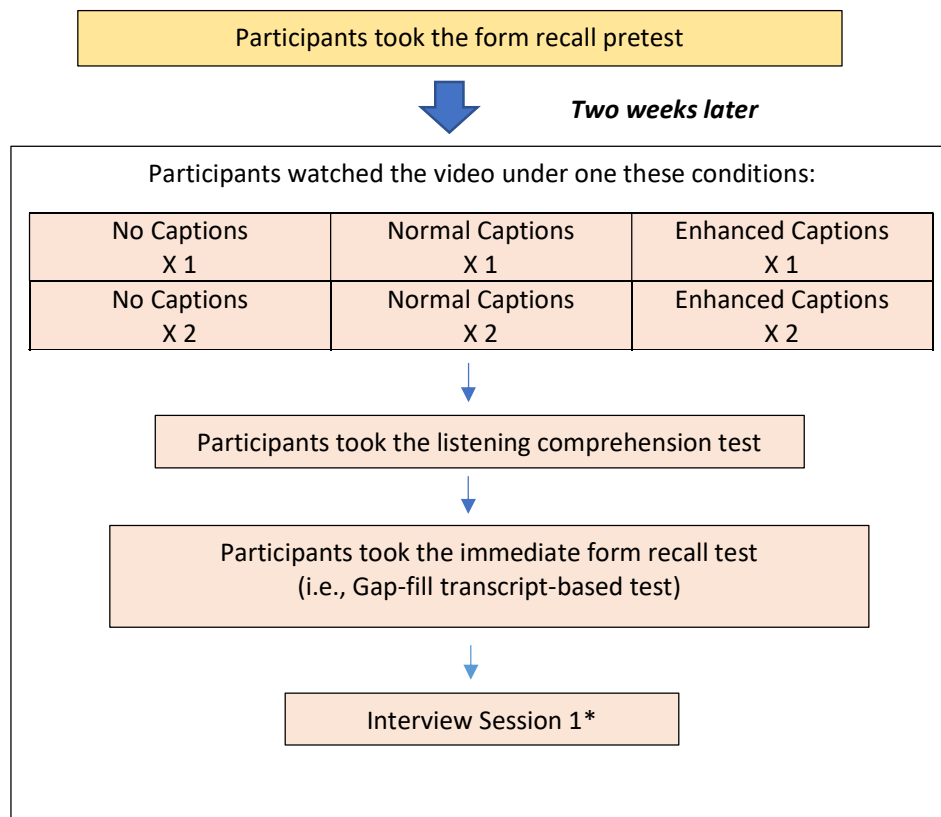
Immediately after watching the video, the comprehension test was administered. This took 25 minutes. While the participants were doing the comprehension test, a Word document that showed the pictures and names of all the major characters was projected. It was felt necessary to provide aid in this way as viewers would have problems trying to recall the names of characters in a show that they had only watched once or twice. Supplying the participants with the names of the characters was not thought to give an unfair advantage to any of the participant groups, since the same was done across treatment conditions. Following the comprehension test, the MWE immediate form recall posttest was administered, which took 45

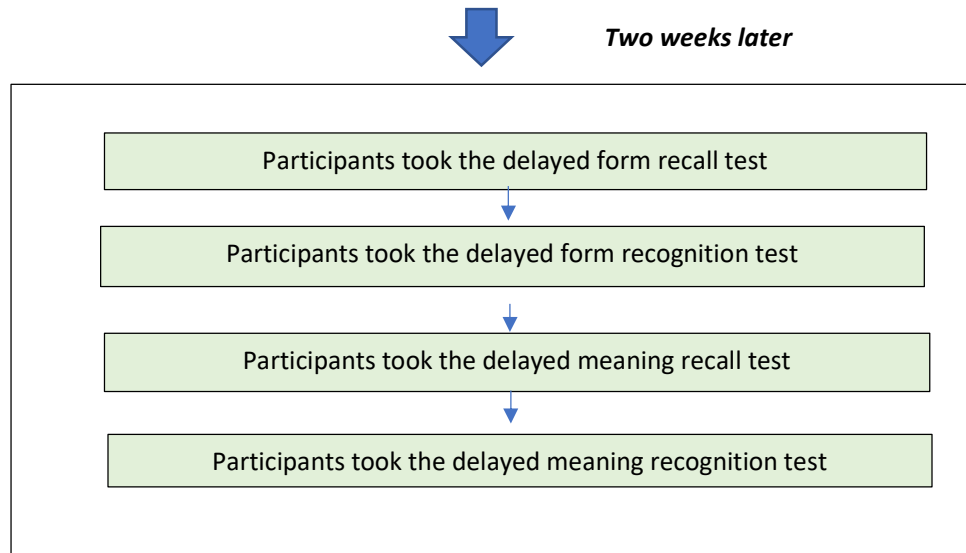
minutes. Due to the length of this transcript-based test, the participants were allowed to skip to the gaps in the text, the location of which was clearly signposted in the margin. Two weeks later, the four delayed posttests that tapped into different aspects of MWE knowledge were carried out.

**2.6.2 Study 2: The intentional learning conditions**

Study 2 followed a similar procedure to Study 1, with a couple of exceptions. Firstly, a different video was used (Video 2). Next, prior to watching the video, the participants were forewarned about the tests that would follow. They were informed that the tests would be in the same format as Study 1, and that they would also be tested again two weeks later. Finally, an interview session was carried out immediately after the immediate posttest. Ten participants participated in each of the interview sessions (i.e., after each of the treatment conditions) on a voluntary basis. These informal short interviews were carried out to ascertain that the learners enjoyed and could follow the input videos. The procedure of the two studies is visually depicted in Figure 6 below.

**Figure 6. Procedure of Study 1 and Study 2**





*\*Conducted in Study 2 only*

## 2.7 Overall schedule

Both the studies were conducted throughout one academic semester. The two stages were separated by two weeks. Table 8 shows the schedule for the study.

**Table 8. Schedule for both Study 1 and Study 2**

Study	Date	Task
	30 October 2017 – 3 November 2017	Briefing and collection of consent forms
<b>Study 1: Incidental learning</b>	6 November 2017 – 10 November 2017	Form recall pretest Vocabulary Levels Test
	19 November 2017 – 24 November 2017	Treatment (Viewing of Video 1) Comprehension test Immediate form recall posttest (Gap-fill transcript-based test)
	3 December 2017 – 8 December 2017	Delayed form recall posttest Delayed form recognition posttest Delayed meaning recall posttest Delayed meaning recognition posttest
	17 December 2017 – 22 December 2017	Form recall pretest Vocabulary Size Test

<b>Study 2: Intentional learning</b>	1 January 2018 – 5 January 2018	Treatment (Viewing of Video 2) Comprehension test Immediate form recall posttest (Gap-fill transcript-based test) Interview session
	14 January 2018 – 19 January 2018	Delayed form recall posttest Delayed form recognition posttest Delayed meaning recall posttest Delayed meaning recognition posttest

## 2.8 Scoring criteria and scoring procedure of the MWE tests

### 2.8.1 Scoring of gap-fill tests

The form recall tests (i.e., pretest, immediate and delayed tests) are gap-fill tests with exactly the same missing parts of the target MWEs to be supplied by the learners. The same scoring criteria and procedure were thus used so that the scores were readily comparable. Two scoring procedures were used at each test time: strict and lenient. Having a lenient scoring procedure allows for partial credit to be given to small gains (Nation & Webb, 2011). Taking a page from Webb and Kagimoto (2009) and Webb and Kagimoto (2011), under the lenient scoring system, responses that contained a minor mistake were marked as correct and received a full score. A minor mistake was defined as responses that are:

- (1) Used in singular but should have been plural
- (2) Used in plural but should have been singular
- (3) Wrongly spelled but (i) is a close approximation of the answer, and (ii) does not constitute a new word
- (4) A wrong part of speech/wrong inflection.

An exception, however, was made for the response “*bare with me*” (as opposed to the correct target item “*bear with me*”). Although this response clearly violated rule 3(ii) above because it constituted a new word, the raters felt that it was necessary to award partial credit, as there were a few participants who wrote both words, but crossed out one of them. The participants clearly had some knowledge of the MWE, but were unsure of the correct spelling. Thus, an exception was made for this particular target item.



While the above scoring procedure was relatively straightforward for a target MWE in which only one content word needed to be completed, it was not the case for target items in which two or three content words of the MWEs needed to be supplied. For these items, it was decided that partial credit would only be awarded for two-gap responses in which only one of the words supplied by the learner contained a minor mistake. For instance, for the MWE “*make the most of*” which appeared as “*m\_\_\_\_\_ the m\_\_\_\_\_ of*” in the tests, “*makes the move*” received 0.5 points, while “*meet the moon*” received 0 points.

For an item which required participants to fill in three gaps, a two over three rule was imposed. This means that partial credit was only given for responses in which two out of three gaps were correct or contained a minor mistake. For instance, for the MWE “*take it lying down*”, which appeared as “*t\_\_\_\_\_ it ly\_\_\_\_\_ d\_\_\_\_\_*” in the tests, the response “*take it lying dead*” (two out of three words are correct) received 0.5 points, while “*turn it lying dead*” (one out of three is correct) received 0 points.

In the strict scoring system, the responses had to be free of mistakes in order to get credit. Partial credit was not awarded in the strict scoring system, which meant that responses which contained a minor mistake were marked as incorrect and received a score of 0. The scoring criteria are further illustrated with examples in Table 9 below.

**Table 9. Scoring criteria for the form recall tests**

<b>Response</b>	<b>Score (under lenient scoring system)</b>	<b>Score (under strict scoring system)</b>	<b>Example</b>
1. All the words are correctly spelled	1	1	<i>Bring out the <u>big guns</u></i> <i>Take it <u>lying down</u></i> <i>Have a <u>beef</u></i>
2. One/two/three words contain(s) a minor spelling mistake that  (i) Does not affect pronounceability	1	0	<i>Bring out the <u>big guns</u></i> <i>Tak it <u>lyieng down</u></i>

(ii) Does not create a different word*			
(iii) Still resembles the target item *except for “bare with me”			
3. Contains mistake(s) in which the response is singular but should have been plural	1	0	<i>Bring out the big <u>gun</u></i>
4. Contains mistake(s) in which the response is plural but should have been singular	1	0	<i>Have a <u>beefs</u></i>
5. Contains mistake(s) in which the response is the wrong part of speech/wrong inflection	1	0	<i><u>Brings/brought/bringing</u> out the big guns</i>
6. For a two-word response: One word contains a minor mistake, while the other contains a major mistake (1/2 correct)	0.5	0	<i><u>Spread</u> the <u>weather</u></i>
7. For a three-word response: Two words contain a minor mistake, while one word contains a major mistake (2/3 correct)	0.5	0	<i><u>Tear</u> it <u>lying</u> <u>down</u></i>
8. For a three-word response: Only one word contains a minor mistake; the other two contain a major mistake (1/3 correct)	0	0	<i><u>Tear</u> it <u>lying</u> <u>dead</u></i>

9. All the words in the response contain a major spelling mistake that:			
(i) Does not resemble the target item	0	0	<i>Break out the <u>black</u> gum</i>
(ii) Affects the pronounceability			
(iii) Constitutes a new word			
10. There is no response	0	0	

Under the lenient scoring procedure, each test was scored by two independent raters, who were postgraduate students in the area of applied linguistics. Prior to independent scoring, the raters were trained by the researcher. The scoring protocol was described in detail and was illustrated with authentic examples. Next, the raters practised scoring the same set of test papers which had been identified by the researcher as a mixed collection of papers that were very easy, fairly easy, and very hard to rate. Scores awarded on these sets of papers were then discussed, and any discrepancies were resolved before the raters started scoring the rest on their own. Inter-rater reliability for all tests was measured using Spearman rank correlation. In Study 1, the inter-rater reliability was .97, .95 and .97 for the form recall pretest, immediate posttest and delayed posttest, respectively. As for Study 2, the inter-rater reliability was .96, .94 and .97 for the pretest, immediate posttest and delayed posttest respectively.

In the strict scoring procedure, there was no subjectivity involved as credit was only given to completely correct answers. Hence, strict scoring was done by the researcher herself.

### ***2.8.2 Scoring of multiple-choice tests (MWE form recognition and MWE meaning recognition)***

Both the delayed recognition tests were in multiple-choice format and were scored dichotomously. A full score of 1 was awarded for the correct answer, while 0 was given to incorrect answers and for items that received no responses. As there was no subjectivity involved, these tests were scored by the researcher.

### 2.8.3 Scoring of the meaning recall test

The nature of meaning recall test necessitated subjective scoring. Thus, detailed scoring criteria were created so as to ensure high inter-rater reliability. Only lenient scoring was applied for this test. The scoring criteria were as follows. A full score was awarded for answers that were completely correct. As participants were also allowed to write in L1 or their national language (Malay), correct translations were also awarded a full score. Responses that were synonymous to the correct answer, for example “*support*” as opposed to the answer key “*cheer*”, were also awarded a full score. Drawing on the scoring criteria used by Winke et al. (2010), a partial score was also awarded for responses that were from the same semantic field or shared semantic features. To illustrate, a response such as “*sorry not the same opinion*” which was an authentic answer by a participant for the target item “*beg to differ*”, received a partial score as it was semantically related to the answer key “*to firmly disagree with someone*”. Finally, responses or translations that shared no semantic features with the key answers, as well as items that were not attempted received a score of 0. Table 10 below provides further details on the scoring criteria, accompanied by examples.

Analogous to the gap-fill tests, the raters were briefed on the scoring criteria, and a norming session was conducted. There were two raters for each of the meaning recall test (i.e., for Study 1 and 2). The raters, who were Malaysian postgraduate students, scored a set of papers containing responses that were very easy, fairly easy and very hard to score. Scores awarded on these sets of papers were then discussed, and any discrepancies were resolved before the raters started scoring the rest on their own. In Study 1, the inter-rater reliability was .79, while in Study 2, the inter-rater reliability was .76. It is acknowledged that the inter-rater reliability is less than desirable especially when compared to the inter-rater reliability of the form recall tests. There was, however, more subjectivity involved in scoring the meaning recall tests.

**Table 10. Scoring criteria for the delayed meaning recall posttest**

Response type	Score	Example
1. Exact answer according to the answer key	1	<i>Q: (Someone's) hands are tied</i> <i>A: to not be able to help</i>

2. Correct translations	1	<i>Q: Talk some sense into (someone)</i> <i>A: memberi nasihat</i>
3. Correct synonyms in English or Malay	1	<i>Q: Root for (someone)</i> <i>A: cheer for someone/supporting someone</i>
4. Answers or translations that are not synonyms, but from the same semantic field	0.5	<i>Q: Beg to differ</i> <i>A: sorry not the same opinion</i>
5. Answers or translations that share some semantic features with the answer key (in other words, responses that are incomplete)	0.5	<i>Q: Put (someone) on the spot</i> <i>A: forcing someone</i>  <i>Q: Kill (someone) with kindness</i> <i>A: Be kind to someone</i>
6. Answers or translations that share no semantic feature with the answer key	0	<i>Q: (Someone's) hands are tied</i> <i>A: Very busy</i>  <i>Q: Look out for someone</i> <i>A: Be careful with someone</i>
7. No response given	0	

Note: "Q" refers to the MWE, "A" refers to the example of the response received

#### **2.8.4 Scoring criteria and scoring procedure of the listening comprehension test**

All items in the comprehension tests were scored dichotomously. In Part 1, which was in the format of multiple-choice, the participants received a full score for circling the correct answer, and 0 for circling the incorrect answer or for failing to provide a response. Similarly,

in Part 2, where participants had to circle whether a statement was True or False, the participants received a 1 for circling the correct answer, and 0 for circling the incorrect answer or for failing to provide a response.

The next chapter details how the data for all the tests in Study 1 were analysed, along with the results of all the analyses.

## CHAPTER THREE:

### STUDY 1 (INCIDENTAL LEARNING) – ANALYSIS AND RESULTS

This chapter focuses on the results of the MWE tests administered in Study 1. The chapter starts by outlining the research questions addressed in this study. This is followed by explanations of the statistical tools used to analyse the scores on each test. The subsequent sections (Section 3.4 to 3.8) present the outcome of each analysis. The chapter then concludes with a summary of the findings.

#### 3.1 Research questions and organisation of the chapter

The aim of Study 1 was two-fold. As stated under the Methodology section, Study 1 sought to investigate whether caption condition (i.e., uncaptioned, normal captions and typographically-enhanced captions) had an effect on incidental MWE learning through L2 viewing. Further, this study sought to investigate whether repetition modulated the effects of caption condition. Of further interest was whether repetition influenced incidental uptake of MWEs to the same degree under all three caption conditions.

To answer these research questions, incidental MWE learning was measured using one pretest, one immediate posttest and four delayed posttests. The five posttests provided insights into the effects of the variables on multiple aspects of the MWE knowledge, namely:

- (i) short-term form recall
- (ii) long-term form recall
- (iii) long-term form recognition
- (iv) long-term meaning recall
- (v) long-term meaning recognition

The subsequent sections are organised accordingly—by looking at the effects of the variables on each of the aspects of MWE knowledge listed above. In addition, as the effects of the variables on listening comprehension were also assessed, the results of the comprehension test are presented under Section 3.8.

#### 3.2 Explanations of statistical tools and terms

The data from all the tests were analysed by a number of statistical models on the R platform (Version 3.4.4, R Core Team, 2018). The statistical models used depended on the nature of the dependent variable of the test. The statistical models along with the relevant terms are explained below. The summary of the models used is also presented in Table 11.

##### Statistical models

- (i) Form recall pretest, immediate posttest and delayed posttest

Two different types of analyses were performed using the data for the form recall tests. The first analysis concerned participants' learning gains from the pretest to the immediate posttest, and from the pretest to the delayed posttest. These data were analysed using general linear models as there was only one data point for each participant, and there were no random effects included. These models were fitted and run using the *lm* function in the *base* package (R Core Team, 2018). The second type of analysis performed for the form recall data was item-level analysis. The dependent variable for this analysis was participants' scores on the individual target items. The item-level analysis was performed for the data based on the lenient scoring system only, as the gains analysis showed similar trends for both the lenient and strict scoring data (see Sections 3.4.1 and 3.4.2). Under the lenient scoring system, the responses in the form recall tests were scored using a three-point scale (i.e., 0, 0.5 and 1, see Section 2.8.1 for more details) by two independent raters. Since the inter-rater reliability was high, the average score from the two judges were used, resulting in five possible scores for each individual item (i.e., 0, 0.25, 0.5, 0.75 and 1). The dependent variable was then a limited number of *ordered* values. As such, cumulative link mixed models (*clmm*) were used. Initial analysis was performed using linear mixed models (*lmer*), but the analysis suggested violation of normality of residuals. Thus, the function *clmm* from the *ordinal* package (Christensen, 2019) was applied.

- (ii) MWE form recognition test, MWE meaning recognition test and the comprehension test

As these tests had a binary variable, they were analysed using Generalised Linear Mixed Model. The function *glmer* from the *lme4* package (Bates, Maechler, Bolker & Walker, 2015) was used.

- (iii) MWE meaning recall test

The meaning recall test was also scored by two judges using a three-point scale (i.e., 0, 0.5 and 1), which resulted in five possible scores (i.e., 0, 0.25, 0.5, 0.75 and 1). Therefore, these data were analysed using the *clmm* function in the *ordinal* package (Christensen, 2019).



**Table 11. Summary of statistical models used for each test**

Test	Statistical model
Form recall pretest, immediate posttest and delayed posttest	<ul style="list-style-type: none"> <li>• General linear models—for learning gains analysis</li> <li>• Cumulative link mixed models—for item-level analysis</li> </ul>
Delayed form recognition posttest	Generalised linear mixed models (mixed effects logistic regression)
Delayed meaning recall posttest	Cumulative link mixed models
Delayed meaning recognition posttest	Generalised linear mixed models (mixed effects logistic regression)
Comprehension test	Generalised linear mixed models

### 3.3 Model development and comparison

The first step of each analysis was to build a model that included the core independent variables that were of central importance, namely (i) caption condition (three levels: uncaptioned, normal captions and enhanced captions), (ii) number of viewings (two levels: once or twice) and (iii) vocabulary size test (VST) score (centred around the mean value). The VST score was included as literature has shown that learners with a larger vocabulary size tend to understand reading and listening texts better than learners with a smaller vocabulary size (Elgort & Warren, 2014; Noreillie, Kestemont, Heylen, Desmet, & Peters, 2018; Schmitt, Jiang, & Grabe, 2011; Stæhr, 2009). Additionally, prior knowledge was also included as a fixed effect as it has been known to positively influence vocabulary learning (e.g., Montero Perez et al., 2014; Peters et al., 2016; Toomer & Elgort, 2019). Prior knowledge was included by taking learners' form recall pretest score as a proxy of their prior knowledge of form and meaning of the target items. The decision to include both VST score and pretest score was also driven by the possibility of a Matthew effect taking place. The Matthew effect has been recognised in economics, politics and other fields. In the field of education, the Matthew effect is used to describe a phenomenon in which the weaker students stand to learn less and lag behind the more proficient learners (e.g., Stanovich, 1986). The Matthew effect has also been observed in the context of vocabulary learning through reading (e.g., Elgort, Perfetti, Rickles & Stafura, 2015; Horst, Cobb, & Meara, 1998) as well as L2 viewing (e.g., Feng & Webb, 2019; Montero Perez et al., 2014; Peters et al., 2016; Peters & Webb, 2018; Puimège & Peters, 2019), as

learners with a bigger vocabulary size learned more words incidentally. As such, the inclusion of both variables as fixed effects seemed fitting. Multicollinearity between these two factors was checked and not found in any of the analyses.

With the exception of the analysis on form recall learning gains and content comprehension, all the other analyses included the development and comparison of models that included a large number of added covariates. These models were compared to take into account the properties of the target items which might have affected learners' scores in the respective tests. The covariates included depended on the MWE tests. Table 12 summarises the covariates included in each analysis. Note that for lemmatised corpus frequency, MI score, figurativeness and compositionality, the values presented in the table are the range of mean values for all target items. The exact values for each target item are outlined under Section 3.4.3 (Table 22). See Methodology (Section 2.4.3) for descriptions of how these values were obtained.

**Table 12. Added covariates included in the model for all tests (for Study 1)**

Covariates	Range <sup>a</sup>	Tests <sup>b</sup>			
		Form recall tests	Form recognition test	Meaning recall test	Meaning recognition test
Number of words in the string (Length A)	2 - 5	✓	×	×	×
Number of words with missing letters (Length B)	1 - 3	✓	×	×	×
Number of letters provided, i.e., the amount of hints provided (Letter A)	1 - 5	✓	×	×	×
Number of letters required to receive partial credit (Letter B)	2 - 11	✓	×	×	×
Lemmatized phrasal frequency in COCA (log-transformed)	19 - 4071 (1.30 - 3.61)	✓	✓	✓	✓

MI score (log-transformed)	0.35 - 14.51 (0.36 - 1.19)	✓	✓	✓	✓
Figurativeness	2.6 -7.0	✓	✓	✓	✓
Compositionality	2.6 - 6.0	✓	✓	✓	✓
Ambiguity <sup>c</sup>	1 - 2	✓	✓	✓	✓

<sup>a</sup> The second column shows the range of the variable. The adjusted range after transformation is presented in parentheses

<sup>b</sup> ✓ indicates that the covariate was included in the analysis, ✗ indicates that the covariate was not included in the analysis

<sup>c</sup> Ambiguity is a categorical covariate with two levels

For models which only included the core variables, the model comparison process started with the most complex model, which included all the core independent variables and their interactions as fixed effects. The significance level of a term was tested by comparison of models with and without the term. The comparison returns a likelihood ratio statistic which is reported in the form of  $(\chi^2(n1) = n2, p = n3)$ , where  $n1$  = degrees of freedom,  $n2$  = likelihood ratio statistics, and  $n3$  = the  $p$ -value, which needs to be less than .05 for the result to be significant. As regards the models which included the added covariates, the model comparison process also started with the most complex model (i.e., all core variables and their interactions), from which each term (i.e., the core variable or their interactions) was removed at a time, using backwards stepwise approach. The difference, however, was that all the added covariates were included in *every* model. Once the interim best-fit model was reached, all the terms were excluded one by one to check whether any predictor was redundant (i.e., a variable whose absence in the model did not lead to any significant difference) and could thus be excluded. Once these were excluded, we are left with the final best-fit model. For the *clmm* and *glmer* models, the random effect structure included just intercepts for Subject and Item.

### 3.4 Form recall - Results

As previously mentioned, two different statistical analyses were carried out to analyse the MWE form recall data. The first set of analysis was conducted to assess the effects of the variables in questions on the absolute learning gains between the pretest and immediate posttest, as well as between the pretest and the delayed posttest. Next, item-level analysis was

conducted to assess how the properties of the target items affected participants' score on the target items in the form recall tests. The following section looks at each statistical analysis in turn.

**3.4.1 Analysis of learning gains between the form recall pretest and immediate posttest (i.e., short-term gains)**

As mentioned under the Methodology section, two different scoring systems were used to score the form recall tests. The descriptive statistics based on both scoring systems are shown in Table 13. As the pretest data for both scoring systems were non-normally distributed with unequal variance, the data were first transformed using Tukey Ladder of Powers before running a one-way non-parametric ANOVA (Kruskal-Wallis test). The Kruskal-Wallis test revealed no significant differences between the groups in their scores in the pretest (strict scoring system: Kruskal-Wallis  $\chi^2(5) = 7.37, p = .19$ ; lenient scoring system: Kruskal-Wallis  $\chi^2(5) = 4.58, p = .47$ ).

**Table 13. Descriptive statistics for the form recall pretest, immediate posttest and delayed posttest based on both scoring systems (for Study 1)**

Condition	Pretest		Immediate posttest		Delayed posttest	
	Mean (SD)					
Enhanced captions X 1	1.79 (1.30)	0.58 (0.96)	6.15 (3.51)	4.53 (3.45)	3.38 (2.41)	2.06 (2.08)
Normal captions X 1	3.82 (3.61)	1.91 (2.52)	7.34 (4.81)	6.00 (5.07)	5.54 (4.28)	4.00 (4.21)
Uncaptioned X 1	2.78 (3.41)	3.31 (1.40)	3.55 (3.85)	2.47 (3.80)	3.72 (3.82)	2.40 (3.46)
Enhanced captions X 2	2.01 (2.34)	0.92 (1.78)	9.48 (4.47)	8.53 (4.87)	4.17 (4.66)	3.20 (4.56)

Normal captions X 2	2.68 (2.24)	1.23 (1.38)	8.40 (4.25)	7.19 (4.52)	4.46 (3.82)	3.22 (3.34)
Uncaptioned X 2	2.92 (2.47)	1.22 (1.77)	3.89 (3.21)	2.56 (3.13)	3.24 (2.91)	1.67 (1.85)

\*Note: Maximum possible score for each test is 18. The shaded columns indicate the values under the strict scoring system.

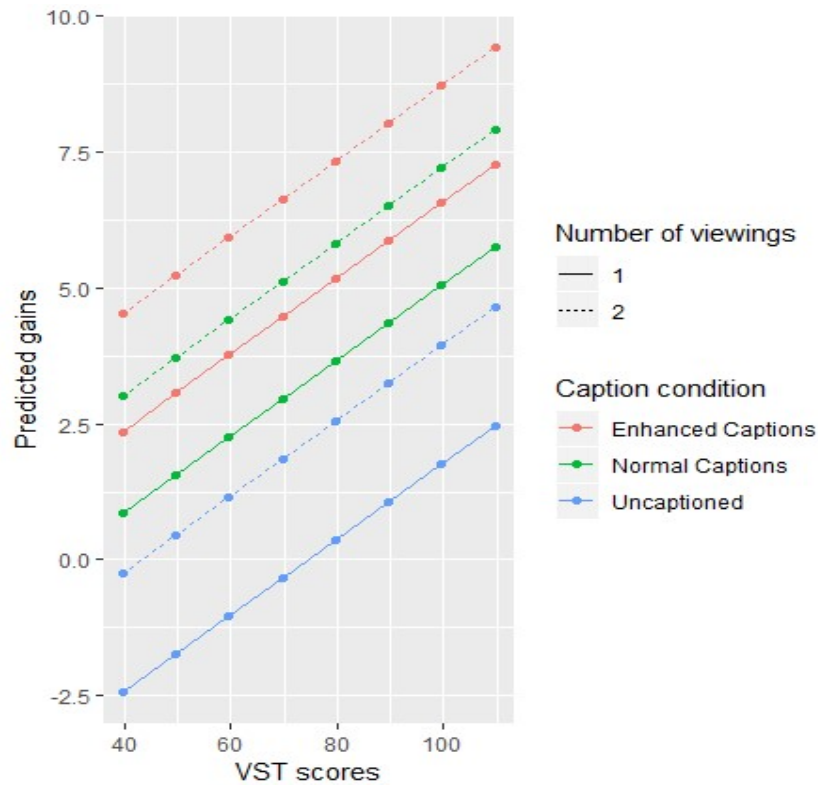
To determine the effects of the core variables on the absolute short-term learning gains, the response variable “short-term learning gains” was first created. This was done by subtracting participants’ total score in the pretest from their total score in the immediate posttest. The absolute short-term mean gains based on both scoring systems for all six conditions are shown in Table 14.

**Table 14. Absolute short-term gains based on both scoring systems (for Study 1)**

Condition	Strict	Lenient
	Mean (SD)	
Enhanced Captions X 1	3.91 (2.89)	3.84 (3.40)
Normal Captions X 1	3.07 (2.75)	3.65 (3.32)
Uncaptioned X 1	0.77 (1.81)	1.07 (1.49)
Enhanced Captions X 2	6.71 (4.38)	6.92 (4.67)
Normal Captions X 2	5.34 (3.44)	5.63 (3.79)
Uncaptioned X 2	0.97 (2.20)	1.33 (2.03)

The short-term gains data were fitted and run using the *lm* function according to the procedure detailed under Section 3.3. The fixed effects included caption condition, number of viewings, VST score, participants’ form recall total pretest score, as well as the interactions between the terms. The results based on the lenient scoring data revealed no interactions between the variables. Further, pretest score was not a significant predictor of short-term gains ( $p = .19$ ). The results of the regression analysis showed that caption condition, number of viewings and VST score explained 36% of the variance in the learners’ gains from the pretest to the immediate posttest ( $R^2 = .36$ ,  $F(4, 117) = 18.1$ ,  $p < .001$ ). Plot 1 below shows the predicted short-term gains of all six conditions, for all the VST score ranges. Caption condition

was a significant predictor of short-term gains ( $p < .001$ ). As shown in Plot 1 and the output of the best-fit model in Table 15, the predicted short-term gains under the normal captions and enhanced captions conditions were higher than the uncaptioned conditions. Multiple comparison using the *emmeans* function (Lenth, 2018) was also carried out with the Bonferroni method  $p$ -value adjustment. The results revealed that the gains under the enhanced and normal captions were indeed significantly higher compared to the uncaptioned conditions (both  $p$ -values  $< .0001$ ). While the gains under the normal captions were lower compared to the enhanced captions conditions, the difference was only marginally significant ( $p = .06$ ). Besides caption conditions, number of viewings ( $p < .001$ ) was also found to be a significant predictor of learning gains, with viewing twice leading to bigger gains. The VST score also had a positive effect on learning gains ( $p < .001$ ). That is, participants with higher VST scores had greater learning gains. From Table 15, it is also clear that the predicted gains for participants (with the mean VST score) who watched the video once under the captioned conditions (i.e., EC1 = 4.46; NC1 = 2.95) were bigger than the gains for participants who watched the video twice without captions (i.e., UC2 = 1.84). Multiple comparisons, however, revealed that only the difference between EC1 and UC2 was significant.



**Plot 1. Predicted short-term gains under the lenient scoring system (Study 1)**

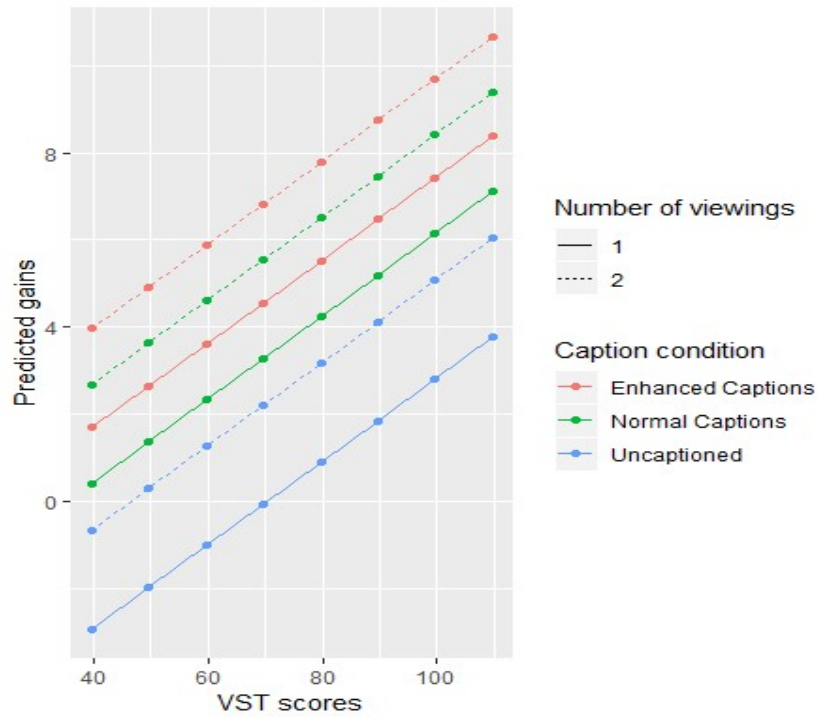
**Table 15. Output of the best-fit model for predicting short-term gains based on the lenient scoring system (Study 1)**

[model: lenient.imme.gains ~ caption condition+VST.mean+no.of.viewing]

Parameter	Estimate	SE	t	p
Intercept*	-0.33	0.60	-0.54	.59
Caption condition (Normal captions)	3.28	0.69	4.74	< .001
Caption condition (Enhanced captions)	4.79	0.69	6.89	< .001
VST Score (centred)	0.07	0.02	3.80	< .001
Number of viewings (Twice)	2.17	0.55	3.93	< .001

\*Note. Intercept level: Caption condition = Uncaptioned; Number of viewings = Once

The results based on the strict scoring data revealed patterns that are largely similar to those of the lenient scoring data. Short-term gains under this scoring system were also significantly predicted by all three factors, i.e., captioned condition, number of viewings and VST score, and the results of the regression showed that the three factors explained 38% of the variance ( $R^2 = .38$ ,  $F(4, 117) = 18.5$ ,  $p < .001$ ). Plot 2 shows the short-term gains under the strict scoring system for all six conditions and all the VST score ranges. As shown in the output of the best-fit model (Table 16), the predicted learning gains under the normal and enhanced captions conditions were higher compared to the uncaptioned condition. Multiple comparisons with the Bonferroni method  $p$ -value adjustment also confirmed that these differences were significant (both  $p$ -values  $< .0001$ ). The difference between the enhanced and normal captions, however, was not significant ( $p = .19$ ). Besides caption condition, the number of viewings ( $p < .001$ ) was also found to be a significant predictor of learning gains. Further, VST score significantly predicted learning gains ( $p < .001$ ), with an increase of one unit in VST score leading to bigger gains. The pretest score was not a significant predictor of short-term gains ( $p = .95$ ). As shown in Table 16, the predicted gains for participants (with the mean VST score) who watched the video once under the captioned conditions (i.e., NC1 = 3.28; EC1 = 4.56) were bigger than the gains for participants who watched the video twice without captions (i.e., UC2 = 2.21). Multiple comparisons, however, revealed that only the difference between EC1 and UC2 was significant.



**Plot 2. Predicted short-term gains under the strict scoring system (Study 1)**

**Table 16. Output of the best-fit model for predicting short-term gains based on the strict scoring system (for Study 1)**

[model: strict.imme.gains ~ caption condition+VST.mean+no.of.viewing]

Parameter	Estimate	SE	t	p
Intercept*	-0.06	0.64	-0.09	.93
Caption condition (Normal captions)	3.34	0.73	4.60	< .001
Caption condition (Enhanced captions)	4.62	0.73	6.32	< .001
VST Score (centred)	0.10	0.02	4.94	< .001
Number of viewings (Twice)	2.27	0.58	3.92	< .001

\*Note. Intercept level: Caption condition = Uncaptioned; Number of viewings = Once.

### 3.4.2 Analysis of learning gains between the form recall pretest and delayed posttest (i.e., long-term gains)

The absolute long-term gains were obtained by subtracting participants' total score in the pretest from their total score in the delayed posttest. Table 17 shows the mean long-term

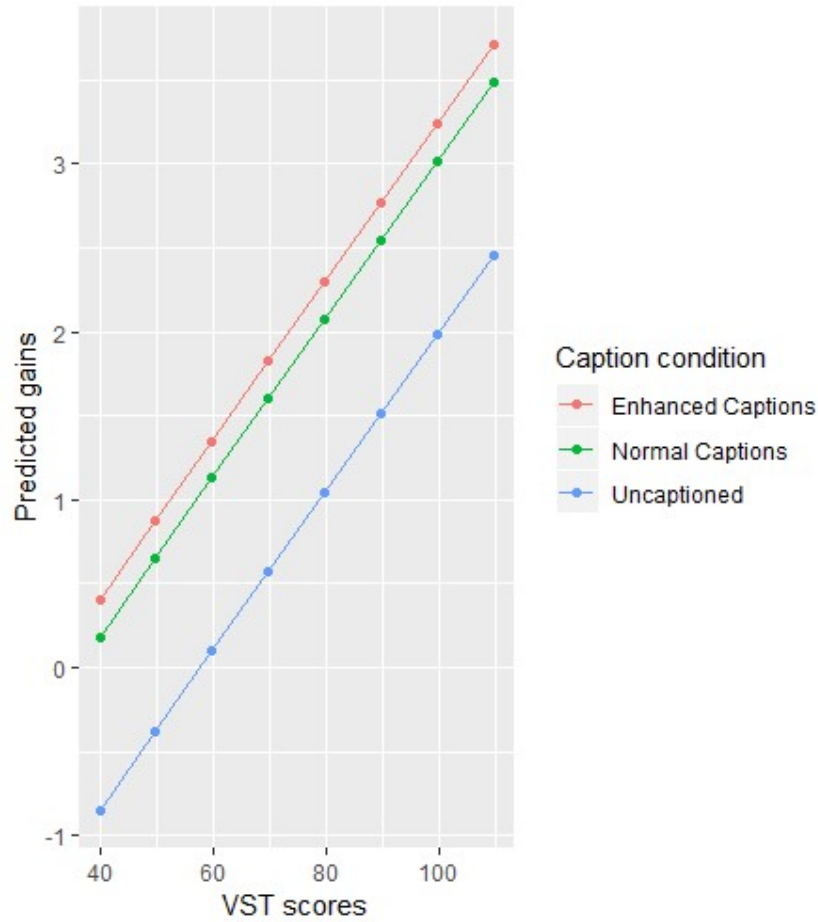


gains for each condition based on both scoring systems. Similar to the short-term gains, data for the long-term gains were also fitted and run using the *lm* function.

**Table 17. Absolute long-term gains based on both scoring systems (for Study 1)**

Condition	Strict	Lenient
	Mean (SD)	
Enhanced Captions X 1	1.08 (2.22)	1.16 (1.42)
Normal Captions X 1	1.50 (1.33)	1.83 (1.99)
Uncaptioned X 1	0.93 (1.70)	1.00 (1.31)
Enhanced Captions X 2	2.16 (2.73)	2.28 (3.03)
Normal Captions X 2	1.93 (2.33)	2.10 (2.49)
Uncaptioned X 2	0.29 (1.42)	0.44 (1.04)

The results of the regression based on the lenient scoring system revealed no significant interactions between the variables. It was also found that caption condition significantly predicted long-term learning gains ( $p = .02$ ). Table 18 below shows the output of the best-fit model, while Plot 3 illustrates the predicted long-term gains for all six conditions and all VST score ranges. As shown in Plot 3 and Table 18, compared to the uncaptioned condition, the predicted long-term learning gains under the normal captions and enhanced captions were higher. Multiple comparisons with the Bonferroni method  $p$ -value adjustment, however, revealed that only the difference between the uncaptioned and enhanced captions was significant ( $p = .02$ ). The difference between the uncaptioned and normal captions was not significant ( $p = .06$ ). Additionally, the difference between the enhanced and normal captions was not significant ( $p = 1.00$ ). The VST score was also found to be a significant predictor ( $p < .001$ ). The two predictors together explained 17% of the variance ( $R^2 = .17$ ,  $F(3, 118) = 7.84$ ,  $p < .001$ ). The number of viewings ( $p = .13$ ) and pretest score ( $p = .35$ ) were not significant, suggesting that these two factors did not exert a strong influence on the long-term learning gains.



