Imagery in L2 Captioned Video: Investigating Incidental Vocabulary Learning from Extensive Viewing as a Function of Modality, Contiguity, and Spacing

Souheyla Ghebghoub
Doctor of Philosophy
Education, University of York
September, 2021

*To my parents, whose **image** inspired eighty five thousand **words.***

Abstract

The aim of this thesis is to study the role of imagery in L2 captioned video by examining modality (Study 1), contiguity (Study 2), and spacing (Study 3) effects in incidental vocabulary learning from extensive TV viewing. An experimental design was employed in which one hundred seventy-three Algerian EFL learners in their third year of the Linguistics Bachelor programme were randomly assigned to either a Control, View, or Non-View group. Treatment participants watched two full-length seasons of documentary series extending to eight viewing hours, over a six-week period of two-week intervals. The View group watched the episodes in the form of L2 captioned video while the Non-View group had the imagery hidden and were therefore exposed to L2 audio and L2 captions only. Four levels of word knowledge were measured: meaning recall and recognition (posttest only) and spoken and written form recognition (pretest-posttest).

Study 1 assessed the effect of obscuring imagery on incidental learning of twenty words using a between-participants design. The results showed successful word learning regardless of the presence of imagery. Study 2 investigated the effect of verbal-visual contiguity (the co-occurrence of a word and its visual referent) on incidental learning of twenty-eight words using a within-participants design (View group only). It introduced contigfrequency, contigduration, and contigratio as three measures of contiguity on two timespans (∓7 seconds and ∓25 seconds) that were longer than those used in previous studies. The results showed that the amount of time visual referents appeared on the screen (contigduration), measured in a ∓25 second timeframe relative to the verbal occurrence, was predictive of learning. These results were more pronounced in the meaning recognition test. Study 3 explored whether words would be learned better when their occurrences were spread across viewing sessions (spaced condition), as compared to appearing within a single session (massed condition) by measuring the incidental learning of eight matched word pairs using a between-items design. It also examined whether learning in these two spacing conditions was influenced by the presence of imagery. The results revealed a positive effect of spaced occurrences in the Non-View group but not the View group, suggesting that a spacing advantage is more likely when fewer cues are available. These results were limited to knowledge of meaning only.

List of Contents

List of Tables

List of Figures

List of Accompanying Material

The experimental materials in PowerPoint and video formats were uploaded to the following pages in Open Science Framework.

Vocabulary Tests (four dependent measures), Open Access:
https://osf.io/yavf7/

Stimuli (introduction to experiment, highlights), private view-only link:
https://osf.io/rhjkv/?view_only=c4fa6e7d3d5949208172be9fbeb78574

Acknowledgements

Gratitude is above all due to The All-Knowing. No one compares to my creator. I thank God for everything.

I am forever grateful to the Algerian Ministry of Higher Education and Scientific Research for offering me a scholarship to pursue a PhD degree in the UK.

There are no words to express my gratitude to my supervisor Dr. Cylcia Bolibaugh. My thank you seems so small compared to what you have done both academically and emotionally. I appreciate your unwavering faith in me, which helped push my boundaries and expand my ideas into this manuscript. Thank you for introducing me to the wonderful world of R and statistics and for the countless hours of reading my thesis which led to significant improvements. I especially thank you for behaving with the utmost courtesy towards me throughout these years. The last year was of great difficulty, and I do not think I would have overcome it if I did not have a supervisor who is as supportive as you.

I am deeply grateful to my thesis committee members. I wish to thank my Thesis Advisory Panel (TAP) member, Dr. Nadia Mifka-Profozic, for her advice and support. I sincerely thank Dr. Michael Rodgers who readily agreed to examine my thesis. Thank you for your valuable feedback.

I thank the Department of Education at the University of York for offering me the opportunity to teach my favourite modules (TESOL Methods and Planning and Communicating Research) during the PhD journey.

I thank the Department of English Language staff at the University of Jijel who were incredibly supportive and amenable to changes in their schedules to meet my research aims. Mainly Dr. Fateh Bounar, Dr. Bakir Benhabiles, Dr. Meriem Bousbae, Dr. Houda Bouhadjar, Dr. Izzedine Fanit, Miss Loubna Hloulo, Mrs Safa Khedimallah, and Mrs Zeyneb Djerrah and those I missed to mention. Thank you all. I wish to thank Mrs Safia Mechtar for helping me to collect more data in Algeria by administering a paper version of the online vocabulary test while I was in the UK. I thank the 323 third-year Algerian students enrolled in the Linguistics Bachelor programme at the University of Jijel in the two academic years 2016 – 2017 and

2017 – 2018. This thesis would not have happened if it had not been for your contribution to its experiment. Thank you enormously.

A sincere thank you is to Dr. Chadia Chioukh, my Master supervisor, for helping to smooth the experiment I conducted. Thank you for always supporting me.
I also thank my colleagues at the Research Centre for Social Sciences, especially Elena. Thank you for the coffee and lunch breaks which had become a great source of emotional support. A special thank you to my cousins Fatima and Myada and my friends in York: Suehyun, Sihem, Şermin, Amel, Amina, and Nihad.
Knowing you're there to cheer me on made difficulties so much easy.

Finally, I acknowledge the support of my family.
Thank you, my parents, for your unconditional and irrevocable love. The trust that you have placed in me, that I could travel abroad and embark on an arduous journey alone, is my most treasured gift. Sorry for the many years I spent away from you.
Also, my siblings Manel, Mokhtar, and Yacine. I will never grow tired of telling you how much I love you. You all make my life's journey meaningful. I appreciate your support throughout the years.

I wish that Myral, Taymou Allah (my niece and nephew who are both the age of my PhD), my niece Maram, and my newborn nephew Iheb would read this one day and just realise how much their coming added meaning to my life.

I am so thankful for the gift of Khalil, my fiancé. You continue to amaze me with every new day. Success means so much to me knowing I have you to share it with.

Declaration

I, Souheyla Ghebghoub, declare that this thesis is a presentation of original work, completed solely by me, under the supervision of Dr Cylcia Bolibaugh. This work has not previously been presented for an award at this, or any other University. All sources are acknowledged as References.

This research was supported by a 1+4 scholarship offered by the Algerian Ministry of Higher Education and Scientific Research

Some of the findings in this thesis have been presented as follows:

Chapter 3

Ghebghoub, S., & Bolibaugh, C. (2019, July 1-3). *Imagery in L2 captioned video: Incidental vocabulary learning from 8 hours of exposure*.[Conference presentation]. Vocab@. Leuven, Belgium.

Ghebghoub, S., & Bolibaugh, C. (2018, July 12-13). *The effect of input modality on incidental EFL vocabulary learning*.[Conference presentation]. 14th BAAL Language Learning and Teaching SIG. Southampton, UK.

Ghebghoub, S., & Bolibaugh, C. (2018, July 9-10). *The effect of input modality on incidental EFL vocabulary learning*. [Conference presentation]. BAAL Vocabulary Studies SIG. London, UK.

Chapter 4

Ghebghoub, S., & Bolibaugh, C. (2019, April). *Incidental Vocabulary learning as a function of verbal-visual contiguity: duration, frequency, and ratio in L2 captioned video*.[Conference presentation]. BAAL Vocabulary Studies SIG. Manchester, UK.

Chapter 1

General Introduction

My interest in this research derives from my personal experience as a second language (L2) learner of English language in the Algerian context, where the English language is not spoken outside the confines of the classroom. Although my formal English language instruction began at 12 years old, I exhibited wide vocabulary range and knowledge of what were supposed to be tough words for a beginner, which my teacher found perplexing. He surmised that I must have been doing something different than my peers out-of-class. He asked whether I had been exposed to extramural English that might explain my relatively advanced vocabulary knowledge. At that age, I did not know about incidental vocabulary learning (I only knew that my vocabulary breadth had annoyed a few of my classmates). It turned out that the source of extramural English was extensive TV viewing. I was exposed to English language TV programs (with Arabic subtitles) at a very young age for being the youngest of a family of six. I had picked up those difficult words unconsciously while viewing TV series with my older siblings.

Eighteen years later, the above experience validates a language learning principle that I have frequently encountered in my research. Recent evidence suggests that incidental extramural exposure is positively associated with L2 vocabulary acquisition (e.g., Leona et al., 2021). Importantly, sustained exposure to L2 TV programs appears to be closely linked with language learning, including in those learners who have not yet received formal instruction (e.g., Puimège & Peters, 2019). However, what is so unique about extensive audio-visual input, mainly, L2 captioned documentary series? How does the presence and absence of on-screen imagery affect incidental L2 word learning? Does the temporal distance of a visual referent from its word form inform its potential for learning? Is it a question of how often a verbal occurrence has a visual referent(s), how long the visual referents last for every verbal co-occurrence, or what the proportion of visual occurrences to verbal occurrences is? Would words be learnt better if they occurred under a spaced or massed condition, and does imagery influence the spacing effect? These are some of the enduring questions in the literature that my thesis addresses through three

different studies in key areas in second language acquisition (SLA) research (Modality, Contiguity, and Spacing).

Chapter 1 (the present chapter) comprises four sections. In the previous lines, I have revealed the personal and academic motives that have driven my research. I will now introduce the reader to previous research on incidental L2 vocabulary learning from viewing audio-visual input. I will then set the scene for the whole thesis by identifying what we know and do not know. Next, I will state all the research questions that the theoretical gaps generate. Finally, the chapter will conclude with an outline of the structure of the thesis that gives a guide to the contents you will encounter throughout the thesis chapters.

1.1 Incidental L2 Vocabulary Learning from Viewing

Vocabulary is the essence of any language. There seem to be as many definitions of vocabulary as there are researchers in the field; however, Harmer (1991) coined one of the most influential definitions. He stated in metaphorical terms: "If language structures make up the skeleton of language, then it is vocabulary that provides the vital organs and the flesh" (p. 153). While the first idea that springs to one's mind when thinking of vocabulary is probably words, vocabulary is much more than just words. It includes lexical chunks that convey single meanings such as collocations; hard work, phrasal verbs; pay off, and idiomatic expressions; put one's mind at ease. Vocabulary researchers adopt the term items instead of words if their study includes multi-word units. The present thesis involves only words, but both terms are used interchangeably throughout the thesis chapters.

Many studies have shown the importance of reading and listening in increasing L2 vocabulary acquisition. Vocabulary develops more often unintentionally through sustained exposure to auditory and written input (Cunningham, 2005), popularly known as incidental vocabulary learning. L2 learners have few opportunities for incidental acquisition, especially in contexts where they do not speak the language outside the classroom, such as in Algeria. Therefore, L2 learners need to be exposed extensively to meaningful L2 input inside or outside the classroom to maximise incidental learning. As put by Hiebert and Kamil: "Not only are students expected to understand words in texts, but also texts can be expected to introduce them to many new words." (2005, p. 1). Several

studies have highlighted the importance of written, spoken, and bimodal input in incidental L2 word development (e.g., Horst, Cobb, & Meara, 1998; Hulme, Barsky, & Rodd, 2019; Pellicer-Sanchez & Schmitt, 2010; Van Zeeland & Schmitt, 2013a). Extensive exposure has particularly been associated with incidental word gains (e.g., Tragant Mestres, Llanes Baró, & Pinyana Garriga, 2018; Webb & Chang, 2015).

In recent years, vocabulary research has been undergoing a viewing turn. The evidence for incidental L2 vocabulary acquisition from viewing is not recent (e.g., Huang & Eskey, 1999; Markham, 1999; Neuman & Koskinen, 1992; Vanderplank, 1988). Nevertheless, a growing list of L2 studies has recently shed light on the critical role of audio-visual input in incidental L2 word acquisition (e.g., Mazahery, Hashemian, & Alipour, 2021; Peters, Heynen, & Puimège, 2016; Rodgers & Webb, 2019). Several factors in TV viewing are known to affect incidental word learning. For example, L2 learners experience a significant amount of audio-visual input, such as films and TV series, which appears to correlate with L2 vocabulary development (Peters, 2018). There is also the motivational factor (Baltova, 1994); TV viewing encourages vocabulary intake for being entertaining and motivating. Most importantly, extensive viewing is in line with the usage-based theory of SLA. This theory emphasises the role of input, experience, and frequency of experience in language learning (Wulff & Ellis, 2018). TV series provide a rich source of information for language learners to process. Based on the usage-based approach, viewing multiple TV episodes could result in frequent encounters of the same word. The varied experiences may help create distinct memory traces which enable fast processing and retrieval of vocabulary.