**Plot 3. Predicted long-term gains under the lenient scoring system (Study 1)**

**Table 18. Output of the best-fit model for predicting long-term gains based on the lenient scoring system (for Study 1)**

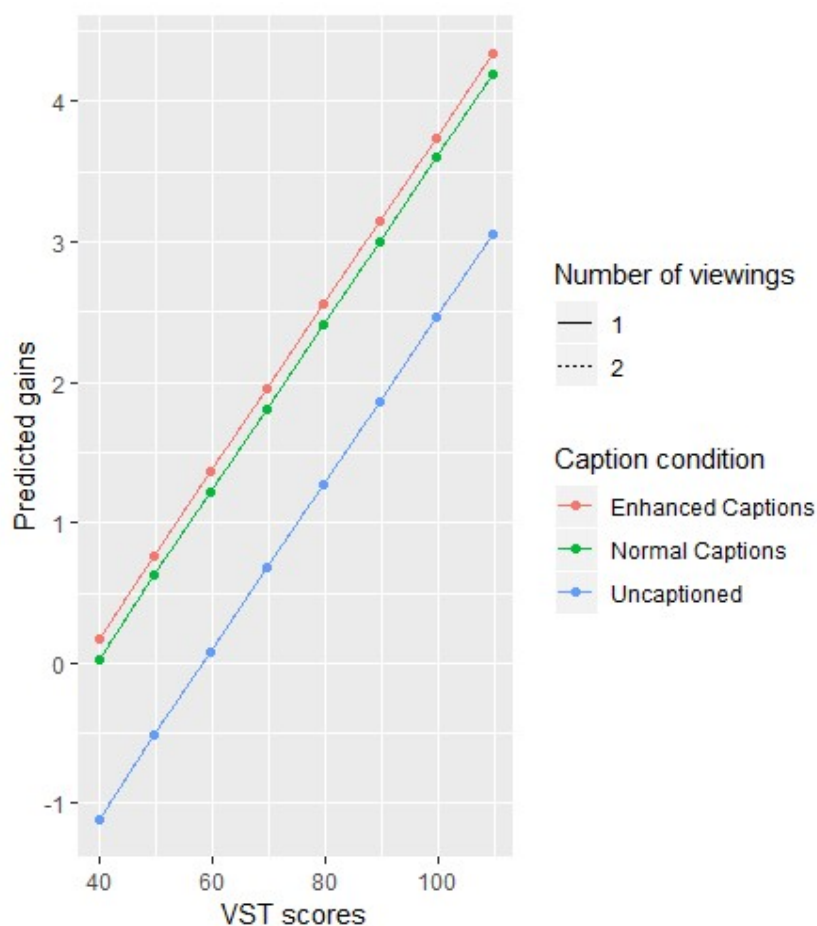
[model: lenient.delayed.gains ~ caption condition+VST.mean]

Parameter	Estimate	SE	t	p
Intercept*	0.57	0.34	1.66	.10
Caption condition (Normal captions)	1.03	0.45	2.29	.02
Caption condition (Enhanced captions)	1.25	0.45	2.77	< .01
VST score (Centred)	0.05	0.01	3.96	< .001

*Note.*\* Intercept level: Caption condition = Uncaptioned.

Similar results were revealed for the data based on the strict scoring system. Long-term gains under this scoring system were also significantly predicted by caption condition ( $p = .01$ )

and VST score ( $p < .001$ ), and these two factors explained 21% of the variance ( $R^2 = .21$ ,  $F(3, 118) = 11.7$ ,  $p < .001$ ). The number of viewings ( $p = .12$ ) and pretest score ( $p = .42$ ) were not found to be significant predictors of long-term gains. Table 19 shows the output of the best fit model, while Plot 4 shows the predicted long-term gains under all six conditions for all VST score ranges. Multiple comparison with Bonferroni correction applied revealed that compared to the uncaptioned conditions, gains under the enhanced captions were significantly higher ( $p = .01$ ). The difference between the uncaptioned and normal captions, however, fell just below significance level ( $p = .05$ ). The difference between the enhanced and normal captions was also not significant ( $p = 1.00$ ).



**Plot 4. Predicted long-term gains under the strict scoring system (Study 1)**

**Table 19. Output of the best-fit model for predicting long-term gains based on the strict scoring system (for Study 1)**

[model: strict.delayed.gains ~ caption condition+VST.mean]

Parameter	Estimate	SE	t	p
Intercept*	0.68	0.34	2.02	.04
Caption condition (Normal captions)	1.14	0.44	2.57	.01
Caption condition (Enhanced captions)	1.28	0.45	2.88	< .01
VST score (centred)	0.06	0.01	5.09	< .001

*Note.* \* Intercept level: Caption condition = Uncaptioned.

In summary, the results based on both scoring systems revealed that learning gains on MWE form recall were influenced by caption condition, number of viewings and participants' vocabulary size. Of the three predictors, however, only the effects of caption condition and vocabulary size were durable, as number of viewings did not significantly predict participants' long-term gains. As the outcome for both scoring systems were similar, the item-level analysis only focused on the data based on the lenient scoring system.

### 3.4.3 Form recall item-level analysis

Item-level analysis was performed for a number of reasons. Firstly, item-level analysis could shed light on how each treatment condition affected the participants' scores on the individual target items. In essence, item-level analysis examined whether the same trends observed in the previous analysis still stood. Next, item-level analysis could assess whether the properties of the target items, such as their length, COCA frequency and MI score influenced participants' score on the target items. For this analysis, the dependent variable was learners' scores on the target item. As mentioned under Section 3.3, for each target item, there were five possible ordered scores, i.e., 0, 0.25, 0.5, 0.75 and 1. Therefore, the *clmm* function was used. Two different sets of models were developed and compared. The first set of models included the core independent variables only, while the next set of models included the core independent variables as well as the item-related variables, which are called added covariates. These models will now be discussed in turn.

*Modelling data with core variables*

The fixed effects included the core independent variables that were of central importance, namely caption condition, number of viewings, test time, VST score (centred) and the interactions between these variables. The removal of each term was then tested according to the procedure detailed under Section 3.3. Random effects included Subject and Item.

Table 20 shows the output of the best-fit model. Model comparison revealed a significant three-way interaction between caption condition, number of viewings and test time ( $\chi^2(4) = 9.97, p = .04$ ). This means that the effect of the number of viewings depended on the caption condition, and the nature of this relationship was different depending on the test time. Planned comparisons were run using the *emmeans* function in the *emmeans* package (Lenth, 2018), with Bonferroni correction applied. The results showed that for both number of viewings, participants in the enhanced captions condition and the normal captions condition were significantly more likely to get a higher score (e.g., 0.75 to 1) on the target items compared to the uncaptioned condition (all  $p$ -values  $< .05$ ). This significant difference, however, was only found for the immediate posttest. The difference between the enhanced captions condition and normal captions condition for any of the tests did not reach statistical significance. Of the three caption conditions, the effect of repetition was only significant for the enhanced caption condition, in that compared to watching once (EC1), watching twice (EC2) led to a higher likelihood of receiving a higher score on the target items ( $p < .0001$ ). However, this difference was only significant for the immediate posttest. Although viewing twice under the normal and unenhanced caption conditions also led to higher scores on the target items compared to viewing once, the difference was not significant.

**Table 20. Output of the best-fit model (with core variables only) for the form recall item-level analysis (for Study 1)**

[model: ordered.score ~

caption.condition\*no.of.viewing\*test.time+VST.mean\*test.time+(1|Subject)+(1|Item)]

Parameter	Estimate	SE	Z	p
Caption condition (Normal captions)	0.57	0.37	1.56	.12
Caption condition (Enhanced captions)	-0.16	0.40	-0.41	.68
Test time (Immediate posttest)	0.29	0.24	1.21	.23
Test time (Delayed posttest)	0.40	0.24	1.66	.10

VST score (centred)	0.05	0.01	7.89	< .001
Number of viewings (Twice)	0.50	0.40	1.27	.21
Test time (Immediate posttest)	0.27	0.32	0.84	.40
X Number of viewings (Twice)				
Test time (Delayed posttest)	-0.18	0.32	-0.58	.56
X Number of viewings (Twice)				
Caption condition (Normal captions)	0.92	0.29	3.17	< .01
X Test time (Immediate posttest)				
Caption condition (Enhanced captions)	1.49	0.32	4.69	< .001
X Test time (Immediate Posttest)				
Caption condition (Normal captions)	0.19	0.29	0.70	.50
X Test time (Delayed posttest)				
Caption condition (Enhanced captions)	0.61	0.33	1.86	.06
X Test time (Delayed posttest)				
Test time (Immediate posttest)	0.01	0.01	2.42	.02
X VST Score (centred)				
Test time (Delayed posttest)	0.01	0.01	1.87	.06
X VST Score (centred)				
Caption condition (Normal captions)	-0.97	0.51	-1.92	.05
X Number of viewings (Twice)				
Caption condition (Enhanced captions)	-0.51	0.53	-0.97	.33
X Number of viewings (Twice)				
Caption condition (Normal captions)	0.69	0.40	1.72	.09
X Test time (Immediate posttest)				
X Number of viewings (Twice)				
Caption condition (Enhanced captions)	1.25	0.43	2.90	< .01
X Test time (Immediate posttest)				
X Number of viewings (Twice)				
Caption condition (Normal captions)	0.36	0.40	0.90	.37
X Test time (Delayed posttest)				
X Number of viewings (Twice)				
Caption condition (Enhanced captions)	0.36	0.44	0.82	.41
X Test time (Delayed posttest)				

X Number of viewings (Twice)

---

*\*Intercept level: Caption condition = Uncaptioned; Number of viewings = Once; Test time = Pretest*

---

The results also revealed a significant interaction between test time and VST score ( $\chi^2(2) = 6.27, p = .04$ ). To compare the VST slopes at each test time, the *emtrends* function was used. As shown in Table 21, the VST slopes at each test time are different, with the slopes at the immediate posttest (0.068) and delayed posttest (0.064) steeper than the pretest (0.054). This means that the effect of VST score was bigger at the immediate and delayed posttest compared to the pretest.

**Table 21. Estimate of the VST slope at each test time**

Test time	VST score trend	SE	Lower CL	Upper CL
Pretest	0.0539	0.0068	0.0405	0.0673
Immediate posttest	0.0677	0.0067	0.0546	0.0808
Delayed posttest	0.0643	0.0067	0.0513	0.0774

#### *Modelling data with added covariates*

To assess the effects of the properties of the target items, complex models that included a number of covariates were fitted and compared. The covariates included the following: (i) number of words in the string (length A), (ii) the number of words with missing letters (length B), (iii) the number of letters provided (i.e., the amount of hints provided) (letter A), (iv) the number of letters required to receive at least partial credit (letter B), (v) lemmatised phrasal frequency, (vi) MI score, (vii) figurativeness, (viii) compositionality and (ix) ambiguity. The lemmatised phrasal frequency and MI score were retrieved from the online Corpus of contemporary American English (COCA) and were log-transformed prior to statistical analysis.

In order to include figurativeness, compositionality and ambiguity in the model, ratings were obtained (see Section 2.4.3). Figurativeness and compositionality were included in the model using the mean rating value. Ambiguity was a categorical factor. Items that were deemed as unambiguous by the majority of the respondents in the rating procedure (i.e., at least six out

of ten) were coded 1, while those that were rated as ambiguous by more respondents were coded 2 (see Table 22).

**Table 22. Summary of added covariates included in the item-level analysis (for Study 1)**

<b>Target MWEs</b>	<b>COCA</b>	<b>MI Score</b>	<b>Figurativeness (mean rating)</b>	<b>Compositionality (mean rating)</b>	<b>Ambiguity</b>
on the same page	796	9.93	4.6	4.8	2
turn a profit	484	10.48	4.0	5.0	1
on (someone's) hands	4071	4.52	4.1	3.4	2
tighten up	263	5.43	3.7	4.8	2
kill someone with kindness	19	11.76	7.0	3.4	1
root for	1845	2.48	6.0	2.6	1
(one's) hands are tied	132	5.55	5.1	5.2	2
slippery slope	676	13.87	3.8	4.3	2
put (someone) on the spot	182	5.04	6.2	3.8	2
talk some sense into (someone)	60	5.98	4.1	6.0	1
beg to differ	288	10.98	3.9	4.9	1
whisked away	264	6.38	5.1	4.4	2
chip in	1279	1.54	5.2	2.8	2
bear with (someone)	517	0.35	5.2	4.7	2
go through a rough patch	28	14.51	4.9	5.2	2
let up on (someone)	106	1.3	5.4	3.7	1
look out for (someone)	2174	3.18	4.0	6.0	2
work (something) out with (someone)	25	8.53	2.6	6.0	2



As the point of interest was to assess the effects of the covariates after accounting for the effects of the core independent variables, the complex models were constructed in a similar way to the previous analysis which only included the core variables. This means that the most complex model included as fixed effects all the core variables, i.e., caption condition, number of viewings, test time, VST score and their interactions. Each term was then removed incrementally. In every model, however, the nine covariates were included. This model development and comparison procedure was chosen as it could assess whether the terms that were previously found significant (in the models with the core variables only) remained significant even after including the added covariates. Further, this model development accounted for the possibility of the added covariates affecting the scores on the target items, after accounting for the effects of the core independent variables. Initially, all nine covariates were included in the most complex model. However, there was an issue with convergence, which was likely due to overfitting, i.e., “having too many variables in the model that essentially make little contribution to predicting the outcome” (Field, Miles & Field, 2012, p. 266). In order for the model to converge, the covariate ‘figurativeness’ had to be removed. The most complex model then included the core variables and their interactions, as well as the eight added covariates.

Results revealed that the best-fit model did not vary significantly from that which only included the core variables. None of the added covariates significantly influenced participants’ score on the target items in the form recall tests. The interaction between captioned condition, number of viewings and test time ( $\chi^2(4) = 9.89, p = .04$ ) remained significant, as did the interaction between VST score and test time ( $\chi^2(2) = 8.07, p = .02$ ). The best-fit model was therefore identical to the previous analysis which only included the core variables.

In summary, the item-level analysis shed light on the ways in which the independent variables affected learners’ scores on the target items. While the two caption conditions led to significantly higher scores compared to the uncaptioned condition under both number of viewings, this was only observed at the immediate posttest. Further, a significant difference between viewing once and twice was only observed for the enhanced caption conditions, and only for the immediate posttest. Additionally, while learners’ scores on the target items were also influenced by their VST scores, the effect of the VST score was found to vary, as it was stronger at the two posttests compared to the pretest. Finally, the added covariates were not found to exert a strong influence on learners’ scores on the target items.

### 3.5 Form recognition - Results

The descriptive statistics for the form recognition test are shown in Table 23. As mentioned in Section 3.3, since the dependent variable was a binary outcome, the function *glmer* (generalised linear mixed model) was used to analyse the data.

**Table 23. Descriptive statistics of the form recognition test (for Study 1)**

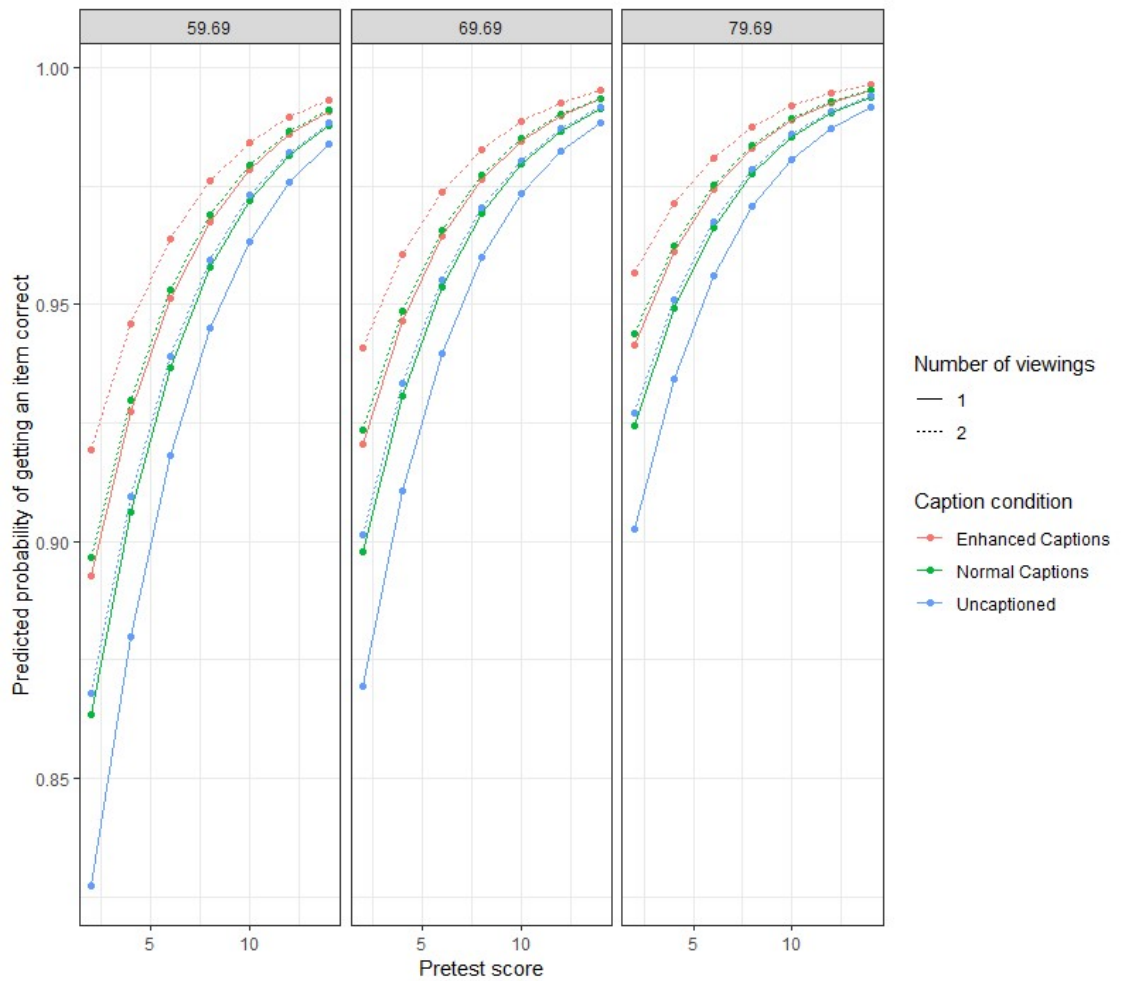
Conditions	Mean	SD
Enhanced Captions X 1	8.41	2.69
Normal Captions X 1	10.21	4.21
Uncaptioned X 1	9.20	3.38
Enhanced Captions X 2	10.09	3.68
Normal Captions X 2	9.91	3.80
Uncaptioned X 2	8.11	3.32

#### *Modelling data with core variables*

The models were fitted and compared according to the procedure outlined in Section 3.3. When a model produced a message warning of failed convergence, the ‘bobyqa’ optimiser was used to produce a stable model. The most complex model included all the core variables i.e., caption conditions, number of viewings, VST score (centred), form recall pretest score, and their interactions as fixed effects. Subject and Item were treated as random effects.

The results revealed that all four fixed effects were significant predictors: VST score ( $\chi^2(1) = 27.6, p < .001$ ), number of viewings ( $\chi^2(1) = 5.33, p = .02$ ), caption condition ( $\chi^2(2) = 9.95, p < .01$ ), and pretest score ( $\chi^2(1) = 32.4, p < .001$ ). No interactions between the predictors were found to be significant. Plot 5 shows the predicted probabilities of getting an item correct for three VST score ranges (i.e., 59.69, 69.69 and 79.69) and all six conditions and pretest score ranges. As shown in the plot, the predicted probabilities of getting an item correct are higher for those who watched the video twice compared to once, under all caption conditions. Further, higher VST scores led to higher probabilities of getting an item correct. The output of the best-fit model shown in Table 24 revealed identical trends. The output includes odds ratios. The odds are defined as the probability of an event occurring divided by the probability of it not occurring (Field, Miles, & Field, 2012, p. 319). As shown in Table 24,

the estimated odds of a correct response for participants who watched the video twice were 1.36 times the odds of those who watched the video once. Further, an increase in one-unit of the VST score resulted in a higher likelihood of getting an item correct. Similarly, the odds of getting an item right increased with higher pretest scores. As for the effects of caption condition, multiple comparisons using the *emmeans* function (Lenth, 2018) with the Bonferroni corrections applied revealed that the odds of a correct response under the enhanced captions were significantly higher than under the uncaptioned condition ( $p < .01$ ). Although the odds of a correct response were also lower under the normal captions compared to the enhanced captions, the difference was not significant ( $p = .26$ ). Similarly, no significant difference between the normal captioned and the uncaptioned condition was found ( $p = .28$ ).



**Plot 5. Predicted probability of getting an item correct in the form recognition test (Study 1)**

**Table 24. Output of the best-fit model (with core variables only) for the form recognition test (for Study 1)**

[model: Score ~

no.of.viewing+VST.mean+pretest.score+caption.condition+(1|Subject)+(1|Item)]

Parameter	Estimate	SE	z	p	OR
Intercept*	-0.83	0.34	-2.42	.02	
Number of viewings (Twice)	0.31	0.14	2.33	.02	1.36
VST Score (centred)	0.03	0.01	5.52	< .001	1.03
Pretest Score	0.21	0.04	5.88	< .001	1.23
Caption condition (Normal captions)	0.28	0.17	1.70	.09	1.32
Caption condition (Enhanced captions)	0.55	0.17	3.21	< .01	1.73

Note: OR = odds ratio. \*Intercept levels: Number of viewings = Once; Caption condition = Uncaptioned

#### Modelling data with added covariates

The next set of models was developed to include covariates that might affect the learners' odds of getting an item correct. The following covariates were included: (i) lemmatised phrasal frequency, (ii) MI score, (iii) figurativeness, (iv) compositionality and (v) ambiguity (see Section 3.3). Of interest here was whether the added covariates had a significant effect on the participants' score in the delayed form recognition test, after accounting for the effects of the core independent variables. Thus, the *glmer* models were compared the same way as the previous analysis: each core variable (i.e., VST score, caption condition, number of viewings, pretest score and the interactions terms) was incrementally removed. The difference is that, for each model, the five covariates were included.

Similar to the analysis that only included the core variables, the results of the complex model showed that all four predictors were significant: pretest score ( $\chi^2(1) = 32.4, p < .001$ ); VST score ( $\chi^2(1) = 27.63, p < .001$ ); caption condition ( $\chi^2(2) = 9.95, p = .007$ ) and number of viewings ( $\chi^2(1) = 5.33, p = .003$ ). The results also revealed that none of the covariates were found to be significant: frequency ( $\chi^2(1) = 0.62, p = .43$ ), MI score ( $\chi^2(1) = 1.98, p = .16$ ), figurativeness ( $\chi^2(1) = 0.76, p = .38$ ), compositionality ( $\chi^2(1) = 1.20, p = .27$ ) and ambiguity ( $\chi^2(1) = 0.95, p = .33$ ). The best fit model was, therefore, identical to that of the previous analysis which only included the core variables. This suggests that participants' scores on the form recognition test were not significantly influenced by the item-related variables.

In summary, participants' long-term form recognition was found to be affected by caption condition, number of viewings, participants' vocabulary size and prior knowledge of the target items. Additionally, the properties of the target MWEs did not affect participants' MWE form recognition.

### 3.6 Meaning recall - Results

The descriptive statistics for the meaning recall test are shown in Table 25. As mentioned in Section 3.3, the data were analysed using the *clmm* (cumulative link mixed model) function in the *ordinal* package (Christensen, 2019). Similar to the form tests, the meaning tests were also analysed in two different ways. The first set of models only included the core variables and their interactions as the fixed effects, while the second set of models included the added covariates.

**Table 25. Descriptive statistics for the meaning recall test (for Study 1)**

Conditions	Mean	SD
Enhanced Captions X 1	4.31	2.95
Normal Captions X 1	6.18	3.23
Uncaptioned X 1	4.93	2.57
Enhanced Captions X 2	4.29	3.12
Normal Captions X 2	5.61	2.87
Uncaptioned X 2	3.91	2.43

*\*Maximum possible total score is 18*

#### *Modelling data with core independent variables*

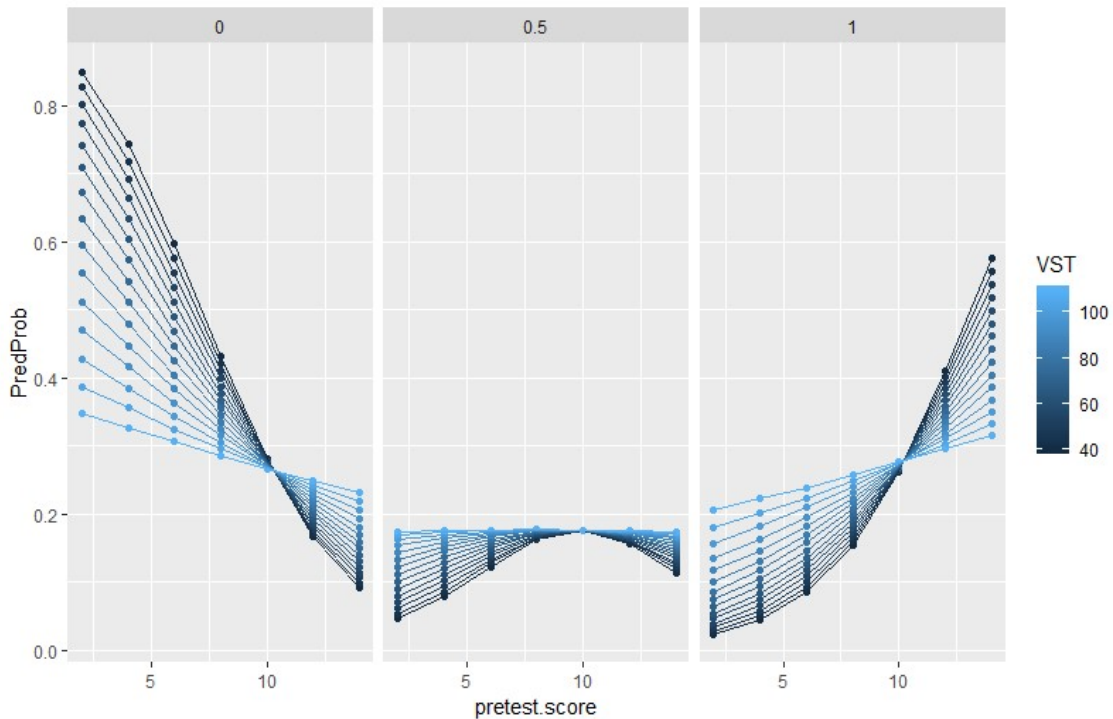
The most complex model started with the inclusion of all the fixed effects, i.e., caption condition, number of viewings, VST score, pretest score and their interactions. Random effects included Item and Subject. The removal of each term was then tested according to the procedure detailed under Section 3.3. The results revealed a significant main effect of VST score ( $\chi^2(1) = 18.2, p < .001$ ) and pretest score ( $\chi^2(1) = 8.9, p < .01$ ). There was also a significant interaction between these two factors ( $\chi^2(1) = 3.84, p = .04$ ). No significant effects emerged for number of viewings ( $\chi^2(1) = 0.02, p = .89$ ) or caption condition ( $\chi^2(2) = 5.75, p = .06$ ).

**Table 26. Output of the best fit model (with core variables only) for the meaning recall test (for Study 1)**

[model: ordered.score ~ VST.mean\*pretest.score+(1|Item)+(1|Subject)]

Parameter	Estimate	SE	z	p	OR
Pretest score	0.21	0.05	4.16	< .001	1.23
VST score (centred)	0.04	0.01	4.98	< .001	1.04
Pretest score X VST score	-0.004	0.001	-2.40	.02	0.99

As shown by the output of the best-fit model in Table 26, the bigger the learners' vocabulary size and pre-existing knowledge, the higher the odds of receiving a higher score on a target item in the meaning recall test. To illustrate the interaction of VST score and pretest score, the predicted probabilities of an average participant receiving each possible score (i.e., 0, 0.25, 0.5, 0.75, 1) on an average item were obtained. As there were far fewer occurrences of participants receiving a score of 0.25 and 0.75 than the other three scores (i.e., 0, 0.5 and 1) they were omitted from the plot. As shown in Plot 6, as the participants' pretest score became higher, so did the predicted probabilities of getting a full score on an average target item. The same applied to the effect of VST score. The effect of pretest score, however, was bigger for those with a lower VST score. This is because for these participants, the increase in pretest score resulted in a dramatic increase in the predicted probabilities of getting a full score on the target items. Conversely, for participants with higher VST scores, the increase in pretest score led to a less dramatic increase in the predicted probabilities of getting an item completely correct. This suggests that for these participants, it was their VST score that predicted the meaning recall score, leaving less room for the influence of a pretest score compared to those with a lower VST score. Although the plot shows that VST score seemed to have an opposite effect after a pretest score of above ten, this is just an artefact of the prediction—only two participants received a pretest score above ten. This suggests that the predicted pattern prior to a pretest score of ten is a more accurate description of the positive effects of both VST score and pretest score on meaning recall.



**Plot 6. The effects of pretest score and VST score on the predicted probabilities of getting a score of 0, 0.5 and 1 on a target item in the meaning recall test (Study 1)**

*Modelling data with added covariates*

Following the analyses of the effects of the core variables, complex models were built to determine if the added covariates had an effect on participants' score on the meaning recall test. The five added covariates were (i) frequency, (ii) MI score, (iii) figurativeness, (iv) compositionality and (v) ambiguity. Model development and comparisons were done in a similar fashion to the previous analysis which only included the core variables; that is, each core dependent variable term was removed incrementally. What is different, however, is that for each model the five covariates were added at the back.

The results revealed that after accounting for the effects of the covariates, the interaction between pretest score and VST score was still significant ( $\chi^2(1) = 5.05, p = .02$ ). As regards the item-related variables, only compositionality was found to be significant ( $\chi^2(1) = 5.71, p = .02$ ). This suggests that after accounting for the effects of the core variables, compositionality of the target items still affected learners' scores in the meaning recall test. As shown in Table 27 the odds of receiving a higher score (e.g., 0.25 to 0.5) were higher for target items that were rated higher on the compositionality scale.

**Table 27. Output of the best fit model (with added covariates) for the meaning recall test (for Study 1)**

[model: ordered.score ~ VST.mean\*pretest.score+comp+(1|Item)+(1|Subject)]

Parameter	Estimate	SE	Z	p	OR
VST score (centred)	0.04	0.01	4.98	< .001	1.04
Pretest score	0.21	0.05	4.16	< .001	1.23
VST score X Pretest score	-0.004	0.001	-2.40	.02	1.00
Compositionality	0.59	0.29	2.01	.04	1.80

Note: OR = odds ratio.

In summary, participants' MWE meaning recall was found to be affected by their VST score and their pretest score. Further, the compositionality of the target MWEs influenced participants' scores in the meaning recall test.

### 3.7 Meaning Recognition - Results

Table 28 shows the descriptive statistics for the meaning recognition test. As mentioned under Section 3.3, the data were analysed using the *glmer* function as the dependent variable was a binary outcome. As in all the other tests, the models which only included the core variables were first built and compared. This was followed by the development of the complex models, in which the added covariates were included.

**Table 28. Descriptive statistics for the meaning recognition test (for Study 1)**

Conditions	Mean	SD
Enhanced Captions X 1	13.75	3.00
Normal Captions X 1	14.30	2.98
Uncaptioned X 1	14.80	2.98
Enhanced Captions X 2	12.96	3.37
Normal Captions X 2	14.95	2.90
Uncaptioned X 2	14.01	3.50

\*Maximum possible total score is 18



*Modelling data with core independent variables*

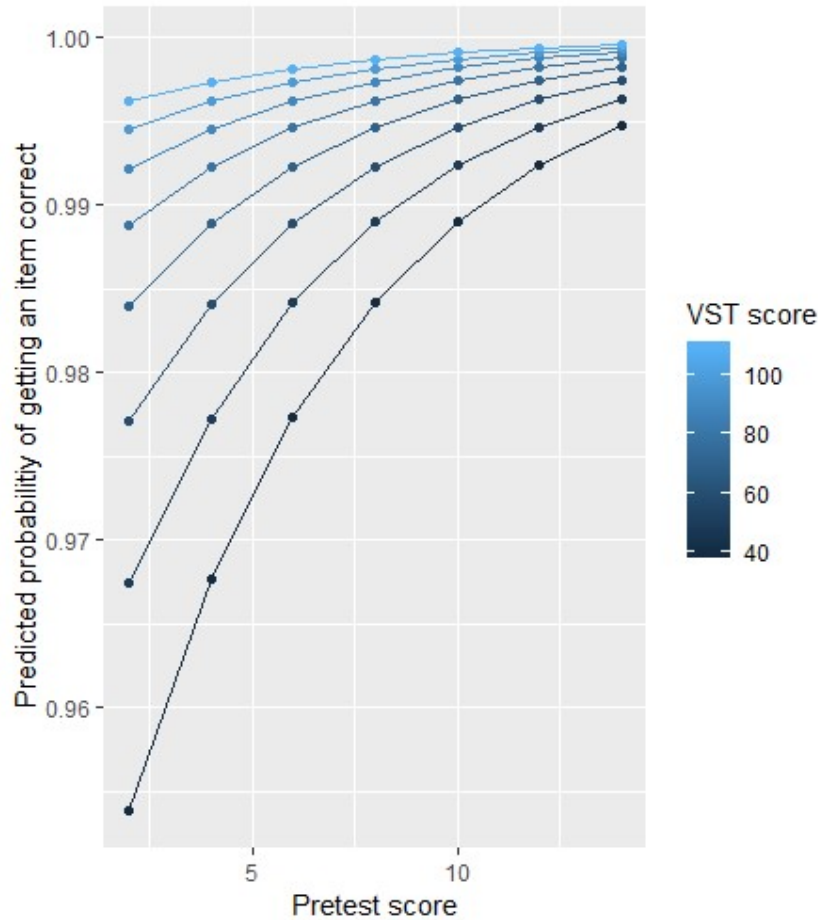
The most complex model included all the fixed effects, i.e., caption condition, number of viewings, VST score, pretest score and their interactions. Random effects included Item and Subject. The models were fitted and compared according to the procedure outlined under Section 3.3. When a model produced a message warning of failed convergence, the ‘bobyqa’ optimiser was used to produce a stable model.

The results revealed a significant effect of VST score ( $\chi^2(1) = 23.3, p < .001$ ) and pretest score ( $\chi^2(1) = 12.5, p < .001$ ). This suggests that learners’ odds of a correct answer in the meaning recognition test increased as their vocabulary size and pre-existing knowledge of the target MWEs became greater (Table 29). This trend can also be seen in Plot 7 below, which shows the predicted probability of getting an item correct for all VST score ranges and pretest total scores. Caption condition ( $\chi^2(2) = 1.49, p = .47$ ) and number of viewings ( $\chi^2(1) = 0.12, p = .73$ ) were not found to be significant, and no significant interactions were found between any of the predictors.

**Table 29. The output of the best-fit model (with core variables only) for the meaning recognition test (for Study 1)**

[model: score ~ VST.mean+pretest.score+(1|Subject)+(1|Item)]

Parameter	Estimate	SE	z	p	OR
Intercept	1.22	0.24	5.12	< .001	
VST score (centred)	0.04	0.01	5.15	< .001	1.04
Pretest score	0.18	0.05	3.80	< .001	1.20



**Plot 7. Predicted probability of getting an item correct in the meaning recognition test (Study 1)**

*Modelling data with added covariates*

Following the analysis of the effects of the core variables, complex models that included the added covariates were built and compared. The five added covariates were (i) frequency, (ii) MI score, (iii) figurativeness, (iv) compositionality and (v) ambiguity. Model development and comparisons were done in a similar way to the previous analysis that only included the core variables and their interaction, that is, each term was removed incrementally. What is different, however, was that for each model the five covariates were included.

In the analysis that included only the core variables, only VST score and pretest were significant. Model comparisons with the inclusion of the added covariates revealed similar results. Once again, both VST score ( $\chi^2(1) = 23.26, p < .001$ ) and pretest score were found to be significant ( $\chi^2(1) = 12.51, p < .001$ ). The effects of two covariates, namely frequency ( $\chi^2(1) = 5.39, p = .02$ ) and MI score ( $\chi^2(1) = 4.12, p = .04$ ) were also significant. This means that

after having accounted for the effects of the pretest score and VST score, recognition of MWE meanings was still significantly influenced by the properties of the target items, specifically by their lemmatised phrasal frequency and MI score. As shown by the output of the best fit model in Table 30, the odds of a correct response on the meaning recognition test were higher if the MI score and frequency increased by one unit.

**Table 30. The output of the best-fit model (with added covariates) for the meaning recognition test (for Study 1)**

[model: score ~ VST.mean+pretest.score+freq+MIscore+(1|Subject)+(1|Item)]

Parameter	Estimate	SE	Z	p	OR
Intercept	-0.91	1.07	-0.89	.40	
VST score (centred)	0.04	0.01	5.15	< .001	1.04
Pretest score	0.18	0.05	3.80	< .001	1.20
Frequency (log 10)	0.47	0.30	1.60	.03	1.60
MI score (log 10)	1.21	0.67	1.80	.04	3.35

In summary, participants' retention of MWE meaning at the recognition level was affected by their vocabulary knowledge. In essence, the better their pretest score and the larger their vocabulary size, the more likely the participants were to recognise the MWE meaning. In addition, the lemmatised phrasal frequency and MI score of the target MWEs also affected participants' recognition of the MWE meanings.

### 3.8 Comprehension test - Results

To assess the effects of the variables in question on participants' comprehension of the video content, their scores on the comprehension test were analysed. Table 31 shows the mean score and standard deviation for each treatment condition. The data were analysed using the *glmer* (generalised linear mixed) function in the *lme4* package (Bates, Maechler, Bolker & Walker, 2015), as the dependent variable was binary.

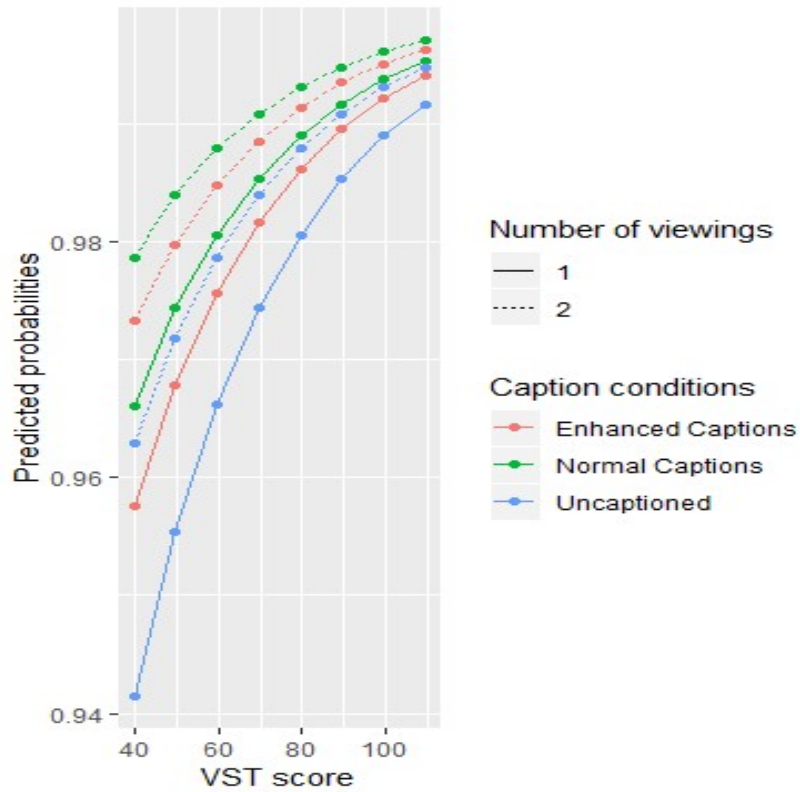
**Table 31. Descriptive statistics for the comprehension test (for Video 1)**

Conditions	Mean	SD
Enhanced Captions X 1	17.44	2.28
Normal Captions X 1	17.87	2.36
Uncaptioned X 1	16.67	1.11
Enhanced Captions X 2	18.55	2.61
Normal Captions X 2	18.00	2.51
Uncaptioned X 2	17.57	1.91

*Note. The maximum possible score was 22*

The fixed effects included the main effects of caption conditions, number of viewings, VST score and their interactions. Random effects included Subject and Item. The models were compared according to the procedure outlined under Section 3.3. The results revealed a significant effect of number of viewings ( $\chi^2(1) = 8.87, p < .01$ ), VST score ( $\chi^2(1) = 25.9, p < .001$ ) and caption condition ( $\chi^2(2) = 8.64, p = .01$ ). Based on the best-fit model, the predicted probabilities of getting an item correct were generated. The predicted probabilities are shown in Plot 8.

Table 32 shows the output of the best-fit model. The output revealed identical trends as the predicted probabilities. The odds of a correct answer for participants who viewed the video twice was 1.62 times the odds of those who viewed the video once. Further, the odds of a correct answer were higher if a participant's vocabulary size increased by one unit. The odds of obtaining a correct answer for the participants in the normal captions condition were 1.77 times the odds of obtaining a correct answer in the uncaptioned condition. Multiple comparisons with Bonferroni correction applied also revealed that the difference was significant ( $p < .01$ ). As for the enhanced captions, the odds of obtaining a correct answer were 1.40 the odds of participants in the uncaptioned conditions, meaning that the odds of getting a correct answer were higher in the former, compared to the latter. The difference, however, failed to reach significant level ( $p = .17$ ). Further, while the odds of a correct answer were higher in the normal captions compared to the enhanced captions, the difference was not significant ( $p = .41$ ). This suggests that the presence of normal captions facilitates comprehension, but the addition of enhanced captions reduced the benefits of captions on comprehension.



**Plot 8. Predicted probability of getting an item correct in the comprehension test (Study 1)**

**Table 32. The output of the best-fit model for the comprehension test (for Video 1)**

[model: comp.v1 ~ VST.mean+caption.condition+no.of.viewing+(1|Subject)+(1|Item)]

Parameter	Estimate	SE	z	p	OR
Intercept*	1.63	0.40	4.11	< .001	
VST score (centred)	0.03	0.01	5.29	< .001	1.03
Caption condition (Normal captions)	0.57	0.19	2.98	< .01	1.77
Caption condition (Enhanced captions)	0.34	0.19	1.78	.07	1.40
Number of viewings (Twice)	0.48	0.16	3.03	< .01	1.62

\*Intercept level: Caption condition = Uncaptioned; Number of viewings = Once

### 3.9 Summary of findings

The first study of the research project was carried out to investigate if there was an effect of caption condition on incidental MWE learning. Another objective of this study was to assess if repetition also had an effect, and to what extent it influenced MWE learning under the three caption conditions. To this end, data from one pretest, one immediate posttest and four delayed posttests were analysed. The four tests provided insights into the effects of the variables in question on multiple aspects of MWE knowledge. The findings revealed that the effects of the variables varied according to the type of MWE knowledge the test was meant to measure.

Firstly, form recall under the incidental learning condition was significantly influenced by caption condition, number of viewings and VST score. Where short-term gains are concerned, the captioned conditions proved to be superior as gains under both the normal and enhanced captions conditions were significantly higher compared to the uncaptioned conditions. Although the short-term gains under the enhanced captions conditions were higher compared to the normal captions conditions, the difference was just marginally significant. The effect of caption condition was also durable, as it emerged as a significant predictor of long-term gains. Specifically, the long-term gains under the enhanced captions were significantly higher than the uncaptioned condition. There was, however, no significant difference between long-term gains under the enhanced and normal caption conditions. Similarly, the difference between the normal captions and uncaptioned conditions was not significant. At the item-level analysis, the superiority of the captioned conditions was also revealed, as both the enhanced and normal captions conditions had significantly higher probabilities of getting a higher score on items in the immediate form recall posttest compared to the uncaptioned conditions. Further, this trend was observed for both number of viewings. Similar to the findings of the gains analyses, the item-level analyses showed that the enhanced captions conditions did not lead to significantly better performance compared to the normal captions condition.

Apart from caption condition, form recall was also influenced by number of viewings. Compared to viewing once, viewing twice led to significantly bigger short-term gains under all caption conditions. It was also revealed that viewing once under the normal and enhanced captions led to bigger short-term gains compared to viewing twice without any captions. Significant difference, however, was only found between EC1 and UC2. Further, the effect of repetition was not found to be durable, as long-term gains were not influenced by number of viewings. The item-level analysis also showed that repetition only influenced learners' scores

on the target items in the immediate posttest, and the significant difference was only found for the enhanced captions condition (i.e., EC2>EC1). As for the effects of VST score, it was revealed that as learners' vocabulary size became larger, so did their short-term and long-term gains.

As for form recognition under the incidental learning condition, this aspect of knowledge was significantly predicted by caption condition, number of viewings, VST score and pretest score. In essence, learners' odds of obtaining a correct answer in the form recognition test became higher as their VST score and pretest score increased. Further, participants in the viewing twice conditions were also more likely to obtain a correct answer in the form recognition test, compared to those who watched the video once. As for the effects of caption condition, normal captions did not lead to significantly higher odds of a correct answer compared to the uncaptioned conditions. The odds of getting an answer correct under the enhanced captions, however, were significantly higher compared to the uncaptioned conditions. This suggests that compared to watching an uncaptioned video, watching a video with enhanced captions under incidental learning conditions may lead to better form recognition. Parallel to the findings of the form recall tests, analysis of the form recognition test showed that enhanced captions did not lead to superior performance compared to normal captions.

Interestingly, neither caption conditions nor number of viewings had a significant effect on MWE meaning knowledge. Instead, meaning recall and meaning recognition were predicted by learners' vocabulary knowledge. An increase in the pretest score and VST score led to better performance in both the meaning recall as well as meaning recognition tests. Apart from that, item-related variables were also found to influence participants' performance in the meaning tests. For instance, the odds of receiving a higher score in the meaning recall test were higher for items that were rated higher on the compositionality scale. This suggests that participants were more likely to recall the meanings of items that were more compositional. As for meaning recognition, the odds of a correct answer are much higher when the MI score and corpus frequency of the MWEs increased.

With regard to the comprehension of the video content, the effects of all three variables, i.e., caption condition, number of viewings and VST score, were significant. Viewing twice led to higher odds of obtaining a correct answer in the comprehension test, as did an increase in the vocabulary size score. As for the effect of caption condition, the odds of a correct answer were significantly higher under the normal captions compared to the uncaptioned condition. No significant difference, however, was found between the uncaptioned and enhanced captions

condition. This suggests that the addition of typographically enhanced captions may reduce the benefits of captioning for comprehension. As such, the results of the comprehension test hints at a trade-off between a positive effect of typographically enhanced captions on the learning of MWEs and a negative effect on comprehension.

Overall, while the presence of captions did foster MWE knowledge, it appeared that it only exerted a strong influence on knowledge of MWE form. The same applied to the effects of repetition. Although repeated viewing led to bigger short-term form recall gains, and better form recognition, it did not emerge as a significant predictor of both meaning recall and meaning recognition. Finally, incidental acquisition of MWE knowledge of both form and meaning was influenced by learners' vocabulary knowledge, as in each test, VST score and/or pretest score emerged as a significant predictor. These results will be discussed (see Chapter Five) in more depth in light of the findings of Study 2, which will be addressed next.



## **CHAPTER FOUR:**

### **STUDY 2 (INTENTIONAL LEARNING) – ANALYSIS AND RESULTS**

This chapter presents the results of the tests administered in Study 2. The chapter starts by outlining the research questions this study intended to answer. This is followed by a short description of the statistical tools and terms. The subsequent sections (Section 4.2 to Section 4.7) will then describe the results of each analysis. The chapter concludes with a summary of the findings.

#### **4.1 Research questions and organisation of the chapter**

The study of the intentional learning conditions—operationalised as informing the students that a vocabulary test would follow—was carried out with a couple of objectives in mind. Firstly, this study investigated whether caption condition (i.e., uncaptioned, normal captions and enhanced captions) had an effect on intentional MWE acquisition through L2 viewing. Another aim of this study was to assess whether repetition, i.e., repeated viewing under intentional learning conditions modulated the effects of caption condition. Additionally, this study sought to determine whether repetition influenced intentional uptake of MWEs to the same degree under all three caption conditions.

Similar to the incidental learning study, intentional MWE learning was measured using one pretest, one immediate posttest and four delayed posttests. These five tests provided insights into the effects of the variables on multiple aspects of MWE knowledge. The subsequent sections are organised in a similar fashion to the previous chapter. The effects of the variables in question on the form tests will be discussed first, followed by the results of the meaning tests. After that, the results of the comprehension test are presented.

#### **4.2 Explanations of statistical tools and terms**

The data from all the tests were analysed on the R platform (Version 3.4.4, R Core Team, 2018). As the tests were all in the same format as those in Study 1, the statistical models and processes were identical to the ones described under Section 3.3. With the exception of the form recall gains analysis and the analysis of the comprehension test, all the other analyses included the comparisons of models that also included the added covariates. The added covariates refer to variables that were item-related. Table 33 summarises the covariates included in each analysis. Note that for lemmatised corpus frequency, MI score, figurativeness and compositionality, the values presented in the table are the range of mean values for all

target items. The exact values for each target item are outlined under Section 4.3.3 (Table 43). See Methodology (Section 2.4.3) for descriptions of how these values were obtained.

**Table 33. Added covariates included in the model for all tests (for Study 2)**

Covariates	Range <sup>a</sup>	Tests <sup>b</sup>			
		Form Recall Tests	Form Recognition Test	Meaning Recall Test	Meaning Recognition Test
Number of words in the string (Length A)	2-5	✓	×	×	×
Number of words with missing letters (Length B)	1-3	✓	×	×	×
Number of letters provided, i.e., the amount of hints provided (Letter A)	1-5	✓	×	×	×
Number of letters required to receive partial credit (Letter B)	1-9	✓	×	×	×
Lemmatized phrasal frequency in COCA (log-transformed)	10 – 12,154 (1.04 – 4.08)	✓	✓	✓	✓
MI Score (log-transformed)	0.43 – 15.32 (0.16 - 1.21)	✓	✓	✓	✓
Figurativeness	4.2 - 6.6	✓	✓	✓	✓
Compositionality	1.6 - 6.5	✓	✓	✓	✓
Ambiguity <sup>c</sup>	1 - 3	✓	✓	✓	✓

<sup>a</sup> The second column shows the range of the variable. The adjusted range after transformation is presented in parentheses

<sup>b</sup>✓ indicates that the covariate was included in the analysis, × indicates that the covariate was not included in the analysis

<sup>c</sup> Ambiguity is a categorical covariate with three levels

### 4.3 Form recall - Results

Analogous to Study 1, the form recall tests were analysed in two different ways to assess the effects of the variables on the learning gains as well as on the item-level. The following sections look at each analysis in turn.

#### 4.3.1 Analysis of learning gains between the form recall pretest and immediate posttest (i.e., short-term gains)

The form recall tests were marked using two scoring systems. Table 34 shows the descriptive statistics of the form recall tests based on both the lenient and strict scoring systems. As the pretest data were non-normally distributed with unequal variance, the data were first transformed using Tukey Ladder of Powers before running a one-way non-parametric ANOVA (Kruskal-Wallis test). The Kruskal-Wallis test revealed no significant differences between groups in their scores in the pretest (lenient scoring system: Kruskal-Wallis  $\chi^2(5) = 4.92, p = .42$ ; strict scoring system: Kruskal-Wallis  $\chi^2(5) = 1.66, p = .89$ )

**Table 34. Descriptive statistics for the form recall pretest, immediate posttest and delayed posttest based on both scoring systems (Study 2)**

Condition	Pretest		Immediate posttest		Delayed posttest	
	Mean* (SD)					
Enhanced Captions X 1	2.53 (2.18)	1.20 (1.77)	6.30 (4.22)	6.47 (3.83)	3.36 (2.26)	1.95 (2.11)
Normal Captions X 1	3.72 (3.80)	2.83 (4.02)	5.75 (4.53)	5.26 (4.69)	4.29 (4.42)	3.17 (4.08)
Uncaptioned X 1	2.27 (3.22)	1.47 (3.09)	3.35 (3.85)	2.13 (3.58)	3.53 (3.64)	2.33 (3.48)
Enhanced Captions X 2	2.63 (1.91)	1.52 (1.83)	8.43 (4.87)	7.60 (5.14)	5.19 (4.20)	4.28 (4.33)

Normal Captions X 2	2.89 (2.33)	1.57 (2.31)	6.29 (4.55)	5.48 (4.49)	3.54 (2.88)	2.52 (2.59)
Uncaptioned X 2	2.80 (2.95)	1.89 (3.05)	2.93 (2.87)	2.11 (2.79)	3.29 (2.89)	1.95 (2.93)

\*Maximum possible total score is 20. The shaded columns indicate the values under the strict scoring system.

To determine the effects of the variables on the absolute short-term learning gains, the response variable “immediate learning gains” was first created by subtracting participants’ total score in the pretest from their total score in the immediate posttest. The absolute short-term mean gains for each caption condition under both the lenient and strict scoring systems are shown in Table 35.

**Table 35. Absolute short-term gains based on both scoring systems (for Study 2)**

Condition	Strict	Lenient
	Mean (SD)	
Enhanced Captions X 1	3.78 (3.91)	5.12 (3.77)
Normal Captions X 1	2.03 (1.70)	2.30 (1.72)
Uncaptioned X 1	1.08 (1.67)	0.67 (1.40)
Enhanced Captions X 2	5.80 (4.33)	6.08 (4.74)
Normal Captions X 2	3.42 (3.03)	3.91 (3.01)
Uncaptioned X 2	1.63 (0.13)	0.21 (2.66)

The fixed effects included in the analysis were caption condition, number of viewings, VST score and the participants’ form recall pretest total score, as well as the interactions between these terms. The models were fitted and run using the *lm* function according to the procedure detailed under Section 3.3. The results based on the lenient scoring data revealed no interactions between the variables. Further, number of viewings ( $p = .07$ ), pretest score ( $p = .85$ ) and VST score ( $p = .22$ ) were not significant predictors of short-term gains. Caption condition ( $p < .001$ ), however, was found to be significant and it explained 24% of the variance ( $R^2 = .24$ ,  $F(2, 123) = 19.41$ ,  $p < .001$ ). As shown in Table 36, the estimated short-term gains (for participants with the mean VST score) under the normal captions and enhanced captions

conditions were higher than the uncaptioned conditions, by 2.16 and 4.35 points respectively. Multiple comparisons using the *emmeans* function (Lenth, 2018) was also carried out with the Bonferroni method *p*-value adjustment. The results revealed that compared to the uncaptioned conditions, short-term gains were significantly higher under the normal captions ( $p < .01$ ) and enhanced captions condition ( $p < .001$ ). Further, short term gains under the enhanced captions were also significantly higher than the normal captions conditions ( $p < .01$ ).

**Table 36. Output of the best-fit model for predicting short-term gains based on the lenient scoring system (for Study 2)**

[model: lenient.imme.gains ~ caption condition]

Parameter	Estimate	SE	T	p
Intercept*	0.55	0.53	1.04	.30
Caption condition (Normal captions)	2.16	0.69	3.11	< .01
Caption condition (Enhanced captions)	4.35	0.70	6.21	< .001

\*Intercept level: Uncaptioned

Analysis of the data based on the strict scoring system revealed similar results. Caption condition was the only predictor found significant, and it explained 30% of the variance ( $R^2 = .30$ ,  $F(2, 119) = 25.44$ ,  $p < .001$ ). Table 37 shows the estimated short-term gains (for participants with the mean VST score). Multiple comparisons with Bonferonni correction applied revealed that gains under the enhanced captions were significantly higher compared to the normal captions ( $p < .01$ ) and uncaptioned conditions ( $p < .0001$ ). Additionally, gains under the normal captions were also significantly higher than the uncaptioned conditions ( $p < .001$ ). The other three variables, i.e., number of viewings ( $p = .09$ ), pretest score ( $p = .17$ ) and VST score ( $p = .15$ ), were not found to be significant. In short, under both scoring systems, short-term gains were only influenced by caption condition.

**Table 37. Output of the best-fit model for predicting short-term gains based on the strict scoring system (for Study 2)**

[model: strict.imme.gains ~ caption condition]

Parameter	Estimate	SE	t	p
Intercept*	0.41	0.55	0.75	.46

Caption condition (Normal captions)	2.70	0.73	3.71	< .001
Caption condition (Enhanced captions)	5.28	0.74	7.12	< .001

\*Intercept level: Uncaptioned

#### 4.3.2 Analysis of learning gains between the form recall pretest and delayed posttest (i.e. long-term gains)

Participants' long-term gains were obtained by subtracting their total score in the pretest from their total score in the delayed posttest. Table 38 shows the long-term gains for each condition under both scoring systems. Similar to the short-term gains analysis, the models were fitted using the *lm* function according to the procedure outlined under Section 3.3. The fixed effects also included caption condition, number of viewings, VST score, form recall pretest score and their interactions.

**Table 38. Absolute long-term gains based on both scoring systems (for Study 2)**

Condition	Strict	Lenient
	Mean (SD)	
Enhanced Captions X 1	0.84 (1.70)	0.75 (1.80)
Normal Captions X 1	0.61 (1.52)	1.17 (0.33)
Uncaptioned X 1	1.27 (1.16)	0.86 (0.74)
Enhanced Captions X 2	2.56 (3.24)	2.76 (3.56)
Normal Captions X 2	0.68 (1.61)	0.96 (1.52)
Uncaptioned X 2	0.49 (1.43)	0.05 (2.34)

The results of long-term gains analysis based on the lenient scoring system revealed similar findings to the short-term gains analysis: of the three variables, only caption condition was found to be a significant predictor, but it explained only 6% of the variance ( $R^2 = .06$ ,  $F(2, 123) = 4.11$ ,  $p = .02$ ). The effects of number of viewings ( $p = .13$ ), VST score ( $p = .07$ ) and pretest score ( $p = .58$ ) were not found to be significant. Table 39 shows the predicted long-term gains for participants with the mean VST score. Multiple comparisons with the Bonferroni correction applied revealed that the difference between the uncaptioned conditions and the

normal captions was not significant ( $p = 1.00$ ). Similarly, while gains under the uncaptioned condition were lower compared to the enhanced captions conditions, the difference failed to reach significance ( $p = .12$ ). Long-term gains under the enhanced captions, however, were significantly higher compared to the normal captions conditions ( $p = .03$ ).

**Table 39. Output of the best-fit model for predicting long-term gains based on the lenient scoring system (for Study 2)**

[model: lenient.delayed.gains ~ caption condition]

Parameter	Estimate	SE	t	p
Intercept*	0.83	0.35	2.38	.02
Caption condition (Normal captions)	-0.19	0.46	-0.41	.68
Caption condition (Enhanced captions)	0.97	0.46	-2.09	.04

\*Note. Intercept level: Uncaptioned

Similar results were found in the analysis of the long-term gains based on the strict scoring system. Only caption condition emerged as a significant predictor, and it explained 6% of the variance ( $R^2 = .06$ ,  $F(2, 123) = 5.21$ ,  $p < .01$ ). The other three variables, i.e., number of viewings ( $p = .07$ ), VST score ( $p = .11$ ) and pretest score ( $p = .06$ ) were not significant. Table 40 shows the output of the best-fit model. Multiple comparisons, with Bonferroni correction applied revealed that the difference between the normal captions and uncaptioned conditions was not significant ( $p = 1.00$ ). It was further revealed that long-term gains under the enhanced captions conditions were significantly higher compared to the normal captions ( $p = .03$ ) and the uncaptioned conditions ( $p = .01$ ).

**Table 40. Output of the best-fit model for predicting long-term gains based on the strict scoring system (for Study 2)**

[model: strict.delayed.gains ~ caption condition]

Parameter	Estimate	SE	t	p
Intercept*	0.41	0.38	1.08	.28
Caption condition (Normal captions)	0.23	0.50	0.45	.65
Caption condition (Enhanced captions)	1.45	0.51	2.87	< .01

\*Note. Intercept level: Uncaptioned

In summary, under the intentional learning condition, form recall was only influenced by caption condition. Number of viewings, pretest score and VST score did not emerge as significant predictors of form recall. The analyses so far have only shed light on the effects of the variables on the learning gains. The next section details the results of the item-level analysis. As the outcome of the gains analysis was similar under both scoring systems, the item-level analysis only focused on the data based on the lenient scoring system.

#### **4.3.3 Form recall item-level analysis**

As previously mentioned in Section 3.3, the objectives of item-level analysis were two-fold. Firstly, item-level analysis investigated whether the trends observed for the gains analyses still persisted at the item-level. In other words, item-level analysis could assess whether the variables affected the participants' performance on the target items the same way they did on participants' performance as a whole. Next, item-level analysis also provided insights into how the properties of the target items influenced participants' scores on the target items.

To this end, the data for the pretest, immediate posttest and delayed posttest were analysed using the Cumulative Link Mixed Model (*clmm*) function in the *ordinal* package (Christensen, 2019). Two different models were developed. The first set of models included the core independent variables only, while the next set of models included both the core variables as well as the item-related variables, which are called covariates. These models will be discussed in turn.

##### *Modelling data with the core independent variables*

The first set of models only included the core independent variables that were of central importance, namely caption condition, number of viewings and VST score, as well as test time. Also included in the most complex model were the interactions between the variables. The random effects included Subject and Item. The models were developed and compared according to the procedure outlined under Section 3.3.

The results revealed a significant three-way interaction between caption condition, number of viewings and test time ( $\chi^2(4) = 15.3, p < .01$ ). This means that the effects of number of viewings depended on the caption condition, and that the nature of this relationship was different depending on the test time. Planned comparisons were run using the *emmeans* function in the *emmeans* package (Lenth, 2018), with Bonferroni corrections applied. The



results showed that under all caption conditions, there was no significant difference between viewing once and twice, for all the test times. As for the effects of caption condition, the enhanced captions condition had a significantly higher score than the uncaptioned condition but only for the immediate posttest ( $p < .05$ ). Further, this significant difference was found for both number of viewings ( $p < .05$ ). Although the enhanced caption condition also had a higher score than the normal caption condition in the immediate posttest, the difference did not reach significance. Similarly, the difference between the normal caption and uncaptioned condition was not significant for any of the form recall tests.

**Table 41. Output of the best-fit model (with core variables only) for the form recall item-level analysis (for Study 2)**

[model: ordered.score ~ caption.condition\*no.of.viewing\*test.time+ caption.condition\*VST.mean+ (1|Subject)+(1|Item)]

Parameter	Estimate	SE	z	p
Caption condition (Normal captions)	0.67	0.40	1.67	.10
Caption condition (Enhanced captions)	0.63	0.41	1.55	.10
Number of viewings	0.86	0.42	2.04	.05
Test time (Immediate)	0.72	0.25	2.89	< .01
Test time (Delayed)	0.77	0.25	3.10	< .001
Vocabulary size test (Centred)	0.06	0.01	4.78	< .0001
Caption condition (Normal captions) X VST Score	0.01	0.01	0.79	.43
Caption condition (Enhanced captions) X VST Score	-0.02	0.02	-1.55	.12
Caption condition (Normal captions) X Number of viewings (Twice)	-0.90	0.53	-1.68	.10
Caption condition (Enhanced captions) X Number of viewings (Twice)	-0.82	0.53	-1.51	.13
Caption condition (Normal captions) X Test time (Immediate)	0.13	0.30	0.41	.68
Caption condition (Enhanced captions) X Test time (Immediate)	0.69	0.31	2.26	.02

Caption condition (Normal captions) X Test time (Delayed)	-0.45	0.30	-1.46	.14
Caption condition (Enhanced captions) X Test time (Delayed)	-0.32	0.31	-1.02	.31
Test Time (Immediate) X Number of viewings (Twice)	-0.68	0.32	-2.11	.04
Test Time (Delayed) X Number of viewings (Twice)	-0.52	0.32	-1.64	.10
Caption condition (Normal captions) X Test time (Immediate) X Number of viewings (Twice)	1.11	0.40	2.74	< .01
Caption condition (Enhanced captions) X Test time (Immediate) X Number of viewings (Twice)	1.39	0.41	3.40	< .001
Caption condition (Normal captions) X Test time (Delayed) X Number of viewings (Twice)	0.45	0.41	1.10	.27
Caption condition (Enhanced captions) X Test time (Delayed) X Number of viewings (Twice)	1.17	0.41	2.86	< .01

\*Intercept level: Caption condition = Uncaptioned; Number of viewings = Once; Test time = Pretest

The results also revealed a significant interaction between caption condition and VST score ( $\chi^2(2) = 6.58, p = .04$ ). To compare the VST slopes for each caption condition, the *emrends* function was used. As shown in Table 42, the VST slopes for each caption condition were different, with the slopes for the normal captions (0.067) and the uncaptioned condition (0.055) steeper than the enhanced caption condition (0.032). This means that the effect of VST score was bigger for the participants in the normal captions and uncaptioned condition compared to the enhanced captions condition.

**Table 42. Estimate of the VST slope for each caption condition**

Test time	VST score trend	SE	Lower CL	Upper CL
Enhanced captions	0.032	0.01	0.012	0.051
Normal captions	0.067	0.009	0.049	0.086
Uncaptioned	0.055	0.012	0.033	0.078

*Modelling data with added covariates*

The effects of the properties of the target items were analysed by fitting and comparing models that included a number of covariates. The initial complex model included all the covariates, namely: (i) number of words in the string (length A) (ii) the number of words with missing letters (length B), (iii) the number of letters provided (i.e., the amount of hints provided) (letter A), (iv) the number of letters required to receive at least partial credit (letter B), (v) lemmatised phrasal frequency in the COCA (frequency), (vi) MI score, (vii) figurativeness, (viii) compositionality and (ix) ambiguity. The lemmatised phrasal frequency and MI score were log-transformed prior to statistical analysis (see Section 3.3). Figurativeness and compositionality were included in the model using the mean value. Ambiguity was a categorical factor. Target items that were deemed as unambiguous by more respondents were coded as 1, while those rated as ambiguous by more respondents were coded as 2. Two items, however, were rated by five respondents as ambiguous while the other five respondents rated these items as unambiguous. As such, these items were not coded as either 1 or 2, and no ambiguity value was attached them. Table 43 summarises the values of the added covariates.

Although the initial complex model included all the covariates, some had to be dropped as there was an issue with convergence, which may have arisen from overfitting, i.e., “having too many variables in the model that essentially make little contribution to predicting the outcome” (Field, Miles & Field, 2012, p. 266). In order for the model to converge, the covariates frequency, figurativeness and MI score had to be excluded. Apart from these three covariates, all the other covariates were included in the complex model, from which the core variables (i.e., caption condition, number of viewings, VST score, test time and the interactions) were removed incrementally, according to the procedure described under Section 3.3.

**Table 43. Summary of added covariates included in the item-level analysis (for Study 2)**

<b>Target MWES</b>	<b>COCA</b>	<b>MI Score</b>	<b>Figurativeness (mean rating)</b>	<b>Compositionality (mean rating)</b>	<b>Ambiguity</b>
whip up (something)	1349	7.61	5.0	3.3	2
unsung hero	236	15.32	5.9	4.8	1
make the most of (something)	1352	2.62	5.1	6.5	1
spread the word	754	3.21	4.7	5.8	
the dust has settled	44	7.61	5.0	4.5	2
screw with (someone/something)	156	5.25	5.3	3.3	2
take (something) lying down	14	4.75	5.6	3.6	2
step (something) up	7217	8.82	5.6	4.7	2
have a beef with (someone/something)	58	3.88	6.3	1.8	2
invest in (someone)	12154	5.78	4.3	6.3	1
have a run-in with (someone)	14	3.46	5.4	3.4	
in the bag	830	3.21	4.3	3.3	2
rat out (someone)	121	8.77	6.6	1.9	1
hopped up	210	8.82	4.3	3.5	2
(someone's) day in the sun	96	2.62	5.3	3.7	2
play the ( ) card	98	0.43	5.0	4.0	2
bring out the big guns	10	2.21	5.1	5.0	2
hit the streets	239	3.21	5.7	1.6	2
small potatoes	157	11.35	4.2	2.7	2
caught up in (something)	6656	8.82	4.6	5.4	2

The results revealed two significant interactions, both of which were also found significant in the previous analyses. The first was a three-way interaction between caption condition, number of viewings and test time ( $\chi^2(4) = 14.3, p < .01$ ), while the second was a two-way interaction between caption condition and VST score ( $\chi^2(2) = 6.35, p = .04$ ). After accounting for these interactions, none of the effects of the added covariates were not found to be significant, suggesting that they did not exert a strong influence on learners' scores on the target items. The best-fit model is hence identical to the earlier analysis that did not include the added covariates.

In summary, the analysis of the form recall data revealed that the gains between the pretest and immediate and delayed posttest were only affected by caption condition. The effect of caption condition was also observed at the item-level as the enhanced captions led to significantly higher scores on the target items (in the immediate posttest) compared to the uncaptioned conditions. Additionally, the covariates that were included in the item-level analysis were not found to influence form recall.

#### 4.4 Form recognition – Results

Table 44 shows the descriptive statistics for the delayed form recognition test. To assess whether the variables in question affected participants' score in the test, the data were analysed using the *glmer* (generalised linear mixed model) function in the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015). The *glmer* function used as the dependent variable was a binary outcome. Two different sets of models were developed and compared. The first set of models only included the core variables, while the second set included both the core variables and item-related variables as added covariates.

**Table 44. Descriptive statistics of the form recognition test (for Study 2)**

Conditions	Mean	SD
Enhanced Captions X 1	9.05	3.47
Normal Captions X 1	10.01	4.61
Uncaptioned X 1	9.47	3.25
Enhanced Captions X 2	10.40	4.08
Normal Captions X 2	9.83	4.05
Uncaptioned X 2	8.50	3.17

*\*Maximum possible total score is 20*

### Modelling data with core variables

The most complex model included caption condition, number of viewings, VST score (centred around the mean value), pretest score and their interactions as fixed effects. The random effects included Subject and Item. The models were developed and compared according to the procedure outlined under Section 3.3.

The results revealed that of the four variables, only pretest score was significant ( $\chi^2(1) = 66.6, p < .001$ ). As shown by the output in Table 45, an increase in the pretest score led to a significant increase in the odds of obtaining a correct answer in the form recognition test. The effects of number of viewings ( $\chi^2(1) = 0.57, p = .45$ ), caption condition ( $\chi^2(2) = 2.66, p = .26$ ) and VST score ( $\chi^2(1) = 2.57, p = .11$ ) were not significant, suggesting that these factors did not influence MWE form recognition.

**Table 45. Output of the best-fit model (with core variables only) for the form recognition test (for Study 2)**

[model: score ~ pretest.score + (1|Subject) + (1|Item)]

Parameter	Estimate	SE	z	p	OR
Intercept	- 0.96	0.26	-3.69	< .001	
Pretest score	0.31	0.03	11.72	< .001	1.36

### Modelling data with covariates

The next set of models were developed to include the core variables of interest as well as the added covariates. The following covariates were included: (i) lemmatised phrasal frequency, (ii) MI score, (iii) figurativeness, (iv) compositionality and (v) ambiguity (see Section 3.3). The objective of building these models was to assess whether the covariates had a significant effect on participants' score in the form recognition test, after accounting for the effects of the core variables. As such, the models were compared in a similar way as the previous analysis, in that each core variable was incrementally removed. In each model, however, the five added covariates were included. The random effects included Subject and Item. The models were fitted and compared using the *glmer* function according to the procedure outlined under Section 3.3. When a model produced a message warning of failed convergence, the 'bobyqa' optimiser was used to produce a stable model.

Similar to the results of the model that only included the core variables, only pretest score was found to be significant ( $\chi^2(1) = 66.6, p < .001$ ). The best-fit model is hence identical to the earlier analysis that did not include the added covariates. The effects of VST score ( $\chi^2(1) = 3.56, p = .06$ ), caption condition ( $\chi^2(2) = 2.66, p = .26$ ) and number of viewings ( $\chi^2(1) = 0.57, p = .44$ ) were not significant. Further, none of the covariates were found to be significant. In summary, participants' recognition of MWE form was found to be affected by participants' pretest scores only.

#### 4.5 Meaning recall - Results

Table 46 shows the descriptive statistics for the delayed meaning recall test. The nature of the data was ordinal, as there were only five possible scores (i.e., 0, 0.25, 0.5, 0.75 and 1) awarded for each response. As such, the data were analysed using the *clmm* function in the *ordinal* package (Christensen, 2019). The model development procedure started off with the core variables only, followed by the inclusion of the added covariates of secondary interest.

**Table 46. Descriptive statistics for the meaning recall test (for Study 2)**

Conditions	Mean	SD
Enhanced Captions X 1	2.73	2.57
Normal Captions X 1	4.68	3.88
Uncaptioned X 1	4.37	2.91
Enhanced Captions X 2	3.44	2.69
Normal Captions X 2	4.29	2.75
Uncaptioned X 2	3.34	2.59

\*Maximum possible total score is 20

#### *Modelling data with core variables*

The fixed effects included caption condition, number of viewings, VST score, pretest score and their interactions. Subject and Item made up the random effects. The models were fitted and run according to the procedure outlined in Section 3.3. The results revealed a significant effect of pretest score ( $\chi^2(1) = 18.9, p < .001$ ) and VST score ( $\chi^2(1) = 11.1, p < .001$ ). There was also a significant interaction between these two factors ( $\chi^2(1) = 6.33, p =$

.01). No other terms and interactions were found to be significant. As shown by the output of the best-fit model in Table 47, an increase in the pretest score led to a significant increase in the probability of receiving a higher score (e.g., 0.25 to 0.5) in the meaning recall test. The same applied to the effects of VST score, as an increase in one unit of the VST score also led to higher odds of receiving a higher score on a target item in the meaning recall test.

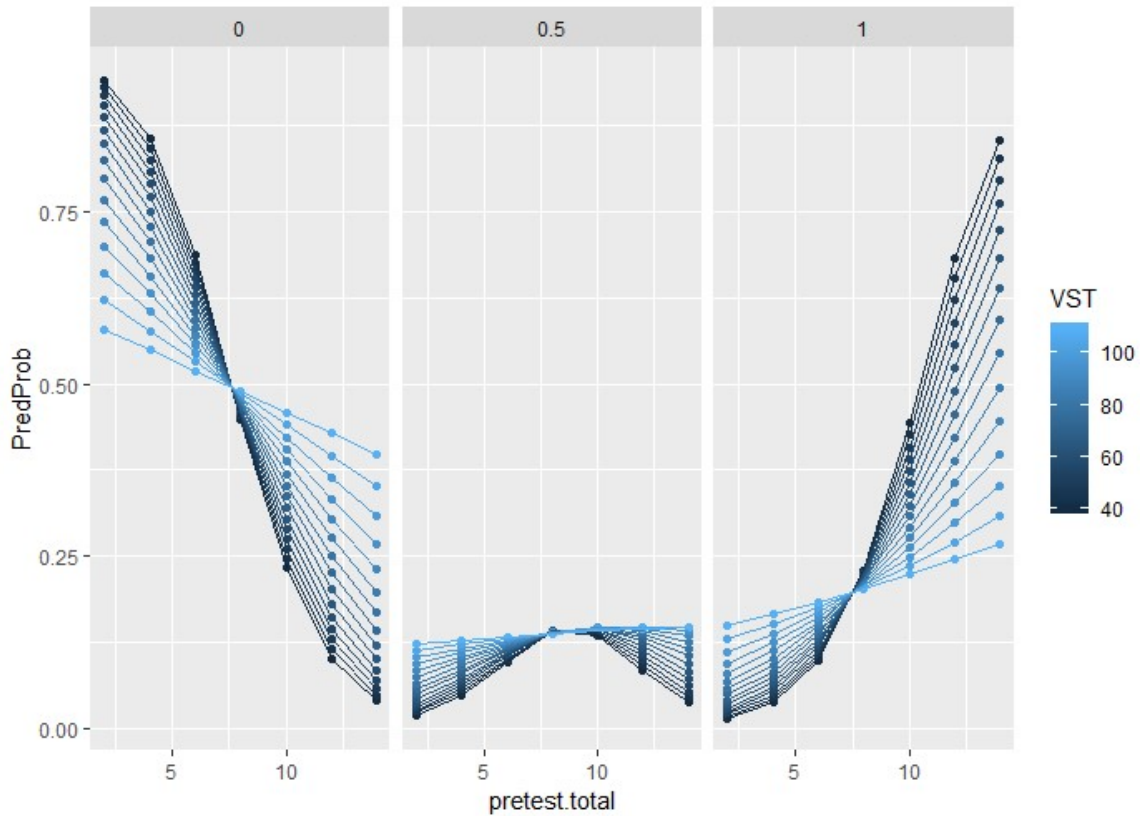
**Table 47. Output of the best fit model (with core variables only) for the meaning recall test (for Study 2)**

[model: ordered.score ~ pretest.score\*VSTmean+(1|Item)+(1|Subject)]

Parameter	Estimate	SE	z	p	OR
Pretest score	0.31	0.06	5.54	< .001	1.36
VST score	0.04	0.01	4.50	< .001	1.04
Pretest score X VST score	-0.006	0.002	-3.02	< .01	0.99

To illustrate the interaction between the pretest score and the VST score, the predicted probabilities of receiving each score (i.e., 0, 0.25, 0.5, 0.75 and 1) were generated and plotted. As there were fewer instances in which the target items were awarded 0.25 and 0.75 compared to the other three scores (i.e., 0, 0.5 and 0.75), the predicted probabilities of receiving these two scores were omitted from the plot. The plot of the predicted probabilities (Plot 9) reflects the output of the best-fit model: as the pretest score and VST score increased, so did the predicted probabilities of receiving a full score on an average target item in the meaning recall test. For participants with a lower VST score, however, an increase in the pretest score resulted in a more dramatic increase in the probability of getting a full score on the target items. This is different from those with higher VST scores, as an increase in pretest score did not affect their predicted probabilities of receiving a full score as strongly as it did for their counterparts with lower VST scores. This suggests that for the participants with a lower VST score, it is their pretest score that predicted their meaning recall. Although VST score is seen to have an opposite effect after a pretest score of above 7.5, this is just an artefact of the prediction—only five participants received a pretest score above 7.5. This suggests that the predicted pattern prior to a pretest score of 7.5 is a more accurate description of the positive effect of both VST score and pretest score on meaning recall.





**Plot 9. The effects of pretest score and VST score on the predicted probabilities of getting a score of 0, 0.5 and 1 on a target item in the meaning recall test (Study 2)**

*Modelling data with added covariates*

The analysis of the effects of the core variables was followed by the analysis that included the added covariates. The added covariates were: (i) frequency, (ii) MI score, (iii) figurativeness, (iv) compositionality and (v) ambiguity (see Section 3.3). They were added to assess if they had an effect on participants' scores on the meaning recall test, after accounting for the effects of the core independent variables. The random effects included Subject and Item. Each core term was removed in a stepwise procedure. In each model, the five covariates were included.

Similar to the analysis that only included the core variables, the results revealed a significant interaction between pretest score and VST score ( $\chi^2(1) = 8.16, p < .01$ ). None of the covariates i.e., frequency ( $\chi^2(1) = .65, p = .42$ ), MI score ( $\chi^2(1) = .53, p = .47$ ), figurativeness ( $\chi^2(1) = 1.72, p = .19$ ), compositionality ( $\chi^2(1) = .61, p = .43$ ) and ambiguity

( $\chi^2(1) = .92, p = .33$ ) were found significant. The best-fit model is hence identical to the earlier analysis that did not include the added covariates.

To summarise, participants' MWE meaning recall was found to be affected by their vocabulary knowledge, i.e., their pretest score and VST score. Further, none of the item-level covariates were found to influence participants' score in the delayed meaning recall test.

#### 4.6 Meaning recognition - Results

Table 48 shows the descriptive statistics for the meaning recognition test. The data were analysed using the *glmer* (generalised linear mixed model) function in the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015). Similar to all the other tests, two different sets of models were developed and compared. The first set of models included the core variables only, while the second set included both the core variables as well as the item-related variables as added covariates.