Viewing research attention has centred on investigating the impact of L1 subtitled and L2 captioned videos (e.g., Majuddin, Siyanova-Chanturia, & Boers, 2021; Sinyashina, 2020b; Teng, 2021). Despite a consensus among researchers that imagery in audio-visual input reinforces L2 vocabulary acquisition, there has been little quantitative analysis of the extent of this support. What remains unknown is the differential effects of imagery (in extensive TV viewing via L2 captioned video) on learning varied aspects of word knowledge. In addition, despite the importance of spacing, it has received scant attention in viewing research, and it is not known whether the presence of imagery interacts with spacing conditions. We will appreciate how relevant this is in the next section.

1.2 What Has Been Unknown About Imagery in Incidental L2 Vocabulary Learning from L2 Captioned Video?

Research on incidental vocabulary learning from TV viewing has tended to focus on captions and subtitles. Regarding viewing in L2 captioned video format, several studies have linked it with incidental L2 word learning but explained this linkage in terms of the benefits of L2 captions and audio (e.g., Peters et al., 2016). Central to this entire thesis is the concept that not only the bimodal verbal input acts on incidental word learning from L2 captioned video but also that imagery promotes word acquisition. This section provides the knowledge gaps this thesis fulfils in research about incidental L2 word learning from TV viewing in L2 captioned video format, and presents them in the order in which they will be addressed in the thesis.

Study 1 shows that extensive TV viewing has been understudied despite growing interest in studying the effect of multimodal input or TV viewing among vocabulary researchers. Only two studies have attempted to produce quantitative analyses on incidental L2 vocabulary learning from extensive TV viewing that surpasses +7 hr (Pujadas & Muñoz, 2019; Rodgers & Webb, 2019). They based their results on data from many but short sessions. It is unknown whether learners can acquire L2 vocabulary from multiple long sessions that reflect extensive viewing conditions outside the classroom for recreational purposes. In addition, no one, to the best of my knowledge at the time of writing, has studied the impact of the presence and absence of imagery in extensive TV viewing of L2 captioned video, or documentary series in particular, on incidental L2 vocabulary learning.

Until recently, researchers have observed the effects of temporal contiguity between words' verbal forms and their visual referents on vocabulary learning only in non-authentic materials and explicit teaching contexts. Before Study 2 of the present thesis, it has not yet been established whether verbal-visual contiguity in authentic audio-visual input, more precisely, the measures I introduce (contigfrequency, contigduration, and contigratio), affect incidental word acquisition. Which of these measures more strongly contributes to learning is worthy of knowing.

To date, one study brought attention to the effect of verbal-visual contiguity in multimodal input on incidental L2 word learning (Rodgers, 2018) and three other

studies have attempted to test this effect (Ahrabi Fakhr, Borzabadi Farahani, & Khomeijani Farahani, 2021; Peters, 2019; Pujadas Jorba, 2019). The present study extends this research, and addresses limitations in internal validity caused by methodological choices. Studies to date have operationalised verbal-visual co-occurrences dichotomously (whether the target word co-occurred with its visual referent at least once in the overall input). Thus, the studies have not taken account of the frequency of verbal-visual contiguity for every word (i.e., contigfrequency).

Further methodological questions concern operationalisation and timeframes of contiguity. Before the present thesis, there have been no data on the effect of contiguity duration (i.e., contigduration) on incidental L2 vocabulary learning. One study assessed the impact of visual referents durations on word learning from audio-visual input but the study included all referents irrespective of where they are compared to word forms (Pujadas Jorba, 2019). Data were also based on 2 hr 55 min viewing and on nouns only. Evidence of contiguity effect based on up to 8 hr viewing data and for different parts of speech is lacking. In addition, no single study exists which questioned the potential influence of the proportion of verbal occurrences accompanied by a visual referent (i.e., contigratio) on incidental L2 vocabulary learning. Moreover, previous studies have identified the visual referents using ∓ 2 seconds and ∓ 5 seconds timeframes (i.e., the duration between verbal and visual occurrences). It is not yet clear whether contiguity effects can be observed if the visual referent occurs at a relatively longer temporal distance from the word.

Finally, little attention has been paid to spacing in incidental learning contexts before Study 3, despite extensive research on this phenomenon in the area of deliberate L2 vocabulary learning. The few available studies have mostly been concerned with unimodal or bimodal input (e.g., Çekiç & Bakla, 2019). It has not been clear whether words would be better learnt when their occurrences were spaced across extensive viewing sessions instead of being massed within a single session. Two studies examined spacing effects in incidental vocabulary learning from extensive TV viewing, with both finding an advantage for words massed in a single session (Pujadas Jorba, 2019; Rodgers & Webb, 2019). Nonetheless, they did not base their results on experimental manipulation of spaced and massed items. Essentially, research to date has not yet determined whether the presence of imagery influences spacing effects in vocabulary learning.

1.3 Research Questions

On the whole, this thesis aims to answer the following 10 research questions:

Norming Study

1 Which of the pool of 54 initially selected words are unknown to third-year Algerian undergraduates in the BA Linguistics programme?

2 How many words do we need to know to achieve 90% and 95% coverage of the two full-length seasons of the documentary series?

3 What is the estimated English language vocabulary size of third-year Algerian undergraduates in the BA Linguistics programme?

Study 1

4 Does viewing two full-length seasons of L2 captioned documentary series (8 hr) over 2-hour long sessions lead to incidental learning of L2 vocabulary?

5 What is the effect of removing imagery and keeping bimodal input?

Study 2

6 *(Model building)* Is the effect of verbal-visual contiguity on incidental word learning from extensive viewing of L2 captioned documentary series moderated by:

(a). The length of the timeframe within which contiguity was measured?

(b). The inclusion/exclusion of weak visual referents and related forms?

7 *(Main)* What is the effect of three verbal-visual contiguity measures: contigduration, contigfrequency, and contigratio on incidental word learning of different parts of speech from extensive viewing of L2 captioned documentary series?

8 *(Exploring)* What are the relative strengths of the three predictors: contigduration, contigfrequency, and contigratio of incidental vocabulary learning from extensive viewing of L2 captioned documentary series?

Study 3

9 Do repeated occurrences distributed across multiple extensive viewing sessions facilitate incidental L2 vocabulary learning from documentary series compared with repeated occurrences massed within a single session?

10 Does any spacing effect vary as a function of the presence of imagery?

1.4 Structure of Thesis

This thesis draws upon three strands of research, modality, contiguity, and spacing and is divided into six chapters. Following this general introduction, Chapter 2 (the Norming Study) provides a detailed description of the preliminary analyses and decision-making processes to select the audio-visual materials for the experimental studies. Chapters 3, 4, and 5 have a similar format and are entitled modality effects in learning, contiguity effects in learning, and spacing effects in learning, respectively. The chapters begin with an introduction that establishes the importance of the study for SLA research and identifies the knowledge gaps in the field. Section 1 refers to previous research, outlines theoretical foundations, and identifies lack of evidence or inadequacies in former studies. Section 2 then gives a brief overview of the study. The remaining part of the chapters proceeds as follows: Method (Section 3), Analyses (Section 4), Results (Section 5), and Discussion (Section 6). I close each chapter by summarising the study aims, results, and limitations, if any, specific to the study.

Chapter 2

The Norming Study

This chapter will report on a two-part norming study conducted in the academic year 2016 – 2017 on two samples of students with similar characteristics to the population targeted in the thesis. The aims of the norming study were twofold: firstly, to ascertain the degree to which the initially selected pool of target words was known/unknown to participants, to inform the filtering process; secondly, to verify that participants' vocabulary size was adequate to ensure 90 – 95% lexical coverage of the selected materials.

2.1 Part 1 – Target Vocabulary Extraction

The first part of the chapter begins by outlining several factors that informed decisions about selecting the set of video materials that might be useful in addressing upcoming research questions in the thesis. It poses the first research question of this chapter and describes the participants involved to address it. The chapter then describes the instruments and procedures employed to identify and test the pool of potential target words on participants. The remainder of the text presents and discusses the results.

2.1.1 Materials Selection

A survey of authentic video materials likely to contain aural and pictorial input was carried out, and BBC science television series presented by Professor Brian Edward Cox were identified as an initially promising sample. Cox is an English particle physicist in the School of Physics and Astronomy at the University of Manchester, United Kingdom. He has been a popular presenter of many science television series and the author/co-author of 950 scientific publications. Seven series (18 episodes) were selected for initial review, two of which (8 episodes) were ultimately found to be in line with the research aims: authentic materials available in the three modalities; spoken (audio), written (L2 captions), and pictorial (imagery), and target words that have a minimum total frequency ≥ eight distributed across the materials.

The review aimed to identify episodes that could be used to create four different viewing sessions in the exposure phase. Across all sessions, the materials needed to meet the following criteria: (1) the materials needed to contain a minimum of 15 words which were likely to be unknown to participants; (2) each target word needed to be present a minimum of eight times across the four video presentations; (3) exposure to words needed to be spaced as evenly as possible across all four video presentations. These criteria were designed to implement optimum conditions for incidental learning as a function of frequency and spacing as will be explained later.

2.1.2 Extraction Procedure

Transcripts of the series were tracked at two websites (https://subsaga.com and www.tvsubtitles.net) and downloaded as English captioning text files in SubRip format. Timestamps and formatting were removed using an online subtitle tool (https://www.subtit letools.com/convert-subtitles-to-plain-text-online). The transcripts were analysed using *The Compleat Lister* function in the online text analysis tool Lextutor (Cobb, n.d.), which is " the most essential tool in the vocabulary researcher's tool box" (Schmitt, 2010, p. 341). Infrequent words of occurrences ≥ eight (see Section 3.1.6) within an episode or all episodes were then identified. Eventually, 54 potential target words (9 adjectives, 9 adverbs, 12 verbs, and 24 nouns) were selected from 18 episodes of seven BBC documentary series.

2.1.3 Research Question

The norming study aimed to answer the following research question:

- Which of the pool of 54 initially selected words are unknown to third-year Algerian undergraduates in the BA Linguistics programme?

2.1.4 Participants

One hundred fifty participants were recruited from the population of tertiary EFL Algerian learners who were third-year undergraduates in the Linguistics Bachelor programme at the University of Jijel, Algeria, in the middle of 2016-2017 academic year. This study was not performed on the target population because testing threatens its internal validity. Familiarity with target words and awareness of the study purpose are two extraneous factors that may negatively affect the experiment (The Hawthorne Effect by Henry Landsberger; Levitt & List, 2011) The participants were all native Arabic speakers with French as a second language. They had studied

English for a minimum of 9 years and of intermediate English language level. They were initially recruited using a Facebook group which they were requested to join.

2.1.5 Target Items

Fifty-five words appearing in 18 episodes of seven BBC documentary series made up the target items for the norming study – Part 1. The list of items along with their parts of speech and frequency of occurrence is in Table 2.1. The distribution of word occurrences across episodes is provided in Appendix B.

Table 2. 1

54 Potential Target Items and their Frequency of Occurrence in 18 Documentary Episodes

Parts of speech	Item	Freq	Parts of speech	Item	Freq
Nouns					
	cosmos	35		fuse	12
	photon	32		emit	11
	snowflake	29		sculpt	10
	nucleus	19		rotate	8
	iceberg	18		forge	8
	sulphur	16		peer	8
	dust	16		bounce	8
	sphere	15		squash	7
	manatee	14	Adjectives		
	supernova	14		dense	15
	entropy	13		magnificent	12
	particle	12		stellar	10
	spectrum	12		cosmic	10
	iron	11		primordial	9
	constellation	11		denser	9
	symmetry	11		alien	9
	moth	11		faint	8
	temple	11		intricate	7
	hexagon	10	Adverbs		
	horizon	10		incredibly	19
	fusion	10		ultimately	15
	tide	10		eventually	14
	aurora	8		roughly	12
	pile	8		literally	10
Verbs				seemingly	10
	orbit	40		spontaneously	8
	stretch	16		virtually	8
	curve	13		relatively	8
	float	13			

2.1.6 Materials: Vocabulary Knowledge scale

A vocabulary test was designed using a simplified version of the Vocabulary Knowledge Scales (VKS, Wesche & Paribakht, 1996) as a template. Participants were asked to self-report their word knowledge on a 3-point (instead of 5-point) scale. This modification allowed students to quickly indicate how well they knew each of the 54 items of vocabulary.

2.1.7 Procedure

The test was administered online and in paper formats. The online version was administered via Qualtrics (https://www.qualtrics.com/) (n = 96) (see Appendix G). A paper version was administered with the assistance of university lecturers to collect more responses (n = 54). The test took approximately 10 minutes.

2.1.8 Ethical Considerations

All participants in the norming study were given an information sheet (Appendix D). It detailed the nature and aim of the study, the amount of time required by participants to complete the test, and contact information. It also stated that participation is voluntary, that they had the right to withdraw during testing, and that data were anonymous at the point of collection. The Education Research Ethics Committee at the University of York approved the norming study.

2.1.9 Scoring

I used a 3-point scoring scale as follows:

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

At level 3, evidence of knowledge of meaning was required, and a re-assignment of scores determined the score. If participants gave an appropriate synonym or translation, they were awarded 3 points, while if they gave an incorrect response, they were awarded 2 points). Participants were encouraged to provide all known meanings for polysemous words, and these were scored by referring to the target meaning in the video materials. For instance, *squash* occurred as a verb (i.e., *crush*) and can act as a noun (i.e., a sports game); thus, only verb responses were scored as 3.

Score Interpretation

The score allocated to each item reflected the degree to which the word was known. Score 1 indicates that the word is unknown, score 2 refers to written form recognition, and score 3 indicates meaning recall. Hence, participants' scores resulted in a spectrum of words, the middle of which represents frontier[1] words, while its two ends represent unknown and known words.

2.1.10 Results

Unlike the online format, in which the forced response option prompted students to do the test fully, non-responded items were identified in data from the paper version. These items might result from students' inattention, lack of motivation to complete the test, or scepticism about their degree of word knowledge. However, missing data treatment was unnecessary. A Pattern Analysis test was run to enable a summary of the missingness: 28 among 54 items (above half) were subject to missingness, for which 12 students among 150 (8%) were responsible, and a total of 43 missing values out of 8250 (0.5%) were detected, indicating negligible missingness < 2. MCAR test using IBM SPSS Statistics (version 25.0; IBM Corp, 2017) was also run to verify that the values were missing completely at random (MCAR). Little's MCAR test findings were not significant either (sig = 0.873, $p > 0.05$).

Item Analysis

Part 1 of this norming study aimed to determine inclusion/exclusion criteria for the items in the materials that were to be used in the studies that make up this thesis. As such, no analysis of participants' mean vocabulary knowledge is reported here. Mean scores and standard deviations were calculated for each word, and ranked in ascending order. A total of 14 items (1 adjective, 8 adverbs, 1 verb, and 4 nouns) with a mean > 2 were excluded as these reflect knowledge of written form or meaning among the population. The remaining 40 items are presented in Table 2.2, grouped by part of speech. As the table indicates, only one adverb had a mean < 2. That is, participants were familiar with 9 out of 10 adverbs. This part of speech was therefore removed from the study.

---

[1] *Frontier* words are words that are familiar merely in form (Durso & Shore, 1991).

Table 2. 2

Means and Standard Deviations for Words with a Mean < 2

Parts of speech	Words	N	M	SD
Nouns	manatee	150	1.23	0.536
	sulphur	149	1.36	0.658
	hexagon	150	1.36	0.605
	supernova	148	1.36	0.585
	entropy	148	1.4	0.531
	constellation	150	1.41	0.592
	aurora	150	1.51	0.632
	symmetry	149	1.52	0.693
	moth	150	1.53	0.692
	photon	150	1.57	0.669
	spectrum	150	1.61	0.712
	snowflake	150	1.7	0.73
	cosmos	150	1.84	0.828
	pile	147	1.85	0.725
	sphere	148	1.86	0.591
	fusion	145	1.89	0.756
	iceberg	148	1.91	0.828
	tide	148	1.97	0.719
	temple	149	1.97	0.813
	particle	148	1.99	0.675
Verbs	forge	148	1.53	0.622
	orbit	150	1.59	0.604
	squash	150	1.73	0.631
	sculpt	149	1.75	0.770
	fuse	149	1.76	0.644
	peer	149	1.83	0.408
	stretch	149	1.87	0.719
	emit	149	1.87	0.729
	curve	150	1.93	0.656
	bounce	150	1.97	0.699
	rotate	149	1.99	0.805
Adjectives	intricate	149	1.36	0.558
	primordial	150	1.51	0.693
	stellar	150	1.51	0.632
	denser	149	1.75	0.687
	alien	150	1.76	0.598
	cosmic	150	1.83	0.757
	faint	149	1.84	0.508
	dense	150	1.91	0.780
Adverbs	virtually	149	1.99	0.683

Note. N = 40.

Further Considerations

The results also showed that the items' orthographic and phonological neighbours influenced students' written form recognition. This was determined whenever the participant had mistaken the meaning of a target item to the meaning of its orthographic or phonological neighbour (e.g., "فم" which means mouth as an answer to the target item "moth". In this case, a score of 1 was assigned). The answers provided by participants are listed in Table 2.3.

Table 2. 3

Neighbours' Influence on Results

Words	Orthographic neighbours
moth	mouth
emit	omit
fusion	fashion
nucleus	nuclear

Item	Phonological neighbours
symmetry	cemetery
denser	dancer
dense	dance
cosmic	cosmetic
sculpt	scalpel
nucleus	nuclear

2.1.11 Discussion

Which of the pool of 54 initially selected words are unknown to third-year Algerian undergraduates in the BA Linguistics programme?

The 54 items were tested on 150 participants using a simplified version of the Vocabulary Knowledge Scale, in response to the question above. Participants recognised 14 out of 54 items; thus, known items were filtered out from the final set of items.

Item Selection

Two factors informed the final selection of the target words. These are participants' mean scores and words' frequencies of occurrence within and across episodes. The study prioritised for selection three types of items: items with lower means, as they reflected words which participants had less knowledge of; items whose frequencies of occurrence were evenly distributed across episodes; and finally, items that helped to create spaced versus massed word pairs to meet the aims of Study 3. Spaced items are items that reoccur across multiple viewing sessions while massed items are items that reoccur in one single session (see Chapter 5). The consideration of the preceding factors led to the selection of 28 items: 20 spaced words (6 adjectives, 6 verbs, 8 nouns) and 8 massed words (8 nouns).

While the resulting means for the selected words are < 2, the standard deviations indicate that some participants did know the words. Conducting a pre-test is, therefore, a prerequisite for the experimental treatment in the upcoming studies. However, it is fundamental to note that participants in Part 1 of the norming study were halfway through their third-year course, while participants contributing to the three main studies were only beginning the course. Thus, it could be maintained that the target items were more likely to be unknown by participants in the main studies than participants in the norming study. The items' frequencies of occurrence per session are displayed in Table 2.4. Characteristics of items are fully specified in Chapter 3; see Section 3.3.6. The selected words appear in two full-length seasons of documentary series: Wonders of the Universe (Cooter, Lachmann, Holt, & Cox, 2011) and Forces of Nature (Cooter, Dyas, & Cox, 2016); each series includes four episodes. A description of the selected two series is in Section 2.2.1.

2.1.12 Methodological Considerations

The nature of the distribution of word occurrences across the episodes needed further norming procedures. As previously noted, the norming study aimed to select items to be encountered in four different presentations of video materials as part of the treatment phase. Thus, to allow items to be more equally spread across presentations, it was decided that learners would be exposed to two episodes in every presentation session.

Table 2. 4

Target Words with Their Parts of Speech and Frequency of Occurrence in Sessions

		Freq of verbal occurrence				
PoS		U1+F1	U2+F2	U3+F3	U4+F4	Post Highlights
Spaced	alien	—	1	4	4	10
adj	cosmic	2	3	1	4	10
	dense	3	3	7	2	15
	denser	1	2	1	5	09
	faint	4	—	1	3	09
	intricate	3	—	2	2	08
Spaced	emit	1	2	1	7	11
verbs	forge	—	4	3	1	09
	orbit	4	11	16	9	40
	sculpt	7	—	2	1	11
	squash	2	1	2	2	08
	stretch	2	—	6	8	18
Spaced	constellation	1	3	3	4	11
nouns	cosmos	17	1	8	9	35
	particle	1	1	2	8	13
	spectrum	—	5	—	5	12
	sphere	11	—	2	—	16
	supernova	—	9	3	—	15
	temple	1	2	—	5	09
	tide	1	5	4	—	12
Massed	hexagon	10				10
nouns	fusion		10			10
	manatee	14				14
	moth			11		11
	photon				32	32
	pile	8				08
	sulphur			16		16
	symmetry	11				11

Note. PoS = Part of speech; Freq = frequency; Adj = Adjectives; U = Wonders of the Universe; F = Forces of Nature.

Highlights Technique

The study design required a minimum of two occurrences per exposure. As can be seen from Table 2.4, not every item met this requirement. This was rectified by embedding missing word occurrences in created highlights of episodes and presenting them to participants as part of the experiment. A highlight could be defined as a short montage of excerpts from events, episodes of a television series, etc. Highlights can be inserted as the first thing into an episode to update viewers

with the latest developments in the series (i.e., a recap) or as the last thing to get them excited to watch upcoming episodes.

Procedure

Because the television documentaries are episodic, highlights of one or both types were implemented for every presentation session depending on the number of occurrences missing. Small segments consisting of selected scenes in which the missing occurrences appeared were cut from previous or subsequent episodes and combined to make up highlights for the experimental session. The two statements, "*Previously on Wonders of the Universe and Forces of Nature*" and "*Next time on Wonders of the Universe and Forces of Nature*", were voice-overed online by a British professional using Fiver Freelance Services (https://www.fiverr.com/categories/music-audio/voice-overs) and included in the highlights. The videos were designed using two computer software. The selected scenes were extracted by playing the video and recording the computer screen using Screencast-O-Matic (https://www.screencast-o-matic.com). The scenes were then combined using Windows Movie Maker (version 16.4.3528.0331; Microsoft Corporation, 2012). Both software were chosen for their free access and ease of use. Examples of the created highlights are in the accompanying materials.

2.2 Part 2 – Lexical Coverage of the Documentary Series

After selecting the vocabulary test items for the thesis, part 2 of the norming study aimed to verify that participants had sufficient vocabulary to ensure adequate lexical coverage. In light of recent research on the relationship between lexical coverage and viewing comprehension, 95% coverage is above what is necessary for viewing comprehension. Learners require a minimum of 90% coverage for adequate comprehension of documentary series (Durbahn, Rodgers, & Peters, 2020). Much less coverage might be needed for the documentaries selected for this thesis due to the provision of L2 captions, which might assist content comprehension. This second part of the chapter starts with an overview of the selected documentary series. It then specifies two research questions and the recruited participants. The section that follows illustrates the adopted lexical analysis procedures.

2.2.1 The Documentary Series

The television series selected as the input material for the present thesis were the four-episodes documentaries Wonders of the Universe (Cooter et al., 2011) and Forces of Nature (Cooter et al., 2016) (see Figure 2.1). Each episode was one hour long; that is, the total length of the treatment material was about eight hours. Both series were published as a book and DVD within the same broadcasting year.