**Table 48. Descriptive statistics for the meaning recognition test (for Study 2)**

Conditions	Mean	SD
Enhanced Captions X 1	11.60	2.72
Normal Captions X 1	13.71	4.21
Uncaptioned X 1	12.60	3.62
Enhanced Captions X 2	11.68	4.37
Normal Captions X 2	13.22	3.73
Uncaptioned X 2	12.37	4.39

*\*Maximum possible total score is 20*

#### *Modelling data with core variables*

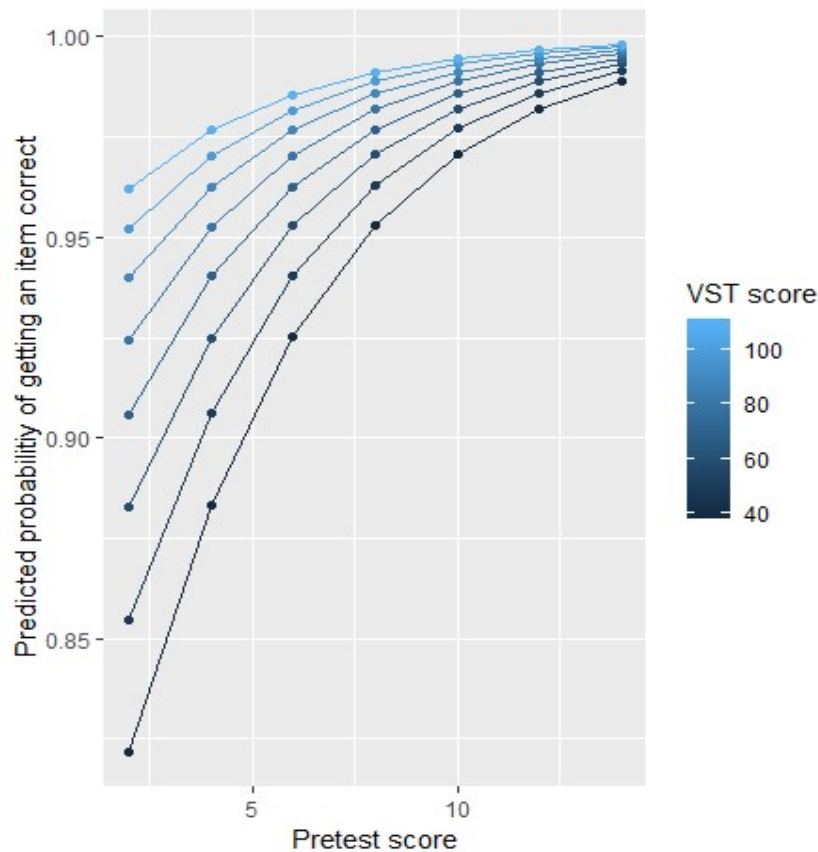
The most complex model included caption condition, number of viewings, VST score, pretest score and their interactions as the fixed effects. The random effects included Subject and Item. The models were fitted and compared according to the procedure outlined under Section 3.3. The results revealed that only VST score ( $\chi^2(1) = 18.2, p < .001$ ) and pretest score ( $\chi^2(1) = 45.4, p < .001$ ) were significant. As shown by the output of the best-fit model in Table 49, an increase in both the pretest score and the VST score led to a statistically higher likelihood

of getting an item correct in the meaning recognition test. This trend can also be seen in Plot 9 below, which shows the predicted probability of getting an item correct for all VST score ranges and pretest total scores. Both caption condition ( $\chi^2(2) = 3.88, p = .14$ ) and number of viewings ( $\chi^2(1) = 0.02, p = .88$ ) were not significant, as were the interactions between the predictors.

**Table 49. The output of the best-fit model (with core variables only) for the meaning recognition test (for Study 2)**

[model: Score~pretest.total+VSTmean+(1|Subject)+(1|Item)]

Parameter	Estimate	SE	z	p	OR
Intercept	0.07	0.23	0.29	.78	
VST score	0.02	0.01	4.42	< .001	1.02
Pretest total	0.25	0.04	6.88	< .001	1.28



**Plot 10. Predicted probability of getting and item correct in the meaning recognition test (Study 2)**

*Modelling data with added covariates*

To investigate whether the properties of the target items affected participants' score in the meaning recognition test, models that included the added covariates were fitted and compared. The added covariates were: (i) frequency, (ii) MI score, (iii) figurativeness, (iv) compositionality and (v) ambiguity (see Section 3.3). The model comparisons were done in a similar way to the previous analyses involving added covariates. That is, each of the core variables was removed from the complex model in a stepwise procedure. In each model, the five covariates were attached. Random effects included Subject and Item.

The results revealed that only the effects of pretest score ( $\chi^2(1) = 47.9, p < .001$ ) and VST score ( $\chi^2(1) = 18.6, p < .001$ ) were significant. No other core variables and interactions between them were found significant. Of the five added covariates, however, three were found to have an effect after accounting for the effects of pretest score and VST score. The three covariates were corpus frequency ( $\chi^2(1) = 7.11, p = .01$ ), MI score ( $\chi^2(1) = 12.1, p < .001$ ), compositionality ( $\chi^2(1) = 8.48, p < .01$ ). Firstly, the odds of a correct response were lower if the item's frequency increased by one unit, as shown in Table 50. Secondly, the odds of a correct response were higher if an item's MI score increased by one unit. Further, the odds of a correct response increased as the item's compositionality rating increased. Figurativeness ( $\chi^2(1) = 0.03, p = .87$ ) and ambiguity ( $\chi^2(1) = 0.06, p = .80$ ) were not found to be significant.

**Table 50. The output of the best-fit model (with added covariates) for the meaning recognition test (for Study 2)**

[model: Score~pretest.total+VSTmean+freq+MIscore+com+(1|Subject)+(1|Item)]

Parameter	Estimate	SE	z	p	OR
Intercept	-1.55	0.71	-2.17	.03	
Pretest score	0.25	0.04	6.88	< .001	1.28
VST score	0.02	0.01	4.42	< .001	1.02
Frequency (log 10)	-0.58	0.21	-2.67	< .01	0.56
MI score (log 10)	1.57	0.68	2.29	.02	4.81
Compositionality	0.45	0.13	3.49	< .001	1.57

In summary, MWE meaning recognition was influenced by learners' pretest score and VST score. Further, after accounting for the effects of the learners' vocabulary knowledge, the

compositionality, frequency and MI score of the target items were also found to affect participants' responses in the meaning recognition test.

#### 4.7 Comprehension test - Results

Table 51 shows the descriptive statistics for the comprehension test. As the dependent variable was binary, the comprehension test data were analysed using the *glmer* function in the *lme4* package (Bates, Maechler, Bolker & Walker, 2015) with the *family = binomial* option specified. The fixed effects included caption condition, number of viewings, VST score and the interactions between them. Random effects included Subject and Item. The models were developed and compared according to the procedure outlined under Section 3.3.

**Table 51. Descriptive statistics for the comprehension test (for Video 2)**

Conditions	Mean	<i>SD</i>
Enhanced Captions X 1	13.47	2.65
Normal Captions X 1	15.30	2.55
Uncaptioned X 1	12.47	3.02
Enhanced Captions X 2	13.72	4.08
Normal Captions X 2	15.96	2.77
Uncaptioned X 2	11.05	3.39

*Note.* The maximum possible score was 22

The results (Table 52) revealed a significant interaction between caption condition and VST score ( $\chi^2(2) = 12.5, p < .01$ ). This means that the effects of caption condition on participants' comprehension test score depended on their VST score. No other terms were found to be significant.

**Table 52. The output of the best-fit model for the comprehension test (for Video 2)**

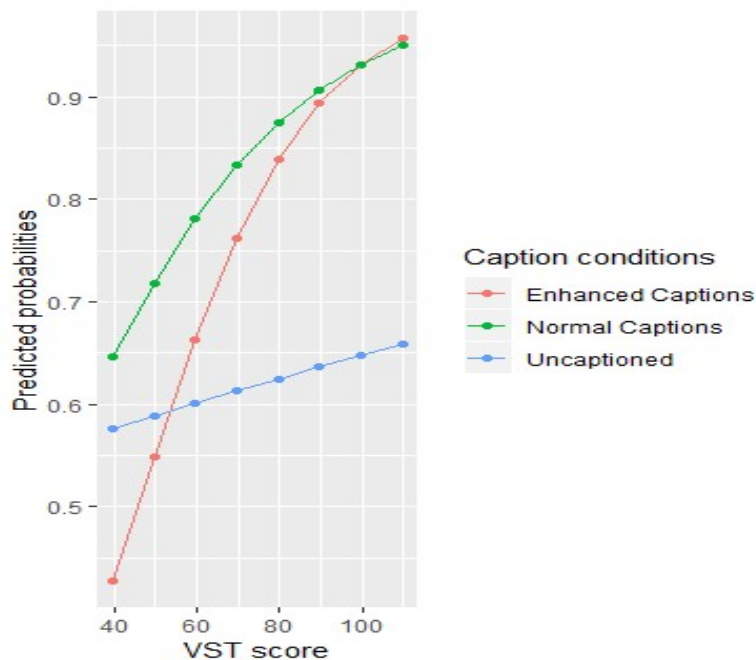
[model: comp.v2 ~ caption.condition\*VST.mean+(1|Subject)+(1|Item)]

Parameter	Estimate	<i>SE</i>	<i>Z</i>	<i>p</i>
Intercept	0.46	0.27	1.72	.09
Caption condition (Normal captions)	1.15	0.19	6.13	< .001

Caption condition (Enhanced captions)	0.70	0.19	3.71	< .001
VST score (Centred)	0.01	0.01	0.56	.58
Caption condition (Normal captions)	0.03	0.01	2.30	.02
X VST Score				
Caption condition (Enhanced captions)	0.04	0.01	3.41	< .001
X VST Score				

\*Intercept level: Caption condition = Uncaptioned

The *emtrends* function in R (emmeans package) was used to compare the slope of the VST score between the caption groups, applying the Bonferroni correction. A significant difference in the slope of VST was found between the enhanced captions and the uncaptioned condition ( $p < .10$ ), but not between the normal captions and the enhanced caption condition ( $p = .68$ ) or between the normal captions and uncaptioned condition ( $p = .06$ ). Plot 11 below shows the predicted probabilities of getting a comprehension item correct for all the VST mean score range and the caption conditions. The plot suggests that the normal caption conditions had the highest probabilities of getting a item correct, at least for participants in the VST mean range between 39.69 to 89.96 (mean VST score = 69.69).



**Plot 11. Predicted probability of getting an item correct in the comprehension test (Study 2)**

#### 4.8 Summary of findings

Study 2 was carried out with a view to investigating whether caption condition had an effect on intentional MWE learning. Another aim of this study was to assess whether there was an effect of repetition, and to investigate to what extent it influenced MWE learning under all three caption conditions. To answer these research questions, one pretest, four immediate posttests and one delayed posttest that tapped into different aspects of MWE knowledge were administered. The data from these tests were then analysed by a number of statistical models in R. The core independent variables of primary interest included caption condition, number of viewings, VST score and pretest score. In addition, the data on the comprehension test were also analysed to assess if comprehension was influenced by any of the core independent variables. Several observations could be made about the findings.

The first observation that stood out was the fact that the number of viewings was not found to be a significant predictor of any of the aspects of MWE knowledge. Further, number of viewings was also not found to predict learners' comprehension. This suggests that repetition is not a very influential factor under intentional learning conditions. As for caption condition, its effect was only found to be significant for form recall knowledge. Specifically, the short-term gains were significantly higher in the enhanced and normal captions compared to the uncaptioned condition. Further, both types of gains were also significantly higher in the enhanced captions compared to the normal captions condition. The superiority of the enhanced captions condition was also shown through the item-level analysis, which revealed that the likelihood of a better score on the target items in the immediate posttest was significantly higher under the enhanced caption condition compared to the uncaptioned condition. The item-level analysis also revealed that the effect of caption condition was influenced by learners' VST score, in that it had a bigger effect for participants in the normal captions condition, followed by the uncaptioned and enhanced caption conditions. Where learning gains are concerned, however, VST score did not emerge as a significant predictor, suggesting that form-recall gains were comparatively similar for learners of all vocabulary sizes.

Although caption condition was found to significantly influence form recall, it did not emerge as a significant predictor of form recognition. In fact, the only variable found to influence form recognition was pretest score, as the bigger the learners' pretest score, the higher their odds of obtaining a correct answer in the form recognition test. Put differently, knowledge

of MWE form at the level of form recall was mainly affected by caption condition, while at the level of form recognition, only prior knowledge mattered.

With regard to MWE meaning, of the four core variables, only the variables related to learners' vocabulary knowledge appeared to have a significant effect. This is because for both meaning recall and meaning recognition, only pretest score and VST score emerged as significant predictors. Specifically, in the meaning recall analyses, it was shown that the likelihood of learners receiving a higher score increased as their VST score and pretest score increased. The same was found for the meaning recognition test. Learners' performance was positively influenced by their vocabulary knowledge, as greater prior knowledge and vocabulary size corresponded to a higher likelihood of obtaining a correct answer in the meaning recognition test. Apart from participants' vocabulary size, factors relating to the properties of the target phrases were found to influence participants' knowledge of MWE meaning. At the recognition level, for instance, after accounting for the effects of learners' vocabulary knowledge, learners' likelihood of obtaining a correct answer in the meaning recognition test was still influenced by three item-related variables, namely frequency, MI score and compositionality.

Overall, under the intentional learning condition, knowledge of MWE form and meaning was mostly influenced by learners' vocabulary knowledge. It is only at the level of form recall that caption condition played a significant role. The roles of both caption condition and vocabulary size score were also important for comprehension, suggesting that the presence of captions was beneficial for both MWE form uptake and content comprehension. In sum, the findings of this study are markedly different from that of Study 1. The next chapter discusses the differences and similarities of the two studies in detail, and posits reasons for these different trends.



## **CHAPTER FIVE: DISCUSSION AND IMPLICATIONS**

Two studies, differentiated at the level of learning conditions, i.e., incidental and intentional learning, were conducted to answer three research questions. The first research question sought to investigate whether caption condition had an effect on incidental and intentional uptake of MWEs through audio-visual input, i.e., L2 viewing. Under each study, the effects of L2 viewing under three caption conditions, i.e., enhanced captions, normal captions and uncaptioned, were investigated. The second research question sought to find out whether repetition, operationalised as repeated viewing, had an effect on the uptake of MWEs under both learning conditions. Finally, the third research question set out to answer whether repeated viewing influenced incidental and intentional MWE uptake to the same degree under all three caption conditions. This chapter provides a detailed discussion of the research findings of both studies, which were presented in Chapter Three and Chapter Four. An overview of the findings of both studies is presented in Table 53.

As both studies sought the answers to the same research questions, the findings are discussed together. The discussion is divided into sections according to the multiple aspects of MWE knowledge. Section 5.1 compares the findings of the MWE form recall tests of both studies. Next, Section 5.2 looks at results of the MWE form recognition tests. Sections 5.3 and 5.4 discuss the findings of the MWE meaning recall and meaning recognition tests, respectively. Section 5.5 details the findings of the comprehension tests of the two studies. Section 5.6 summarises the key findings of both studies. Section 5.7 looks at implications for teaching. Finally, Section 5.8 discusses the limitations of the studies and includes suggestions for future research.

**Table 53. Overview of findings: Significant predictors and interactions found for each MWE test under both studies**

MWE test (Type of analysis)	Study 1 (Incidental learning)		Study 2 (Intentional learning)	
	Core variable	Item-related variable	Core variable	Item-related variable
Form recall (short-term gains)	Caption condition No. of viewings VST score		Caption condition	
Form recall (long-term gains)	Caption condition VST score		Caption condition	
Form recall (item-level analysis)	Caption condition X no. of viewings X test time  VST score X test time	None	Caption condition X no. of viewings X test time  Caption condition X VST score	None
Form recognition	Pretest score VST score No. of viewings Caption condition	None	Pretest score	None

Meaning recall	Pretest score X VST score	Compositionality	Pretest score X VST score	None
Meaning recognition	VST score Pretest score	Frequency MI score	VST score Pretest score	Frequency MI score Compositionality

## **5.1 Form recall under the incidental and intentional learning conditions**

### ***5.1.1 Factors that predict form recall under the incidental learning condition***

Knowledge of MWE at the level of form recall was influenced by different factors depending on the learning condition. The first significant predictor of form recall under the incidental learning condition was caption condition. As shown through the gains analyses, both caption types led to significantly larger gains compared to the uncaptioned condition. Under the lenient scoring system, for example, the predicted short-term gains for participants with the mean VST score under the normal caption conditions were 2.95 and 5.12 for viewing once and twice, respectively. In contrast, participants with the mean VST score under the uncaptioned condition who viewed the video once did not make any gains (i.e., -0.33), while those who watched it twice had predicted short-term gains of 1.84. The effects of normal captions, however, was not durable as the difference between the long-term gains between the normal captions and uncaptioned conditions was not significant. In contrast, the enhanced captions conditions proved to be superior compared to the uncaptioned condition as the former led to significantly higher short-term and long-term gains compared to the latter. Further, the item-level analysis revealed that compared to the uncaptioned condition, the enhanced captions had a significantly higher likelihood of receiving a better score on the items in the immediate posttest. This finding is encouraging, as it shows the direct effect of typographic enhancement on immediate form recall. While the short- and long-term gains also provided evidence of the positive effect of typographically-enhanced captions, it could be argued that these gains may have been confounded by the differing test formats of the pretest and immediate and delayed posttests (see Limitations, Section 5.8).

The superior performance of the captioned groups in the form recall tests lends support to previous viewing studies that found evidence in favour of captioned viewing over uncaptioned viewing for single-word acquisition (e.g., Huang & Eskey, 1999; Markham, 1999; Neuman & Koskinen, 1992; Sydorenko, 2010; see Montero Perez et al., 2013 for meta-analysis). Additionally, the results of the form recall tests also align with the theory of multimedia learning (Mayer, 2009) and Paivio's (1986, 2006) dual coding theory, which propose that learning is more effective with both words and pictures compared to when words or pictures alone are present. Based on the Noticing Hypothesis (1994), which posits that more noticing leads to more learning, it was anticipated the enhanced captions condition would also outperform the normal captions condition. Previous reading studies also provided grounds to

suspect that typographically enhanced MWEs would attract more attention and in turn promote the recall of MWEs under incidental learning conditions (e.g., Choi, 2018; Sonbul & Schmitt, 2013; Szudarski & Carter, 2016; Toomer & Elgort, 2019). This prediction, however, was not borne out in the form recall tests. For instance, while the gains under the enhanced captions were numerically higher compared to the normal captions, this difference fell just below the significance level. There are three possible reasons that could explain why the results of the present study are in contrast with the previous reading studies. Firstly, unlike taking in written input, viewing a video does not provide much chance for learners to fixate on novel lexical items for long, including the typographically enhanced ones. When reading, for example, learners could go back and forth between the enhanced and unenhanced items. This opportunity is not afforded when learners view a video in real time, especially in the case of the present study, where pausing was not an option. Secondly, it is also possible that noticing alone is not enough. As suggested by Schmidt (1995, 2001), noticing occurs at a low level of awareness. Therefore, although typographic enhancement may have heightened the learners' attention to the target items, the learners may not be aware of what exactly they were paying attention to. Consequently, the enhanced items may not be internalised and may leave weaker memory traces. Additionally, compared to written input, audio-visual input offers additional modes that may be more interesting than words alone, such as moving images. As such, learners would then need to spread their attention between the different modes. In short, the present study suggests that where knowledge of MWE form recall is concerned, the effects of watching a video with typographically enhanced captions and normal captions may be similar. Nevertheless, it is encouraging to observe that viewing a captioned video once does indeed result in MWEs being acquired incidentally. It is important to note, however, that the learning gains depended not only on the caption condition, but also on learners' vocabulary size.

Indeed, vocabulary size was found to predict both the short-term and long-term gains. This suggests that L2 viewing does not benefit all learners equally. A recent viewing study by Puimège and Peters (2019) also found that while incidental learning of MWEs at the level of form recall does occur through watching an uncaptioned video, it is also predicted by learners' vocabulary size. Similar to their study, the present study found that the bigger the learners' vocabulary size, the bigger their learning gains in form recall knowledge. As emphasized above, the predicted short-term gains (under the lenient scoring system) showed that for learners who viewed the uncaptioned video once, those with a VST score smaller than the mean value (i.e., 69.69) recorded no incidental short-term gains. In addition, the effect of vocabulary

size applied to all caption conditions. This finding thus lends support to the Matthew effect, which posits that learners with a larger vocabulary size tend to learn more new words compared to those with a smaller vocabulary size (Penno, Wilkinson, & Moore, 2002; Stanovich, 1986). Moreover, this finding corroborates previous viewing studies on single words (e.g., Peters et al., 2016), as well as reading studies that have found that vocabulary size influences incidental learning of MWEs at the level of form recall (e.g., Peters, 2016).

Apart from caption condition and vocabulary size, form recall under the incidental learning condition was also predicted by repeated viewing or repetition. Under each respective caption condition, learners who watched the video twice had larger predicted short-term gains compared to those who watched the video once. The positive effect of repetition was also evident at the item-level, as viewing twice led to a higher likelihood of receiving a higher score on the target items in the immediate posttest. It has to be pointed out, however, that while repetition under all caption conditions led to higher scores on the target items, only the enhanced caption condition reached statistical significance (i.e.,  $EC2 > EC1$ ). Webb et al.'s (2013) reading-while-listening study also found that number of encounters had a positive effect on form recall. However, it is important to highlight that in their study, the number of repetitions ranged from one encounter to 15 encounters. It should also be pointed out that Webb et al. (2013) did not carry out a form recall pretest, so the basis of comparison was the mean scores on the immediate posttest, rather than the gains in score from form recall pretest to posttest. Further, repetition was operationalised differently in their study. In their study, repetition was operationalised as multiple encounters with the target collocation within the same text.

The way repetition was operationalised in Durrant and Schmitt's (2010) reading study is more similar to the present study. This is because in one of their three conditions—the verbatim repetition—the learners were exposed to the target collocations twice, embedded in the same sentence each time. This condition was found to lead to the best recall of the collocations, compared to the varied repetition and single repetition conditions. Thus, the present study lends support to their findings, in that repeated viewing leads to better short-term MWE form recall compared to single viewing. The positive effects of repetition on MWE form recall can also be explained using two theories, namely the usage-based approach and the instance-based approach. According to these theories, for language acquisition to happen, type and token frequency is of vital importance (Wulff, 2018). Additionally, each encounter with a target item creates memory traces within a contextualized episode and these memory traces

may then help resonate or reactivate knowledge for the subsequent encounters (Bolger et al., 2008).

Contrary to the result of the present study, Pellicer-Sánchez's (2017) reading study did not find an effect of repetition on incidental MWE form recall, as the differences between four and eight exposures fell short of significance. Analogous to Webb et al.'s (2013) study, repetition in Pellicer-Sánchez's study was also operationalised as repeated encounters within the same text. The lack of significant effect of repetition on form recall in Pellicer-Sánchez's study might be attributed to a few reasons. Apart from the obvious difference in terms of input modality, the nature of the target items in her study is different from the present study. Familiar words, such as the ones that make up the target items in the present study, attract little attention (Godfroid et al., 2013), compared to completely novel words such as the ones in Pellicer-Sánchez's study. As such, it stands to reason that the pseudowords in her study might have drawn the learners' attention, rendering the effect of repetition less strong. Another plausible reason for the lack of significant effect for repetition in Pellicer-Sánchez's results might be the timing of the test. In Webb et al. (2013), Durrant and Schmitt (2010) and the present study, form recall was tested immediately after treatment. Pellicer-Sánchez, however, tested form recall one week after treatment. This may have further weakened the effect of repetition. A similar finding was observed in the present study: the effect of repetition was not found to be durable, as it did not predict long-term gains. This suggests that a difference of one encounter may not have been sufficient to create a long-lasting effect on participants' form recall. Engendering MWE form recall knowledge that is resistant to attrition may require more than two viewings, but this then requires a greater investment of class time unless shorter video materials were used.

Although repetition benefitted participants under all caption conditions, the predicted gains revealed patterns similar to the ones observed for those who viewed the video once, as viewing twice under both captioned conditions (i.e., EC2 and NC2) led to significantly higher short-term gains compared to the uncaptioned condition (i.e., UC2). The benefit of captioning is emphasized even more when comparing the short-term gains made between viewing once and viewing twice. This is because the predicted short-term gains (under the lenient scoring system) for participants who watched the video *once* with normal and typographically-enhanced captions were larger compared to the participants who watched the video *twice* without captions. To illustrate, the predicted short-term gains for participants with a VST mean score under EC1 and NC1 were higher than UC2 by about 2.62 and 1.11 points respectively.

Having said that, only the difference between EC1 and UC2 was found to be significant. The difference between NC1 and UC2 failed to reach significance level. This suggests that rather than investing extra time for repeated uncaptioned viewing, learners might be better off (at least with regard to MWE uptake) viewing a video once with the presence typographically-enhanced captions, as the learning gains are bigger in terms of MWE knowledge at the level of form recall. This aligns well with the suggestion made by Szudarski and Carter (2016) that “input enhancement might reduce the number of encounters needed to make L2 collocational gains” (p. 261). It has to be acknowledged, however, that the recommendation by Szudarski and Carter (2016) was made in relation to the comparison between enhanced and unenhanced lexical items. In their reading study, repetition (which they termed as input flood) coupled with typographic enhancement resulted in significantly better form recall compared to repetition alone. In the present study, no significant difference was found between viewing a video twice with typographically enhanced captions (i.e., EC2) and with normal captions (i.e., NC2). Similar trends were found for groups that watched the video once only (i.e., no significant difference was found between EC1 and NC1). Therefore, the same plausible reasons for the lack of significant difference may also apply in this case. To reiterate, the modality of input as well as the presence of noticing at a low level of awareness may have rendered the effects of typographical enhancements weaker.

After accounting for the effects of repetition, caption condition and learners’ vocabulary size, analysis which also included item-related variables was carried out to assess whether they affected form recall. The analysis revealed that none of the added covariates predicted form recall. It has to be conceded, however, that the number of items was small (see Limitations, Section 5.8) and that, for some of the predictors, the range was too small. The inclusion of length as a covariate based on the number of content words with missing letters, for example, only ranged from one to three words. As such, this small difference may not have affected form recall as much as the core variables.

To sum up, incidental MWE acquisition at the level of form recall was predicted by caption condition, number of viewings and vocabulary size. These factors influenced short-term gains, as well as learners’ likelihood of receiving a higher score in the immediate posttest. Of these three factors, however, the effect of repetition was not durable as it did not significantly influence long-term gains or participants’ likelihood of getting a higher score in the target items in the delayed posttest. In contrast, the effects of caption condition and vocabulary size were durable as they influenced long-term gains in MWE form recall.



### ***5.1.2 Factors that predict form recall under the intentional learning condition***

The analysis of MWE form recall under the intentional learning conditions revealed trends that are different from those of the incidental learning conditions. When learners were forewarned of forthcoming tests, only caption condition predicted both their short-term and long-term gains. Similar to findings under the incidental learning condition, both the normal and enhanced captions conditions significantly outperformed the uncaptioned conditions. Additionally, at the item-level, the enhanced captions condition was significantly more likely to get a higher score compared to the uncaptioned condition in the immediate posttest. Contrary to the findings under the incidental learning condition, the enhanced caption conditions led to *significantly* higher gains compared to the normal captions conditions. This adds to the body of evidence that typographic enhancement is effective for MWE acquisition under intentional learning conditions. This finding resembles Peters (2012), in which typographic enhancement was employed as a within-subject factor, so the participants saw half of the single words and MWEs typographically-enhanced, while the other half appeared unenhanced. Prior to reading the text, one group was instructed to pay attention to new vocabulary in general, while the other was specifically informed to pay attention to the MWEs. The immediate posttest revealed that the participants recalled the form of the typographically-enhanced words and MWEs more than the unenhanced ones. Interestingly, however, in the immediate posttest, the group that was specifically instructed to focus more on the MWEs performed as well as their counterparts who did not receive the specific instruction. Peters (2012) therefore argued that “typographical salience may be a successful method to effectively focus students’ attention on target items because it is more specific and more guiding compared to a general instruction” (p. 76). The findings of the present study concur with this notion because, despite receiving specific instruction about forthcoming tests on MWEs, the learning gains under the normal caption conditions were still significantly lower than the enhanced captions conditions.

The fact that the enhanced captions performed significantly better than the normal captions under intentional learning but not under the incidental learning may be explained using theories in the areas of attention and awareness. According to Tomlin and Villa (1994), detection alone, without awareness, can lead to acquisition. They further argue that chances of detection may be increased with the use of input enhancement. This could be why the use of typographic enhancement in the first study led to higher gains in form recall compared to the normal captions, but the difference was only marginally significant. As speculated in the earlier section, although noticing leads to more learning, noticing occurs at a low level of awareness

(Schmidt, 1995, 2001). This is also known as noticing at the level of attention (Godfroid et al., 2010). There is also noticing at the level of awareness, where learners may realise what it is that they are paying attention to (Godfroid et al., 2010). This is reflected in the present study because in Study 2, the learners knew the items they would be tested on, and how they would be tested owing to their experience in Study 1. Put differently, the presence of awareness could be the reason why the short-term gains under the enhanced captions were significantly higher than the normal captions in Study 2, but not in Study 1. However, this has to remain speculation as inferences can only be made based on the pretest and posttest scores, and there is no behavioural data such as eye-tracking and stimulated recall .

Although the normal captions did not outperform the enhanced captions, the normal captions did hold an advantage over the uncaptioned conditions, as shown through the superior short-term gains under the former. This adds weight to the supposition that captioned viewing leads to more learning compared to uncaptioned viewing. As posited by Vanderplank (2016), the presence of captions can help learners segment the speech stream they are listening to and thus discern the spoken words. This may explain why the normal caption condition had significantly higher short-term form recall scores compared to the uncaptioned condition. The orthographic representation of the words might have helped the learners pick up the words and fill in some of the blanks in the form recall test. The success rate of the normal caption condition in recalling the MWE forms, however, was lower than the enhanced caption condition possibly due to the lack of demarcation of the MWEs. Despite being aware of the tests, in the absence of typographic enhancement, the learners under the normal captions conditions may have been unaware that the single words were constituents of MWEs. In contrast, their counterparts in the enhanced caption condition had the advantage of seeing these MWE word strings owing to the visual salience brought about by typographic enhancement. It is also important, however, to note that the superiority of normal captions over uncaptioned viewing was only observed for the short-term gains. The difference in long-term gains between these conditions was marginal and did not reach statistical significance. This suggests that where the retention of MWE form is concerned, the effect of normal captions and uncaptioned viewing may be similar. In short, although pre-empting the possibility of an MWE test may result in learners paying closer attention to an input video, caption condition still matters as it influences learners' uptake of the lexical composition of the phrases.

Contrary to the results of the incidental learning condition, repetition under the intentional learning condition did not predict form recall. Both types of gains were similar for

those who watched the video once and their counterparts who watched the video twice. This also applied across the three caption conditions. This suggests that test anticipation reduces the need for repeated viewing. Further, the item-level analysis showed that under all caption conditions, there was no significant difference between viewing once and twice, for any of the form recall tests. Comparisons of these results with existing studies, however, could not be made owing to the lack of studies that look at how repetition affects MWE learning under intentional learning conditions. In one intentional learning study by Peters (2014), for example, a large and durable effect of repetition on verb-noun and adjective-noun collocations was found in both the unannounced form recall immediately and delayed posttest. However, unlike the present study, participants in Peters's study were exposed to the target items in differing frequencies (i.e., 1, 3, or 5) through focus-on-form activities. The effects of repetition therefore may have been more pronounced as compared to the present study. In summary, where MWE learning through watching audio-visual materials is concerned, the present study provides evidence that with test awareness, viewing a video once seems to be just as effective as viewing it twice for promoting MWE form knowledge at the level of form recall.

As discussed in the previous section, the effect of vocabulary size on MWE form recall under the incidental learning condition was strong, i.e., the larger the participants' vocabulary size, the larger their learning gains (both short-term and long-term). Under the intentional learning condition, however, vocabulary size was not found to be a significant predictor of learning gains. Analyses revealed that under all three caption conditions, learners with smaller vocabulary sizes made similar short-term and long-term gains in MWE form recall knowledge to their more proficient peers. This suggests that as far as learning gains were concerned, the vocabulary size and the Matthew effect in the second video seemed to matter much less compared to the first video. Further, at the item-level analysis, the effect of vocabulary size was found to depend on the caption condition. Vocabulary size was found to have a bigger effect on the normal caption and uncaptioned condition compared to the enhanced caption condition. This could be because for learners in the enhanced caption condition, it is the presence of typographic enhancement that influenced their likelihood of getting a higher score on the target items, leaving less room for the influence of their vocabulary size.

Similar to the findings of the incidental learning condition, none of the item-related variables included in the models emerged as a significant predictor of form recall under the intentional learning condition either. Again, it is possible that the small number of target items coupled with the relatively narrow range of some of the predictors may have prevented

potential trends from emerging. Additionally, this result has to be interpreted in light of the possibility of over-fitting. As mentioned in the results chapters, some potential predictors had to be dropped in order for the models to converge. For the form recall models under the intentional learning condition, for example, frequency, figurativeness and MI score had to be removed before the models could converge.

### ***5.1.3 Summary of the factors the predict form recall under both learning conditions***

In summary, when comparing the acquisition of MWEs at the level of form recall under the incidental learning and intentional learning conditions, several key differences are identified. Firstly, when learners are not forewarned of a test, repeated viewing is necessary to improve learning gains. Watching an input video with the aid of enhanced captions once, however, proves to be more beneficial than watching an uncaptioned video twice. In contrast, when there is an expectation of a test, viewing a video twice seems unnecessary since watching it once brings about comparable learning gains (at least as far as MWE form uptake is concerned).

The second difference relates to the effect of learners' vocabulary size. The present study suggests that when learners are not aware that they will be tested on novel MWEs, the more proficient learners gain more MWE knowledge of form compared to the less proficient learners. On the other hand, test anticipation seems to help lower-proficiency learners to overcome their disadvantage as their learning gains are similar to those of their peers with larger vocabulary sizes.

In terms of similarities, the most striking one is the role of caption condition in influencing MWE form recall. Previous reading studies provided grounds to suspect that typographic enhancement would boost form recall when learners are not expecting a test. It is conceivable, however, that had learners been aware of a test, the effect of caption conditions might have been diluted. The finding of the second study, however, revealed that test awareness is not sufficient, and that typographic enhancement is still needed to help learners notice MWEs. Without typographic enhancement, MWEs might not be focused on because a large proportion of MWEs are made up of comparatively frequent words (Martinez & Murphy, 2011; Nation, 2013). Further, as shown in the literature, even proficient learners find it difficult to identify MWEs they encounter in written input (Eyckmans, Boers, & Stengers, 2007). This could be why typographic enhancement proved to be superior than the normal captions condition (at least where learning gains are concerned) under the intentional learning

conditions, as it induces noticing of relevant lexical units, and may help learners take in the MWE boundaries and their precise lexical compositions. Further, as discussed above, noticing of the lexical units in Study 2 could have been at the level of awareness, which may have made the role of typographic enhancement even more prominent, compared to under the incidental condition. As suggested by the results of Study 1, without test announcement, typographically enhanced captions may lead to similar gains as normal captions due to noticing that is only at the level of attention. While the current findings suggest that incidental and intentional acquisition of MWE forms through audio-visual materials is possible provided that it is facilitated by the use of captions and repeated viewing, it is important to note that the gains were relatively small. To illustrate, under the lenient scoring system, the predicted incidental short-term gains for learners with the mean VST score (i.e., 69.69) under the ‘best’ condition (i.e., watching the sitcom episode twice with typographically-enhanced captions, or EC2), was only about 6.63 points. Besides the UC1 condition, which did not record any short-term learning gains, learners with the mean VST score under the remaining four conditions had predicted short-term gains ranging between 1.84 (UC2) to 5.12 (NC2) points. While the gains may be on the small side, it is typical for studies on incidental vocabulary learning to reveal small gains even after a relatively substantial amount of study time (Webb, 2019). For example, in a longitudinal viewing study, Rodgers (2013) found that only 6.4 out of the 28 unknown single words were learned after seven hours of L2 viewing. Nevertheless, these incidental gains, are “central to lexical development” (Webb, 2019, p. 232), even though they are relatively small.

Although learners were aware of the MWE tests, short-term gains under the intentional learning condition were not substantially higher than the incidental learning gains. In fact, under the enhanced caption conditions, the predicted short-term gains were only 4.9 points. Comparable gains were predicted regardless of the number of viewings and vocabulary size as these two variables were not found to be significant. These relatively small short-term gains suggest that substantial gains from L2 viewing may require more than the presence of typographically-enhanced captions, repeated viewing and test awareness. Nevertheless, these factors do influence MWE form knowledge, not only at the level of form recall, but also form recognition. The next section looks more closely at MWE form recognition under both learning conditions.

## **5.2 Form recognition under the incidental and intentional learning conditions**

### ***5.2.1 Factors that predict form recognition under the incidental learning condition***

As was the case for MWE form recall, MWE form recognition was influenced by different factors, depending on the learning conditions. Under the incidental learning condition, four factors influenced form recognition. The first factor was caption condition. Once again, the superiority of typographically-enhanced captions over the uncaptioned condition was demonstrated, as the uncaptioned condition showed lower odds of a correct answer in the form recognition test. Further, this difference was significant. The odds of a correct answer in the normal caption condition were also lower than in the enhanced caption condition. However, this difference did not reach statistical significance. Previous studies such as Sonbul and Schmitt (2013) and Toomer and Elgort (2019), in which typographic enhancement was employed as a within-subject factor, have reported the advantages of enhanced MWEs over unenhanced MWEs for form recognition in the context of incidental MWE acquisition through reading. It is then surprising that in the current study the enhanced caption conditions did not significantly outperform the normal caption conditions. This trend, however, is similar to the trend found for the form recall tests. Therefore, similar explanation may be offered for the lack of significant difference. It could be speculated that unlike reading texts, real-time viewing does not allow much time for learners to fixate on salient MWEs. This could be why the odds of a correct response under the normal caption condition were not significantly worse than the enhanced captions condition, as the time spent looking at the MWEs were the same under the two caption conditions. The results of Montero Perez et al.'s (2014) viewing study lends weight to this supposition, as similar trends in their form recognition results were found. In their study on the effects of captioning on incidental vocabulary learning and video comprehension, four conditions differing in the amount and saliency of captions were put to test. One group watched the input videos without any captions, while the other three groups watched them with full captions, keyword captions and full captions with highlighted keywords, respectively. While the results revealed that the captioned groups significantly outperformed the control group in the form recognition test, no significant differences between the full captions condition and the full captions with highlighted keyword condition were found. This is similar to the findings of the form recognition test in the present study, in which the difference between the normal captions and uncaptioned condition fell below the significance level. A few differences between their study and the present study have to be pointed out, however. Firstly, although MWEs were included in Montero Perez et al.'s study, the target items were predominantly

single words. Secondly, form recognition was tested immediately after treatment. Additionally, the format of the form recognition test in Montero Perez et al. is different from the present study, in that their participants were given a selection of words and were then required to tick the words that they thought had appeared in the input clips. Nevertheless, Montero Perez et al. and the present study suggest that as far as form recognition is concerned, the effect of typographic enhancement in the context of L2 viewing may be similar to normal captions. However, it has to be emphasized that typographic enhancement still holds an advantage over the uncaptioned condition, which suggests that long-term retention of form recognition may be better when viewing an L2 video with the presence of typographically-enhanced captions, compared to viewing with an absolute absence of captions.

The second predictor found to positively influence MWE form recognition under incidental learning was repetition. Specifically, the findings from the form recognition test showed that learners who viewed the input video a second time had higher odds of obtaining the correct answer in the form recognition test compared to those who only viewed the video once. This is consistent with Webb et al.'s (2013) reading-while-listening study, which found that gains in knowledge of collocational form increased as the number of encounters increased. It has to be noted, however, that in their study, the form recognition test was carried out immediately after treatment. The durability of the effect of repetition is therefore unknown. The findings of the present study thus show that where incidental MWE acquisition through audio-visual input is concerned, the effect of repetition on form recognition is durable. This is different from the findings of Pellicer-Sánchez's (2017) reading study, in which no significant difference was found in the delayed form recognition posttest (i.e., choosing the right collocate among five options) between the participants who had eight encounters and the target collocation and their peers who had four encounters. Although this is contradictory to the findings of the present study, it could be that a difference of one encounter may lead to a more pronounced difference in the ability to correctly recognise MWE forms. In other words, while the learners in Pellicer-Sánchez's study (2017) had more chances (i.e., as many as four encounters) to familiarise themselves with the novel words, the learners under the viewing once conditions in the present study only had one encounter with novel MWEs. As such, the familiarity with the novel word strings may not have sunk in yet, and a second encounter may have given the learners a chance to close the gap in their knowledge since the attention given to the novel MWEs may not have diminished as much from the first encounter. As shown by previous eye-tracking studies such as those by Pellicer-Sánchez (2016), Mohamed (2018) and

Elgort et al. (2018), attention to new lexical items diminishes after five or six encounters as familiarity sets in. This could explain why form recognition did not get significantly better in eight encounters compared to four encounters as in Pellicer-Sánchez (2017). As mentioned earlier, however, it is difficult to make a valid comparison between the present study and the earlier studies, owing to the many differences such as modality of input and the operationalisation of repetition. Nevertheless, the findings of the current study provide more evidence in support of the role of repetition for MWE form recognition. Further, the current study provides evidence that repetition in the context of audio-visual input, under the incidental learning condition, predicts delayed MWE form recognition. In addition, a difference in *one* viewing resulted in a higher likelihood of MWE forms being recognised correctly for all caption conditions.