Figure 2. 1

Covers of the Selected Documentaries

Apart from meeting the criteria set in Section 2.1.1, many factors formed the basis for selecting the material. Professor Cox is English, making written and spoken input more likely to conform to the default English language in the EFL Algerian curriculum. Secondly, Cox won a string of awards for his excellent work in communicating science to the general public. The sales of his series-based science books exceeded 1.3 million copies. Both series are factual, entertaining, engaging, and presented in a way accessible to the general audience and were considered among the BBC's most-watched programmes. They were also received as enjoyable by participants of the pilot study. The episodes revolve around the same theme, meeting the narrow viewing principle (Rodgers & Webb, 2011) that advocates the use of successive topic-related episodes. Finally, the channels that broadcast the series are unavailable in Algeria, reducing participants' probability of watching them. The rationale for the choice of documentaries is elaborated in Section 3.1.1.

Wonders of the Universe

Produced by BBC, Discovery Channel, and Science Channel, the series was first broadcast in the United Kingdom weekly from 6 to 27 March 2011 and averaged 6 million viewers a week. In every episode, Cox visits some of the most dramatic parts of planet earth to address profound questions about ourselves and the universe. A total of 95% of Google users liked the series, and 3840 IMBd users gave it an average vote of 8.9/10 (https://g.co/kgs/y5VrBK). Cox was named 'Best Presenter' at the Royal Television Society Awards (2011) and 'Best Performer' in a non-acting role at the Broadcasting Press Guild Awards (2011). The running time of its episodes, Destiny, Stardust, Falling, and Messengers, range from 58 minutes and 03 seconds to 59 minutes and 05 seconds, with the opening narration and closing credits excluded.

Forces of Nature

The series was co-produced by BBC Studios, PBS and France Télévisions and premiered in the United Kingdom weekly from 4 to 25 March 2016. This series combines stories about people in different parts of the globe with a deeper understanding of the natural forces that form the universe. It averaged 8.2/10 based on 496 IMBd users (https://g.co/kgs/Eb6caF). Director Cooter won 'Best Science Documentary' at the 45th Grierson British Documentary Awards (2017). The running time of its episodes, The Universe in a Snowflake, Somewhere in Spacetime, The Moth and the Flame, and The Pale Blue Dot is in the range of 58 minutes and 25 seconds to 58 minutes and 40 seconds, excluding the repeated opening narration and closing credits. The overall running time for the two series is 7 hours, 48 minutes, and 38 seconds. Details for episodes are shown in Table 2.5.

Table 2. 5

Length of Each Episode in the Two Documentary Series

	Wonders of the Universe				Forces of Nature			
Duration	EP1	EP2	EP3	EP4	EP1	EP2	EP3	EP4
Total	59:05	58:32	58:48	58:03	58:32	58:40	58:25	58:33
Opening narration	01:21	01:15	01:19	01:19	01:05	01:20	01:02	01:16
Closing credit	00:25	00:25	00:25	00:25	00:25	00:25	00:25	00:25

Note. EP = episode; Duration is in minutes and seconds.

2.2.2 Research Questions

To determine how far/close is the estimate of vocabulary size of third-year Algerian undergraduates in the BA Linguistics programme to the requisite vocabulary knowledge for 90% and 95% coverage of the selected series, Part 2 of the norming study addressed the following two research questions:

1. How many words do we need to know to achieve 90% and 95% coverage of the two full-length seasons of the documentary series?
2. What is the estimated English language vocabulary size of Algerian undergraduates in the third year of the BA Linguistics programme?

2.2.3 Participants

There were forty participants in Part 2 of the norming study. Participants were tertiary EFL Algerian learners who had recently finished their third year in the Linguistics Bachelor programme at the University of Jijel, Algeria (See Section 2.1.4). They were randomly selected from the same population of participants in Part 1 of this study but by the end of the 2016-2017 academic year.

2.2.4 Vocabulary Profile of Episodes

Transcripts of the episodes were analysed to estimate the number of words needed to achieve 90% and 95% coverage of the two selected documentary series. This section details the choices made in the lexical analysis of the materials. Previous studies have not made explicit decisions regarding how to deal with different features of captions to establish a valid lexical analysis.

Several sources of captions were reviewed. The most accurate transcripts based on the DVDs' captions (to be employed in the experiment) were found at https://isubtitles.org/wonders-of-the-universe/english-subtitles/794018 and https://www.fmsubs.com/tvseries/Forces-of-Nature-with-Brian-Cox/40603/ for Wonders of the Universe and Forces of Nature, respectively. Firstly, transcripts were proofread while watching their equivalent episodes before the analysis stage. A handful of spellings in few transcripts were automatically highlighted and changed to British English spelling in the Microsoft Word document. Timestamps and formatting were removed using the previously used online subtitle tool; a total of 153,126 extraneous characters were cleaned up. In addition, captions included many

types of non-speech information (NSI), which were treated differently depending on their anticipated effects on the analysis.

First, the analysis included three types of NSI. These are sounds effects, e.g., (HORSE WHINNYING); paralanguage which involves sounds emanating from speaker's vocal cords, e.g., (EXCLAIMING); and speaking manner identifiers, e.g., Look! (GASPS) Ooh! These are immediately recognised from the audio and are only meant for the deaf and hard of hearing. However, they were included because they constitute multiple words of disparate corpus frequencies that were to be processed. In contrast, language identifiers, e.g., (PARA KAPOONI SPEAKING OWN LANGUAGE) and speaker identifiers (as shown in Figure 2.2), were regarded as definite redundant words and were therefore excluded from the analysis. Including these two types in the analysis would have misrepresented the materials' running words. Using Microsoft Word "Replace All" function, a total of 671 words were deleted from transcripts.

Figure 2. 2

An Example DVD's Caption Showing Speaker's Identity

Moreover, a few episodes were found to contain a handful of open captions. Unlike captions for the hard of hearing (also called closed captions), open captions are an actual part of imagery in the video. In these series, they are translations for speakers who do not speak the English language. In Part 1 of this study, the analysis of episodes' transcripts aimed to extract from the series potential English language target words available in both written and spoken forms. As such, the issue of open captions was not addressed at that stage. However, the current analysis needs to include English language texts that will assist participants to comprehend the

material, regardless of the text's modality. Open captions were tracked, and a total of 2,141 words were added to the transcripts while watching episodes in a DVD Viewing Room at the University of York.

Finally, open captions suggest that part of the input will be in both English (open captions) and another language (audio). When such bilingual input is involved, the extent to which the overall study input amounts to an English language listening-while-reading could be questioned. Therefore, the number of words in open captions was calculated and compared to the total number of running words for each presentation session. The bilingual text amounts to less than one-twentieth of the overall material, except for the second session, which is slightly higher (6,94%), as shown in Table 2.6. Overall, the figures indicate that the materials represent an English language input for both listening and reading.

Table 2. 6

Open Captions' Word Percentage in Relation to the Total Running Words Per Session

Running words	U1+F1	U2+F2	U3+F3	U4+F4
Total	11,093	11,083	10,865	10,305
Open captions	509	770	495	367
Open captions %	4.58	6.94	4.55	3.56

Note. U = Wonders of Universe; F = Forces of Nature

Transcripts of the episodes were analysed using *Compleat Web VP* function from the *Lextutor*. The British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) are two corpora on which the text analysis could be based. The new combined BNC-COCA has been particularly useful in recent research for being more representative of word frequency. Hence, BNC-COCA-25 frequency analysis was employed, which breaks down texts into up to twenty-five 1,000-word frequency levels. In the analysis output text, *Compleat Web VP* replaces figures by the word *number*, contractions by constituent words (e.g., didn't => did not), and 'a' and 'I' are treated as words. In addition, words that are beyond the 25,000-word frequency level are classified by the tool as *Off-List.*

2.2.5 Vocabulary Size Test

The Vocabulary Size Test (VST) was used to estimate participants' vocabulary size (Nation & Beglar, 2007). The test was developed to accurately, reliably, and comprehensively measure written receptive English vocabulary of selected frequency levels (Beglar, 2010, p. 103). It is freely available to the general public from Nation's website at https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-tests or Cobb's at https://www.lextutor.ca/tests/vst/.

The VST monolingual 14,000 format was used. The test is available in the 14,000- and 20,000- word levels. The former was selected because tertiary EFL learners are unlikely to master words beyond this level. The monolingual version was used because the version is not available in participants' L1, Arabic. Also, the bilingual version is beneficial for lower proficiency EFL learners (Nation, n.d.). The VST was validated and found highly reliable (Beglar, 2010). It is more congruent with the high proficiency of tertiary EFL learners (Janebi Enayat, Amirian, Zareian, & Ghaniabadi, 2018, p. 12). It is straightforward to fill and easy to administer.

Procedure

The test consisted of 140 items presented in a multiple-choice format. Participants were required to select the best definition of each item from four choices, every 10 items represent knowledge of a 1,000-word frequency level based on the BNC lists. The paper version was administered along with refreshments. The test was conducted in one sitting and lasted for a maximum of 40 minutes (the time limit was not imposed). Figure 2.3 shows an example item from the 5,000-word level.

Figure 2. 3

An Example Item from the 5,000-Word Level Used in VST

Fifth 1000

1. DEFICIT: The company had a large **deficit**.
   a. spent a lot more money than it earned
   b. went down a lot in value
   c. had a plan for its spending that used a lot of money
   d. had a lot of money in the bank

Note. Adapted from "A vocabulary size test," by ISP Nation and D. Beglar, 2007, *The Language Teacher, 31,* Appendices.

Scoring

Participants' total score on the 140 items was multiplied by 100 to obtain their total vocabulary size. That is, a participant who responded correctly on 45 out of 140 items would have an estimated vocabulary size of 4,500- most frequent word families (i.e., 45 x 100).

2.2.6 Results

The requisite Vocabulary Size for 90% and 95% Coverage of Documentary Series

The vocabulary profile of the eight combined episodes was calculated; the results are shown in Table 2.7. The table displays the number of tokens, types, word families, and Off-List words per 25,000-word frequency levels. The words in the Off-List category were found to include proper nouns. These were manually identified and added to the analysis results. The cumulative coverage with and without proper nouns was calculated. It is worth noting that a variety of words such as lifetime, timescale, gunshot, snowman, seafood, forever, humankind, viewpoint, moonlight, which are likely to be known by participants, were classified by the analysis tool as Off-List. Thus, the figures were approximated. Furthermore, a double-check analysis excluding sounds effects and paralanguage captions was carried out, and results were fairly similar to those in Table 2.7.

The eight episodes of the series consisted of a total of 43346 words. As shown in Table 2.7, the first 1,000-word frequency level accounted for most tokens (81.63%), the second for 6.31 %, the third for 4.91 %, and the fourth for 2.31%. The remainder accounted for less than 0.90 % each. The percentage of proper nouns which occurred in the series was 1.26%. Suppose proper nouns were counted as easily understood. In that case, lexical coverage that is slightly below 90% (89.20) requires knowledge of only 2,000- most frequent word families, while 95% coverage might require knowledge of a little less than 4,000- most frequent word families. With proper nouns excluded, knowledge of 3,000 and 4,000- most frequent word families is requisite to achieve over 90% and just above 95% lexical coverage, respectively.

Table 2. 7

Vocabulary Profile of the Series: Tokens, Types, Word Families, and Cumulative Coverage with and without Proper Nouns by Twenty Five 1,000- word Frequency Levels

BNC word frequency level	Tokens		Types		Word families		Cumulative coverage %	
	No of words	%	No of words	%	No of words	%	without PNs	With PNs
1,000	3538	81.63	1647	38.52	811	34.13	81.63	82.89
2,000	2733	6.31	805	18.83	469	19.74	87.94	89.20
3,000	2129	4.91	555	12.98	363	15.28	92.85 [a]	94.11 [a]
4,000	1000	2.31	323	7.55	224	9.43	95.16 [b]	96.42 [b]
5,000	380	0.88	155	3.62	120	5.05	96.04	97.30
6,000	218	0.50	114	2.67	104	4.38	96.54	97.80
7,000	263	0.61	94	2.20	81	3.41	97.15	98.41 [c]
8,000	90	0.21	47	1.10	42	1.77	97.36	98.62
9,000	126	0.29	53	1.24	46	1.94	97.65	98.91
10,000	73	0.17	38	0.89	34	1.43	97.82	99.08
11,000	36	0.08	19	0.44	17	0.72	97.90	99.16
12,000	35	0.08	20	0.47	19	0.80	97.98	99.24
13,000	22	0.05	8	0.19	7	0.29	98.03 [c]	99.29
14,000	5	0.01	4	0.09	4	0.17	98.04	99.30
15,000	23	0.05	6	0.14	5	0.21	98.09	99.35
16,000	6	0.01	6	0.14	6	0.25	98.10	99.36
17,000	26	0.06	8	0.19	7	0.29	98.16	99.42
18,000	14	0.03	2	0.05	2	0.08	98.19	99.45
19,000	9	0.02	7	0.16	6	0.25	98.21	99.47
20,000	2	0.00	2	0.05	2	0.08	—	—
21,000	1	0.00	1	0.02	1	0.04	—	—
22,000	21	0.05	2	0.05	2	0.08	98.26	99.52
23,000	0	0.00	0	0.00	0	0.00	—	—
24,000	2	0.00	2	0.05	2	0.08	—	—
25,000	2	0.00	2	0.05	2	0.08	—	—
PNs	548	1.26	261	6.11	—	—	—	—
Off-list	199	0.46	95	2.22	—	—	—	—
Total	43346	≈100	4275	≈ 100	+2376	—	≈ 100	≈ 100

Note. PNs = Proper nouns.

[a] Reaching 90% coverage

[b] Reaching 95% coverage

[c] Reaching 98% coverage

The Estimated English Language Vocabulary Size of the Target Population

The vocabulary size of the target population was approximated using the VST results. Participants' scores ranged from 3500 to 9200 with a mean score of 5842 (*SD* = 1535.88). As Table 2.7 shows, just 3000- most frequent word families is more than enough to reach 90% coverage of the selected series. Consequently, the requisite coverage of the series was well within the capacities of target learners.

2.2.7 Discussion

Part 2 of the norming study aimed to investigate whether the population targeted in this thesis has the vocabulary knowledge critical to comprehend the selected science series.

How many words do we need to know to achieve 90% and 95% coverage of the two selected documentary series?

To estimate the vocabulary size necessary for viewing comprehension, lexical analysis of the text in the episodes using *Compleat Web VP* function from the *Lextutor* was conducted. The results that 95% coverage require knowledge of 4,000-most frequent word families, and a little less than this level with proper nouns included, are slightly higher than the 2,000 – 3,000 (plus proper nouns) results found for spoken short stories (Van Zeeland & Schmitt, 2013), spoken academic English (Dang & Webb, 2014), and songs (Tegge, 2017; Romanko, 2017). However, they are relatively consistent with previous results from films (Nation, 2006; Webb & Rodgers, 2009a) and science talks in general (Nurmukhamedov, 2017).

Though it is not necessary to reach 98% coverage to ensure comprehension of the series under review, it is interesting to note from the results that it is highly significant for the learner to know proper nouns to reach this ideal lexical coverage. Assuming that proper nouns are known, the two full-length seasons of these science documentary series demonstrated lexical demands that are as low as 7,000- most frequent word families. Although the result is in line with findings for spoken text (6,000-7,000; Nation, 2006) and British movies (7,000; Webb & Rodgers, 2009a), studies of a similar genre (i.e., science register), showed higher demands with proper nouns included 9,000 (Coxhead & Walls, 2012) and 10,000 (Nurmukhamedov, 2017). If proper nouns are unknown, 7,000 to 13,000- most frequent word families are needed for a lexical coverage between 97% and 98%. This result is consistent

with results for life and medical sciences (13,000) and physical sciences (10,000) (Dang & Webb, 2014) as well as corpora of different genres (5,000 to 10,000; Webb and Rodgers, 2009). Hence, knowledge of proper nouns is crucial for advanced comprehension of the selected materials.

What is the estimated English language vocabulary size of third-year Algerian undergraduates in the BA Linguistics programme?

In answer to the second question, participants had a minimum mean score of 3,500-most frequent word families. The results of this analysis were then compared to the vocabulary profile of the series. The result was that the coverage of the series was well within the learners' capacities.

The experimental research was carried out at the beginning of the academic year. In contrast, Part 2 of the norming study was conducted at the end of the academic year, and participants were, therefore, at an advanced English language level. Despite this limitation, the resulting vocabulary size was large enough to expect the knowledge of a minimum of 4,000- most frequent word families from the part of the target population. Furthermore, the presence of L2 captions is expected to facilitate participants' comprehension. Overall, the results prove the adequacy of the selected materials to the population targeted in the thesis.

2.3 Conclusion

To ensure the reliability and validity of the studies in this thesis, a two parts norming study was conducted. In the first part, eight episodes of two BBC documentary series, *Wonders of the Universe* and *Forces of Nature*, were selected to serve the aim of the thesis. A total of 54 potential words were initially extracted from 18 episodes of seven series. Using a modified VKS, the words were tested on 150 participants selected from the population of tertiary Algerian EFL learners in their third year of the Linguistics Bachelor programme at the University of Jijel, Algeria during the 2016-2017 academic year. Score 2 and 3 referred to written form recognition and meaning recall, respectively. A total of 28 items were finally selected: 20 spaced words (6 adjectives, 6 verbs, 8 nouns) and 8 massed words (8 nouns).

The second part of the norming study aimed to determine whether vocabulary in the selected documentary series allows a 90% to 95% coverage from Algerian

third-year university EFL learners. Transcripts of episodes were analysed to establish their lexical demands. The results showed that 3,000, 4,000 and 7,000-most frequent word families (plus proper nouns) provide over 90%, 95%, and 98% coverage of the transcripts. The vocabulary size of 40 participants of the same population was then estimated using VST. Results have shown that the transcripts were within the learners' lexical competence and that the selected series represents a suitable material for the target population.

The findings in this norming study are subject to three limitations. First, it is unfortunate that a norming study of this type of research cannot be performed directly on the target population due to time constraints. At the time of decision making, the target participants in the thesis were enrolled in the second-year undergraduate programme. Furthermore, the small sample size (i.e., forty) in Part 2 makes the findings less generalisable to the target population. A final limitation in this study and similar studies that cannot use pseudowords is the inclusion of cognates. Among the 28 selected words in the study, 18 were cognates with words in French, the second language of target learners. As will be explained in Chapter 3, Section 3.1.6, selecting cognates is inevitable in studies of this type. Notably, Part 1 results illustrate that the cognate status of words does not always inform its degree of knowledge in L2 learners, since many cognates were unknown by most participants.

Chapter 3

Study 1. Modality Effects in Learning

Modality is the way of representing information in a medium (Bernsen, 2008) and it could be visual, aural, or read. Multimodality refers to the integration of multiple modalities to represent an input that is appealing to learners' different sensory modalities. In a multimodal learning-based approach, input modalities differ in their expressive strengths and learners' perceptual, cognitive, and affective systems (Tzovaras, 2008, p. 24). The present chapter discusses the potential strengths of imagery as a visual representation within the context of extensive viewing of L2 captioned documentary series. It consists of an empirical study based on a direct comparison between two combinations of input modalities, one with imagery and one without imagery.

Within the last two decades, learning from audio-visual input in general and L2 captioned video, in particular, has become an increasingly important area of focus in the field of second language acquisition. Vocabulary researchers especially documented the significant learning effect of viewing L2 captioned video on the incidental acquisition of words. However, a limited number of studies have examined vocabulary learning outcomes from extensive viewing, and these have tended to focus on the benefits of bimodal input. Only a few researchers have explored how imagery contributes to vocabulary learning from viewing. This study seeks to obtain data that will help to address these research gaps. It first investigates whether viewing two full-length seasons of L2 captioned documentary series benefits incidental learning of L2 vocabulary at four levels of word knowledge, meaning recall and recognition, and spoken and written form recognition. The study then further explores the role of imagery in producing this learning effect by comparing a View condition to a Non-View condition.

3.1 Visual Modality Effect in Incidental L2 Vocabulary Learning from Extensive Viewing of L2 Captioned Video

This review will first demonstrate the potential benefits of extensive viewing on incidental vocabulary learning and the few prominent studies highlighting this theme. It will be argued that research is warranted to investigate the influence of +7 hr viewing of L2 captioned documentary series for recreational purposes (i.e., in conditions resembling out-of-class viewing) on incidental acquisition of L2 vocabulary. Section two, 'The Value of Bimodal Input', will describe the evidence base for the effect of listening-while-reading on incidental L2 vocabulary learning. Section three will argue that imagery has a significant role in incidental vocabulary learning from L2 captioned video. It will discuss and identify the few studies that addressed this role through comparing View (audio + caption + imagery) to Non-View (audio + caption) conditions. The reader will then encounter the theoretical underpinnings that inform how the presence and the lack of imagery affect incidental learning of different aspects of word knowledge from L2 captioned video. Finally, the review will end with a description of the dependent measures and the word-related variables considered in this thesis.

3.1.1 Extensive Viewing of Documentary Series

The present section will highlight the need to explore L2 vocabulary learning outcomes from sustained exposure to multimodal input for recreational purposes. It will address specifically the narrow viewing principle; that viewing multiple videos of related content is beneficial for word learning, and the benefits of watching documentaries. It will then review studies based on their adopted exposure length. The final subsection will draw attention to two methodological considerations: interstudy interval and session length.

Watching television such as films and series is one of the world's favourite pastimes. This fact has been more apparent as Netflix, the most popular TV streaming service, surpassed 200 million subscribers after the COVID-19 pandemic hit the world in early 2020. This massive viewership may indicate that people usually experience prolonged exposure to TV programs in their leisure time, which suggests the greater influence watching television could exert on language learners. It has become increasingly evident that exposure to second language vocabulary in

the form of multimodal input, mainly L2 captioned video, improves vocabulary retention (e.g., Majuddin, Siyanova-Chanturia, & Boers, 2021; Markham, 1999; Montero Perez, Van Den Noortgate, & Desmet, 2013; Price, 1983; Teng, 2020; Vanderplank, 1988, 2016). Nonetheless, the majority of studies fail to capture the potential to acquire words longitudinally. What we know about incidental vocabulary learning from multimodal input is primarily based on studies offering minimal input (e.g., Aini, Jelani, & Boers, 2018; Peters et al., 2016; Peters & Webb, 2018; Winke, Gass, & Sydorenko, 2010, Peters, 2019). Most of these studies have restricted input to one hour or less. In real-life scenarios, however, learners would continue watching TV on a weekly or even daily basis. This viewing could be of selected films or continuous episodes of favourite TV shows or documentary series, of which seasons could last over a few months or even years. Hence, further interventions that accurately reflect real exposure conditions are needed to gain deeper insights into incidental vocabulary learning.

Webb (2015) defined extensive viewing as "regular silent uninterrupted viewing of L2 television inside and outside of the classroom"(p. 159). Some of the documented merits of extensive viewing include developing listening skills, increasing motivation to learn, and extending vocabulary knowledge (Rodgers, 2016) due to the constant repetition of words across and within viewing sessions. Given these potential benefits, extensive viewing has been proposed as an approach to L2 vocabulary learning by using television programs as the core material (Webb, 2015; Webb & Rodgers, 2009b). Webb (2015) discussed two forms of extensive viewing programs: classroom-based extensive viewing and out-of-class extensive viewing, and suggested that the latter should be the more promising.

Extensive Viewing for Recreational Purposes

Based on three concepts, the present study attempts to provide empirical evidence for vocabulary acquisition through extensive TV viewing as a recreational activity. These concepts are out-of-class extensive viewing (Webb, 2015), incidental learning, and invisible learning that dictates that: "we learn more, and do so invisibly, when we separate structures of control that restrict freedom and self-determination from learning experiences" (Moravec, 2016).

Firstly, out-of-class viewing is one of the most popular leisure activities among L2 learners and has been shown to positively correlate with vocabulary knowledge (e.g., Lindgren & Muñoz, 2013; Peters, 2018). The present study treats successive episodes of documentary series as extensive stretches of contextualised language input that support vocabulary learning as an incidental outcome of input comprehension. Within this framework, viewing comprehension could be supported by post-viewing tasks such as comprehension questions, episodes reviews, role plays, as long as these tasks do not interfere with learners' enjoyment nor comprehension.