Apart from caption condition and repetition, MWE form recognition under the incidental learning condition was also influenced by learners' vocabulary size. The odds of a correct response in the recognition test were higher for participants with larger vocabulary sizes. As mentioned in previous sections, it is well established in the literature that learners with a larger vocabulary size stand to gain more from incidental viewing (e.g., Feng & Webb, 2019; Montero Perez et al. 2014; Peters et al., 2016; Peters & Webb, 2018). Further, vocabulary size is highly correlated with listening success (Stæhr, 2009). This may explain why vocabulary size was a significant predictor for both form recall and form recognition. This is in line with two other viewing studies on single word acquisition. In Montero Perez et al. (2014) discussed above, vocabulary size positively correlated with the immediate recognition of the form of the target items. As mentioned, however, they tested form recognition by asking participants to indicate whether they thought the target items appeared in the input videos. The same test format was used to assess immediate form recognition in Peters et al. (2016), which also found that learners' vocabulary size correlated positively with learners' ability to recognise more target items. In short, the findings of the current study provide further evidence that, even in the context of audio-visual input, the effects of vocabulary size extend beyond single word recognition, as learners with a larger vocabulary size are better at recognising the correct MWE forms compared to their lower-proficiency peers.

The other learner-related factor that predicted MWE form recognition was pre-existing knowledge. In many of the previous studies mentioned above, an indication of learners' prior vocabulary knowledge was inferred from their scores in a vocabulary test. In Peters et al. (2016), for example, learners' scores in a frequency-based multiple-choice meaning



recognition test represented their prior vocabulary knowledge. The same test was used in Peters and Webb's (2018) study to account for prior vocabulary knowledge. Although such vocabulary tests provide us with an indication of learners' mastery of the words according to the frequency bands, they do not give us direct evidence of how much learners know about the *actual* target items. Therefore, in the present study, including learners' form recall pretest score is thought to provide a clearer picture of the prior knowledge of the target MWEs. This is because form recognition was not pre-tested. In other words, learners' form recall pretest knowledge was taken as a proxy of their pre-existing form recognition knowledge. This was thought to be reasonable as receptive knowledge usually precedes productive knowledge (Schmitt, 2010). Further, multicollinearity between the form-recall pretest score and participants' VST score was not found in any of the analysis, lending further support to the proposition that both factors explained the variance. The findings of the current study suggest that the higher the learners' scores in the form recall test, the higher their chances of recognising the MWE forms correctly in the delayed posttest. This is in line with the findings of previous viewing studies that found evidence of the influence of prior vocabulary knowledge, inferred from learners' vocabulary test scores, on the recognition of novel words (e.g., Peters et al., 2016; Peters & Webb, 2018).

Although the present study found an effect of prior knowledge on form recognition, this was not the case for Toomer and Elgort's (2019) conceptual replication of Sonbul and Schmitt's research (2013). Similar to the present study, the latter found that as learners' proficiency increased, the learners produced more correct answers in the form recognition test. In Toomer and Elgort, however, neither vocabulary size nor prior knowledge predicted learners' form recognition. Why these two factors emerged as significant predictors of form recognition in the present study but not in Toomer and Elgort (2019) can only be speculation. Besides the obvious difference such as modality of input, there are a couple of other possible reasons for this incongruent finding. One factor might be that the target collocations in Toomer and Elgort are shorter in length compared to the target items in the incidental phase of the present study, which ranged from two to five words. Another, as acknowledged by the authors, is that their target items are fairly transparent in meaning. In the present study, the sentential context in the form recognition test provided clues to the figurative meaning of the MWEs, so this might have invited more context comprehension. Context comprehension relates to vocabulary size (Liu & Nation, 1985). Therefore, this could be why vocabulary size emerged as a significant predictor of form recognition in this study as the MWEs needed to be matched

with the figurative meanings. Another obvious difference between the present study and Toomer and Elgort is that in the latter, the target collocations appeared nine times. While these reasons may have given rise to the incongruent findings, they must remain speculative. What is important here is that the present finding adds to the growing body of evidence in the context of L2 viewing that vocabulary knowledge, be it vocabulary size or prior knowledge of actual MWE items, positively influences MWE form recognition.

In regard to item-related factors, none of them predicted MWE form recognition under the incidental learning condition. This is similar to what was found in the form recall item-level analysis, where none of the item-related predictors emerged as significant predictors (see Section 5.1.1). As productive knowledge is more difficult to enhance compared to receptive knowledge, it is somewhat expected that the effects of the item-related variables would also be insignificant for form recognition, as they did not influence form recall either. However, it has to be conceded that the test format may have influenced the learners' performance. Specifically, the options provided and the strength of the distracters may have rendered the participants less sensitive to the item-related variables. Additionally, the small sample size might be another reason why the item-variables did not emerge as significant predictors of form recognition.

In summary, in the present study, incidental MWE learning at the level of form recognition was predicted by learner-related variables such as vocabulary size and pre-existing knowledge, as well as the variables under investigation, i.e., caption condition and repetition.

### ***5.2.2 Factors that predict form recognition under the intentional learning condition***

In contrast to the findings under the incidental learning condition, MWE form recognition under the intentional learning condition was predicted by one learner-related factor only, which was pre-existing knowledge. Learners who did better in the form recall pre-test had a better likelihood of obtaining a correct answer in the form recognition test. This suggests that even with awareness of forthcoming vocabulary tests, learners' ability to recognise the MWEs was influenced by their pre-existing knowledge. It should be noted, however, that this is somewhat expected, because if the learners knew an item at pre-test, then it is highly likely that they would still know it at the delayed posttest. In other words, getting an item right in the posttest does not necessarily mean the learners remember it from the input video.

Unlike the incidental learning condition, vocabulary size did not predict form recognition under the intentional learning condition. This suggests that learners of all vocabulary sizes had a similar probability of recognising MWE forms. This finding, however,

is not surprising, considering that vocabulary size did not predict either short-term or long-term gains in form recall knowledge. Put differently, given that receptive aspects of vocabulary are easier to master than productive aspects (Nation, 2001), if learners of lower-proficiency levels were able to perform as well as their more proficient peers in the form recall, then the same would be expected in the form recognition test. Further, a similar finding was shown in Peters's (2016) intentional learning study, where vocabulary size positively correlated with learners' scores in the form recall test, but was not found to influence form recognition. It should be noted, however, that although the learners in her study took unannounced posttests, their treatment involved vocabulary activities meant to draw learners' attention to collocations. In other words, Peters and the present study show that under intentional learning conditions, characterised by explicit focus-on-form activities or the presence of test announcement, MWE form recognition appears not to be influenced by vocabulary size.

Similar to the lack of the effect of vocabulary size, repetition's lack of effect on form recognition was also expected, since it did not predict form recall either. Once again, this suggests that when learners are aware of impending vocabulary tests, repetition seems to make less of a difference as viewing once or twice leads to similar gains in both MWE form recall and form recognition. While there are previous intentional learning studies that did find an effect of frequency of occurrence on word learning, these studies incorporated vocabulary-focused activities in their design (e.g., Folse, 2006; Laufer & Rozovski-Roitblat, 2011). This may then render the effect of repetition stronger than in the present study.

As caption condition was found to be a significant predictor of form recall, it is rather unexpected that form recognition under the same learning condition, i.e., intentional learning, was not influenced by caption condition. In Montero Perez et al.'s (2015) viewing study mentioned earlier, the results were similar to the present study, in that visual salience had an effect on form recognition under the incidental learning condition. This study should be mentioned again here, as they found that caption type had an effect on both the intentional groups (i.e., learners who were forewarned of the vocabulary test) and the incidental groups, suggesting that visual salience has an effect on form recognition even when learners are aware of the test. Although this contradicts the finding of the present study, it has to be reiterated that visual salience in their study was created by having isolated keywords appearing as captions. Further, there was no condition in which the participants watched the input video without captions. Therefore, due to this difference and the others mentioned earlier (e.g., L2, target items, length of video and format of test), a direct comparison should not be made.

Interestingly, form recognition under the intentional learning condition was not predicted by learners' VST score. This is different from the results of the incidental learning condition, where form recognition was predicted by vocabulary size. It could be hypothesised that under the intentional learning conditions, the knowledge of impending tests coupled with the phonological representation of the MWEs may suffice in order for the MWE forms to be learned at a similar extent for participants of all vocabulary size. This is likely the case, given that the target MWEs consisted of words that were all likely to be familiar to the learners, as emphasized before. Further, vocabulary size was not found to predict MWE gains in form recall knowledge either. In other words, this suggests that when learners are aware of impending tests, their vocabulary size is unlikely to influence the uptake of the MWE forms at the level of recognition.

As was the case with form recognition under the incidental learning condition, no item-related variable was found to predict this aspect of knowledge under the intentional learning condition either. This finding suggests that form recognition, or at least the test format used to assess form recognition in the present study, renders the effects of the item-related variables as non-significant. In essence, with or without the awareness of MWE test, the present study found that learners' ability to recognise the correct MWE form was not significantly influenced by factors such as length, corpus frequency, MI score, figurativeness, literality and compositionality.

### ***5.2.3 Summary of the factors that predict form recognition under both learning conditions***

To summarise, MWE form recognition was influenced by factors that were not related to the properties of the target items. Specifically, under the incidental learning condition, bigger vocabulary size and pre-existing knowledge of the target MWEs led to a higher likelihood of recognising the MWE forms correctly. In addition, repeated viewing and the presence of captions, especially typographically-enhanced captions, resulted in better form recognition than an uncaptioned video. When learners anticipated the MWE tests, however, the only factor that predicted their form recognition was their pre-existing knowledge of the target items. The results of the form recognition tests, however, should be interpreted with caution. For one thing, the form recognition test was only administered two weeks after the treatment. For another, to avoid test effects, the learners were not pre-tested on this aspect of knowledge. This was why their form recall pre-test was used as a proxy for their pre-existing knowledge of the target items. A pretest and immediate posttest would have provided a clearer picture of the learning

gains. This is a limitation of the current study. The same applies to the meaning tests, which were only administered as delayed posttests. These are the tests that we will turn to next.

### **5.3 Meaning recall under the incidental and intentional learning conditions**

#### ***5.3.1 Factors that predict meaning recall under both learning conditions***

The findings of the meaning recall tests revealed a few key similarities for both learning conditions. Firstly, participants' scores in the test under both learning conditions were generally low. Under the incidental learning condition, for example, participants' mean total scores under viewing once ranged from 4.31 (EC1) to 6.18 (NC1), while the score under repeated viewing ranged from 3.91 (UC2) to 5.61 (EC2). Similarly, low scores were observed under the intentional learning conditions, as viewing once led to a total mean score ranging from 2.73 (EC1) to 4.68 (NC1), while the scores under viewing twice ranged between 3.34 (UC2) to 4.29 (NC2).

Participants' overall poor scores could be the reason why neither repetition nor caption condition emerged as significant predictors of meaning recall under either learning conditions. These two factors will be discussed in turn. Firstly, the fact that typographic enhancement did not result in superior meaning recall compared to the other two caption conditions is similar to Montero Perez et al.'s (2014) findings. In their study, a significant effect of types of captioning on meaning recall was not found either. This means that regardless of the conditions (i.e., full captions, full captions with highlighted keywords, or no captions), no difference was found in the meaning recall posttest. The authors offered a few possible explanations, which could also apply to the present study. One of the factors relates to the fact that captioning does not provide "concrete information on meaning" (p. 134). A similar finding was reported in Montero Perez et al.'s (2015) eye-tracking study. Indeed, typographic enhancement may have attracted attention to the target forms, but did not induce the kind of cognitive processing that resulted in the meaning of the words being learned. Szudarski and Carter's (2016) study reported similar findings: typographic enhancement resulted in the acquisition of form recall and recognition, but did not influence meaning recall. They hypothesised that enhanced MWEs might not have been processed "robustly enough to make semantic gains" (p. 259). As noted by Boers (2019), while typographic enhancement can "orient a reader's attention to language forms, it cannot ensure on its own that the learner will grasp the meaning of those forms" (p.148).

It is reasonable to assume that greater attention induced by typographic enhancement may result in learners putting more effort in interpreting the meaning. However, in the case of

real-time authentic input such as the one used in the present study, the target MWEs were encountered very briefly, and only once or twice. As posited by Hulstijn (1992), one encounter is unlikely to result in a strong form-meaning link. Further, inferring word meaning from context is a difficult and often unsuccessful process (Lin & Nation, 1985). This is bound to be the case for the target MWEs too, as the meanings of these expressions do not follow straightforwardly from adding up the meanings of the constituent words (e.g., Boers, Eyckmans, & Stengers, 2007). Another reason for the lack of an effect of the caption condition in the meaning tests in the current study may also be that the learners were exposed to the meanings of the phrases in the first delayed posttest (i.e., contextualised form recall test) and the second delayed posttest. This may have created a possible test effect, and diluted any differential effects of caption condition. In short, while the present finding suggests that typographic enhancement alone may not be enough to enhance MWE knowledge at the level of meaning recall, it has to be interpreted in light of the possible limitation pertaining to test ordering and test effects.

The fact that repetition did not emerge as a significant predictor may be explained along similar lines. Firstly, literature has established that learners are slower to acquire meanings compared to forms. In a study on single-words, for example, Pellicer-Sánchez (2016) demonstrated that meaning recall was the most difficult aspect of knowledge to acquire. Further, for both L1 and L2 participants, meaning recall was also the aspect with the lowest degree of certainty. Pellicer-Sánchez's more recent (2017) study on learning collocations incidentally from reading echoed this finding: meaning recall was the component with the lowest degree of certainty. Secondly, earlier reading studies on incidental single word acquisition demonstrated that different aspects of vocabulary require a different number of encounters (Pellicer-Sánchez & Schmitt, 2010; Webb, 2007). This is perhaps why a difference of one encounter failed to create differential effects in meaning recall. More encounters may be needed to enhance this aspect of MWE knowledge. In Webb et al.'s (2013) reading-while-listening study, for example, learners who experienced 15 encounters had significantly higher scores on the productive form and meaning test compared to those who had ten, five or one encounter. Similarly, learners who had ten encounters also had significantly higher scores compared to those who had five or one encounter. This suggests that a difference of five encounters may be needed to create a significant difference in meaning recall. Additionally, while the results of Webb et al.'s (2013) productive meaning recall test were based on the immediate posttest, the present study only administered the meaning recall test two weeks after

treatment. This may have diluted the effects of repetition even further. Szudarski and Carter's (2016) study provides a more comparable result with findings that support the present study. As is the case of the present study, they tested meaning recall two weeks after treatment. Their results revealed that when compared to a control group which only participated in the pretest and posttest, even a difference of six and 12 occurrences of collocations did not lead to significantly better meaning recall. Again, this underscores the fact that a greater number of encounters may be needed to enhance the knowledge of MWE meaning under both learning conditions. Alternatively, combining repetition with form-focused exercises or instruction may be more effective. As noted by Szudarski and Carter, promoting this aspect of MWE knowledge may require a "more explicit treatment" (p. 259). In Peters et al. (2016), however, frequency of an item's occurrence significantly predicted meaning recall. This contradicts the present finding. However, this may be partly due to the different way frequency of occurrence is operationalised in Peters et al.'s study. Unlike the present study, their study operationalised frequency of occurrence as the number of times the target words appeared in the video. Further, the frequency of the target words ranged from one to five, which may have then rendered the effect of frequency of occurrence stronger for meaning recall compared to the present study. This also means that the target items were encountered in different contexts, some of which may be more informative than others. While amount and contextual clues were not included in the analysis in the present study, some of the MWEs were indeed encountered in contexts that were not as informative, or contexts that require inferencing to get to the figurative meaning. An example of such MWE was "*rooting for you*". This MWE was embedded in the following context:

Eddie: You waiting for your report card?

Dave: Yeah, and a birthday card from my dad.

Dave (asking the postman): Is there anything else for me?

Postman: Buddy, we talked about this. I would lead with the card. I'm rooting for you.

As can be seen from the context, the learners would also have to know the meaning of "*lead with*" in order to get to the meaning of "*rooting for*". Further, the scene does not contain strong visual clues that hint at the figurative meaning. Therefore, re-encountering the same MWEs in the same context may be helpful for the development of form knowledge (as shown through the results of the form recall test for Study 1) as the context is already familiar on the second encounter, so more attention can be given to the MWEs. However, if the content is

uninformative as exhibited by the example above, then encountering it twice may make little difference in establishing form-meaning mappings. In short, in the context of repeated encounters through watching the same video, the present study suggests that a difference of one encounter does not lead to better MWE meaning recall, regardless of the learning conditions.

The two variables that did emerge as a significant predictor of meaning recall under both conditions were learners' pre-test score and their vocabulary size. As mentioned earlier, meaning recall was not pre-tested. As such, learners' scores in the form recall pre-test were taken as a proxy of their pre-existing knowledge. It may be argued that learners' knowledge of form does not necessarily entail knowledge of meaning. However, owing to the fact that the form recall pretest contained contextualised clues that point to the meanings of the MWEs, it provides a reason to suspect that correct answers in the pre-test were guided by the knowledge of both form and meaning. Therefore, including learners' form recall pretest score as a measure of their pre-existing knowledge of meaning in the present study was thought to be reasonable. The findings under both learning conditions revealed that the bigger their pre-existing knowledge, the more likely the participants would recall the MWE meanings correctly.

The effect of pre-existing knowledge on the meaning recall test, however, was influenced by learners' vocabulary size. While an increase in both variables led to a higher probability of an upwards shift in score (e.g., 0.25 to 0.50) on a target item in the meaning recall test, certain patterns were observed under both learning conditions. Firstly, for the more proficient learners (i.e., with higher VST scores), while an increase in pre-existing knowledge led to greater probabilities of recalling the meaning correctly, the rate at which the probabilities increased was not as dramatic compared to the less proficient learners (i.e., with lower VST scores). Put differently, for the less proficient learners, an increase in their pretest score corresponded to a more substantial increase in the predicted probabilities of getting an item correct as compared to their more proficient counterparts. In a recent reading study, Peters and Webb (2018) also found that for learners who had a greater knowledge of single words (as inferred through the frequency-based vocabulary test), the odds of a correct response on the meaning recall were higher compared to the peers who had less prior knowledge. Three other viewing studies lend support to this finding. In Montero Perez et al. (2014, 2015), discussed above, vocabulary size also had a significant and large effect on learners' meaning recall of the target words (which also included MWEs). Similarly, Peters et al.'s (2016) study on the effects of captioning and subtitles found that the higher the number of words known in the vocabulary



test, the higher the odds of recalling the meaning of the novel words. In the context of incidental MWE learning from audio-visual material, Puimège and Peters (2019), mentioned earlier, also found that vocabulary size significantly predicted MWE meaning recall. Unfortunately, the results of their meaning recall test were not further analysed as they were thought to have been contaminated by a test effect. In sum, the present study adds to the body of evidence for the positive role of vocabulary knowledge, i.e., prior knowledge and vocabulary size, in enhancing the knowledge of MWEs at the level of meaning recall under both incidental and intentional learning conditions. It should be noted, however, that correct answers in the test may reflect not only learners' ability to recall the meanings of the phrases from watching the input videos, but also their ability to guess the meanings from the constituent words. There is grounds to suspect this, since compositionality was also found to exert an influence, as explained in the next paragraph.

In regard to the effects of the item-related variables, compositionality was found to affect meaning recall under the incidental learning conditions, but not under the intentional learning conditions. It should first be reiterated that the target items in each study, (i.e., in each video) were different, with different amount and quality of contextual clues, which may have given rise to the different results. In other words, it is acknowledged that the item-related variables found significant may not only be due to the learning conditions, but also the different target items in each study. Results of the meaning recall analysis in Study 1 revealed that an increase in compositionality, resulted in a higher likelihood of an item receiving a higher score (e.g., from 0.5 to 0.75). This reflects the findings of Boers, Lindstromberg and Webb (2014). In their exploration of the effects of alliteration under the incidental learning condition, they found that while the L2 learners' gains at the level of meaning knowledge was influenced by the presence of alliteration, the most influential factor was the degree of semantic transparency of the target phrases. It should be noted, however, that participants' performance in the meaning recall test may be influenced by their ability to guess the figurative meaning of the MWEs based on the constituent words. Further, due to the small number of items as well as participants' poor scores in the meaning recall test, it is hard to generalise the effect of compositionality on MWE meaning recall.

### ***5.3.2 Summary of factors that predict meaning recall under both learning conditions***

In summary, meaning recall under both learning conditions was found to be affected by learner-related variables such as prior knowledge and VST score. Further, the present finding

also suggests that after accounting for the effects of the learner-related variables, meaning recall may be affected by compositionality. Although none of the treatment effects, i.e., caption condition and repetition, influenced meaning recall, it has to be conceded that productive knowledge is more difficult to acquire than receptive knowledge. Additionally, the fact that the meaning recall tests were administered two weeks after treatment may have weakened the treatment effects. Importantly, the findings suggest that acquisition of the meaning aspects of MWEs may require more than repeated viewings and the use of typographic enhancement. This is because the results of the meaning recognition test, which will be addressed next, reveal patterns that are largely similar.

#### **5.4 Meaning recognition under the incidental and intentional learning conditions**

##### ***5.4.1 Factors that predict meaning recognition under both learning conditions***

As was the case for MWE meaning recall, the variables that predicted MWE meaning recognition under both learning conditions were learners' vocabulary size and prior knowledge only. An increase in both led to higher odds of a correct response in the meaning recognition test. The finding that vocabulary knowledge positively influenced meaning recognition is in line with previous viewing studies. For example, in the previously mentioned study on the effects of caption types by Montero Perez et al. (2014), a large effect of vocabulary size was found for meaning recognition. Unlike the present study, however, their study also found a large effect size of caption type on meaning recognition. While they found no evidence that typographic salience (i.e., keyword captions and full captions with highlighted keywords) led to better meaning recognition than full captions, they did find that typographic salience promoted better meaning recognition when compared to no captions. It has to be pointed out, however, that in their study meaning recognition was tested immediately after treatment. Further, the target words were mostly single words. These two reasons could be why the effect of caption type on meaning recognition was more pronounced in their study compared to the present study.

Peters and Webb's (2018) viewing study had parallel findings on the role of vocabulary knowledge in enhancing the meaning recognition of single words. Similar to the present study, they found that the bigger the learners' prior vocabulary knowledge (inferred from the participants' results in a frequency-based vocabulary test), the better the odds of a correct response in the immediate meaning recognition test. Contrary to the present study, however, their study also found that an increase in frequency led to higher odds in recognising meaning

correctly. As mentioned earlier, the fact that repetition is operationalised differently in their study may be a reason for the contradictory finding. Additionally, the effects of repetition were only observed in the immediate posttest, as the delayed meaning recognition test was not further analysed. The effect of repetition in their study may then be stronger as the time lapse between the treatment and the meaning recognition test was only a week, whereas in the present study, meaning recognition was tested two weeks after treatment. Furthermore, the frequency of occurrences of the single words in their study ranged between one to six occurrences. Hence, a difference of one encounter, such as in the case of the present study, may have weakened the effects of repetition. The findings of Peters et al. (2016) are similar to Peters and Webb (2018), in that meaning recognition was found to be influenced by frequency of occurrence *and* vocabulary size test score, with an increase in both leading to a higher increase in the odds of a correct response. The findings can then be explained along similar lines: repetition may have exerted a stronger influence on the meaning recognition of the single words due to how it was operationalised and the fact that meaning recognition was tested immediately after treatment.

In the context of MWE acquisition through written input, there is a limited number of studies that include a meaning recognition test. In studies that do include one, the format is different from the present study. In Webb et al. (2013), for example, receptive knowledge of meaning and form was tested by requiring participants to write the L1 meanings of the L2 collocations. For this test, meaning recognition was found to be significantly better for participants who had 15 encounters with the target collocations, compared to their peers who had ten, five, one or zero encounters, respectively. Again, the fact that frequency of occurrence played a positive role in their study may be attributed to the fact that it ranged from zero to 15, as opposed to one or two encounters, as is the case in the present study.

In terms of item-related variables, two variables were found to be significant predictors of meaning recognition under both learning conditions. The first factor is mutual information (MI). Although there are studies that looked at the influence of MI score on L2 MWE learning, to the best of my knowledge, none has looked at its effect on meaning recognition. For instance, in Nguyen and Webb (2017), MI score was included as a predictor of the Vietnamese learners' knowledge of the collocation forms. Their study found that Vietnamese EFL students tend to know collocations with lower MI scores, compared to those with higher MI scores. This is reminiscent of Durrant and Schmitt's study (2009), in which non-native writers were found to use collocations with very high t-scores, i.e., very frequent collocations, and underuse collocations with high MI scores. In Puimège and Peters viewing study, in which MI score was

included as a predictor in the form recall test, a negative correlation between MI score and learning was also found. Items with high MI scores were not learned well. Both Nguyen and Webb (2017) and Puimège and Peters (2019) posited that the negative correlation found in their respective studies is likely due to the fact that phrases with high MI scores tend to be made up of low-frequency words. While there is substantial evidence of the effect of MI score on the learning or knowledge of the forms of MWEs, not much is known about its effect on the learning of the meanings of MWEs. In the present study, higher MI score was associated with higher odds of a correct answer in the meaning recognition test, and this was found for both learning conditions. The results, however, have to be interpreted cautiously. Firstly, it has to be conceded that the strength of the distracters may have influenced the results, in that for some items the correct answer may be more easily guessed than the others. Secondly, although the MWEs with a high MI score are made up of lower-frequency words, the meaning of these phrases as a whole could be more easily deduced compared to the MWEs that are made up of high-frequency words. For example, although the item “*unsung hero*” has a MI score of 15.32, it was rated by native speakers as 4.8 on the compositionality scale. In contrast, the MI score for “*hit the streets*” is 3.21, but was only rated as 1.6 on the compositionality scale. This means that the learners may be able to deduce the meaning of “*unsung hero*” more easily than “*hit the streets*”. Put differently, other factors have to be considered when determining whether MI score really influences the ease of MWE meaning inference. Due to the small number of the target items, the evidence for the effect of MI score on meaning recognition is not compelling. Further, while the relationship between MI score and meaning recognition may be positive in the present study, the results may have been influenced by a test effect, owing to learners’ exposure to the embedded MWE meanings in the form recall and form recognition delayed posttests. This applies to all other item-related variables that were found to affect meaning recognition, too.

The second item-related variable that affected meaning recognition under both learning conditions was corpus frequency. Studies on the effect of corpus frequency have shown that it correlates with learners’ knowledge of MWEs. For instance, in Schmitt and Redwood’s (2011) investigation of L2 learners’ productive and receptive knowledge of phrasal verbs, the participants were found to have better existing knowledge of higher-frequency collocations than lower-frequency ones. The effect of corpus frequency, however, is moderate. Similar results were found in González Fernández and Schmitt (2015), whereby corpus frequency was found to have an influence on learners’ existing knowledge of collocations. This effect was,

however, slight, which prompted the authors to suggest that it cannot be used as the major predictor of collocation learning. In these two studies, i.e., Schmitt and Redwood, and González Fernández and Schmitt, knowledge of MWEs was assessed using a form recall and form recognition test. Similarly, although Puimège and Peters (2019) found a small correlation between corpus frequency and MWE learning, they did not include a test on meaning recognition. Thus, not much is known about how well corpus frequency predicts learners' knowledge of MWEs at the level of meaning recognition. In the present study, the effect of corpus frequency on meaning recognition was significant under both learning conditions. The direction of the effect, however, was different. For the items under the incidental learning conditions, an increase in one unit of corpus frequency led to higher odds of obtaining a correct answer, whereas under the intentional learning condition, an increase in one unit of corpus frequency resulted in lower odds of getting an item correct in the meaning recognition test. It is unclear why the effect of corpus frequency was different under the two learning conditions. Aside from the obvious fact that the two sets of target items are different with a different range of corpus frequencies, all six conditions also performed rather similarly in the test under both learning conditions. To illustrate: the mean total score under the incidental learning conditions (i.e., Study 1) ranged from 12.96 to 14.95, while in the intentional learning conditions (i.e., Study 2) it was 11.60 to 13.71. These scores are vastly different from the meaning recall scores. Therefore, it is hard to strongly conclude that corpus frequency does indeed influence meaning recognition, as the results of the present study may have been influenced by the strength of the distractors in the multiple-choice test as well as test effect. As such, the effect of corpus frequency on meaning recall has to be interpreted in light of these possible reasons.

Compositionality was also found to predict meaning recognition, but only under the intentional learning condition. The inconsistent finding could be due to the fact that in the latter, the range in the compositionality ratings for the target MWEs was much wider than in the former. To illustrate, the mean ratings for the target items in the first video range between 2.6 to 6, while in the second video the mean ratings are between 1.6 to 6.5, suggesting that for the latter, the items vary more in terms of whether or not the meanings of the target items follow straightforwardly from the meanings of the constituent words. This may be the reason why compositionality emerged as a significant predictor of meaning recognition under the intentional learning condition only. As posited by Boers (2019), “compositionality is associated with transparency of meaning, and transparency of meaning is naturally one of the factors likely to influence the learnability of a given MWI” (p. 145). Indeed, according to the

result, an increase in compositionality led to a higher increase in the likelihood of recognising the MWE meanings correctly. This makes sense as a higher compositionality rating means that there is a clearer link between the meanings of the individual words and the figurative meaning of the phrase. The chances of the learners recognising the meanings of the phrases were then higher.

#### ***5.4.2 Summary of the factors that predict meaning recognition under both learning conditions***

In summary, MWE meaning recognition was found to be affected by learner-related variables such as vocabulary size score and prior knowledge. Additionally, after accounting for the effects of these variables, meaning recognition was still affected by various item-related variables, such as corpus frequency, MI score and compositionality. As noted under the meaning recall section, the lack of treatment effects on the acquisition of meaning in the present study may be attributed to the difficulty in engendering this aspect of knowledge, as well as test effects brought about by the order of the delayed posttests. While the treatment conditions were not found to influence the acquisition of MWE meaning, they did enhance content comprehension. This is the test that will be addressed next.

### **5.5 Factors that influence listening comprehension**

To ascertain that the learners engaged with the input video and that they were processing it for content (especially in the first study), a comprehension test was administered immediately after they viewed each video. It should be reiterated, however, that the learners received forewarning about the comprehension test before both videos. This means that the learners were most likely actively trying to comprehend the input videos. Having said that, it was expected that being forewarned of MWE tests would perhaps make learners focus more on potential MWEs, and make comprehension of the content of Video 2 matter less. Based on the learners' performance on the comprehension test, however, it seems safe to say that the learners engaged with the content of the Video 2 quite well, though not as well as with Video 1. To illustrate, for Video 1, the total mean score ranged between 16.67 to 18.55 (out of the maximum possible score of 22), whereas in Video 2, the total mean score ranged between 11.05 to 15.96 (out of the maximum possible score of 20). The small differences between the conditions, however, means that the predictors that emerged as significant may need to be interpreted cautiously.

Firstly, caption condition emerged as a significant predictor of comprehension in both videos. Previous studies have furnished evidence of the superiority of captioned video over uncaptioned video for comprehension (e.g., Gass et al., 2019; Winke et al., 2013; Montero Perez et al., 2013 for meta-analysis). The findings of the present study lend further support for this, as, for the first video, the odds of getting an item correct in the comprehension test for the normal caption conditions were higher than for the uncaptioned conditions. Further, this difference was statistically significant. Similarly, in the second video, participants under the normal captions condition had a higher probability of getting an item correct in the comprehension test, compared to those under the uncaptioned video.

As viewing with the presence of normal captions led to better comprehension compared to viewing an uncaptioned video, it was expected that enhanced captions would also be as beneficial as normal captions. Surprisingly, however, analysis of the results of Video 1 showed there was no significant difference between viewing without captions and with enhanced captions. This suggests that typographic enhancement may reduce the benefit of captioning on comprehension. Similar finding was observed for Video 2. This is because for most of the participants (i.e., VST score range 39.69 to 89.96), the predicted probabilities of getting an item correct were higher under the normal caption conditions compared to the enhanced caption conditions. This hints at a possible trade-off between a positive effect of typographically enhanced captions on the learning of MWEs and a negative effect on comprehension. This is also reminiscent of Choi's (2018) findings. In his study of the processing and learning of enhanced English collocations, Choi's (2018) found that while the group that saw the enhanced collocations recalled more of the target collocations, they did worse in the recall of the unenhanced content words. Specifically, the group that saw the baseline text, i.e., text without any enhancement, recalled almost 50 percent more of the unenhanced text compared to their peers who saw textually-enhanced collocations. This prompted Choi to suggest that the use of textual enhancement results in a trade-off between the learning of enhanced MWEs and unenhanced text. The findings of the present study suggests that the same might be true in the content of L2 viewing. That is, although the typographically enhanced captions led to significantly higher form recall gains compared to the uncaptioned condition, typographic enhancement did not seem to boost content comprehension. In fact, comprehension seemed to be better when normal captions were used. This finding aligns well with the cognitive load theory (e.g., Sweller, 1994) and the theory of limited processing capacity (Ellis, 2011; VanPatten, 1996), which is the underlying idea of TOPRA (Barcroft, 2002). These theories posit that the working memory is a limited capacity processing system and that the human

attentional resources are finite. As such, it is conceivable that in the present study, attention that could have been allocated to comprehending the input videos was usurped by the MWE learning process imposed by the typographically enhanced captions. This may explain why the enhanced captions conditions did better than the normal captions conditions in the form recall tests in Study 2, but did not outperform the normal captions condition in the comprehension test.

It has to be acknowledged, however, that the superior performance of the participants under the normal captions condition may be attributed to their bigger vocabulary size. The VST mean score for the participants under this condition were 73.04 and 71.09, under viewing once (NC1) and twice (NC2) respectively, while for the enhanced caption conditions, the mean VST scores were 68.37 and 65.92 respectively, for the one viewing (EC1) and two viewings (EC2) conditions. As shown in the literature, learners with a larger vocabulary size are better at understanding reading and listening texts compared to learners with a smaller vocabulary size (Laufer & Ravenhorst-Kalovski, 2010; Noreillie, Kestemont, Heylen, Desmet, & Peters, 2018; Schmitt, Jiang, & Grabe, 2011; Stæhr, 2009).

That vocabulary size influences learners' comprehension of audio-visula input is further emphasized by the fact that VST score was found to be a significant predictor. In essence, an increase of one unit in VST score resulted in higher odds of getting an item correct in the comprehension test for the first video. This finding is in line with a recent study by Durbahn et al. (2020), who also found that the more words known in a text, the higher the learners' viewing comprehension. Additionally, for the second video, VST score was found to modulate the effect of caption condition.

Another significant predictor of comprehension was number of viewings. This finding is supported by the literature on the effects of repeated listening on L2 listening comprehension. For example, previous studies have also shown that compared to first listening, listening for the second time significantly increases content comprehension (Lund, 1991; Sakai, 2009). Further, repeated viewing has been shown to lead to better understanding compared to other methods that enhance listening comprehension such as activation of background knowledge and previewing of comprehension question (Chang & Read, 2006). In addition, the present study lends support to Nguyễn's (2017) viewing study, which found that compared to one viewing, viewing the same TED talk again leads to significantly better comprehension.

The effect of repeated viewing on comprehension, however, was only found for the first video. As for the second video, repetition was not a significant predictor of comprehension. The lack of effect of repetition on all the MWE tests in Study 2 may be accounted for by the



learners' expectation of the MWE tests. As regards the comprehension test, however, learners were informed that questions on video content would follow in both studies. By Hulstijn's (2001) definition, comprehension of both videos was then tested under intentional learning conditions. It is therefore unclear why repetition had an effect on comprehension of the first video, but not the second video. Furthermore, feedback from research participants as well as the comparison of the lexical coverage in both videos indicated that the second video was harder to comprehend than the first video. It is then even more surprising that repetition had an effect on the first video, rather than the second video. It could be speculated that going through the first experience of watching the video once and being tested for their comprehension in Study 1 may have made the learners in the EC1, NC1 and UC1 groups (i.e., the conditions in which the video was only viewed once) pay closer attention to the second video. The knowledge that these learners gleaned from Study 1, in terms of test format and the nature of the questions they were expected to answer, may have prepared them enough to rival their peers who had the advantage of watching the video a second time. Alternatively, upon learning that they would be tested on the MWE knowledge aspects, learners in the EC2, NC2 and UC2 groups may have used more of their attentional resources on the target items, instead of the video content. This may then have caused their performance on the comprehension test to be similar, but not superior to their peers who only viewed the video once. As said, however, this must remain speculation.

In summary, the present study presents evidence that comprehension is facilitated by the presence of normal captions (as suggested by Montero Perez et al., 2013). The present study also suggests that the use of typographically enhanced captions does not necessarily lead to superior comprehension. In fact, there is a possibility that compared to enhanced captions, normal captions might be better for content uptake. As regards the effects of repetition on comprehension, however, no strong conclusion could be made. Owing to the mixed findings, whether it is worth investing extra classroom time for repeated viewing cannot be answered conclusively. Finally, the fact that vocabulary size determines comprehension highlights the importance of lexical coverage as a criterion in choosing a video as input for both MWE learning and comprehension.

## **5.6 Summary of all findings**

The main aim of the research project was to assess the effects of caption condition and repetition on incidental and intentional MWE learning. The answer to the question, however,

is not straightforward as it is clear from the discussion above that different aspects of MWE knowledge were influenced by different factors. Further, whether a variable exerted a strong influence also depended on the learning condition. Several key patterns could be observed. Firstly, under both learning conditions, caption condition had an effect at the level of MWE form recall and MWE form recognition only. Under the incidental learning condition, for instance, the superiority of the captioned conditions over the uncaptioned condition was observed through the significantly higher short-term gains under the former. Further, long-term gains the enhanced captions were also significantly higher than the uncaptioned condition. The typographically enhanced captions, however, did not lead to significantly higher short-term and long-term gains compared to the normal captions. This suggests that where incidental learning of forms is concerned, the use of normal captions may be just as effective as typographically enhanced captions. The effect of typographic enhancement on form recall under the intentional learning condition, however, seems to be stronger as both types of gains under this condition were significantly higher compared to the normal captions. The inconsistent findings between Study 1 and Study were attributed to the different levels of noticing that each learning condition may have induced. Interestingly, however, caption condition was not a significant predictor of form recognition under the intentional learning condition. The lack of effect of caption condition on form recognition was attributed to the fact that receptive knowledge is easier to enhance than productive knowledge.

Secondly, while repeated viewing was found to enhance form recall and form recognition, this was only observed under the incidental learning condition. Repeated viewing had no effect on any of the aspects of MWE under the intentional learning condition, suggesting that test awareness renders a second viewing redundant for the development of MWE knowledge. In contrast, under the incidental learning condition, repeated viewing led to better knowledge of MWE forms. Having said that, the effect of repetition was only observed for the short-term gains, suggesting that the effect of repetition may not be durable for MWE learning (at least in the context of incidental MWE form recall through L2 viewing). Additionally, while repeated viewing facilitated form recall knowledge under all caption conditions, the extent to which knowledge of form was enhanced under each caption condition varied. Similar to the trend observed under viewing once, viewing twice with normal and typographically-enhanced captions yielded significantly higher short-term gains compared to the uncaptioned condition. While the results suggested that repeated viewing may be necessary to reap bigger gains, it was

also revealed that watching a video *once* with typographically-enhanced captions led to larger predicted short-term gains in form recall compared to viewing an uncaptioned video *twice*.