Secondly, the current study investigates to what extent "minimal users" who view TV for recreational/comprehension purposes rather than for performing specific pre-determined tasks could gain vocabulary knowledge incidentally "aided by an intelligent use of inference" (Seibert, 1945, p. 296). Of particular relevance to the present study design is a longitudinal but qualitative study by Vanderplank (2019) that explored students' attitudes and learning behaviour in extensive informal viewing. What distinguishes this study from others is that participants had control over which programme to watch and how to watch it. Vanderplank invited 36 L2 learners to participate in a self-paced viewing programme as part of the EURECAP Project. Participants were requested to watch at least one film every week over six weeks, then over 12 weeks, while writing diaries on their viewing behaviour. The films included optional captions and were selected from a range of their own preferred films. Prolonged exposure to films revealed three types of viewers: Minimal users who minimised interruption to maximise enjoyment, evolving users who adjusted to the use of captions and adopted strategies over time to maximise both language learning and enjoyment, and maximal users who intentionally watched films to learn the language.

I put forward the view that extensive viewing maximises opportunities for serendipitous discoveries, that may represent the core foundation for incidental vocabulary learning. To this end, viewers may effectively learn vocabulary from extensive viewing without necessarily demonstrating explicit vocabulary learning strategies. Many Modern Language students among Vanderplank's participants, who were expected to demonstrate higher intrinsic and extrinsic motivation to learn, were identified as minimal users who "… reported a strong preference for watching

films ‘as films’, a reluctance to treat films as language learning resources” (p. 418). It may be true that “spending hours letting captioned films run and pausing only occasionally may not benefit a second language viewer who has the goal of developing their language knowledge and skills” (p. 418). It should, however, be considered that there is a variety of ways in which students deal with unknown words, and taking notes and looking up their meanings in the dictionary is just one of them. We should not leave out of consideration that no two students are the same and that learning is the natural outcome of disparate complex strategies available to the learners and which they may use either consciously or unconsciously.

It is unknown whether the “minimal users”, who favour enjoyment over interruption, include active thinkers who effectively use cognition to gain knowledge and understanding. As Seibert (1945) maintained: “a large part of the new acquisitions have been made mainly [incidentally] through the art of inferring the meanings of the unknown words from the context” (p. 296). Thus, students may frequently implement internal mental processes, such as contextual guessing from the multimodal input, which is conducive to incidental vocabulary learning. This use could make them appear more inclined to enjoy the films and less dependent on dictionaries and note-taking to understand and remember words. Although all support strategies are strongly linked to successful learning, contextual guessing and reliance on memory involve more complex thinking processes compared to dictionary use, by which fewer efforts are made. Hence, it could be that learners who experience a higher cognitive involvement load are more likely to retain words compared to learners who passively look at meanings (Hulstijn & Laufer, 2001).

While learning can be attained incidentally as a result of enjoyment, enjoyment can be lost due to imposed learning, such as having to chase answers to pre-determined vocabulary tasks. Recent evidence suggests that pre-teaching items prior to viewing benefits vocabulary acquisition (e.g., Mazahery et al., 2021; Suárez & Gesa, 2019). Nevertheless, Mishan (2005) cautioned against an overemphasis on pre-teaching because it may “… induce learners to listen [to] them, which can interfere with their comprehension of the whole.” (p. 217). Rather, multimodal input can by itself be an effective medium for the illustration of new word meanings.

Finally, the current study's advocacy of extensive viewing for entertainment and enjoyment concords with the concept of invisible learning (Romaní & Moravec, 2011). Out-of-class extensive viewing removes structures of control which in turn opens up opportunities to acquire vocabulary. As Moravec put it, there is: "the false assumption that students will not learn unless they are told what to learn…. Learning may blossom when we eliminate authoritarian control or direction of a learning experience by an 'other'".

In sum, this section highlighted that the current study's argument for extensive viewing for entertainment and enjoyment concords with three concepts. These are out-of-class extensive viewing, incidental learning, and invisible learning.

Narrow Viewing

Viewing successive episodes of television programs such as documentary series, which are the focus of the current study, might optimise extensive viewing outcomes. A potentially beneficial genre of viewing, known as narrow viewing (Rodgers & Webb, 2011), often involves episodes that feature the same characters (e.g., drama series) or share a similar genre and revolve around a central theme. Rodgers and Webb indicated that television programs expose L2 learners to repeated occurrences of both low- and high-frequency words. More importantly, in contrast to watching films and single episode programs, narrow viewing material consists of few word families. Narrow viewing further enables learners to become better acquainted with the material's vocabulary, characters, and narrative, facilitating comprehension that supports acquiring unknown words (Rodgers, 2016). These proposed benefits draw on research on narrow reading (Krashen, 1981), which has shown a lexical advantage from reading texts of similar genres or topics (e.g., Schmitt & Carter, 2000). In line with the narrow viewing principle, it could also be argued that learners should benefit from the vocabulary in documentary series if episodes share the same scriptwriter and presenter (as is the case in the present study). Recent evidence supports this, suggesting that learners are more likely to acquire words from texts by the same author (Chang & Renandya, 2019). A search through the literature revealed that most researchers studying extensive learning from viewing have adhered to narrow viewing principle, by presenting episodes of the same TV series.

Documentary Series

While viewing research has employed thus far various genres of TV programs, the rationale for choices of the genre is not usually well explained in the literature. In this chapter, I argue that the impact of watching complete documentary series on incidental L2 vocabulary learning is understudied. Research has shown potential for documentary series to facilitate foreign vocabulary learning; however, the available lines of evidence have been based on minimal input (e.g., Alshumrani, 2019; Feng & Webb, 2020; Peters et al., 2016; Peters & Webb, 2018). As mentioned above, a literature search revealed only a few studies that investigated the effect of sustained exposure to multimodal input on incidental L2 word learning. These studies have been limited to drama TV series. To the best of my knowledge, no single study exists which examines such learning effects from large amounts of input from documentary series.

A survey of documentary audience viewing habits in Canada suggests that regular documentary viewing is rising (Hot Docs, 2018). Of the 3607 respondents, 72% watch documentaries no less than twice a month (3 times ≥ for 35%). Due to more documentaries being on offer, 55% of audiences watched more documentaries than three years earlier. As for viewership sources, more than 94% of participants who responded to the survey indicated that they watched documentaries at home. Commenting on these results, the authors pointed out:

> "Documentary viewership remains strong. Interestingly, the way viewers are consuming the content is changing. Movie theatres and film festivals, for example, are losing ground to the myriad ways one can easily – and relatively cheaply – stream an ever-growing amount of documentary content from the comfort of one's home" (p. 9).

Television remains the most popular device to watch a documentary. Regarding streaming platforms, Netflix came first at 72% viewership, while YouTube came second at 54%. The study also showed that popular social media platforms had increased the likelihood of documentary viewing. Facebook and Twitter accounts have become "a viable promotional and discussion vehicle for documentaries" (p. 10), with 73% of users among respondents indicated that they tended to search about newly released documentaries shared by their friends, and

53% of Facebook users tended to post, like, and share information about a documentary.

In sum, viewing studies have employed different genres of TV viewing; however, they have not specified the reason for their selection, for instance, the reason for choosing drama TV series over films or documentary series. In this chapter, I argue that an observed increase in desire for documentary content and a paucity of evidence for vocabulary learning from successive documentary viewing sessions support my interest in researching incidental vocabulary learning from this valuable source of input.

Length of Exposure

Based on my vocabulary literature search, I identified three categories of TV viewing empirical studies which varied remarkably in length (+3 hr, +5 hr, and +7 hr). While there has been little agreement on which duration constitutes extensive exposure, the present study postulates that the optimal use of this type of viewing is the longest (e.g., +7 hr). Extensive viewing is, by definition, a type of viewing that offers a substantial amount of input; hence, the more input the student receives, the greater the learning outcome. This suggestion aligns with meaning-focused input, one of four strands of an effective language programme (Nation, 2007). The strand involves learning receptively through extensive reading, listening, and viewing. The duration also accords with Peters' finding (2018) that 40% of participants reported watching L2 programs and films several times a week (i.e., extensively) in their leisure time.

Most extensive viewing studies fall into the +3 hr viewing category. The studies yet differ in design and objectives. Some studies were oriented towards out-of-class viewing. In a recent study, participants of B1 to B2 level of English language were requested to watch L2 captioned episodes of an American documentary series (3 hours) over three weeks without controlling for the viewing time (Sinyashina, 2020a). The study aimed to compare between incidental + intentional (N = 17) and intentional + incidental (N = 13) learning practices for learning 16 words. Overall, descriptive data showed remarkable vocabulary gains for both groups by the end of the study period. In a similar study of a relatively longer TV exposure (4 hr 30 min), students of BA level were divided into three

groups of 30 participants each and were required to watch episodes of a British comedy series at home as part of their homework (Zarei, 2009). The study examined the effect of different types of soundtracks and subtitling on L2 vocabulary learning using a 40-item multiple-choice test. The results showed a significant effect for viewing L2 captioned videos on vocabulary learning.

The most recent study was similar to Zarei's in exposure length (4 hr 40 min) and in one focus (subtitles) (Mazahery et al., 2021). The study examined various types of subtitling as well as pre-teaching on vocabulary learning from an American crime TV series. The average viewing length was 20 minutes. The study was conducted on intermediate Iranian EFL learners assigned to four groups of 20 participants each. The authors reported that extensive exposure to the series increased participants' form recognition and meaning recall at the posttest. Another study that had a pre-teaching focus investigated the effect of sustained exposure to L2 captioned TV series on learning 40 pre-taught words using a between-subjects design (Suárez & Gesa, 2019). The author used episodes from a TV series totalling 3 hr 16 min of viewing over 11 weeks, with an average viewing length of 24 min 30 s per week. Written form and meaning recall results from the 117 Catalan-Spanish bilinguals revealed an influential role for viewing L2 captioned video in this formal context of vocabulary instruction for less proficient learners. It should be noted that the context of this study was less incidental given that target words were pre-taught and focused on in post-viewing.

Fewer studies fall into the +5 hr category. One study consisted of a more extended viewing period (5 hr 25 min) than the four studies reviewed earlier. Nonetheless, the longitudinal data were obtained for the similar purpose of assessing other aspects related to the development of word knowledge (Frumuselu, De Maeyer, Donche, & Colon Plana, 2015). The researchers explored the effects of L1 and L2 captioned videos on L2 learners' acquisition of informal words from episodes of An American comedy series. They employed a between-subjects design to 40 university students of multiple English language proficiency (L1 group, N = 18; L2 group, N = 22). Viewing sessions lasted 25 minutes each and were conducted over seven weeks. Meaning recognition and recall findings from pretests and posttests demonstrated more improved performance in the L2 captions group than the L1 captions group. The studies discussed so far mostly lacked control groups or

statistical data intended to answer specific research questions regarding the impact of extensive viewing on vocabulary learning.

An exclusive focus on the influence of TV viewing on vocabulary learning is warranted. One recent study fulfilled this purpose (Sinyashina, 2020b) using an exposure duration similar to that of Frumuselu et al. (2015). The author assigned Spanish participants who were intermediate L2 learners of English language to either an Intentional (N = 11), Incidental (N = 12), or Control (N = 9) group. Incidental group participants were requested to watch a TV series at home at their own pace over three weeks. Episodes consisted of 10 target words of 1 to 5 verbal frequency range. One-week delayed tests revealed a lack of acquisition of words following the viewing. Sinyashina speculated that +5 hr exposure might have been insufficient to achieve vocabulary gains. She also attributed the negative results to differences in the selection of input and words as well as participants' related variables, compared to other studies. For instance, target words occurred infrequently in the input, which might have been less comprehensible for the target sample.

The studies described above are subject to certain limitations. They were carried out on a small group of learners (e.g., Frumuselu et al., 2015). Some others lacked information on when exactly participants watched episodes, which could have affected acquisition measures (e.g., Sinyashina, 2020a). To illustrate, viewing occurred at random; thus, it was unknown whether participants viewed the overall episodes in one day, over days, or weeks. Also, if watched in one day, it was unknown whether it was the first day or last day of the intervention. Moreover, learning was not entirely incidental in some studies (e.g., Mazahery et al., 2021). Pre-teaching makes vocabulary gains more likely to result from an eagerness to reinforce knowledge of pre-taught words rather than the outcome of a need to comprehend input. The present study aims to examine incidental vocabulary acquisition from TV viewing as an extra-mural activity.

Another shortcoming of previous research into extensive television viewing and vocabulary learning is the limited input. While studies in the +3 hr and +5 hr categories are many and few, respectively, studies in the +7 hr category are rare. Two studies that fall under the latter category were identified. Rodgers & Webb (2019) examined the potential of viewing 10 episodes of an American TV series

(over 10 weeks) in enhancing knowledge of 60 target words. Participants were undergraduates of pre-intermediate to intermediate English language proficiency. The strengths of this study included the large sample size; there were 73 participants in the control group and 187 in the viewing group. Moreover, target words occurred from 5 to 54 times, which is a diverse frequency range. In addition, a one-week interval was generally allowed between viewing sessions. Each session had an average viewing length of 42 min 49 s while viewing time was consistent across all participants. Vocabulary knowledge was measured at the level of a tough multiple-choice test, in which distractor and target items matched in aspects of form and meaning, and a sensitive test, in which they did not match. The study found that viewing the 10 episodes series resulted in the acquisition of 25% of target words.

Another study that falls into the + 7 hr category was undertaken over eight months (Pujadas & Muñoz, 2019). The researchers investigated the impact of viewing 8 hr 35 min of L2 captioned and L1 subtitled American TV series on beginners' acquisition of L2 vocabulary. They performed a pretest-posttest designed experiment with 80 Catalan-Spanish bilinguals of beginner English language proficiency and whose age range fell between 13 and 14 years old. Participants watched episodes over 24 viewing sessions of about 20 minutes length. A total of 120 target items occurred through the episodes, which were viewed either with L2 captions and focused instruction (N = 22), L2 captions and non-focused instruction (N = 22), L1 subtitles and focused instruction (N = 19), or L1 subtitles and non-focused instruction (N = 17). The results revealed significant vocabulary gains for all groups, suggesting that extensive viewing of L2 captioned video develops vocabulary knowledge.

The previous study (Pujadas & Muñoz, 2019) highlights the potential trade-off between the number of target items and the frequency of occurrence when designing a study. While the authors targeted a high number of items (N = 120), most of these (75%) occurred only 2 to 5 times throughout the intervention, while 14 occurrences marked the maximum frequency of occurrence. A high number of words tested is indeed pivotal to generalise results; nevertheless, for reliable learning of different lexical aspects, words need to be met several times (e.g., Pellicer, 2016). Hence, it could be postulated that the quality of occurrences equally matters. The number of words targeted in the present study is not ideal relative to the previous

investigations (i.e., Pujadas & Muñoz, 2019; Rodgers & Webb, 2019); nonetheless, 60% of words occurred from 10 to 17 times, while 20% were in each of 24 to 40 and 8 to 9 ranges of occurrences.

Methodological Perspectives in Viewing Research

As indicated above, recent developments in the field of incidental vocabulary learning from video have led to a particular interest in vocabulary acquisition from prolonged exposure to multimodal input. The present section introduces two essential viewing aspects pertaining to methodology, which current research discussions might have overlooked or have not explicitly addressed.

Interstudy Interval.

A critical factor in extensive viewing research is the interstudy interval. Previous studies implemented a one-week interval. The present study adopts two weeks as an interstudy interval to consider both extensive viewing and spaced learning. As its name implies, extensive viewing is viewing that occurs repeatedly at short intervals. An optimal extensive-viewing programme is one with intervals that allow word forms to be revisited before they are forgotten. On the other hand, there is evidence to maintain that the longer the spacing interval, the more likely input would be retained (Bahrick, Bahrick, Bahrick, & Bahrick, 1993; see Study 3, Chapter 5, Section 5.1.3 for a review). As such, two weeks were used on the postulate that it is neither too short nor too long to account for extensive viewing and spaced learning.

Session Length.

Another vital factor in extensive TV viewing research is the length of the viewing session itself. In this study, I suggest that an optimal duration for sessions should generally be no less than an hour. The suggestion is premised on the fact that the ecological validity of research designs for studies on extensive viewing is higher when the design includes multiple long sessions. Thus far, there is no consensus on the optimal length of sessions in extensive viewing research, and it seems unlikely that it would be determined scientifically due to uncertainty regarding what is perceived as extensive to different learners. However, though researchers suggest that prolonged exposure is most effective, they have been mainly interested in viewing sessions that were shorter than 40 minutes. In fact, the average viewing length per session in the previously reviewed studies ranged from 20 to 30 minutes,

with only Rodgers and Webb (2019) reaching an average running time of 42 minutes.

The present study implements viewing sessions that are two hours long each (i.e., two episodes) for various reasons. Firstly, studies have usually mentioned avoidance of viewer's fatigue and loss of interest as reasons for their selection, when in fact, TV viewing is an activity that is considered engaging and entertaining. It should not necessarily be compared to regular classroom tasks that can cause frustration. More important than the session length may be the type of input. Students' perception of the viewing length is dependent on what they view. Their attention is likely to be sustained if they are presented with highly motivating input, and their brain would make them feel as if less time has passed than actually has (e.g., Soares, Atallah, & Paton, 2016). For instance, a high grossing film could make viewing appear shorter to students than watching a political debate. Thirdly, irrespective of the context to which previous studies attempted to generalise (i.e., classroom-based or out-of-class viewing), their research designs reflected a classroom-based approach that kept input to a minimum per session. The current study aims to explore the potential of real-life learning experiences that occur informally at home. The aim can only be achieved by implementing long sessions which reflect more out-of-class viewing.

There is a range of evidence to suggest that long viewing sessions are typical. Students are usually adjusted to lengthier format, since most films are 80 to 120 minutes long (Jarząbek, 2018). Although 45 minutes is considered to be the maximum for most episodes of American TV series, the advent of view-on-demand and boxsets means that viewers are less likely to watch only one episode per day. A survey on television consumption in seven countries by Statista (2018) revealed that viewers in the United Kingdom aged between 18 and 24 years old watched 3 hours daily. The minimal daily viewing in this age group was 1.28 hours in Japan, without including the amount of viewing taking place via other streaming options such as Netflix and YouTube. In addition, episodes of French-made TV series are usually 50 minutes long (Morin, 2015), while episodes in Turkey, which is second in worldwide TV series distribution after the United States (Bhutto, 2019), typically run for a minimum of 2 hours and are broadcast weekly.

Overall, incidental learning effects from actual home viewing has received scant attention in research. Studies have been largely restricted to practices that fit into the classroom. In the present study, I use two-hour long viewing sessions for the reasons discussed above. The results will likely inform future classroom practices, including implementing out-of-class viewing activities as part of the EFL programme.

In conclusion, this chapter has so far focused on the potential of extensive documentary viewing in general and related aspects. The following two sections will help the reader appreciate the value of bimodal input and imagery, before discussing the theoretical principles for the consequences of keeping and obscuring imagery in L2 captioned video.

3.1.2 The Value of Bimodal Input

Researchers of the effect of input on vocabulary learning do not emphasise on the mechanisms leading to vocabulary learning from bimodal input. The function of this section is to highlight the theoretical foundation that explains the positive impact of L2 bimodal input on L2 vocabulary acquisition. I will first provide a historical overview of bimodal input research. I will then examine the research evidence for its effect on L2 vocabulary learning. The section will end with a discussion of a number of factors that characterise bimodal input and which can be considered in this study as conducive to L2 vocabulary learning.

Historical Overview

The term bimodal input is often referred to as the state of having two modes of input, mainly, spoken and written input. In the literature on bimodal input, two expressions are used, often interchangeably, by researchers in the field: reading-while-listening and listening-while-reading. Since captions in L2 captioned video refer to a supplementary material to audio-visual input, with audio representing the primary input, the expression reading-while-listening is used in this thesis when I refer to this type of bimodal input.

Early research on bimodal input was mainly rooted in attempts to improve L1 reading skills and reading fluency. Chomsky's research in 1976 is among the earliest of these studies. In an attempt to increase reading fluency, she asked third

grade slow readers to read texts from a book while listening to their audio recordings simultaneously. This strategy helped improve confidence and motivation in readers and increase their fluency "with apparent ease" (Chomsky, 1978). Chomsky's results inspired other fellow researchers to use bimodal input as a remedy for poor readers. Examples include Carbo's "Making Books Talk to Children" (1981), where she introduced ways to record storybooks in order for readers to listen while reading.

Findings from other studies were, however, at odds with Chomsky's. One pretest-posttest designed study compared the effects of two modes of input: reading-only (n = 10) and reading-while-listening (n = 10), in increasing L1 reading fluency for third-grade students of varying reading ability levels (Rasinski, 1990). Both practices improved reading speed and word recognition, based on results from reading two passages of 100 words each. However, no significant difference was marked between the two practices. Likewise, a later study found that repeated-reading and listening-while-reading worked equally well to increase the reading fluency of adults with deficits in L1 reading skills (Winn, Skinner, Oliver, Hale, & Ziegler, 2006). The authors operationalised reading fluency as words correctly read per minute and errors per minute. Although this study was limited by its low sample size (n = 12), it was in line with previous findings that reading-while-listening may not be as good as reading aloud since the former necessitates a faster reading rate (Skinner et al., 1993). Overall, early examples of research into the influence of bimodal input on L1 reading have shown mixed results.

In contrast, results from L2 reading studies have generally been promising. One study on young learners lacked a favourable outcome for bimodal input (Tragant Mestres et al., 2018). Nevertheless, a similar study reported superior reading fluency and comprehension scores for bimodal input than unimodal input (Llanes, Tragant, Pinyana, Cerviño-Povedano, 2016).

Adults seem to show benefits from bimodal input. For example, a study compared the effect of two modes of input: listening-only and reading-while-listening, on university learners' listening gains and perceptions (n = 84) (Chang, 2009). Results showed that learners gained 10% more with bimodal input and showed a strong preference towards it. Subsequent studies have shown that, compared to unimodal input, reading-while-listening significantly facilitates reading

and listening fluency (e.g., Chang & Millett, 2014) as well as reading comprehension (Chang & Millett, 2015).

These positive results of adults' L2 reading and listening skills suggest that bimodal input may also assist L2 vocabulary learning. As L2 learners develop their speed of lexical access (fluency) and comprehension in the language, they can cover more text and acquire automaticity in word decoding and recognition which may eventually lead to further vocabulary learning. A brief account of incidental L1 vocabulary learning research is provided next, before proceeding to review L2 studies.

Incidental L1 Vocabulary Learning Research

There is an ongoing debate about the relative importance of reading-while-listening in the vocabulary learning literature, as studies seem to offer contradictory findings. In general, early examples of research include studies that attempted to assist L1 poor readers. For instance, Meyer (1982) assessed whether listening to recorded tapes while following the print with a finger and reading along with the tape contributed to vocabulary acquisition (and text comprehension). Based on evidence from disabled adults (n = 20) who participated in a pretest-posttest designed experiment, she reported that the practice was an effective vocabulary building strategy. In contrast, opposite results were found in a study with children (Dowhower, 1987). For seven weeks, elementary students (n = 17) read five story texts in one of two conditions of repeated-reading: reading-only and listening-while reading, using a pretest-posttest design. Dowhower measured word recognition (accuracy): the number of words identified correctly for each test passage (in addition to reading rate, comprehension, and prosodic measures). No significant difference in the number of words recognised was detected. This result has been replicated in subsequent studies (e.g., Reitsma, 1988) but not in others (e.g., Valentini, Ricketts, Pye, & Houston-Price, 2018) where a superiority for reading-while-listening practice has been obtained.