Although vocabulary size and prior knowledge were predictors of secondary interest, they were found to influence many aspects of MWE knowledge under the incidental learning condition. In essence, the bigger the learners' vocabulary knowledge, as measured using their VST score, the bigger their incidental form recall gains score. Additionally, the participants performed better in the form recognition test as their VST score and/or pretest score increased. Interestingly, however, in Study 2, VST score was not found to influence learners' form recall gains, or their performance in the form recognition test. This suggests that with test awareness, MWE form uptake may be enhanced to a similar extent for learners of all vocabulary sizes. Another similar finding for both learning conditions was that besides the item-related variables, meaning recall and recognition were only influenced by learners' vocabulary knowledge. Neither caption condition or repetition influenced this aspect of MWE knowledge, highlighting the fact that visual salience as well as encountering target items once or twice is not sufficient to help learners grasp the MWE meanings, especially in the case of the non-transparent MWEs. As mentioned earlier, however, the results of the meaning tests must be interpreted with caution, due to potential test effects.

All in all, the findings of the present study are largely consistent with the literature on both incidental and intentional vocabulary acquisition in the context of L2 viewing. Firstly, the presence of captions facilitates the uptake of lexical items. Secondly, typographic enhancement has the potential to boost the learning of MWEs owing to the explicit demarcation of the target items. The present findings, however, suggest that without test announcement, normal captions may be as beneficial as enhanced captions at least where the learning of MWE forms is concerned. It is under the intentional learning condition that typographically enhanced captions appear to lead to superior knowledge of form, compared to normal captions. Next, while both normal captions and typographically enhanced captions could lead to better knowledge of MWE form, they do not result in the uptake of meaning and this was borne out in the meaning tests. Similarly, repeated viewing fosters the incidental learning of MWE knowledge of form, but appears to be insufficient to engender knowledge of MWE meanings. Additionally, test awareness reduces the number of encounters with an MWE needed to obtain knowledge of MWE form. In short, where the use of L2 viewing in vocabulary learning is concerned, the findings support the use of normal and typographically enhanced captions, as well as repetition (under incidental learning conditions) to fuel learners' MWE lexical development at the level

of form recall. Typographically enhanced captions, however, are unlikely to facilitate content and MWE form uptake simultaneously. Therefore, teachers should be informed of not only the potential of typographically enhanced captions for MWE learning, but also the possibility of their adverse effects on content comprehension. The following section details more suggestions on how and when to use typographically-enhanced captions and repeated viewing to bolster learning.

### **5.7 Pedagogical implications**

The findings of the present study have an important bearing on L2 pedagogy. Firstly, the findings indicate that watching a video for its content, without a deliberate focus on the linguistic elements, can be a useful means of enhancing L2 MWE learning. However, two considerations have to be borne in mind: (a) without captions the learning gains are negligible, and (b) even with the aid of captions and typographically-enhanced captions, only certain aspects of MWEs may be enhanced through L2 viewing. Specifically, when teachers use a video with the primary goal of facilitating the understanding of the content, they could use normal captions as they promote learners' recall of the form of unknown MWEs. While enhanced captions also facilitate the uptake of MWE form, they may take away too many attentional resources, leaving learners with few resources for content uptake. As such, if a video is to be primarily processed for content, teachers could first use normal captions. The second viewing could then be dedicated to boosting learners' MWE knowledge through the use of enhanced captions. Further, if learners are forewarned of an MWE test, the use of typographically-enhanced captions for the second viewing may prove to be even more beneficial for the purpose of accelerating the acquisition of MWE form.

Indeed, a second viewing should be advocated for language teachers as it leads to better MWE knowledge under incidental learning condition. This is true especially for uncaptioned viewing. When typographically enhanced captions are used, however, repeated viewing may not be necessary for incidental learning of MWE forms, as the present study found that short-term gains obtained from viewing a video once with enhanced captions were greater compared to gains from viewing an *uncaptioned* video *twice*. Teachers may, of course, choose to play a video twice, as repeated viewing leads to greater gains irrespective of caption condition. However, in the interest of minimising the time investment without compromising the incidental learning gains, teachers could opt for captioned viewing if the video is meant to be

processed for content. Alternatively, subsequent viewings may be assigned as homework to solidify form-meaning connections.

When teachers prepare video-based activities in which learners are explicitly warned about MWEs tests, however, letting learners watch a video once may be sufficient to promote MWE form knowledge. The extra time investment may not make much difference in knowledge gain. What may make a difference is the use of typographic enhancement. The findings suggest that teachers should use typographically-enhanced captions, as learners' awareness of impending tests does not mean that they are able to notice and recognise the unknown MWEs as a whole. As suggested by Peters (2012), prompting learners to merely allocate "attentional resources to FS ... had no effect on their recall because as a technique it was probably not specific and guided enough to effectively focus students on the items to be learned and to induce deep processing" (p. 78). In sum, the use of bolding and underlining of novel MWEs can help learners notice and demarcate these phrases.

The present finding also suggests that L2 viewing may support the learning of MWE forms irrespective of proficiency level. In other words, L2 viewing may lead to an increase in MWE knowledge of form for all proficiency levels. The extent of the learning, however, depends on the caption condition. Further, the findings of the first study suggest that the more proficient students may pick up more MWEs compared to their weaker peers. Consequently, teachers should be mindful of lower proficiency learners, and bolster their learning using more explicit activities so that their learning gains can rival that of their more proficient peers. Additionally, as vocabulary size determines how much learners get out of watching a video, teachers should make it a point to choose input videos with lexical coverage that suits their learners' vocabulary profile, as it impacts listening and viewing comprehension. It has to be conceded, however, that the participants in the present study came from the same student population and there was not an extremely wide variation in proficiency levels. Therefore, the conclusion that all proficiency levels may benefit from L2 viewing is only tentative.

The enhancement of MWE meanings, however, may require different treatments, beyond providing captions and repeated viewing. As suggested by Meunier (2012), "more productively oriented approaches" (p. 122) may be needed for certain aspects of MWEs. As mentioned earlier, on its own, typographic enhancement cannot clarify the meaning of MWEs, especially the semantically opaque ones. Additionally, the present study also suggests that, even with test awareness, the aid of typographic enhancement is not effective in facilitating the

acquisition of MWE meaning. Therefore, teachers may want to include explicit activities that foster form-meaning links, such as translation exercises (e.g., Laufer & Girsai, 2008) and meaning-matching activities (e.g., Szudarski, 2012). As suggested by Szudarski and Carter (2016) “if learners’ productive knowledge of collocations needs to be increased, then input enhancement, input flood and potentially semantic processing could be combined” (p. 261).

In sum, the present study lends some support for the use of captions and typographically-enhanced captions for the purpose of fostering MWE acquisition through L2 viewing. This aligns well with Webb and Rodgers’s (2009a) suggestion that watching TV can be an effective method of learning vocabulary. The addition of captions would be one way to ensure that L2 learners get more out of viewing TV. As such, teachers may benefit from learning how to use readily available software such as *SubtitleEdit* (<https://subtitle-edit.en.softonic.com/>) to include the caption type (i.e., normal or enhanced) that fits the purpose of the activity. Additionally, learners’ awareness of the value of captioned viewing for the learning of novel phrases should also be raised. This is because the option of captioned viewing is available on platforms such as YouTube and Netflix, which are platforms that L2 learners are likely to use for leisure outside the classroom. If extensive L2 viewing is considered as a “useful complement to extensive reading in developing lexical knowledge” (Peters & Webb, 2018, p. 20), then learners stand to gain much more when they are aided by the presence of captions.

### **5.8 Limitations and directions for future research**

Several limitations have to be acknowledged. The first relates to the small sample size, both in terms of number of participants and items. As there were some learners whose VST scores were unavailable, they were excluded from the analysis. This resulted in one of the conditions (i.e., UC1) having only 15 participants in Study 1, which could be considered on the low side. As regards the number of target items in each study, although more target items are desirable, the authentic nature of the video did not allow for this. It should be noted, however, that the number of items in the present investigation is similar to that in earlier studies (e.g., Webb et al., 2013) and more than other comparable studies (e.g., Choi, 2018; Sonbul & Schmitt, 2013; Toomer & Elgort, 2019). There were more target MWEs in the videos, but they appeared in varying frequency. As it is impossible to control this, only MWEs that appeared once were chosen as target items, and repetition was then operationalised as repeated viewing. This way of operationalising repetition was also thought to be better for MWE acquisition (at

least in the early stages), as Durrant and Schmitt's (2010) study showed that verbatim repetition, i.e., encountering an MWE in the same context twice, is better than varied repetition, i.e., encountering an MWE in different contexts. A greater number of target items that occur with a higher frequency can be obtained from using a longer video. However, the participants in the repetition condition would then have to sit through a very long treatment. Therefore, to reduce the risk of fatigue, shorter videos were chosen, which naturally entails fewer target MWEs.

The relatively small number of items also means that the findings related to the effects of the item-related variables cannot be generalised, and should only be considered as exploratory. For instance, although compositionality was found to predict meaning recall, it stands to reason that successful recall may not be solely due to the treatment effect, but learners' ability to guess the figurative meanings of the MWEs from the constituent words. Put differently, it is difficult to determine whether correct responses in the meaning recall test can be attributed to the presence or absence of captions. This is because in the meaning recall test, the learners could see the MWEs, which may have then prompted them to deduce the figurative meaning. As such, a correct response may not necessarily be due to successful retrieval during L2 viewing. Additionally, as emphasized before, the nature of the target items in the present study may have rendered the effects of some item-related variables stronger than others. This also provides strong grounds to suggest that future studies include syntactically and semantically diverse MWEs as target items, in order to assess whether the learnability of MWEs is indeed affected by variables such as the ones investigated in the present study. For instance, it would be interesting to replicate the study with items such as transparent collocations. This is because typographic enhancements may have a greater effect since learners do not need to figure out the meaning of such items, so the challenge only lies in remembering the lexical composition. Also, as acknowledged earlier, having more target items may provide more robust findings as the issue of model overfitting may be avoided.

One could argue that the effects of repetition may be clearer when the difference in viewing is more than once. Having said that, it is more ecologically valid for teachers to repeat an input video twice, especially for longer videos. It is unlikely that teachers would spend a large proportion of their precious classroom time on TV viewing. It is, of course, possible for teachers to assign subsequent viewings as homework. However, learners are unlikely to watch an input video more than twice of their own accord. In fact, this was tested during the piloting stage, in which a condition for three repeated viewings was included. Observation, along with

the pilot participants' feedback, indicated that three viewings was indeed too taxing for the participants.

The next limitation pertains to the incongruence between the format of the form recall immediate posttest, and the form recall pretest and delayed posttest. The format of the form recall immediate posttest tested the learners' episodic recall of the MWEs. The format was decided upon in keeping with the notion that congruent learning task and test condition will lead to better performance (Lotto & de Groot, 1998; Schmitt, 2010). However, this format was different from the form recall pretest and delayed posttest, in which learners were required to insert the newly acquired MWEs in a context not entirely the same as the input video. Although incongruent to the immediate posttest, the format of the pretest was chosen as it was thought to test the learners' existing knowledge of MWEs, without giving away the target items. Put differently, the format of the pretest did not give away the form of the MWEs, thereby not affording the learners an extra learning opportunity. However, it could be argued that the learning gains between the pretest and immediate and delayed posttest might have been confounded by the different test formats. Further, it has to be conceded that the immediate posttest may have solidified what the learners had learned, giving those who learned more during the treatment an advantage in the delayed posttest. Future studies could perhaps have different groups that take the posttests at different times, akin to Peters (2014). Additionally, future studies may also want to include both types of immediate posttests (i.e., episodic form recall test and contextualised form recall test), taken by different groups of participants, to obtain a clearer picture of the extent of MWE learning.

The study was also limited by the fact that form recognition, meaning recall and meaning recognition were not pre-tested. This study placed more emphasis on the aspects of MWE form, as it was unlikely that learners would be able to infer and remember the meaning of new non-transparent MWEs after one or two encounters. It is, however, plausible that the contextualised visual cues facilitate the acquisition of the meaning of MWEs. Therefore, it seemed worth measuring gains made in knowledge of MWE meaning as well. As mentioned earlier, the results of the meaning tests have to be interpreted cautiously, however, as the form recall and form recognition delayed posttests contained contextual clues which hinted at the MWE meanings. This may then have created a test effect. Future studies can make use of the methodology in Peters and Webb (2018) study, in which two different experiments were carried out with different groups of learners, so that one experiment tested form recognition and meaning recall, while the other tested meaning recognition.



Finally, it is important to note that while the present study compared and contrasted the outcome of the incidental and intentional learning conditions, the comparison is not direct. While the participants remained the same, the target MWEs used in the two videos were different. A more direct comparison can be made by exposing learners under both the intentional and incidental conditions to the same target MWEs, akin to what was done in Montero Perez et al. (2015). This is something that future studies could look into, although it would necessitate a large sample of participants.

A last research angle that should be noted for future exploration is the ‘sustainability’ of these two methods: test announcement and typographic announcement. The present study suggests that test announcement renders repeated viewing redundant, as it led to a similar increase in the four aspects of MWE knowledge as viewing once. However, how long would the effect of test announcement last? If teachers forewarn the learners each time they watch a video in the interest of greater learning gains, will the learners always allocate as much attentional resources or will the novelty of the treatment wear off? Similarly, it would be interesting to investigate if the use of typographic enhancement reaches a saturation point or a point at which learners become oblivious to it, due to constant exposure. Further, as noted by Boers (2019), “enhancing too much of a text defeats the purpose of making selected items stand out” (p. 148). Simply put, longitudinal studies are needed to assess the long-term effectiveness of these interventions.



## CONCLUSION

This research project was motivated by the realisation that little is known about the potential use of audio-visual materials, or L2 viewing, as a means of enhancing MWE knowledge. Despite the meteoric rise of studies concerning MWEs in the past decade, more focus had been given to pedagogic interventions by means of written input, and to a limited extent, bimodal input. It is only recently that the focus has shifted to multimodal input, i.e., L2 viewing. Guided by the fact that L2 viewing has been shown to facilitate the uptake of single words (e.g., Peters & Webb, 2018), this research project sought to investigate whether the same applied for MWE learning.

By drawing on previous intervention studies, this project explored the ways in which L2 viewing could be further optimised for MWE learning. The first way involved the manipulation of L2 audio-visual input through the use of captions and typographically-enhanced captions. In the context of single word acquisition, previous studies have consistently furnished evidence for the superiority of captioned viewing over uncaptioned viewing for single word learning (e.g., Montero Perez et al., 2013). This prompted the inclusion of both caption conditions in the research project to assess if the same would be true for MWE learning. Another means that has been proven to stimulate independent learner uptake of MWEs in the context of written input is typographic enhancement. The use of typographic enhancement has been reported to be favourable for MWE learning, as it helps learners notice the salient phrases and facilitates their intake more than the unenhanced ones (e.g., Bishop, 2004; Boers et al., 2017; Choi, 2018). This provided a strong reason to believe L2 viewing with the aid of typographically-enhanced captions would benefit L2 learners the same way. Thus, this created the third caption condition in the research project.

Another factor that has been firmly established as beneficial for both single word and MWE learning is repetition. Previous MWE studies on the effects of repetition, however, have utilised materials that are modified so as to include the necessary number of target items required to enable multiple encounters (e.g., Webb et al., 2013; Pellicer-Sánchez, 2017). As learners are unlikely to be exposed to such modified input, the ecological validity of such texts is then in question. Further, it would require a great deal of resourcefulness on the teacher's part to create or modify an authentic L2 video to include a substantial amount of repeated MWEs. Thus, an alternative way to investigate repetition in conjunction with unmodified authentic L2 video would be through repeated viewing, which is how repetition was

operationalised in the research project. The investigation into the effects of repetition was not only to compare MWE learning between one viewing and two viewings, but also to assess whether repeated viewing modulated MWE learning under the respective caption conditions. For instance, it was of interest to find out if the use of typographic enhancement might eliminate the need for repetition, as suggested by Szudarski and Carter (2016).

Another observation made from previous studies was the fact that the effects of typographic enhancement and repetition have been predominantly investigated in the context of incidental learning. Although deliberate learning has been said to require fewer repetitions for comparable gains (Webb & Nation, 2017), whether this applies to MWE learning through L2 viewing is not known. Similarly, the investigation into the effects of typographic enhancement has mostly been confined to the context of incidental learning. It is thus unclear whether typographic enhancement is still necessary when learners know that they will be tested on their MWE knowledge. This provided the impetus for the research project, in which the incidental and intentional learning conditions were created to assess if they modulated the effects of typographic enhancement and repetition on MWE learning. Following Hulstijn's (2001) distinction, the two conditions were differentiated by the presence of test announcement.

The findings, gleaned from the data obtained through multiple tests on different aspects of MWE knowledge, showed trends that align with the literature on pedagogic interventions for MWE learning, as well as on vocabulary learning in general. Firstly, the presence of captions and typographically-enhanced captions facilitated the uptake of MWE forms. This is because both the incidental and intentional short-term form recall gains were significantly higher under the two caption conditions compared to the uncaptioned condition. The effects of typographic enhancement on incidental form recall, however, were similar to the normal captions. This lack of significant difference was attributed to the possibility that noticing at a low level of awareness may have weakened the effects of typographic enhancement. In contrast, under the intentional learning conditions, typographically enhanced captions proved to be superior than normal captions as short-term gain under the former were significantly higher compared to the latter. This corroborates previous reading studies that found that typographic enhancement leads to the noticing and uptake of salient items (e.g., Boers et al., 2017; Choi, 2018). The present study, however, provides reason to believe that test announcement may strengthen the effects of typographic enhancement by virtue of noticing at the level of awareness. Additionally, the results of the present study also lend weight to the

supposition that MWEs might go unnoticed (e.g., Martinez & Murphy, 2011) even under intentional learning conditions, and that unknown MWEs are not recognised as readily as unknown words (e.g., Bishop, 2004). As such, test awareness alone may not be sufficient, as learners may still need to rely on typographic enhancement in order to recognise and attend to MWEs. Also, that caption condition matters for form recognition lends further support for the use of captions in L2 videos intended for meaning-focused activities. While caption condition was not found to predict form recognition under the intentional learning condition, this could be due to the fact that receptive knowledge precedes productive knowledge. Put differently, caption condition did not exert such a strong influence on form recognition as it did on form recall, as the former is more easily enhanced compared to the latter.

Another compelling reason to use typographically-enhanced captions lies in the fact that they reduce the number of viewings needed to make incidental short-term gains in form recall knowledge. This is evidenced by the greater short-term gains made from viewing a video once under the enhanced captions, compared to the gains made from repeated uncaptioned viewing. It is encouraging to find that time investment could be reduced by the use of typographically-enhanced captions. Having said that, however, repetition was shown to lead to bigger short-term gains for all caption conditions, so assigning subsequent viewings as homework for the purpose of greater incidental MWE gains should be encouraged. This is especially since the effect of repetition on form recall was not found to be durable under incidental learning. Therefore, subsequent viewings may be needed to leave stronger memory traces that are less prone to attrition. Under the intentional learning condition, however, the effect of repeated viewing appears to be redundant, for any of the MWE aspects. This aligns well with the supposition that fewer repetitions are needed for deliberate learning, in the sense that when learners are aware of the test, viewing a video for the second time leads to similar gains as viewing it once only.

In spite of their promising potential as a means of boosting MWE form knowledge, it appeared that captioned viewing and repetition are not enough for the development of the meaning aspect of MWE knowledge. In essence, these two factors did not emerge as significant predictors of meaning recall or meaning recognition, under both learning conditions. Again, this finding is in line with previous studies, which also found that MWE meaning acquisition is not influenced by repetition, even when coupled with typographic enhancement (e.g., Szudarski & Carter, 2016). This finding also underscores the fact that typographic enhancement on its own cannot elucidate the meanings of the MWEs. Additionally,

encountering a word once or twice is not enough to establish a strong form meaning link in the meaning lexicon (Hulstijn, 1992), especially in the case of non-transparent novel phrases. Importantly, this points to the need for captioned and repeated viewing to be complemented by form-focused activities or explicit instruction as a way to bolster learning and create initial form-meaning mapping.

The final important finding pertains to the fact that vocabulary knowledge influences the amount of MWE learning that takes place. This is especially so when learners are not aware that they will be tested. The Matthew effect found in the incidental short-term and long-term gains analysis, for instance, emphasizes that while L2 viewing benefits learners of all vocabulary sizes, the more proficient learners stand to gain more than their less proficient counterparts. Thus, teachers need be aware of these less proficient learners, and ensure that they do not lag further behind their more proficient peers by scaffolding their knowledge through pre- or post- activities that could solidify their learning. Additionally, it is crucial that the input video chosen matches learners' vocabulary profile, as comprehension was found to be influenced not only by caption condition and number of viewing (at least based on the results of Video 1), but also by learners' vocabulary size. The findings from the comprehension test also shed light on the possible drawback of using typographic enhancement. This is because although typographically enhanced captions facilitate the uptake of MWE form, this appeared to be at the expense of content uptake. As such, it is suggested that typographically enhanced captions should only be used when content comprehension is of secondary interest. Alternatively, typographically enhanced captions may be used during the second viewing, after the learners have viewed the input video with the help of normal captions.

All in all, the findings of this research project highlight the fact that the acquisition of MWE knowledge is a complex process, and one that requires a number of factors beyond just a combination of typographic enhancement and repetition. Although these two factors are proven to be facilitative (at least for MWE form acquisition), the gains were not substantial, suggesting that supplementary activities may be needed for more robust and durable gains at all levels of MWE competence. Finally, the findings could be considered only as the first step towards gaining a better understanding of the potential value of L2 viewing as a method of expanding learners' phraseological knowledge. Further research is needed to establish the role of these interventions in a wider range of ESL and EFL contexts.

## REFERENCES

- Aitken, K. G. (1978). Measuring listening comprehension in English as a second language. *TEAL Occasional Papers*, 2.
- Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259-302). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.
- Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, 52(2), 323-363. doi:10.1111/0023-8333.00186
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. Retrieved from <http://CRAN.R-project.org/package=lme4>
- 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. doi:10.1017/S0142716402004022
- Bishop, H. (2004). The effect of typographic salience on the look up and comprehension of unknown formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 227-248). Amsterdam: John Benjamins.
- Boers, F. (2019). Factors affecting the learning of multiword items. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 143-157). New York, NY: Routledge.
- Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb–noun collocations. *Language Teaching Research*, 18(1), 54-74. doi:10.1177/1362168813505389
- Boers, F., Demecheleer, M., He, L., Deconinck, J., Stengers, H., & Eyckmans, J. (2017). Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*, 27(2), 448-469. doi:10.1111/ijal.12141
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, 10(3), 245-261. doi:10.1191/1362168806lr195oa
- Boers, F., Eyckmans, J., & Stengers, H. (2007). Presenting figurative idioms with a touch of etymology: More than mere mnemonics? *Language Teaching Research*, 11(1), 43-62. doi:10.1177/1362168806072460

- Boers, F., & Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition*. Hampshire, UK: Palgrave Macmillan.
- Boers, F., Lindstromberg, S., & Webb, S. (2014). Further evidence of the comparative memorability of alliterative expressions in second language learning. *RELC Journal*, 45(1), 85-99. doi:10.1177/0033688214522714
- Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes*, 45(2), 122-159. doi:10.1080/01638530701792826
- Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics*, 18, 171-191.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing*, 15(2), 119-157. doi:10.1177/026553229801500201
- Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition*, 24(2), 215-221. doi:10.1017/S0272263102002061
- Bybee, J. L., & Hopper, P. J. (2001). *Frequency and the emergence of linguistic structure*. Amsterdam and Philadelphia: John Benjamins Publishing.
- 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. doi:10.2307/40264527
- Choi, S. (2018). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, 21(3), 403-426. doi:10.1177/1362168816653271
- Christensen, R. H. B. (2019). ordinal - Regression Models for Ordinal Data. R package version 2019.12-10. Retrieved from <http://www.cran.r-project.org/package=ordinal/>
- Cieślicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research*, 22(2), 115-144. doi:10.1191/0267658306sr263oa
- Columbus, G. (2010). Processing MWU: Are MWU subtypes psycholinguistically real. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 194-210). London, UK: Continuum.
- Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32, 45-61. doi:10.1017/S0267190512000074



- Davies, M. (2008). *Corpus of Contemporary American English (COCA)*. Retrieved from: <https://corpus.byu.edu/coca/>
- De Ridder, I. (2002). Visible or invisible links: Does the highlighting of hyperlinks affect incidental vocabulary learning, text comprehension, and the reading process? *Language, Learning & Technology*, 6(1), 123-146.
- Díaz-Cintas, J., & Ramael, A. (2014). *Audiovisual translation: Subtitling*. Oxfordshire, England: Routledge.
- Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System*, 88. doi:10.1016/j.system.2019.102166
- Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching (IRAL)*, 47, 157-177.
- Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, 26(2), 163-188. doi:10.1177/0267658309349431
- Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, 40(2), 341-366. doi:10.1017/S0272263117000109
- Elgort, I., Perfetti, C. A., Rickles, B., & Stafura, J. Z. (2015). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and event-related brain potential (ERP) indicators of learning. *Language, Cognition and Neuroscience*, 30(5), 506-528. doi:10.1080/23273798.2014.942673
- Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, 64(2), 365-414. doi:10.1111/lang.12052
- Ellis, N. C. (2011). Implicit and explicit SLA and their interface. In C. Sanz & R. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 35-48). Washington, DC: Georgetown University Press.
- Ellis, R. (1993). The structural syllabus and second language acquisition. *TESOL Quarterly*, 27(1), 91-113. doi:10.2307/3586953
- Ellis, R. (1999). Factors in the incidental acquisition of second language vocabulary from oral input. In R. Ellis & S. Fotos (Eds.), *Learning a second language through interaction* (pp. 35-61). Amsterdam, The Netherlands: John Benjamins.
- Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, 54(2), 227-275. doi:10.1111/j.1467-9922.2004.00255.x

- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29-62. doi:10.1515/text.1.2000.20.1.29
- Eyckmans, J., Boers, F., & Stengers, H. (2007). Identifying chunks: Who can see the wood for the trees. *Language Forum*, 33(2), 85-100.
- Feng, Y., & Webb, S. (2019). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, 1-25.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London, UK: Sage.
- Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, 40(2), 273-293. doi:10.2307/40264523
- Frisbie, D. A., & Becker, D. F. (1991). An analysis of textbook advice about True-False tests. *Applied Measurement in Education*, 4(1), 67-83. doi:10.1207/s15324818ame0401\_6
- 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., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language, Learning & Technology*, 23(2), 84-104.
- Gieve, S., & Clark, R. (2005). 'The chinese approach to learning': Cultural trait or situated response? The case of a self-directed learning programme. *System*, 33(2), 261-276. doi:10.1016/j.system.2004.09.015
- Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, 35(3), 483-517. doi:10.1017/S0272263113000119
- Godfroid, A., & Schmidtke, J. (2013). What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports and vocabulary learning scores. *Noticing and second language acquisition: Studies in honor of Richard Schmidt*, 183-205.
- González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have?: The effects of frequency and amount of exposure. *International Journal of Applied Linguistics*, 166(1), 94-126. doi:10.1075/itl.166.1.03fer
- Haladyna, T. M. (1992). The effectiveness of several multiple-choice formats. *Applied Measurement in Education*, 5(1), 73-88. doi:10.1207/s15324818ame0501\_6
- Hildyard, A., & Olson, D. R. (1978). Memory and inference in the comprehension of oral and written discourse. *Discourse Processes*, 1(2), 91-117.

- Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 47-69). Hove: Language Teaching Publications.
- 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, I. S. 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. doi:10.2190/RG06-LYWB-216Y-R27G
- Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 113-125). London: Macmillan.
- 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, UK: Cambridge University Press.
- Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal, 80*(3), 327-339. doi:10.1111/j.1540-4781.1996.tb01614.x
- Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition, 24*(4), 541-577. doi:10.1017/S0272263102004023
- Kim, Y. (2006). Effects of input elaboration on vocabulary acquisition through reading by Korean learners of English as a Foreign Language. *TESOL Quarterly, 40*(2), 341-373. doi:10.2307/40264526
- Ko, M. H. (2012). Glossing and second language vocabulary learning. *TESOL Quarterly, 46*(1), 56-79. doi:10.1002/tesq.3
- Kreiter, C. D., & Frisbie, D. A. (1989). Effectiveness of multiple true-false items. *Applied Measurement in Education, 2*(3), 207-216. doi:10.1207/s15324818ame0203\_2

- 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, UK: Multilingual Matters.
- 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. doi:10.3138/cmlr.59.4.567
- Laufer, B., & Girsai, B. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, 29(4), 694-716. doi:10.1093/applin/amn018
- 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., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research*, 15(4), 391-411.
- Lee, S. K. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning*, 57(1), 87-118.
- Lenth, R. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package. doi:<http://CRAN.R-project.org/package=emmeans>
- Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning*, 47(3), 467-505. doi:10.1111/0023-8333.00017
- Li, J., & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case study approach. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 22-46). New York, NY: Continuum.
- Lin, P. M. S. (2014). Investigating the validity of internet television as a resource for acquiring L2 formulaic sequences. *System*, 42(1), 164-176. doi:10.1016/j.system.2013.11.010
- Lin, P. M. S. (2018). Formulaic language and speech prosody. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 78-93). New York, NY: Routledge.
- Lin, P. M. S., & Siyanova-Chanturia, A. (2014). Internet television for L2 vocabulary learning. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 149-158). London, UK: Routledge.

- Liu, N., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal*, 16(1), 33-42. doi:10.1177/003368828501600103
- Lotto, L., & de Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, 48(1), 31-69. doi:10.1111/1467-9922.00032
- Lund, R. J. (1991). A comparison of second language listening and reading comprehension. *Modern Language Journal*, 75(2), 196-204. doi:10.1111/j.1540-4781.1991.tb05350.x
- Macis, M. (2018). Incidental learning of duplex collocations from reading: Three case studies. *Reading in a Foreign Language*, 30(1), 48-75.
- Macis, M., & Schmitt, N. (2017). The figurative and polysemous nature of collocations and their place in ELT. *ELT Journal*, 71(1), 50-59. doi:10.1093/elt/ccw044
- Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals*, 32(3), 321-328.
- Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, 45(2), 267-290. doi:10.5054/tq.2011.247708
- Mayer, R. E. (2009). *Multimedia learning*. Cambridge, UK: Cambridge University Press.
- Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics*, 32, 111-129. doi:10.1017/S0267190512000128
- Mohamed, A. A. (2018). Exposure frequency in L2 reading. *Studies in Second Language Acquisition*, 40(2), 269-293. doi:10.1017/S0272263117000092
- Mohd Jelani, N. A., & Boers, F. (2018). Examining incidental vocabulary acquisition from captioned video: Does test modality matter? *ITL-International Journal of Applied Linguistics*, 169(1), 169-190.
- Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning and Technology*, 18(1), 118-141.
- Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *Modern Language Journal*, 99(2), 308-328. doi:10.1111/modl.12215
- Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41(3), 720-739. doi:10.1016/j.system.2013.07.013

- Mueller, R. A. G., & Gibbs, R. W., Jr. (1987). Processing idioms with multiple meanings. *Journal of Psycholinguistic Research*, 16(1), 63-81. doi:10.1007/BF01067751
- 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* (Vol. 10, pp. 3-13). Amsterdam: John Benjamins.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59-81.
- Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Nation, I. S. P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts. Retrieved from <http://www.vuw.ac.nz/lals/staff/paulnation/nation.aspx>
- Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle, Cengage Learning.
- Neuman, S., & 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. doi:10.2307/747835
- Nguyễn, C. Đ. (2017). *Fostering incidental vocabulary uptake from audio-visual materials: The role of text comprehension*. (Unpublished doctoral dissertation). Victoria University of Wellington, Wellington, New Zealand.
- Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 21(3), 298-320. doi:10.1177/1362168816639619
- Noreillie, A., Kestemont, B., Heylen, K., Desmet, P., & Peters, E. (2018). Vocabulary knowledge and listening comprehension at an intermediate level in English and French as foreign languages: An approximate replication study of Stæhr (2009). *ITL - International Journal of Applied Linguistics*, 169(1), 214-233.
- OECD. (2009). *Communications Outlook 2009*. Paris: OECD.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.
- Paivio, A. (2014). *Mind and its evolution: A dual coding theoretical approach*. New York, NY: Psychology Press.

- Park, H., Choi, S., & Lee, M. (2012). Visual input enhancement, attention, grammar learning, & reading comprehension: An eye-movement study. *English Teaching*, 67(4), 241-265.
- Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading. *Studies in Second Language Acquisition*, 38(1), 97-130. doi:10.1017/S0272263115000224
- Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, 21(3), 381-402. doi:10.1177/1362168815618428
- Pellicer-Sánchez, A., & Boers, F. (2018). Pedagogical approaches to the teaching and learning of formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 153-173). New York, NY: Routledge.
- 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.
- Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology*, 94(1), 23-33. doi:10.1037/0022-0663.94.1.23
- Peters, E. (2009). Learning collocations through attention-drawing techniques: A qualitative and quantitative analysis. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language* (pp. 194-207). Basingstoke, UK: Palgrave Macmillan.
- Peters, E. (2012). The differential effects of two vocabulary instruction methods on EFL word learning: A study into task effectiveness. *International Review of Applied Linguistics in Language Teaching*, 50(3), 213-238. doi:10.1515/iral-2012-0009
- Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research*, 18(1), 75-94. doi:10.1177/1362168813505384
- Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, 20(1), 113-138. doi:10.1177/1362168814568131
- Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *International Journal of Applied Linguistics*, 169(1), 142-168.
- Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, 63, 134-148. doi:10.1016/j.system.2016.10.002

- Peters, E., Noreillie, A. S., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to foreign language input on learners' vocabulary knowledge in two languages. *Language Learning*, 69(3), 747-782. doi:10.1111/lang.12351
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 40(3), 551-577. doi:10.1017/S0272263117000407
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18(1), 1-28.
- Price, K. (1983). Closed-captioned TV: An untapped resource. *MATESOL Newsletter*, 12, 1-8.
- Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, 47(4), 424-438. doi:10.1080/09571736.2019.1638630
- Qi, Y., & Ding, Y. (2011). Use of formulaic sequences in monologues of Chinese EFL learners. *System*, 39(2), 164-174. doi:10.1016/j.system.2011.02.003
- R Core Team. (2018). R: A Language and environment for statistical computing (Version 3.4.4) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org>
- Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, 17(2), 219-240. doi:10.2307/3586651
- Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions*. (Unpublished doctoral dissertation). Victoria University of Wellington, Wellington, New Zealand.
- Rodgers, M. P. H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, 45(4), 689-717. doi:10.5054/tq.2011.268062
- 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.
- Rott, S. (2007). The effect of frequency of input-enhancements on word learning and text comprehension. *Language Learning*, 57(2), 165-199. doi:10.1111/j.1467-9922.2007.00406.x
- Roy Morgan Research. (2015). *Delhi to Dunedin, Beijing to Ballarat: The time spent with TV, internet, newspapers and radio across Asia*. Retrieved from:



<http://www.roymorgan.com/findings/6277-time-spent-with-television-radio-internet-newspapers-across-asia-december-2014-201506090624>

- Sakai, H. (2009). Effect of repetition of exposure and proficiency level in L2 listening tests. *TESOL Quarterly*, 43(2), 360-372. doi:10.1002/j.1545-7249.2009.tb00179.x
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11, 11-26.
- Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In *Attention and awareness in foreign language learning* (Vol. 9, pp. 1-63). Honolulu, HI: University of Hawai'i at Manoa.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge, UK: Cambridge University Press.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York, NY: Palgrave Macmillan.
- Schmitt, N., Dörnyei, Z., Adolphs, S., & Durow, V. (2004). Knowledge and acquisition of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 55-86). Amsterdam, The Netherlands: John Benjamins.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in text and reading comprehension. *Modern Language Journal*, 95(1), 26-43. doi:10.1111/j.1540-4781.2011.01146.x
- Schmitt, N., & Redwood, S. (2011). Learner knowledge of phrasal verbs: A corpus-informed study. In F. Meunier, S. de Cock, G. Gilquin, & M. Paquot (Eds.), *A taste for corpora: In honour of Sylviane Granger* (pp. 173-209). Amsterdam, The Netherlands: John Benjamins.
- 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. doi:10.1177/026553220101800103
- Sharwood-Smith, M. (1993). Input enhancement in instructed SLA: Theoretical Bases. *Studies in Second Language Acquisition*, 15(2), 165-179. doi:10.1017/S0272263100011943
- Shin, D., & Nation, I. S. P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT Journal*, 62(4), 339-348. doi:10.1093/elt/ccm091

- Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27(2), 251-272. doi:10.1177/0267658310382068
- Siyanova-Chanturia, A., & Van Lancker Sidtis, D. (2018). What online processing tells us about formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 31-61). London, New York: Routledge.
- Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, 63(1), 121-159. doi:10.1111/j.1467-9922.2012.00730.x
- 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. doi:10.1017/S0272263109990039
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360-407. doi:10.1598/RRQ.21.4.1
- Stengers, H., Boers, F., Housen, A., & Eyckmans, J. (2010). Does 'chunking' foster chunk-uptake? In S. D. Knop, F. Boers, & A. D. Rycker (Eds.), *Fostering language teaching efficiency through cognitive linguistics* (pp. 99-117). Berlin, Germany: Mouton de Gruyter.
- Stengers, H., Boers, F., Housen, A., & Eyckmans, J. (2011). Formulaic sequences and L2 oral proficiency: Does the type of target language influence the association? *IRAL - International Review of Applied Linguistics in Language Teaching*, 49(4), 321-343. doi:10.1515/iral.2011.017
- Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55(4), 661-699. doi:10.1111/j.0023-8333.2005.00320.x
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295-312. doi:10.1016/0959-4752(94)90003-5
- Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, 14(2), 50-73.
- Szudarski, P. (2012). Effects of meaning- and form-focused instruction on the acquisition of verb-noun collocations in L2 English. *Journal of Second Language Teaching Research*, 1(2), 3-37.

- Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, 26(2), 245-265. doi:10.1111/ijal.12092
- Titone, D. A., & Connine, C. M. (1994). Comprehension of idiomatic expressions: Effects of predictability and literality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5), 1126-1138. doi:10.1037/0278-7393.20.5.1126
- Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, 16(2), 183-203. doi:10.1017/S0272263100012870
- Toomer, M., & Elgort, I. (2019). The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt (2013). *Language Learning*, 69(2), 405-439. doi:10.1111/lang.12335
- Tulving, E. (1993). What is episodic memory? *Current Directions in Psychological Science*, 2(3), 67-70. doi:10.1111/1467-8721.ep10770899
- Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, 69(3), 559-599. doi:10.1111/lang.12343
- van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457-479. doi:10.1093/applin/ams074
- Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT Journal*, 42(4), 272-281.
- Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching*, 43(1), 1-37. doi:10.1017/S0261444809990267
- Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching*. London, UK: Palgrave Macmillan.
- VanPatten, B. (1996). Input processing and grammar instruction: Theory and research. In Norwood, NJ: Ablex.
- Watanabe, Y. (1997). Input, intake, and retention effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, 19(3), 287-307. doi:10.1017/S027226319700301X

- Webb, S. (2002). *Investigating the effects of learning tasks on vocabulary knowledge*. (Unpublished doctoral dissertation). Victoria University of Wellington, Wellington, New Zealand.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(2), 45-65. doi:10.1093/applin/aml048
- Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 175-184). New York, NY: Routledge.
- Webb, S. (2019). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225-239). London, UK: Routledge.
- Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*, 43(1), 55-77. doi:10.5054/tj.2010.215611
- Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics*, 32(3), 259-276. doi:10.1093/applin/amq051
- Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford, UK: Oxford University Press.
- Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, 63(1), 91-120. doi:10.1111/j.1467-9922.2012.00729.x
- Webb, S., & Rodgers, M. P. H. (2009a). Vocabulary demands of television programs. *Language Learning*, 59(2), 335-366. doi:10.1111/j.1467-9922.2009.00509.x
- Webb, S., & Rodgers, M. P. H. (2009b). The lexical coverage of movies. *Applied Linguistics*, 30(3), 407-427. doi:10.1093/applin/amp010
- Weir, C. J. (1993). *Understanding and developing language tests*. New York, NY: Prentice Hall.
- Winke, P. (2013). The effects of input enhancement on grammar learning and comprehension. A modified replication of Lee (2007) with eye-movement data. *Studies in Second Language Acquisition*, 35(2), 323-352. doi:10.1017/S0272263112000903
- 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.
- Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *Modern Language Journal*, 97(1), 254-275. doi:10.1111/j.1540-4781.2013.01432.x

- Wong, W. (2003). Textual enhancement and simplified input: effects on L2 comprehension and acquisition of non-meaningful grammatical form. *Applied Language Learning*, 13(2), 17-46.
- Wray, A. (2002). *Formulaic language and the lexicon*. New York, NY: Cambridge University Press.
- Wulff, S. (2018). Acquisition of formulaic language from a usage-based perspective. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 19-37). New York, NY: Routledge.
- Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of Vocabulary Levels Tests. *System*, 35(2), 182-191. doi:10.1016/j.system.2006.12.009

## APPENDICES

### *Appendix 1: Participant profile*

<b>Condition</b>	<b>Code</b>	<b>Gender</b>	<b>Age</b>	<b>VST score (out of 140)</b>	<b>First language</b>
Normal captions X 1	ONC001	F	19	62	Mandarin
Normal captions X 1	ONC002	F	19	55	Mandarin
Normal captions X 1	ONC003	M	18	108	Mandarin
Normal captions X 1	ONC004	F	19	65	Mandarin
Normal captions X 1	ONC005	F	18	67	Mandarin
Normal captions X 1	ONC006	F	20	75	Mandarin
Normal captions X 1	ONC007	M	18	100	Malay
Normal captions X 1	ONC008	M	18	68	Mandarin
Normal captions X 1	ONC009	F	20	74	Mandarin
Normal captions X 1	ONC010	F	19	69	Mandarin
Normal captions X 1	ONC011	F	19	77	Mandarin
Normal captions X 1	ONC012	F	18	49	Malay
Normal captions X 1	ONC013	M	21	38	Mandarin
Normal captions X 1	ONC014	F	21	82	Malay
Normal captions X 1	ONC015	M	21	61	Malay
Normal captions X 1	ONC016	M	19	60	Mandarin
Normal captions X 1	ONC018	M	18	102	Mandarin
Normal captions X 1	ONC019	M	19	98	Mandarin
Normal captions X 1	ONC021	F	23	49	Mandarin
Normal captions X 1	ONC022	M	19	78	Mandarin

Normal captions X 1	ONC023*	M	19	89	Malay
Normal captions X 1	ONC024	M	19	82	Mandarin
Normal captions X 1	ONC027	M	20	66	Mandarin
Normal captions X 1	ONC028	M	20	95	Mandarin
Uncaptioned X 1	OUC001	M	18	76	Mandarin
Uncaptioned X 1	OUC004	M	17	65	Mandarin
Uncaptioned X 1	OUC005	F	18	62	Mandarin
Uncaptioned X 1	OUC006	M	18	86	Malay
Uncaptioned X 1	OUC007	F	18	83	Malay
Uncaptioned X 1	OUC008	M	17	112	Mandarin
Uncaptioned X 1	OUC009	M	18	53	Mandarin
Uncaptioned X 1	OUC010	M	20	75	Mandarin
Uncaptioned X 1	OUC011	M	19	61	Mandarin
Uncaptioned X 1	OUC012	M	20	90	Mandarin
Uncaptioned X 1	OUC013	M	18	94	Mandarin
Uncaptioned X 1	OUC014	M	19	77	Mandarin
Uncaptioned X 1	OUC015	M	18	69	Mandarin
Uncaptioned X 1	OUC016	f	18	51	Mandarin
Uncaptioned X 1	OUC017	F	18	62	Mandarin
Enhanced captions X 1	OEC001	F	19	51	Mandarin
Enhanced captions X 1	OEC003	M	20	71	Mandarin
Enhanced captions X 1	OEC004	F	19	70	Mandarin
Enhanced captions X 1	OEC005	F	20	84	Mandarin
Enhanced captions X 1	OEC006	F	19	74	Mandarin

Enhanced captions X 1	OEC007	M	19	67	Mandarin
Enhanced captions X 1	OEC009	F	19	58	Mandarin
Enhanced captions X 1	OEC010	M	20	56	Mandarin
Enhanced captions X 1	OEC011	F	20	70	Mandarin
Enhanced captions X 1	OEC012	F	20	103	Mandarin
Enhanced captions X 1	OEC013	F	20	58	Mandarin
Enhanced captions X 1	OEC014	F	19	74	Mandarin
Enhanced captions X 1	OEC015	F	21	64	Malay
Enhanced captions X 1	OEC016	F	19	70	Mandarin
Enhanced captions X 1	OEC017	F	19	63	Mandarin
Enhanced captions X 1	OEC018	F	19	63	Mandarin
Enhanced captions X 1	OEC019	F	20	70	Mandarin
Enhanced captions X 1	OEC020	F	19	63	Mandarin
Enhanced captions X 1	OEC021	F	20	70	Mandarin
Enhanced captions X 1	OEC022*	F	20	80	Mandarin
Enhanced captions X 2	TEC001	M	18	73	Mandarin
Enhanced captions X 2	TEC002	F	18	53	Mandarin
Enhanced captions X 2	TEC003	F	19	80	Mandarin
Enhanced captions X 2	TEC004	F	18	94	Mandarin
Enhanced captions X 2	TEC005	F	18	80	Mandarin
Enhanced captions X 2	TEC006	M	18	41	Mandarin
Enhanced captions X 2	TEC007	M	18	77	Malay
Enhanced captions X 2	TEC008	M	18	42	Mandarin
Enhanced captions X 2	TEC009	M	18	70	Mandarin



Enhanced captions X 2	TEC010	F	18	68	Mandarin
Enhanced captions X 2	TEC011	F	19	75	Mandarin
Enhanced captions X 2	TEC012	F	19	77	Mandarin
Enhanced captions X 2	TEC013	F	18	44	Mandarin
Enhanced captions X 2	TEC014	M	20	87	Mandarin
Enhanced captions X 2	TEC015	M	18	68	Mandarin
Enhanced captions X 2	TEC016	M	18	62	Mandarin
Enhanced captions X 2	TEC017	F	18	70	Mandarin
Enhanced captions X 2	TEC018	F	18	75	Mandarin
Enhanced captions X 2	TEC019	F	19	101	Mandarin
Enhanced captions X 2	TEC020	F	18	51	Mandarin
Enhanced captions X 2	TEC021	F	18	42	Mandarin
Enhanced captions X 2	TEC022	F	17	61	Mandarin
Enhanced captions X 2	TEC023	F	18	44	Mandarin
Enhanced captions X 2	TEC024	M	22	61	Mandarin
Enhanced captions X 2	TEC025	F	18	52	Mandarin
Uncaptioned X 2	TUC001	F	18	70	Mandarin
Uncaptioned X 2	TUC002	M	19	53	Mandarin
Uncaptioned X 2	TUC003	F	18	94	Mandarin
Uncaptioned X 2	TUC004	F	18	65	Mandarin
Uncaptioned X 2	TUC005	M	18	84	Mandarin
Uncaptioned X 2	TUC006	M	20	78	Mandarin
Uncaptioned X 2	TUC007	F	18	65	Mandarin
Uncaptioned X 2	TUC008	F	18	72	Mandarin

Uncaptioned X 2	TUC009	M	20	69	Mandarin
Uncaptioned X 2	TUC010*	F	18	60	Mandarin
Uncaptioned X 2	TUC012	M	18	55	Mandarin
Uncaptioned X 2	TUC013	F	18	91	Mandarin
Uncaptioned X 2	TUC014	F	18	49	Mandarin
Uncaptioned X 2	TUC015	F	18	50	Mandarin
Uncaptioned X 2	TUC016	M	18	68	Mandarin
Uncaptioned X 2	TUC017	F	19	63	Mandarin
Uncaptioned X 2	TUC018	F	18	61	Mandarin
Uncaptioned X 2	TUC019	M	18	43	Mandarin
Uncaptioned X 2	TUC020	F	18	65	Mandarin
Normal captions X 2	TNC001	F	18	75	Malay
Normal captions X 2	TNC002	F	19	73	Mandarin
Normal captions X 2	TNC003	M	18	81	Mandarin
Normal captions X 2	TNC004	F	18	60	Mandarin
Normal captions X 2	TNC005	F	18	83	Mandarin
Normal captions X 2	TNC006	F	18	82	Mandarin
Normal captions X 2	TNC007	F	18	76	Mandarin
Normal captions X 2	TNC008	F	18	78	Mandarin
Normal captions X 2	TNC009	F	19	82	Mandarin
Normal captions X 2	TNC010	F	18	64	Mandarin
Normal captions X 2	TNC011	F	18	71	Mandarin
Normal captions X 2	TNC012	F	19	62	Mandarin
Normal captions X 2	TNC013	F	18	57	Mandarin

Normal captions X 2	TNC014	M	19	60	Mandarin
Normal captions X 2	TNC015	F	18	67	Mandarin
Normal captions X 2	TNC016	M	18	79	Mandarin
Normal captions X 2	TNC018	F	18	44	Mandarin
Normal captions X 2	TNC019	F	18	87	Mandarin
Normal captions X 2	TNC020	F	18	53	Mandarin
Normal captions X 2	TNC021	F	18	81	Mandarin
Normal captions X 2	TNC022	F	20	67	Mandarin
Normal captions X 2	TNC023	M	19	82	Mandarin
Normal captions X 2	TNC024*	F	19	50	Mandarin

*Note: M = male; F = female; \* refers to participants whose data were not included in Study 1 (due to absence of pretest)*

## *Appendix 2: Information sheet*



### **Research Project: The effects of captioning on the acquisition of multiword units INFORMATION SHEET FOR PARTICIPANTS**

Thank you for your interest in this project. Please read this information before deciding whether or not to take part. If you decide to participate, thank you. If you decide not to take part, thank you for considering my request.

#### **Who am I?**

My name is Elvenna Majuddin and I am a Doctoral student in School of Linguistics and Applied Language Studies at Victoria University of Wellington. This research project is work towards my Doctoral thesis.

#### **What is the aim of the project?**

This project looks at whether captions, or on-screen words in the same language as the audio, can help learners pick up multiword units such as idioms (e.g. *kick the bucket*) and collocations (e.g. *strong tea*). This project also aims to find out if enhanced captions, such as the bolding and underlining of words (e.g. **kick the bucket**), and multiple viewings (watching the videos twice and three times) can further help the learners remember the multiword units and the content of the video. This research has been approved by the Victoria University of Wellington Human Ethics Committee (0000024621).

#### **How can you help?**

If you agree to take part you will first take two diagnostic vocabulary tests. Each of the vocabulary tests will take about 30 minutes. After two weeks have passed, you will start Trial 1 of the study. First, you will take four multiword unit tests. These tests will take about 45 minutes. You will then watch a video that is about 15 minutes long. Next, you will take a comprehension test. This takes about 20 minutes. After that, you will take the four multiword unit tests again. Two weeks later, the same multiword unit tests will be carried out again.

Trial 2 will be conducted two weeks after that. The second trial is similar to the first trial. This means that you will watch a different video, but still take the four multiword unit tests three times: (i) two weeks before watching the video, (ii) immediately after watching the video, and (iii) two weeks after watching the video. You will also take a comprehension test after watching the video.

In summary, if you decide to participate, you will need to be present for two trials that will take place throughout Semester 2, 2017/2018. Trial 1 will take place in Week 1, Week 3, Week 5 and Week 7;

while Trial 2 will take place in Week 10, Week 12 and Week 14. All these will be conducted outside your normal class time, in a designated classroom in Tunku Abdul Rahman University College, Kota Kinabalu.

Participating in this research project is voluntary. It has no impact on your classes and the results do not affect your course grades in any way. You can withdraw from the study by contacting me any time before the 1<sup>st</sup> of December 2017. If you withdraw, the information you provided will be destroyed or returned to you.

If you decide to participate in this research project, you will receive a small gift as a token of appreciation. Additionally, you will receive your diagnostic vocabulary test results and practical advice on vocabulary learning. If it is indicated in the consent form that you would like to receive the results of the study, a summary will be sent to you via e-mail at the end of the research project.

#### **What will happen to the information you give?**

This research is confidential. This means that the researchers named below will be aware of your identity but the research data will be aggregated and your identity will not be disclosed in any reports, presentations, or public documentation.

Only my supervisors and I will know the results of your tests. The test papers will be kept securely and destroyed 5 years after the research ends.

#### **What will the project produce?**

The information from my research will be used in my PhD dissertation. The dissertation will be submitted for marking to the School of Linguistics and Applied Language Studies, Victoria University of Wellington, and then deposited in the University Library. Articles might be submitted for publications in scholarly journals. The results of the study might also be presented in academic conferences.

#### **If you accept this invitation, what are your rights as a research participant?**

You do not have to accept this invitation if you don't want to. If you do decide to participate, you have the right to:

- choose not to answer any question;
- withdraw from the study before the 1<sup>st</sup> of December 2017;
- ask any questions about the study at any time;
- be able to read any reports of this research by emailing the researcher to request a copy.

#### **If you have any questions or problems, who can you contact?**

If you have any questions, either now or in the future, please feel free to contact either:

##### **Student:**

Name: Elvenna Majuddin  
University email address:

##### **Supervisor:**

Name: Anna Siyanova  
Role: Primary Supervisor

elvenna.majuddin@vuw.ac.nz

School: Linguistics and Applied Language

Phone: +64-4-463 5922

University email address:

anna.siyanova@vuw.ac.nz

#### **Human Ethics Committee information**

If you have any concerns about the ethical conduct of the research you may contact the Victoria University HEC Convener: Associate Professor Susan Corbett. Email [susan.corbett@vuw.ac.nz](mailto:susan.corbett@vuw.ac.nz) or telephone +64-4-463 5480.

*Appendix 3: Consent form*



SCHOOL OF LINGUISTICS & APPLIED LANGUAGE STUDIES Te Kōwhiri Reo  
Incorporating DEAF STUDIES RESEARCH UNIT ENGLISH LANGUAGE INSTITUTE NEW ZEALAND DICTIONARY CENTRE  
VICTORIA UNIVERSITY OF WELLINGTON, PO Box 600, Wellington 6140, New Zealand  
Phone +64-4-463 5900 Fax +64-4-463 5904 Email [ls@vuw.ac.nz](mailto:ls@vuw.ac.nz) Web [www.victoria.ac.nz/ls](http://www.victoria.ac.nz/ls)

Consent Form

**Consent to participate in:**

Research project: *The effects of captioning on the acquisition of multiword units.*

Researcher: *Elvenna Majumdar, School of School of Linguistics and Applied Language Studies, Victoria University of Wellington.*

**Participant's agreement:**

I have read and understood the participant information sheet. I have had the opportunity to ask questions about the study, and had them answered satisfactorily. I understand that I may withdraw from the project (or any information I have provided) before 1<sup>st</sup> of December, 2017 without having to give reasons.

I understand that any information I provide will be kept confidential to the researcher and supervisors. I understand the any published results will not use my name, and that no opinions will be attributed to me in any way that can identify me.

I understand that the materials will be destroyed at the end of the research project unless I indicated that I would like them to be returned to me. I understand that if I volunteer for the follow-up interview, the tape recordings of our interaction will be wiped out after the research project has been completed, unless I indicated that I would like them to be returned to me. I also understand that my participation in this project has no impact on my classes and the results do not affect my course grades in any way.

Name: ..... [redacted] .....

Signature: ..... [redacted] .....

Date: 6/11/17

E-mail (if you would like to receive a summary of the findings):

..... [redacted] .....

***Appendix 4: Form recall pretest (Study 1 – The incidental learning condition)***

Name: \_\_\_\_\_

Programme: \_\_\_\_\_

In each of the following question, there is one phrase with missing letters. Look at the context and fill in the blanks with the missing letters.

1. Parents should be on the s\_\_\_\_\_ p\_\_\_\_\_ about raising their children. Parents should have a similar understanding about what to expect from their children.
  
2. He has been in business for five years, but has not yet recovered his cost or tu\_\_\_\_\_ a pr\_\_\_\_\_.
  
3. If the police does not control the crowd, they will have a violent disturbance on their ha\_\_\_\_\_.
  
4. You'll be out on the st\_\_\_\_\_ if the rent isn't paid.
  
5. Please ke\_\_\_\_\_ an e\_\_\_\_\_ on your brother. You need to watch him so he does not go missing.
  
6. Adam is on cl\_\_\_\_\_ ni\_\_\_\_\_ after the birth of his son. The new member of family brings him so much joy.
  
7. My first priority is to ti\_\_\_\_\_ up on discipline. More rules need to be introduced in this school.
  
8. While most coaches can be tough, ours k\_\_\_\_\_ his players with k\_\_\_\_\_. Our coach gets what he wants by being extremely nice to his players.
  
9. I'm sure you will win. Everyone is r\_\_\_\_\_ for you, they really want you to win.
  
10. I'd like to help you but you missed the deadline. I'm afraid my h\_\_\_\_\_ are t\_\_\_\_\_. You should have completed your assignment before 1<sup>st</sup> of August.



11. When the mother found an empty bottle of beer in her teenage son's bedroom, she warned him that he was on a sl\_\_\_\_\_ sl\_\_\_\_\_ towards alcoholism.
12. We couldn't f\_\_\_\_\_ out where all the money had gone.
13. Sarah p\_\_\_\_\_ the boss on the sp\_\_\_\_\_ by asking him when he was going to give us a pay raise.
14. Jack won't listen to me, but somebody needs to t\_\_\_\_\_ some s\_\_\_\_\_ into him before he gets hurt.
15. John: I think Max is sure to win.  
Amy: I b\_\_\_\_\_ to d\_\_\_\_\_. I don't think he has a chance.
16. I'm feeling a bit un\_\_\_\_\_ the we\_\_\_\_\_. I think I'm getting a cold.
17. Two days after they were married, Aaron and Lauren were wh\_\_\_\_\_ away to Sweden for their snow-filled honeymoon.
18. Each of us ch\_\_\_\_\_ in \$50 to take our parents out to dinner.
19. If you just b\_\_\_\_\_ with me for a few more minutes, we will have all the paperwork finished.
20. My best friend moved away five years ago, but we still k\_\_\_\_\_ in t\_\_\_\_\_ through Facebook.
21. Andy's g\_\_\_\_\_ through a ro\_\_\_\_\_ pa\_\_\_\_\_ at the moment because his wife wants a divorce.
22. Though Jacob complained bitterly, his mother would not l\_\_\_\_\_ up on him. She refused to put less pressure on him until his grades improved.

23. My older brother always l\_\_\_\_\_ out for me when we were kids. He was my constant protector.

24. Danny: Which dress do you like best?

Nina : I don't know, I can't m\_\_\_\_\_ up my m\_\_\_\_\_.

25. I think I can w\_\_\_\_\_ s\_\_\_\_\_ out with Karen and find a common ground so that both of us are satisfied.

*Appendix 5: Immediate form recall posttest (Study 1 – The incidental learning condition)*

Name: \_\_\_\_\_

Programme: \_\_\_\_\_

The following is the transcript of the video you just watched. There are 20 phrases that contain words with missing letters. Try and recall what you saw in the video, and fill in the blanks with the missing letters.

Part 1

Eddie: As a kid, there are things that terrify you.  
(Narrating)

Principal: Parents, thank you all for coming. We know you're very busy, but we felt we needed to bring this matter to your immediate attention. This man has been seen in the area handing out drugs to children and telling them they're stickers.

Yes, you in back.

Jessica: When do report cards come out?

Eddie: In my family, that's what terrified us.  
(Narrating)

Eddie: My mom was singing at the restaurant because my dad thought it would help business.  
(Narrating)

Louis: Well, happy anniversary. Married 45 years, I can't imagine it.

Jessica: Hey, Nancy! That coupon is expired! You have to check the dates.

Louis: Literally, I can't imagine it

Nancy: How did you see that?

Jessica: I see everything.

Jessica: Okay, just one sprig of parsley per plate, please. It is a garnish, not a salad.

Louis: Okay, thank you, thank you. The voice of an angel, the incredible singing lady.

What are you doing?

Jessica: The restaurant has been open almost a month, and we have been losing money every day. We have (1) to ti \_\_\_\_\_ up around here, or we will never (1)  
(2) tu \_\_\_\_\_ a pr \_\_\_\_\_. (2)

Louis: Look, if you're going to help with the restaurant, we need to get (3) on the s \_\_\_\_\_ p \_\_\_\_\_, and my page says the key to success is to (4) k \_\_\_\_\_ (3)  
them with k \_\_\_\_\_. I've always said.. (4)

Jessica: Shh..Did you hear that? Mitch, did you just take a crouton from the salad bar?

Mitch:            \*shakes head\*

Jessica:           Are you saying I did not hear a crunch? You are trying to dissolve the crouton.

Louis:             Mitch, why don't you go see if any of the servers need help, okay?

Jessica:           \*sees a customer pouring more black pepper\* Ohh, hell no!

## **Part 2**

Eddie:             You waiting for your report card?

Dave:              Yeah, and a birthday card from my dad.

Eddie:             I'm going to change all my C's to B's before my mom comes home.

Dave:              Sweet.

Eddie:             Yeah. I'm not making the same rookie mistake I made last year.

*(flashback to last year)*

Jessica:           Put your gym clothes in the hamper and take out the trash.

Eddie:             Okay, I will. Love you!

Jessica:           You love me? What are you hiding?

Eddie:             There's nothing in my backpack.

Jessica:           Ohhh, I know. \*Discovers hidden report card\* B minus!

*(end of flashback)*

Dave:              Is there anything else for me?

Postman:           Buddy, we talked about this. I would lead with the card. I'm **(5)** r\_\_\_\_\_ for you.           **(5)**

Emery:             Mom!

Evan:              We got report cards!

Jessica:           Ohh, okay. What is this?

Emery:             Our school doesn't give out grades.

Evan:              It fosters unhealthy competition.

Jessica:           Plus/minus rainbow? One leprechaun? Two clouds... that seems bad!

Evan: No, clouds are good. Clouds are rain.

Jessica: What are you, a farmer?

Eddie: Dominoes! Straight A's, moms! The one good thing about moving here is I have no friends and no distractions. That's why I got a... Grandma, what does Fonzie say?

Grandma: A!

### **Part 3**

Jessica: School is too easy! You need to make school more challenging, or else my son will fall behind.

Principal: I'm sorry. There's not much I can do about a straight-A student.

Jessica: Well, is there extra school? Where is the closest CLC?

Eddie: Oh, no.

Principal: What's CLC?

Eddie: Chinese learning center, an after-school program for reading, math, science, and violin.  
*(narrating)*

Principal: I'm sorry. We don't have a Chinese learning center here.

Eddie: Yes!

Principal: But we do have an after school program called animal encounters, where we bring in farm animals for the children to play with. In fact, I have a baby chick in my lap right now.

Jessica: Please put your lap chicken away.

### **Part 4**

Jessica: I don't understand these people. It's like success is not important to them. Nancy, no wasting napkins!

Louis: But the lady at table six asked for it.

Jessica: She ordered a salad. She does not need an extra napkin.

Nancy: My (6) h\_\_\_\_\_ are t\_\_\_\_\_.

Louis: Jessica, we talked about this.

(6)

Jessica: It's a (7) sl\_\_\_\_\_ sl\_\_\_\_\_. First, an employee pops a crouton. Next thing you know, they are wheeling ovens out of the kitchen. (7)

Same with the customer. First, they get extra napkin. Next thing you know, they run out on the bill.

Same with the boys. First, they have baby chickens in their lap. Next thing you know, they are homeless!

Louis: You know, just because they don't offer CLC here doesn't mean you can't provide it for the boys.

Jessica: Well, I'm not a teacher.

Louis: No, but you are very smart. I mean, have you ever met a "wheel of fortune" puzzle you couldn't (8) f\_\_\_\_\_ out? (8)

Jessica: I always guess them.

Louis: Mm-hmm. I mean, you're way smarter than the teachers they had at CLC.

Jessica: Some of them were not impressive.

Louis: Mm-hmm.

Louis: Oh, never mind. You can't teach the boys CLC. I mean, what would I do at the restaurant? You are so valuable here.

Jessica: I really am.

Louis: Yeah. But we do have a crisis (9) on our ha\_\_\_\_\_. Our boys did get straight A's. (9)

## Part 5

Dave: Oh, hey, man. Look what my mom got me for my straight C's. You want to play?

Eddie: Does the yellow man like dumplings?

Dave: *\*looks confused\**

Eddie: We do, we do. Sorry. I (10) p\_\_\_\_\_ you in an awkward sp\_\_\_\_\_. Let me go get my breakaway pants so I can rip them off. (10)

Jessica: We're doing CLC.

Eddie: But there aren't any in Orlando!

Jessica: Well, there is now... homeschool CLC. I will be your teacher.

## Part 6

- Louis: Wooh, rough day at the ranch.
- Jessica: Is Nancy messing up the iced teas again? I told her it is 4 cups of water to 1 tea bag.
- Louis: It is so hard to not have you there. But this is best for the boys. Our sacrifice will be worth it.
- Evan: Mom, Eddie's refusing to use the bathroom pass.
- Eddie: I'm not trying to use a pass to whiz in my own house.
- Jessica: Dinner's almost ready. You finish your CLC?
- Eddie: You know I didn't.
- Jessica: Dinner after CLC.
- Eddie: Dad, mom is crazy. Can't you **(11)** t\_\_\_\_\_ some s\_\_\_\_\_ into her? **(11)**
- Louis: Uhh, well, you know, it's, uhh, tough for all of us. I'm dying without her at the restaurant. Nobody's having a good time. Love you!
- Eddie: Love you? My dad never said that. My family loved each other. We just didn't say it. We showed our love through criticism and micro management, so if you said "love you," you were probably hiding something.  
(narrating)

## Part 7

- Louis: Okay. Here you are.
- Customer: Oh, I'm okay. I don't need any extra napkins. Thank you.
- Louis: I **(12)** b\_\_\_\_\_ to d\_\_\_\_\_. **(12)**
- Nancy: Hey, we have a jukebox?
- Louis: I guess so. Hey, Mitch, go ahead and plug her in.
- Mitch: Really?
- Louis: Happy employees equal happy customers.
- Mitch: Ahh, it's like we've been **(13)** wh\_\_\_\_\_ away to Trinidad. **(13)**
- Eddie: *\*looks through the window\** Ohh, hell no!

## Part 8

- Eddie: Dad doesn't want you at the restaurant!
- Jessica: What? That's not true. Your father is struggling without me, but he is making a sacrifice for your education.
- Eddie: Ask him yourself!
- Louis: Of course I want you here. It is awful without you. I am very good. Eddie's just making up excuses, probably to get out of CLC. I will double his workload. Ohh, no. Got to run. I see Mitch eyeing the croutons. Love you!  
*(on the phone)*
- Eddie: Oh, hell no!

## Part 9

- Louis: Everybody having a good time?
- Jessica: Boys, welcome to your new CLC classroom.
- Louis: Jessica, what brings you to Cattleman's Ranch?
- Jessica: I realize... why not have the boys do CLC here? That way, I can also help you at the restaurant. Unless you don't want me here.
- Louis: Oh, no. Of course I want you here.
- Jessica: Good! Boys, CLC, this booth. If Mitch steals two croutons per hour and a 3-pound box costs \$5, how many days till we are **(14)** on the st\_\_\_\_\_? **(14)**
- Evan: How many croutons per box? What are we paying in rent? How can I solve for "x" when I don't know "z"?
- Jessica: Nancy, what did I tell you about extra napkins? Also, why are we out of tea bags?
- Louis: I don't want you here.
- Jessica: I knew it, as soon as you expressed emotions with words!
- Louis: Why do you micro manage everything?
- Jessica: Why am I the only one who cares that we are losing money? I'm just **(15)** I\_\_\_\_\_ out for our family. **(15)**
- Louis: So am I, but I'm doing it in a nicer way. You should treat people the way you want to be treated.
- Jessica: You think people are inherently good, but they are not.
- Customer: I'm sorry. Could I just get my check, please?



Louis: No, I'm sorry. My son got straight A's, and I told my wife I love her. Please **(16)** b\_\_\_\_\_ with us. **(16)**

We're **(17)** g\_\_\_\_\_ through a ro\_\_\_\_\_ pa\_\_\_\_\_. **(17)**

Jessica: Look at that table. They are about to dine and dash. They've ordered the most expensive items on the menu. They've been there a long time, and they haven't paid. And that one guy is pumping up his sneaker.

Louis: I'll take care of this.

Jessica: How? By showering them with napkins? No, you make them pay, then you kick them out.

Louis: I will handle this my way.  
Hello.

Customer 1: We're not done eating yet.

Louis: I know. Look... I know what you guys are doing here. You're going to dine and dash because you think it's fun and nobody gets hurt. But this is a family business.

Louis: The point is, we're not rich. In fact, we're struggling just to break even. So if you guys order more than you can pay for, we can **(18)** w\_\_\_\_\_ s\_\_\_\_\_ out. **(18)**

Customer 1: Yeah, okay. I... we're sorry. We're just being stupid.

Customer 2: And we can cover it. I got my dad's credit card.

Louis: Thanks, guys. I appreciate it.  
*(to the customer)*

*Louis:* I took care of it. I talked to them the way that I would want to be talked to, and they got it. People are good.  
*(to Jessica)*

Jessica: Are you sure about that?

Louis: *\*discovers that the customers dined and dashed\** I'll be in my office.

Mitch: Excuse me, Mrs. Huang? We were just talking, and we'd like to cover their check.

Jessica: What?

Nancy: The table that ran out... we want to **(19)** ch\_\_\_\_\_ in and pay their bill. **(19)**

Mitch: Yeah, Louis would do it for us. I mean, he's a great boss.

## Part 10

Customer 1: You hit us with your car.

Jessica: You hit my car with your bodies. My husband is a good man. He believes in the good of people. I don't. But all his employees respect and admire him, which makes you start to wonder.

Customer 2: My body feels cold.

Jessica: Ohh, it's shutting down. My point is, I don't want my husband to lose his faith in people.

Customer 1: Okay. We'll pay our check.

Jessica: You are going to do a lot more than that.

*(back at the restaurant)*

Customer 1: After we left, we thought about what you said, and we realized you were right.

Customer 2: Yeah, what we did was wrong.

Louis: Really?

Customer 3: Yes. You were right to see the good in us.

Louis: Did you have something to do with this?

*(to Jessica)*

Jessica: What? Me? No. Okay, well, I'm going to go make dinner now... sliced beef with peppers and onions, maybe some rice, also. Love you!

## Part 11

Eddie: I missed all the fun.

Louis: Well, you know, most moms don't care enough about their kids to tutor them for two hours a day.

Eddie: It was three hours.

Louis: Three hours?! My God! Look, I'm not going to lie. Your mom is tough. And she's never going to **(20)** \_\_\_\_\_ up on you or any of us, but it's because she cares.

**(20)**

Eddie: Yeah.

Louis: You think you can beat your old man?

Eddie: Dad, you serve people all day. You ready to get served?

Emery: Mom said we can play.

Evan: Nobody block my shots.

Emery: Come on, come on. Eddie, pass.

Eddie: That's how it was in my family. We didn't do sappy "I love you's." We didn't have to.  
(*narrating*) We just showed it.

Jessica: Looks like we'll be going for academic scholarships.

*Appendix 6: Delayed form recall posttest (Study 1 – The incidental learning condition)*

Name: \_\_\_\_\_

Programme: \_\_\_\_\_

**Below are 20 sentences that contain incomplete phrases. You have seen the 20 phrases in the video you watched two weeks ago. Try and recall what you saw in the video, and fill in the missing letters to create a complete phrase.**

**Example:**

**(0) Question: The exam was a pi\_\_\_\_\_ of c\_\_\_\_\_. I did not even have to think hard.**

**Answer : The exam was a piece of cake. I did not even have to think hard.**

1. Though Jacob complained bitterly, his mother would not l\_\_\_\_\_ up on him. She refused to put less pressure on him until his grades improved.
  
2. John: I think Max is sure to win.  
  
Amy: I b\_\_\_\_\_ to d\_\_\_\_\_. I don't think he has a chance.
  
3. I'd like to help you but you missed the deadline. I'm afraid my h\_\_\_\_\_ are t\_\_\_\_\_. You should have completed your assignment before 1<sup>st</sup> of August.
  
4. He has been in business for five years, but has not yet recovered his cost or tu\_\_\_\_\_ a pr\_\_\_\_\_.
  
5. Each of us ch\_\_\_\_\_ in \$50 to take our parents out to dinner.
  
6. I think I can w\_\_\_\_\_ s\_\_\_\_\_ out with Karen and find a common ground so that both of us are satisfied.
  
7. Parents should be on the s\_\_\_\_\_ p\_\_\_\_\_ about raising their children. Parents should have a similar understanding about what to expect from their children.
  
8. We couldn't f\_\_\_\_\_ out where all the money had gone.
  
9. You'll be out on the st\_\_\_\_\_ if the rent isn't paid.

10. If you just b\_\_\_\_\_ with me for a few more minutes, we will have all the paperwork finished.
11. While most coaches can be tough, ours k\_\_\_\_\_ his players with k\_\_\_\_\_. Our coach gets what he wants by being extremely nice to his players.
12. Andy's g\_\_\_\_\_ through a ro\_\_\_\_\_ pa\_\_\_\_\_ at the moment because his wife wants a divorce.
13. If the police does not control the crowd, they will have a violent disturbance on their ha\_\_\_\_\_.
14. Sarah p\_\_\_\_\_ the boss on the sp\_\_\_\_\_ by asking him when he was going to give us a pay raise.
15. When the mother found an empty bottle of beer in her teenage son's bedroom, she warned him that he was on a sl\_\_\_\_\_ sl\_\_\_\_\_ towards alcoholism.
16. My older brother always l\_\_\_\_\_ out for me when we were kids. He was my constant protector.
17. Jack won't listen to me, but somebody needs to t\_\_\_\_\_ some s\_\_\_\_\_ into him before he gets hurt.
18. I'm sure you will win. Everyone is r\_\_\_\_\_ for you, they really want you to win.
19. Two days after they were married, Aaron and Lauren were wh\_\_\_\_\_ away to Sweden for their snow-filled honeymoon.
20. My first priority is to ti\_\_\_\_\_ up on discipline. More rules need to be introduced in this school.

*Appendix 7: Delayed form recognition posttest (Study 1 – The incidental learning condition)*

Name: \_\_\_\_\_

Programme: \_\_\_\_\_

**Below are 20 phrases taken from the video you saw two weeks ago. Please fill out the missing words with the most suitable choice among the four options provided.**

**Example: The exam was a p\_\_\_\_\_ of c\_\_\_\_\_**

- a) *pinch of cake*
- b) *pitch of cake*
- c) *piece of cake*
- d) *pint of cake*

**Answer: The exam was a p\_\_\_\_\_ of c\_\_\_\_\_**

- a) *pinch of cake*
- b) *pitch of cake*
- c) *piece of cake*
- d) *pint of cake*

1. I'd like to help you but you missed the deadline. I'm afraid my h\_\_\_\_\_ are t\_\_\_\_\_.  
You should have completed your assignment before 1st of August.

- a) heels are tied
- b) hands are tied
- c) hands are tired
- d) heels are tired

2. Andy's g\_\_\_\_\_ through a ro\_\_\_\_\_ pa\_\_\_\_\_ at the moment because his wife wants a divorce.

- a) going through a rough path
- b) getting through a road path
- c) getting through a root path
- d) going through a rough patch

3. Sarah p\_\_\_\_\_ the boss on the sp\_\_\_\_\_ by asking him when he was going to give us a pay raise.

- a) put the boss on the spot
- b) placed the boss on the space
- c) pushed the boss on the spot
- d) picked the boss on the spin

4. My first priority is to ti\_\_\_\_\_ up on discipline. More rules need to be introduced in this school.

- a) tick up
- b) tidy up
- c) tighten up
- d) tip up

5. Two days after they were married, Aaron and Lauren were wh\_\_\_\_\_ away to Sweden for their snow-filled honeymoon.
- a) whisked away  
b) whirled away  
c) whiled away  
d) whipped away
6. Though Jacob complained bitterly, his mother would not l\_\_\_\_\_ up on him. She refused to put less pressure on him until his grades improved.
- a) listen up on him  
b) lay up on him  
c) load up on him  
d) let up on him
7. Each of us ch\_\_\_\_\_ in \$50 to take our parents out to dinner.
- a) chucked in  
b) checked in  
c) chipped in  
d) chimed in
8. If the police does not control the crowd, they will have a violent disturbance on their ha\_\_\_\_\_.
- a) on their hats  
b) on their hands  
c) on their hair  
d) on their halls
9. I think I can w\_\_\_\_\_ s\_\_\_\_\_ out with Karen and find a common ground so that both of us are satisfied.
- a) walk somewhere out  
b) wander somewhere out  
c) wear something out  
d) work something out
10. Parents should be on the s\_\_\_\_\_ p\_\_\_\_\_ about raising their children. Parents should have a similar understanding about what to expect from their children.
- a) on the same pace  
b) on the similar path  
c) on the same page  
d) on the smart plan

11. My older brother always l\_\_\_\_\_ out for me when we were kids. He was my constant protector.

- a) lived out for me
- b) looked out for me
- c) listened out for me
- d) leaned out for me

12. He has been in business for five years, but has not yet recovered his cost or tu\_\_\_\_\_ a pr\_\_\_\_\_.

- a) turned a profit
- b) tuned a price
- c) turned a promotion
- d) tuned a project

13. Jack won't listen to me, but somebody needs to t\_\_\_\_\_ some s\_\_\_\_\_ into him before he gets hurt.

- a) tell some suggestions
- b) teach some stuff
- c) talk some sense
- d) tell some stories

14. You'll be out on the st\_\_\_\_\_ if the rent isn't paid.

- a) on the street
- b) on the stool
- c) on the stuff
- d) on the step

15. John: I think Max is sure to win.

Amy: I b\_\_\_\_\_ to d\_\_\_\_\_. I don't think he has a chance.

- a) bet to differ
- b) back to dinner
- c) bet to dismiss
- d) beg to differ

16. If you just b\_\_\_\_\_ with me for a few more minutes, we will have all the paperwork finished.

- a) bear with me
- b) brawl with me
- c) buy with me
- d) break with me



17. We couldn't f\_\_\_\_\_ out where all the money had gone.

- a) fly out
- b) fade out
- c) figure out
- d) fight out

18. When the mother found an empty bottle of beer in her teenage son's bedroom, she warned him that he was on a sl\_\_\_\_\_ sl\_\_\_\_\_ towards alcoholism.

- a) sloppy slide
- b) sleety slope
- c) sleek slab
- d) slippery slope

19. I'm sure you will win. Everyone is r\_\_\_\_\_ for you, they really want you to win.

- a) roaming for you
- b) rooting for you
- c) roaring for you
- d) rocking for you

20. While most coaches can be tough, ours k\_\_\_\_\_ his players with k\_\_\_\_\_. Our coach gets what he wants by being extremely nice to his players.

- a) keeps his players with kindness
- b) knocks his players with kindness
- c) kills his players with kindness
- d) kicks his players with kindness

*Appendix 8: Delayed meaning recall posttest (Study 1 – The incidental learning condition)*

Name: \_\_\_\_\_

Programme: \_\_\_\_\_

**Below are 20 phrases taken from the video you saw two weeks ago. Please write the meaning of the phrases according to the context in the video. You can write as many answers as you can think of.**

**Example:**

**Question: *Take it easy***

**Answer: Relax, be calm, be gentle, do not get excited.**

1. Work something out with someone

\_\_\_\_\_

2. Figure out

\_\_\_\_\_

3. Tighten up

\_\_\_\_\_

4. Talk some sense into someone

\_\_\_\_\_

5. Let up on someone

\_\_\_\_\_

6. Kill someone with kindness

\_\_\_\_\_

7. Someone's hands are tied

\_\_\_\_\_

8. Rooting for someone

\_\_\_\_\_

9. Chip in

\_\_\_\_\_

10. Bear with someone

\_\_\_\_\_

11. Put someone on the spot

---

12. Beg to differ

---

13. Look out for someone

---

14. Going through a rough patch

---

15. Slippery slope

---

16. On the street

---

17. Have something on one's hands

---

18. Whisked away

---

19. On the same page

---

20. Turn a profit

---

*Appendix 9: Delayed meaning recognition posttest (Study 1 – The incidental learning condition)*

Name: \_\_\_\_\_

Programme: \_\_\_\_\_

Below are 20 phrases taken from the video you saw two weeks ago. Choose the most suitable meaning of the phrases, according to the context in the video. You should only choose one answer out of the four options provided.

**Example**

**(0) Question: A piece of cake**

**Answer:**

**(a) very easy**  
**(b) very soft**

**(c) very sweet**  
**(d) very tricky**

1. Have something on one's hands

(a) To have a person or problem that one must deal with

(b) To receive a stolen item

(c) To be responsible for someone's death

(d) To be in the safe care of someone

2. Figure out

(a) To escape from an unpleasant place

(b) To find excuse to not participate in an activity

(c) To price something at a certain amount of money

(d) To think about a problem until a solution is found

3. Tighten up

(a) To be stricter or more serious about something

(b) To make a relationship stronger

(c) To secure the lid or cover

(d) To keep a secret

4. On the street

(a) To march or join a protest

(c) To be homeless

(b) To find something at an unexpected place

(d) To leave someone who needs help

5. Look out for someone

(a) To be careful around someone

(c) To examine someone

(b) To spy on someone

(d) To watch over and care for someone

6. Someone's hands are tied

(a) To be involved in other people's business

(c) To get caught doing something illegal

(b) To not be able to help

(d) To not let go of someone or something

7. Kill someone with kindness

(a) To tell a white lie to a person you love

(c) To get what you want by being very kind to a person

(b) To assist someone to die in the least painful way

(d) To take advantage of a kind person

8. On the same page

(a) To do something in a way that someone else would

(c) To have a similar understanding or way of thinking

(b) To not show any progress

(d) To read something several times

9. Rooting for someone

- |   |  |
|---|--|
| (a) To seek and remove a person who is causing trouble                          | (c) To transfer someone from their hometown          |
| (b) To give encouragement to someone because you want them to achieve something | (d) To teach someone about their culture since young |

10. Going through a rough patch

- |   |   |
|---|---|
| (a) To use the resources that are available | (c) A period of time when one is having a lot of difficulty |
| (b) To drive on a rocky road                | (d) A piece of work that is poorly made                     |

11. Turn a profit

- |  |                                 |
|--|---------------------------------|
| (a) To refuse to accept money            | (c) To go out of business       |
| (b) To earn money after paying for costs | (d) To steal someone's business |

12. Let up on someone

- |                                   |  |
|-----------------------------------|--|
| (a) To tell someone a secret      | (c) To stop hoping that someone will improve     |
| (b) To be disappointed by someone | (d) To reduce the pressure or demands on someone |

13. Slippery slope

- |   |                                      |
|---|--------------------------------------|
| (a) A habit that might develop into something extremely bad | (c) A path that is wet and dangerous |
| (b) A criminal who is difficult to catch                    | (d) A clever and dishonest person    |

14. Put someone on the spot

- (a) To pressure someone to answer a difficult question
- (b) To separate someone from the rest

- (c) To meet someone at the right place and time
- (d) To watch someone perform on the stage

15. Bear with someone

- (a) To suffer because of a mistake made by someone
- (b) To hug someone tightly

- (c) To be patient and wait while someone does something
- (d) To be with an unkind person

16. Whisked away

- (a) To do something too much or in an extreme way
- (b) To walk away from an argument

- (c) To brush something away
- (d) To take someone very quickly from one place to another

17. Talk some sense into someone

- (a) To persuade someone to behave in a reasonable way
- (b) To convince someone to take the blame

- (c) To explain something in great detail
- (d) To encourage someone to try a new activity

18. Beg to differ

- (a) To change one's opinions
- (b) To store something in a different container

- (c) To request for a short break
- (d) To firmly disagree with someone

19. Chip in

- (a) To let something happen regardless of the results
- (b) To contribute money to buy or pay for something

- (c) To break off or break away in small chips
- (d) To have a light snack before the main meal

20. Work something out with someone

- (a) To come to agreement by compromise
- (b) To guide or push something through a barrier

- (c) To find a solution to a mathematical problem
- (d) To prepare something in a very short time



*Appendix 10: Listening comprehension test (Study 1 – The incidental learning condition)*

Name: \_\_\_\_\_

Programme: \_\_\_\_\_

**Part 1**

**The following questions are based on the video you just watched. Circle the correct answer.**

1. Why does Louis want Jessica to sing at the restaurant?
  - A) To help business
  - B) To annoy his employees
  - C) To help her improve her singing
  
2. How does Jessica help the restaurant save money?
  - A) By blocking the pepper shaker
  - B) By charging customers to hear her sing
  - C) By asking her children to work without pay
  
3. What does Eddie plan to do with his report card before his mother comes home?
  - A) Copy his mother's signature
  - B) Hide his report card under his shirt
  - C) Change all the C's to B's
  
4. Why doesn't Emery and Evan's school give out grades?
  - A) Because pictures such as rainbows and clouds are easier to understand
  - B) Because grades create unhealthy competition
  - C) Because children like collecting stickers like clouds and rainbows
  
5. What is the real reason Louis wants Jessica to teach CLC?
  - A) Louis believes that the boys are weak in their studies
  - B) Louis believes that Jessica is smarter than the teachers
  - C) Louis does not want Jessica at the restaurant
  
6. Why does Jessica bring the boys to do CLC at the restaurant?
  - A) Because she wants the boys to see Louis
  - B) Because she wants to teach the boys and help Louis at the same time
  - C) Because she wants to find out if Louis is lying
  
7. How did Jessica know that the three customers were going to dine and dash?
  - A) The customers ordered the most expensive items on the menu
  - B) The customers looked like they did not have money
  - C) The customers did not finish the food they ordered
  
8. Why does Jessica want the dine-and-dash customers to admit they are wrong?
  - A) She does not want the restaurant to lose money
  - B) She does not want Louis to lose his faith in people
  - C) She does not want the customers to repeat the same mistake

9. Why doesn't Jessica plug in the jukebox in the restaurant?
  - A) Because she wants the customers to hear her sing
  - B) Because she doesn't want the employees to be lazy
  - C) Because each song costs the restaurant 2.5 cents
  
10. When Mitch and Nancy offered to cover the dine-and-dash bill, it proves that
  - A) Louis is right to believe in the good of people
  - B) Mitch and Nancy are scared of Jessica
  - C) Mitch and Nancy do not want Louis to be embarrassed
  
11. When Louis told Jessica that Eddie was making excuses to get out of CLC, Jessica wanted to
  - A) Cut Eddie's allowance
  - B) Double Eddie's CLC workload
  - C) Double Eddie's house chores

**Part 2**

**The following statements are based on the video you just watched. If the statement is true, circle the word True. If the statement is false, circle the word False.**

- |  |            |
|--|------------|
| 12. Louis and Jessica have a daughter named Emily.                               | True/False |
| 13. Eddie reveals that Jessica tutors him and his brothers for two hours a day.  | True/False |
| 14. Louis believes that happy employees are good for business.                   | True/False |
| 15. Jessica is tough on the children and employees because she cares.            | True/False |
| 16. Dave received a trampoline for his straight C's.                             | True/False |
| 17. Louis' restaurant is a family business.                                      | True/False |
| 18. When Eddie's family said "I love you", they were probably hiding something.  | True/False |
| 19. Jessica believes that micromanaging is better for business than being nice.  | True/False |
| 20. The after-school programme at Eddie's school allows students to visit farms. | True/False |
| 21. Jessica thinks that success is not as important to other people.             | True/False |
| 22. Dave does not live with his father.  | True/False |

***Appendix 11: Form recall pretest (Study 2 – The intentional learning condition)***

Name: \_\_\_\_\_

Programme: \_\_\_\_\_

In each of the following question, there is one phrase with missing letters. Look at the context and fill in the blanks with the missing letters.

1. He is a role model for other players to l\_\_\_\_\_ up to.
  
2. Jude: Do you want to go swimming with us tomorrow?  
  
Jane: Yes, c\_\_\_\_\_ me in.
  
3. I can wh\_\_\_\_\_ up a meal in no time.
  
4. His jokes are funny, but sometimes he g\_\_\_\_\_ too f\_\_\_\_\_ and hurts people's feelings.
  
5. Teachers are the uns\_\_\_\_\_ he\_\_\_\_\_ of a great writer's success. They are often not noticed or praised for their hard work.
  
6. The exam was a pi\_\_\_\_\_ of c\_\_\_\_\_. I did not even have to think hard.
  
7. We are only in Paris for a day, so let's m\_\_\_\_\_ the m\_\_\_\_\_ of it. We have to enjoy our day as much as possible.
  
8. We have arranged a meeting for next Thursday, so if you see anyone, do sp\_\_\_\_\_ the w\_\_\_\_\_. We have to inform everyone.
  
9. When the score got to 8-2, we knew the game was in the b\_\_\_\_\_. We knew we were going to be the champion.
  
10. Let's just wait until the d\_\_\_\_\_ has s\_\_\_\_\_ before we decide what to do. It's better to make a decision when the situation has calmed down.

11. I used to spend a lot of time in London, but now I only go there on \_\_\_\_\_ in a bl \_\_\_\_\_ m \_\_\_\_\_.
12. Someone has been sc \_\_\_\_\_ with my computer, and now it doesn't work anymore.
13. The gap between the rich and the poor is wide, and the poor aren't going to t \_\_\_\_\_ it ly \_\_\_\_\_ d \_\_\_\_\_. They are going to be more violent.
14. After the recent bombing attacks, airports in major cities around the world have installed more cameras to st \_\_\_\_\_ up security.
15. I don't have anything against advertising, but I do have a b \_\_\_\_\_ with how many bad advertisements there are on TV.
16. We i \_\_\_\_\_ a lot in Tom, so we have every right to expect a lot from him. We devoted a lot of time and effort in training him to be a professional athlete.
17. Jeff passed his test with fl \_\_\_\_\_ co \_\_\_\_\_. His parents were very proud of his achievement.
18. He had a r \_\_\_\_\_ in with his boss. The argument caused him to lose his job.
19. My brother came home drunk so I r \_\_\_\_\_ him out to my mother. I told my mother that he had been sneaking out at night.
20. You have so much energy on your television shows, I always suspect you must be ho \_\_\_\_\_ up on energy drinks.
21. I was so c \_\_\_\_\_ up in my school work that I didn't realise what was happening with my sister.
22. Seeing my best friend again after a long time ma \_\_\_\_\_ my d \_\_\_\_\_. Our reunion made me really happy.

23. The cooking competition is designed to give home cooks their d\_\_\_\_\_ in the s\_\_\_\_\_.  
The home cooks will finally get the attention they deserve.
24. The candidate pl\_\_\_\_\_ the race c\_\_\_\_\_, claiming that she received less attention than  
the Malay candidate simply because she is Chinese.
25. After failing to convince the IT department that new security passwords are needed,  
Mike felt it was time to br\_\_\_\_\_ out the b\_\_\_\_\_ g\_\_\_\_\_. So he called a meeting with  
the Head of the Company.
26. The new Iphone model is confirmed to h\_\_\_\_\_ the str\_\_\_\_\_ at the end of 2017.
27. Last week's rain was sm\_\_\_\_\_ pot\_\_\_\_\_ compared to the thunder storm we had two  
months ago.

*Appendix 12: Immediate form recall posttest (Study 2 – The intentional learning condition)*

Name: \_\_\_\_\_

Programme: \_\_\_\_\_

The following is the transcript of the video you just watched. There are **23 phrases** that contain words with missing letters. Try and recall what you saw in the video, and fill in the blanks with the missing letters.

**Part 1**

Reporter: Well, it's that time of year again. The town of Natesville, famous for its rich soil that for reasons unknown produces the plumpest radishes in the world, is celebrating its annual Thanksgiving Radish Harvest Festival.

Not only is the festival an opportunity for Natesville residents to show off their radishes, it's also an opportunity to show off the radish's many uses as a food, a garnish and as a projectile.

The festival also offers small business owners who are afraid of the Internet the opportunity to advertise.

Rosa: Knock Knock Knock Housekeeping.

Virginia: Call the gals in pink to remove your stink.

Reporter: This week long event also gives some of the city's **(1)** uns\_\_\_\_\_ and unpaid **(1)** he\_\_\_\_\_ their **(2)** d\_\_\_\_\_ in the s\_\_\_\_\_. **(2)**

Barney: Now, as always, it's you kids of Natesville that raise the money for this wonderful festival. The child who sells the most candy gets to be the little pilgrim in the 'We're Thankful for the Radishes Thanksgiving Day' float.

Jimmy: Yeah, I was the little pilgrim once.

Sabrina: Are you kidding me? If you would have **(3)** pl\_\_\_\_\_ the little pilgrim **(3)** c\_\_\_\_\_ when we first met, you would've gotten in my pants a lot sooner.

Jimmy: It did get me a couple of kisses on the playground. It was the first time in my life that I ever achieved anything. I felt special. Everyone was **(4)** l\_\_\_\_\_ up to **(4)** me, and I **(5)** m\_\_\_\_\_ the m\_\_\_\_\_ of it. **(5)**

Jimmy: Now I want Hope to feel the way I felt that day. It could be just the boost her self-esteem needs to finally get her potty trained.

Sabrina: Oh, that would be good. I think she's starting to understand that squatting behind the couch and grunting is losing its charm.

**Part 2**

Jimmy: How many bars do you think we need to sell so Hope can be the little pilgrim and ride on the radish float?

Virginia: I don't know, and if I were you, I wouldn't get too **(6)** i \_\_\_\_\_ in her winning. **(6)**

Sabrina: I hear you guys are candy-selling pros. You have any tips for us?  
Virginia: Absolutely not, we cannot get involved with this.

Jimmy: What's the big deal? Why?

Burt: Don't ask questions.

Maw Maw: Just tell the boy! It's better coming from you.

Virginia: Fine, but I'm warning you, this family did things we are ashamed of. Things you're going to wish you didn't know about.

Virginia: It all happened 20 years ago at the Radish Festival. I had just started working at Knock Knock Knock Housekeeping and Jimmy just started his life-long love affair with fire and explosives.

*(flashback to 20 years ago)*

Barney: Ladies and gentlemen, please excuse my sweating. It must be at least 70 degrees out here. Now, remember, we're going to need every child to participate in the candy drive. So what do you say, parents? Come and get this candy!

Rosa: Look at them. You know those crazy parents will be selling the candy for their kids.

Virginia: Which will teach the kids nothing. I guarantee you, 20 years from now, those brats will still be living at home.

Virginia: Rosa and I were speaking the same language, even though she had a funny accent. We both agreed that we wouldn't be the kind of parents who do everything for their kids.  
*(narrating)*

Young Jimmy: It's hot. Can I get in?

Burt: Have you sold any candy?

Young Jimmy: Not yet.

Burt: Sorry, kid. No cash, no splash. Kiddie pool's for closers.

Young Jimmy: Oh, come on, can't you help me? I want to be the little pilgrim.

Virginia: Sorry, hon. We're doing this for your own good. Work builds character. Ooh, I'm going to see if Carlos will give us some free Astro pops.

Burt: **(7)** C \_\_\_\_\_ me in. I want one. **(7)**

Young Jimmy: Me, too.

Burt: Sorry kid, Astro Pops are for closers.

Virginia: Hey! I thought we said we weren't going to help them sell chocolate!

Rosa: Andele, andele, andele! (*Quick quick quick!*)

Virginia: Jimmy, remember how we said we were finally going to take you to the movies tonight to see Aladdin?

Young Jimmy: Yeah!

Virginia: Well, that's off. We're going to war.

### Part 3

Young Jimmy: You guys suck at this.

Virginia: We tried. We went to every house in the neighbourhood, but no one even answered the door.

Maw Maw: The problem isn't you, Virginia. It's those crummy candy bars. Just because they're brown and filled with nuts doesn't mean they need to taste like crap. Don't worry. Maw Maw's got an idea.

Virginia: It was time for us to **(8)** br\_\_\_\_\_ out the b\_\_\_\_\_ g\_\_\_\_\_. So Maw Maw **(8)**  
*(narrating)* started **(9)** wh\_\_\_\_\_ up a batch of her own special chocolate. Of course, she **(9)**  
 only used the finest of ingredients.

Virginia: She had created the most delicious and addictive food known to man. We called **(10)**  
*(narrating)* it Maw Maw's Magic Brown. Once people got a taste, **(10)** w\_\_\_\_\_ **(10)**  
 sp\_\_\_\_\_. Everyone was hooked. And that started a major turf war.

Virginia: They hit us, but we hit them back harder. When the **(11)** d\_\_\_\_\_ had **(11)**  
*(narrating)* s\_\_\_\_\_ and the candy was gone, it was time to count up the earnings and see  
 who was going to ride that float.

Barney: Looks like Rosa's the winner.

Virginia: Hold on. I've got this contest **(12)** in the b\_\_\_\_\_. **(12)**

Virginia: Seeing Jimmy up on that float made me feel proud that we were able to win the  
*(narrating)* contest for him. But then I realized that even though Jimmy was happy, we made  
 another child very sad, and all of a sudden, I didn't feel so good.

Virginia: At that moment, Rosa and I both realized there are no winners in a candy war.  
 And there's been peace between the families ever since. So if you want Hope to  
 ride on that float, you're going to have to do it without our help.



#### Part 4

- Jimmy: I thought you said "cookie". It sounded like you said "cookie".
- Sabrina: Why would I say, "If we hurry, we can have a cookie in the shower"? Doesn't even make sense.
- Jimmy: That's why I said no.
- Jimmy: Rosa?! (*sees Rosa and Carlos selling candy*)
- Jimmy: Rosa's back in the candy game! She and Carlos are selling chocolate bars out of their ice cream truck.
- Virginia: That dirty, lying, chocolate slinging ho!
- Maw Maw: Fire up the Bunsen burner. We're going to drop some cocoa madness on those chumps.
- Jimmy:  
(*narrating*) Once we learned Rosa was back to dealing candy bars, we knew our only chance at getting Hope to win the contest was to pull Maw Maw out of retirement. When Maw Maw's recipe **(13)** h\_\_\_\_\_ the str\_\_\_\_\_, the Chance family was back on top. **(13)**
- Jimmy:  
(*narrating*) When Rosa's family knew we were back in business, things got rough.
- Rosa: (*after destroying the Chance's chocolate 'lab'*)  
Let this be a lesson. Don't **(14)** sc\_\_\_\_\_ with the Flores family. **(14)**
- Jimmy:  
(*narrating*) And we weren't going to **(15)** t\_\_\_\_\_ that ly\_\_\_\_\_ d\_\_\_\_\_. **(15)**  
(*Virginia, Burt & Jimmy play the violin badly*)
- Jimmy:  
(*narrating*) As sticky as things were getting on the streets, things were starting to melt on the inside, too. While we were fighting mayhem at home, the Flores cartel decided to hit us where it really hurt.
- Burt: (*upon discovering the horse head in his bed*)  
Ahhhhhhh!!!
- Jimmy: Maw Maw **(16)** st\_\_\_\_\_ things up by making the chocolate even stronger. **(16)**  
Nobody could resist our latest strain.

#### Part 5

- Frank: Come on, man, just give me some chocolate. A chunk, a section, a little bitty square.

Jimmy: I can't do it, Frank! Barney said I can't sell in the store.

Frank: Barney also said, if I just wander around doing nothing all day, he'll fire me. That was ten years ago. Give me some. I know you're holding.

Jimmy: Hey, get out of there. Stop it! Get out!

Barney: What do we have here? That's Maw Maw's Magic Brown. It's back on the street? James, my office now!

Barney: You have been detained by a Natesville auxiliary police officer. You have the right to remain silent. You also have the right to leave at any point because said auxiliary officer has no legal authority to detain you.

Jimmy: Wait, so I can just leave?

Barney: Not if you ever want your daughter to ride on top of the "We're Thankful for Radishes" float.

Jimmy: Oh, Barney, come on. Don't make Hope suffer just because you **(17)** have a b\_\_\_\_\_ with me **(17)**

Barney: Oh, this has nothing to do with you. You're **(18)** sm\_\_\_\_\_ pot\_\_\_\_\_. I want the russet. **(18)**

Jimmy: The what?

Barney: The russet. It's the king of potatoes. How do you work in produce and not know that?

Jimmy: Fine. I'll go get you a potato.

Barney: I don't want a potato. I want your mother. 20 years ago, your mother and I **(19)** had a r\_\_\_\_\_ in. **(19)**

*(Flashback to 20 years ago)*

Virginia: Hold on. I've got this contest in the bag.

Barney: Wow, this is a lot of money, Virginia. How'd you sell so much candy?

Virginia: Guess I just have a knack for it.

Barney: I find that hard to believe. I was darn sure your mother cheated, and like any good auxiliary police officer, I was going to keep the pressure on until she admitted it.  
*(narrating)*

Barney: You cheated.

Virginia: Prove it.

*(End of flashback)*

Barney: So, James, I guess the radish is in your basket now. How do you want to play it?

Jimmy: Barney, please. Don't make me **(20)** r\_\_\_\_\_ out my own mom. Snitches get stitches. **(20)**

Barney: I'm asking you to do what's best for you and your daughter. You want her to be the little pilgrim, don't you?

Jimmy: What do I have to do? I don't have to beat up my own mom, do I?

Barney: Absolutely not. You are going to wear a wire.

## **Part 6**

Jimmy: Back at the chocolate factory, things were really out of control. **(21)** Ho\_\_\_\_\_ up on chocolate, nobody had slept for days. And everyone was going more than a little crazy **(21)**

Jimmy: How's it going, Mom?

Virginia: Oh, damn it, Jimmy! I was up to four, and now I've got to start all over again.

Jimmy: Yeah, sorry to interrupt you from counting money from our current candy sales, which we have gotten by cheating when we added a secret addictive ingredient to the chocolate bar.

Virginia: Get out of my face!

Jimmy: So I have to push the big button and the little red button?  
*(Tape recording)*

Barney: The little red button is inside the big button.  
*(Tape recording)*

Virginia: He's wearing a wire. He's wearing a wire!

Virginia: You're going to talk, Jimmy. Who gave you the wire?

Jimmy: Ow, you got chocolate in my eye.

Virginia: Tell us where that wire came from. Tell us! We're not playing around.

Jimmy: It was Barney. He told me he'd rig the contest so that Hope could be the little pilgrim if I helped him prove you've been cheating all these years.

Virginia: Why didn't you just tell us that, you idiot? The only reason we've been doing this is to get Hope on that float.

**Part 7**

Virginia: Fine, you're right! I cheated.

Barney: I knew you were dirty.

Virginia: Not any worse than you, rigging this contest just to get me to confess to rigging the contest.

Barney: Hey, no one's innocent. The real little pilgrim of Natesville was a psychotic dwarf who slaughtered an entire Indian village to get a basket of radishes.

Jimmy: Parents can sometimes **(22)** g\_\_\_\_\_ a little too f\_\_\_\_\_ where their kids are **(22)**  
(narrating) concerned. You never know how much help to give them, and the truth is the kids probably don't even care. It's the adults that get **(23)** c\_\_\_\_\_ up in the **(23)**  
competition.

Rosa: I know you cheated.

Virginia: Prove it.

**Appendix 13. Delayed form recall posttest (Study 2 – The intentional learning condition)**

Name: \_\_\_\_\_

Programme: \_\_\_\_\_

**Below are 23 sentences that contain incomplete phrases. You have seen the 23 phrases in the video you watched two weeks ago. Try and recall what you saw in the video, and fill in the missing letters to create a complete phrase.**

**Example:**

**(1) Question: The exam was a pi\_\_\_\_\_ of c\_\_\_\_\_. I did not even have to think hard.**

**Answer : The exam was a piece of cake. I did not even have to think hard.**

1. After failing to convince the IT department that new security passwords are needed, Mike felt it was time to br\_\_\_\_\_ out the b\_\_\_\_\_ g\_\_\_\_\_. So he called a meeting with the Head of the Company.
2. We i\_\_\_\_\_ a lot in Tom, so we have every right to expect a lot from him. We devoted a lot of time and effort in training him to be a professional athlete.
3. Jude: Do you want to go swimming with us tomorrow?  
  
Jane: Yes, c\_\_\_\_\_ me in.
4. We are only in Paris for a day, so let's m\_\_\_\_\_ the m\_\_\_\_\_ of it. We have to enjoy our day as much as possible.
5. I don't have anything against advertising, but I do have a b\_\_\_\_\_ with how many bad advertisements there are on TV.
6. Teachers are the uns\_\_\_\_\_ he\_\_\_\_\_ of a great writer's success. They are often not noticed or praised for their hard work.
7. Last week's rain was sm\_\_\_\_\_ pot\_\_\_\_\_ compared to the thunder storm we had two months ago.
8. Let's just wait until the d\_\_\_\_\_ has s\_\_\_\_\_ before we decide what to do. It's better to make a decision when the situation has calmed down.

9. My brother came home drunk so I r\_\_\_\_\_ him out to my mother. I told my mother that he had been sneaking out at night.
10. The cooking competition is designed to give home cooks their d\_\_\_\_\_ in the s\_\_\_\_\_. The home cooks will finally get the attention they deserve.
11. You have so much energy on your television shows, I always suspect you must be ho\_\_\_\_\_ up on energy drinks.
12. I can wh\_\_\_\_\_ up a meal in no time.
13. He is a role model for other players to l\_\_\_\_\_ up to.
14. He had a r\_\_\_\_\_ in with his boss. The argument caused him to lose his job.
15. Someone has been sc\_\_\_\_\_ with my computer, and now it doesn't work anymore
16. The new Iphone model is confirmed to h\_\_\_\_\_ the str\_\_\_\_\_ at the end of 2017.
17. The gap between the rich and the poor is wide, and the poor aren't going to t\_\_\_\_\_ it ly\_\_\_\_\_ d\_\_\_\_\_. They are going to be more violent.
18. When the score got to 8-2, we knew the game was in the b\_\_\_\_\_. We knew we were going to be the champion.
19. The candidate pl\_\_\_\_\_ the race c\_\_\_\_\_, claiming that she received less attention than the Malay candidate simply because she is Chinese.
20. After the recent bombing attacks, airports in major cities around the world have installed more cameras to st\_\_\_\_\_ up security.
21. I was so c\_\_\_\_\_ up in my school work that I didn't realise what was happening with my sister.
22. His jokes are funny, but sometimes he g\_\_\_\_\_ too f\_\_\_\_\_ and hurts people's feelings.

23. We have arranged a meeting for next Thursday, so if you see anyone, do sp\_\_\_\_\_ the  
w\_\_\_\_\_. We have to inform everyone.

*Appendix 14: Delayed form recognition posttest (Study 2 – The intentional learning condition)*

Name: \_\_\_\_\_

Class: \_\_\_\_\_

Below are 23 phrases taken from the video you saw two weeks ago. Please fill out the missing words with the most suitable choice among the four options provided.

**Example:** *The exam was a p\_\_\_\_\_ of c\_\_\_\_\_*

- c) pinch of cake                      c) piece of cake*  
*d) pitch of cake                      d) pint of cake*

**Answer:** *The exam was a p\_\_\_\_\_ of c\_\_\_\_\_*

- c) pinch of cake                      c) piece of cake*  
*d) pitch of cake                      d) pint of cake*

1. We have arranged a meeting for next Thursday, so if you see anyone, do sp\_\_\_\_\_ the w\_\_\_\_\_. We have to inform everyone.

- a) spare the word                      c) spare the wisdom  
b) spread the wind                      d) spread the word

2. After failing to convince the IT department that new security passwords are needed, Mike felt it was time to br\_\_\_\_\_ out the b\_\_\_\_\_ g\_\_\_\_\_. So he called a meeting with the Head of the Company.

- a) break out the best game                      c) break out the big game  
b) bring out the broad guns                      d) bring out the big guns

3. Let's just wait until the d\_\_\_\_\_ has s\_\_\_\_\_ before we decide what to do. It's better to make a decision when the situation has calmed down.

- a) dust has settled                      c) daylight has shone  
b) dusk has started                      d) dust has spread



4. When the score got to 8-2, we knew the game was in the b\_\_\_\_. We knew we were going to be the champion.
- a) boot
  - b) bag
  - c) book
  - d) boat
5. The gap between the rich and the poor is wide, and the poor aren't going to t\_\_\_\_\_ it ly\_\_\_\_\_ d\_\_\_\_\_. They are going to be more violent
- a) take it lying down
  - b) try it lying dead
  - c) turn it lying down
  - d) test it lying deep
6. After the recent bombing attacks, airports in major cities around the world have installed more cameras to st\_\_\_\_\_ up security.
- a) step
  - b) strike
  - c) stand
  - d) stop
7. I don't have anything against advertising, but I do have a b\_\_\_\_\_ with how many bad advertisements there are on TV.
- a) beat
  - b) brawl
  - c) beef
  - d) bowl
8. My brother came home drunk so I r\_\_\_\_\_ him out to my mother. I told my mother that he had been sneaking out at night.
- a) ran
  - b) ruled
  - c) ratted
  - d) ripped

9. Last week's rain was sm\_\_\_\_\_ pot\_\_\_\_\_ compared to the thunder storm we had two months ago.

- a) smart pottery
- b) small potty
- c) small potatoes
- d) smart potatoes

10. The new Iphone model is confirmed to h\_\_\_\_\_ the str\_\_\_\_\_ at the end of 2017.

- a) hire the stray
- b) hit the streets
- c) hold the streets
- d) hide the straw

11. He is a role model for other players to l\_\_\_\_\_ up to.

- a) live
- b) let
- c) lean
- d) look

12. We i\_\_\_\_\_ a lot in Tom, so we have every right to expect a lot from him. We devoted a lot of time and effort in training him to be a professional athlete.

- a) invested
- b) instilled
- c) inculcated
- d) intervened

13. The candidate pl\_\_\_\_\_ the race c\_\_\_\_\_, claiming that she received less attention than the Malay candidate simply because she is Chinese.

- a) placed the race card
- b) played the race character
- c) planned the race contest
- d) played the race card

14. Jude: Do you want to go swimming with us tomorrow?

Jane: Yes, c\_\_\_\_\_ me in.

- |         |           |
|---------|-----------|
| a) cut  | c) corner |
| b) call | d) count  |

15. He had a r\_\_\_\_\_ in with his boss. The argument caused him to lose his job.

- |         |         |
|---------|---------|
| a) race | c) run  |
| b) row  | d) ride |

16. I can wh\_\_\_\_\_ up a meal in no time.

- |         |          |
|---------|----------|
| a) whip | c) whiz  |
| b) whop | d) whisk |

17. The cooking competition is designed to give home cooks their d\_\_\_\_\_ in the s\_\_\_\_\_.  
The home cooks will finally get the attention they deserve.

- |                      |                    |
|----------------------|--------------------|
| a) day in the scene  | c) day in the sun  |
| b) day in the street | d) day in the spot |

18. You have so much energy on your television shows, I always suspect you must be  
ho\_\_\_\_\_ up on energy drinks.

- |           |            |
|-----------|------------|
| a) hooked | c) holed   |
| b) hopped | d) hoisted |

19. Someone has been sc\_\_\_\_\_ with my computer, and now it doesn't work anymore.

- a) scamming
- b) scrambling
- c) scouting
- d) screwing

20. His jokes are funny, but sometimes he g\_\_\_\_\_ too f\_\_\_\_\_ and hurts people's feelings.

- a) gives too far
- b) gets too fast
- c) gets too fight
- d) goes too far

21. We are only in Paris for a day, so let's m\_\_\_\_\_ the m\_\_\_\_\_ of it. We have to enjoy our day as much as possible.

- a) meet the most of it
- b) move the most of it
- c) make the most of it
- d) mark the most of it

22. Teachers are the uns\_\_\_\_\_ he\_\_\_\_\_ of a great writer's success. They are often not noticed or praised for their hard work.

- a) unsung heroes
- b) unsaid heroes
- c) unsold heroes
- d) untamed heroes

23. I was so c\_\_\_\_\_ up in my school work that I didn't realise what was happening with my sister.

- a) caught
- b) carried
- c) caged
- d) covered

*Appendix 15: Delayed meaning recall posttest (Study 2 – The intentional learning condition)*

Name: \_\_\_\_\_

Class: \_\_\_\_\_

**Below are 23 phrases taken from the video you saw two weeks ago. Please write the meaning of the phrases according to the context in the video. You can write as many answers as you can think of.**

**Example:**

**Question: *Take it easy***

**Answer: Relax, be calm, be gentle, do not get excited.**

1. Small potatoes

\_\_\_\_\_

2. Hit the streets

\_\_\_\_\_

3. Bring out the big guns

\_\_\_\_\_

4. Play the ... card

\_\_\_\_\_

5. Someone's day in the sun

\_\_\_\_\_

6. Caught up in something

\_\_\_\_\_

7. Hopped up

\_\_\_\_\_

8. Rat out someone

\_\_\_\_\_

9. Have a run-in with someone

\_\_\_\_\_

10. Invest in someone

---

11. Have a beef with someone

---

12. Step something up

---

13. Take something lying down

---

14. Screw with someone/something

---

15. The dust has settled

---

16. In the bag

---

17. Spread the word

---

18. Make the most of something

---

19. Unsung heroes

---

20. Goes too far

---

21. Whip up something

---

22. Count someone in

---

23. Look up to someone

---

*Appendix 16: Delayed meaning recognition posttest (Study 2 – The intentional learning condition)*

Name: \_\_\_\_\_

Programme: \_\_\_\_\_

Below are 23 phrases taken from the video you saw two weeks ago. Choose the most suitable meaning of the phrases, according to the context in the video. You should only choose one answer out of the four options provided.

**Example**

**(0) Question: A piece of cake**

**Answer:**

**(c) very easy**  
**(d) very soft**

**(c) very sweet**  
**(e) very tricky**

1. Unsung heroes

(a) Not noticed or praised for doing hard work

(c) Not having songs written for war heroes

(b) Not selected to fight for one's country

(d) Not having the right skills to survive

2. Count someone in

(a) To be important to someone

(c) To include someone in an activity

(b) To rely or depend on someone

(d) To send an invitation

3. Step something up

(a) To increase the intensity of something

(c) To walk closer to something

(b) To take over a job or a role

(d) To lift something higher

4. Small potatoes

- (a) Something that is not perfect when compared to something else
- (b) Something that is not important when compared to something else
- (c) Something that is not beautiful when compared to something else
- (d) Something that is not smooth when compared to something else

5. Whip up something

- (a) To reverse suddenly
- (b) To put something into good condition
- (c) To strike someone
- (d) To quickly make something to eat

6. Someone's day in the sun

- (a) A day full of exciting activities
- (b) A period of getting attention or appreciation
- (c) A powerful position or situation
- (d) A sunny and enjoyable day

7. Have a run-in with someone

- (a) Have an argument or disagreement with someone
- (b) Have a continuous series of events with someone
- (c) To go for a jog with someone
- (d) To escape from a dangerous place with someone

8. Play the...card

- (a) To have a hidden advantage
- (b) To trick someone into doing something illegal
- (c) To use a particular quality to gain an advantage
- (d) To play game that attracts a large audience



9. Invest in someone

(a) To give good advice to someone

(c) To trick someone into giving something up

(b) To invite someone to join an activity

(d) To use or devote resources to achieve something

10. Make the most of something

(a) To use or enjoy something as much as possible

(c) To create something of good quality

(b) To produce something in a large quantity

(d) To put in a lot of effort

11. Hit the streets

(a) To have a fight on the street

(c) To sell something outside

(b) To march on the street

(d) To be available to buy for the first time

12. Spread the word

(a) To make a supportive remark or favourable recommendation

(c) To speak or discuss with someone

(b) To give a piece of good advice

(d) To tell other people the news

13. Take something lying down

(a) To do one's job poorly

(c) To accept something unpleasant without fighting back

(b) To remain hidden while preparing to attack

(d) To keep quiet and not be noticed

14. In the bag

- (a) Certain to be won or achieved
- (b) All of one's belongings
- (c) To leave someone or something
- (d) To allow someone to take the blame

15. Look up to someone

- (a) To expect someone to supply something
- (b) To find a particular person in a crowd
- (c) To examine someone
- (d) To admire and respect someone

16. Bring out the big guns

- (a) To make use of the most important or powerful people or tool
- (b) To carry the heaviest weapon
- (c) To bring in the most expensive gears
- (d) To ask for help from influential people

17. The dust has settled

- (a) When something is not used for a long time
- (b) When things have calmed down
- (c) When something breaks or gives out
- (d) When dust is brushed or wiped off from something

18. Have a beef with someone

- (a) To have some reason for disagreement or unhappiness with someone
- (b) To strengthen a relationship with someone
- (c) To have a serious discussion with someone
- (d) To have an enjoyable meal with someone

19. Screw with someone/something

- (a) To create confusion
- (b) To cause trouble or problems
- (c) To destroy someone/something
- (d) To avoid someone/something

20. Rat out someone

- (a) To help someone get out of trouble
- (b) To suspect that someone has done something wrong
- (c) To reveal information about someone to a person in authority
- (d) To help someone go into hiding

21. Caught up in something

- (a) To do the work that one should have done
- (b) To be seen doing something illegal
- (c) To learn the news of someone or something
- (d) To be so involved in an activity that other things go unnoticed

22. Hopped up

- (a) To be really angry at something
- (b) To come close to something
- (c) To be excited and full of energy
- (d) To have high hopes

23. Goes too far

- (a) To travel to a far place
- (b) To do something in an extreme way
- (c) To do what is expected
- (d) To cause someone to go higher or move forward

*Appendix 17: Listening comprehension test (Study 2 – The intentional learning condition)*

Name: \_\_\_\_\_

Programme: \_\_\_\_\_

**Part 1**

**The following questions are based on the video you just watched. Circle the correct answer.**

1. What did the high school donate to the Natesville Auxiliary(volunteer) Police?
  - A) Radishes
  - B) Whistles
  - C) Chocolate
  
2. Why did Jimmy want Hope to become the little pilgrim?
  - A) To boost her confidence for potty(toilet) training
  - B) To make her the most popular child
  - C) To prove that Virginia was not cheating
  
3. Why was young Jimmy not allowed in the pool?
  - A) He was being naughty
  - B) He did not sell any chocolate bars
  - C) He did not win the contest
  
4. Why did young Virginia fail to sell the original chocolate bars?
  - A) She did not try hard enough
  - B) Rosa's family stopped her from selling the bars
  - C) The original chocolate bars did not taste good
  
5. At first, Virginia did not want to help Hope win because
  - A) She wanted Rosa's grandson to win
  - B) She did not want to burden Maw Maw
  - C) She knew winning would make another child sad
  
6. When young Virginia just started working with Knock Knock Knock Housekeeping, young Jimmy also started his long-life affair with
  - A) Chocolate
  - B) Fire and explosives
  - C) Growing radishes
  
7. What did Maw Maw do when Burt found the horse's head in his bed?
  - A) She made her chocolate stronger
  - B) She destroyed the Flores family's chocolate bars
  - C) She sold her chocolate bars at a cheaper price
  
8. Why was Virginia angry at Sabrina?
  - A) Sabrina started stealing Maw Maw's Magic Brown
  - B) Sabrina did not help Jimmy win the contest
  - C) Sabrina did not sell enough Maw Maw's Magic Brown

9. What did Barney want Jimmy to do?
- A) Beat his mother
  - B) Wear a wire
  - C) Confess to cheating

**Part 2**

**The following statements are based on the video you just watched. If the statement is true, circle the word True. If the statement is false, circle the word False.**

- |  |            |
|--|------------|
| 10. Burt named the horse Clip Clop.  | True/False |
| 11. Maw Maw thought the cupboard she was hiding in was a toilet.                           | True/False |
| 12. The Chance family played the violin badly.   | True/False |
| 13. Maw Maw's Magic Brown secret ingredient was cocoa powder.                              | True/False |
| 14. The Carlos family hit Burt with water guns.  | True/False |
| 15. As an auxiliary (volunteer) officer, Barney had the legal right to arrest a person.    | True/False |
| 16. Virginia could not stop herself from eating Maw Maw's Magic Brown.                     | True/False |
| 17. Hope was successfully potty trained after becoming the little pilgrim.                 | True/False |
| 18. Dating Sabrina gave Jimmy the confidence to grow both Ricky Radish and Sabrina Squash. | True/False |
| 19. Frank thinks that Barney will not do what he threatens to.                             | True/False |
| 20. Barney was much fatter 20 years ago.   | True/False |

*Appendix 18: Ethic's approval*



**MEMORANDUM**

Phone 0-4-463 5480  
Email susan.corbett@vuw.ac.nz

TO	Elvenna Majuddin
COPY TO	Dr Anna Siyanova
FROM	AProf Susan Corbett, Convener, Human Ethics Committee
DATE	12 May 2017
PAGES	1
SUBJECT	<b>Ethics Approval: 24621</b> The effects of captioning on the acquisition of multiword units

Thank you for your application for ethical approval, which has now been considered by the Standing Committee of the Human Ethics Committee.

Your application has been approved from the above date and this approval continues until 12 May 2020. If your data collection is not completed by this date you should apply to the Human Ethics Committee for an extension to this approval.

Best wishes with the research.

Kind regards

A handwritten signature in black ink that reads "Susan Corbett".

Susan Corbett  
Convener, Victoria University Human Ethics Committee