Incidental L2 Vocabulary Learning Research

L2 research studies on the impact of bimodal input on incidental vocabulary learning have also produced inconsistent results. Similar to L1 research designs, studies have more frequently adopted empirical comparisons on the effect of reading-only,

reading-while-listening, and listening-only. The most recent study revealed equivalent word gains in reading-only, reading-while-listening, or reading with textual input enhancement groups (Vu & Peters, 2020). Participants were beginner and intermediate L2 learners (N = 60) and read four graded readers over four weeks. The authors surmised that the bimodal input might have been against learners' preferences or pushed them into cognitive overload that distracted them from noticing new words. A similar finding was previously reported (Brown, Waring, & Donkaewbua, 2008). The study compared vocabulary learning effects of bimodal input to reading and listening modes in a pretest-posttest design. It implemented three graded readers and included 35 university students of pre-intermediate to intermediate English language level. Meaning recognition and recall tests revealed that the reading-while-listening group significantly outperformed the listening group but not the reading group.

Further null results have been found in studies with children. Examples include an extensive study in which EFL learners were tested on knowledge of meaning recognition of 50 items that were encountered in graded readers (Serrano, Andriá, & Pellicer-Sánchez, 2016). In another study two years later, an advantage for bimodal input over unimodal input in children was not found (Tragant Mestres et al., 2018). However, following an intervention programme with young EFL learners, participants reported having learnt as much vocabulary as in a traditional class or more (Tragant, Muñoz, & Spada, 2016).

Notwithstanding the above unfavourable outcomes, an extensive body of work has demonstrated an advantage for bimodal input. Here I report the design and findings of two studies comparing reading-only and reading-while-listening on incidental L2 vocabulary acquisition. Firstly, the study by Webb and Chang (2012) was based on repeated-reading of 28 short texts. Data were collected from 82 adolescent learners in a pretest-posttest design over a two seven-week period using modified VKS. Form and meaning results indicated that participants in the reading-while-listening group significantly outperformed the reading-only group. Similar results were obtained in a similar study with university L2 learners (N = 60) encountering 24 unknown words in graded readers, based on form, grammar, collocation recognition, and meaning recall results (Teng, 2016).

In addition to the previous approach, some authors have demonstrated the positive influence of reading-while-listening on L2 vocabulary acquisition using a control condition instead of a comparison condition of unimodal input (e.g., Tangkakarn & Gampper, 2020; Webb & Chang, 2015; Webb, Newton, & Chang, 2013). For instance, Webb & Chang (2015) examined the potential of extensive reading-while-listening 10 graded readers by secondary school ESL learners (N = 61) on incidental acquisition of 100 words. Results of meaning recognition pretests and posttests revealed significant gains compared to participants that did not receive the treatment. Additionally, a study adopted a reading-while-listening mode to determine the impact of repetition on incidental acquisition of L2 collocations. The study revealed high vocabulary gains from this source of input (Webb et al., 2013).

There are additional insights from bimodal input research. Data from several studies highlighted the effectiveness of bimodal input in video contexts. For instance, low-intermediate L2 learners watched an audio-visual input and revealed significant word gains when words were presented in bimodal condition than with captions-only or audio-only (e.g., Hsieh, 2020). Studies have also shown that bimodal input positively contributes to learners' perceptions (e.g., Serrano, Andriá, & Pellicer-Sánchez, 2016; Tragant Mestres et al., 2018). Moreover, high language proficiency was found to be significantly associated with greater incidental vocabulary gains (Webb & Chang, 2015), while a high frequency of word occurrences was found to be conducive to the acquisition of word forms (Teng, 2016).

Theoretical Perspectives on the Value of Bimodal Input

Listening can assist reading comprehension in the same way that reading can assist listening comprehension. However, a common practice amongst researchers is to adopt one stance towards bimodal input (i.e., which mode assists the other) depending on the primary modality in the study (e.g., assisted reading, Dowhower, 1987; Webb & Chang, 2012). For instance, in L2 captioned video studies, vocabulary researchers generally argue that captions assist listening and neglect that listening also plays a role in supporting the processing of captions. Authors rarely discuss the mutual assistance in bimodal input from one modality to another (c.f. Charles & Trenkic, 2015). Furthermore, although extensive research has been carried out on bimodal input, there have been few discussions about its theoretical

underpinnings. This section will address possible mechanisms involved in learning vocabulary from this source of input while emphasising that reading and listening contribute interchangeably to incidental vocabulary learning.

Speed of Lexical Access.

Bimodal input can help increase the speed of lexical access, which in turn leads to vocabulary development. It has long been recognised that the gateway to vocabulary learning opens when learners can read or listen to texts fluently (e.g., Grabe & Stoller, 1997). On the other hand, disfluency in reading can be caused by the inability to perceive prosodic and rhythmic aspects of the language since the learner does not know how spoken forms are represented in written texts (Schreiber, 1980). Consequently, it could be argued that bimodal input improves fluency by making prosodic clues accessible to the learner via listening, which helps syntactic processing. This processing, in turn, leads to automaticity in word decoding and eventually word learning.

Sound-script Incongruence.

Sound-script incongruence offers a strong rationale for the use of bimodal input. Learners process words differently according to the word knowledge available. Presenting input in reading-only mode improves learners' knowledge of written form and implies that the teacher expects them to induce the correct spoken forms. Likewise, the listening-only mode infers that learners have to figure out the corresponding orthographic word forms. An interesting comment made by a learner in a previous study was that bimodal input led to the discovery that *sculpture* should be written as such and not as *sculture* (Tragant et al., 2016). Hence, bimodal input reduces vocabulary learning difficulty because learners do not have to induce correspondence rules between letter and sound. This argument will be further elaborated in light of the low letter-to-sound correlation in the English language, by referring to heteronyms and homophones.

English has a relatively low letter-to-sound correlation, which makes words more difficult to spell or pronounce. To exemplify, the sound /i:/ can be written in at least eight ways: thesis, theory, read, receive, screen, key, believe, quarantine. Likewise, the letter 'a' can be pronounced in five ways: (/ae/) as in *sad*; (/ɑː/), *argue*; (/eɪ/), *rate*; /ɔ:/, *fall*; /ə/, *interval*. Thornbury (2002) explained: "while spelling is

fairly law-abiding, there are also glaring irregularities" (p. 27). Salient letters as in *muscle* or consonants' clusters as in *strength* contribute to a word's difficulty. Thus, in bimodal input, spoken form increases the speed of lexical access and contributes to instant word processing in the same way that orthographic forms permit good listening. Simultaneous reading and listening facilitates form recognition and offers the gift of time that is devoted for learning more advanced knowledge such as meaning and use.

Bimodal input also decreases vocabulary learning difficulty that *heteronyms* impose. In linguistics, a heteronym (a type of homograph) is one of two or more words with the exact spelling but differ in meaning and pronunciation. A congruent example of a heteronym is that of stress (the relative emphasis). Stress exists at two levels: the word level is called lexical stress, which is given to word syllables and is the kernel of this discussion; the phrase or sentence level is called prosodic stress, which is given to the whole word and will be addressed later. Heteronyms could be best illustrated with lexical stress: "conflict" could be pronounced (/kɒnflikt/ CONflict/noun) to mean a situation of serious disagreement, with stress on the first syllable. It could also be pronounced (kənflikt/ conFLICT/verb) that is synonymous with "clash", with stress on the second syllable. Other examples of heteronyms include converse, convert, project. The examples show that it is common in English to form a noun from a verb by merely shifting the stress. English words could consequently change their meaning depending on stress position. Indeed, words that share orthography entail a learning burden in a reading-only mode, at the level of both form and meaning. Bimodal input helps to lower this burden with the phonological distinction that listening provides.

In the same vein, a *homophone* is a word that shares pronunciation with another word but has a different spelling or meaning. For example, "review" and "revue" are similar sounding words with different spellings. Another example is write/wright/rite. Even native speakers are not always immune to its caused confusion. Homophones can create frustration in a listening-only mode, which only reading-mode can relieve. As a result, bimodal input is more beneficial for vocabulary learning. Less attention and time need to be devoted to word decoding and recognition and more for processing advanced word knowledge. It offers

students an opportunity to not be troubled by words and be delighted by their complexity instead.

Text and Speech Segmentation.

Support for the role of bimodal input can also be found in its supply of automatic, accurate segmentation of written texts (via listening) and spoken texts (via reading). With regards to the former (i.e., listening), non-fluent L2 readers are confronted with the inability to maintain the integrity of written texts as a result of their tendency to break sentences into incoherent parts and make them meaningless. Audio in bimodal input helps to retain the segmentation by providing semantic wholes that facilitate reading (Brown et al., 2008). In the words of Bisson, Van Heuven, Conklin, & Tunney (2014a), bimodal input "… presumably helps [learners] segment the seemingly uninterrupted flow of words into more manageable chunks" (p. 861).

Reading, on the other hand, helps to segment speech. This further emphasises the previously noted view that the relationship between listening and reading is very likely reciprocal. To begin with, extensive exposure to L2 bimodal input has been found to be positively correlated with the ability to segment speech in adult learners with the aid of reading (Charles & Trenkic, 2015). Unlike in written texts, utterances in speech are not separated by punctuation, making it difficult to discern when a phrase starts and ends. Similarly, words are not separated by blank spaces, making it hard to figure out their first and last letters. This difficulty impedes discourse tracking and segmentation and, as a result, lexical recognition. Therefore, bimodal input is essential to parsing phrases and resolving these difficulties.

Oronyms often constrain the listener's task to determine where words begin and end; these are words or phrases that sound very much the same as another word or phrase but are spelt differently. To put it in Burridge and Stebbins' words (2016): "Because of the seamlessness of speech, sequences of sounds can be divided into words in more than one way; these are oronyms" (p. 203). An example they used to illustrate oronyms is "it's hard to recognise speech / it's hard to wreck a nice beach". Clearly then, an L2 learner who is not familiar with *appoint* could mishear "they will appoint trained teachers" as "they will a point trained teachers", which completely alters the intended meaning of the utterance.

The written text in bimodal input helps reduce *mondegreen*s, which is the act of mistakenly hearing oronyms. It is traced back to a Scottish song lyric "and laid him on the green" which Wright (1954) wrote that she misheard it as "lady Mondegreen" when she was young. These mondegreens occur because hearing is composed of auditory perception (i.e., physics of sound waves entering the ear to reach the brain's auditory cortex) followed by meaning-making (i.e., making sense of the sounds). Communication could break down somewhere in between due to noise or lack of visual cues, especially the speaker's mouth (e.g., radio) (Konnikova, 2014). At times, the impediment lies in the accent of the speaker (e.g., tone, pitch, and pace). Konnikova elaborated:

> "Human speech occurs without breaks: when one word ends and another begins, we don't actually pause to signal the transition . . . you hear a continuous stream of sounds that is more a warbling than a string of discernible words . . . the culprit is the perception of the sound itself: some letters and letter combinations sound remarkably alike, and *we need further cues* [emphasis added], whether visual or contextual, to help us out. In their absence, one sound can be mistaken for the other." (Konnikova, 2014).

Based on the former grounds, I argue that bimodal input is more conducive to vocabulary learning than unimodal input due to its text segmentation benefits. Mainly, it offers a combination of speech and transcription that is likely to decrease the perception of mondegreens and promote understanding. Together, reading and listening permit the proper segmentation of texts which allows for better comprehension, enhanced fluency, and word recognition, all of which are central to incidental vocabulary learning.

Acoustic Variability.

Another factor related to incidental L2 vocabulary learning is acoustic variability which refers to variation in speech as a result of different voices (i.e., speakers). Firstly, several lines of evidence suggest that increased acoustic variability positively affects L2 vocabulary acquisition. Based on these research findings, researchers have called for incorporating acoustic variability in the L2 classroom by presenting new word forms in input-based materials of different speakers (e.g., Barcroft & Sommers, 2005).

Secondly, acoustic variability in bimodal input enables learners to follow written discourse efficiently assisted with voice recognition. Listening offers various voices, and the speaker's voice carries a great deal of input that enables immediate voice decoding and identification. When the input has more than one interlocutor, the speakers' voices can be distinguished. As Plante-Hébert put it: "The auditory capacities of humans are exceptional in terms of identifying familiar voices." (as cited in Universite de Montreal, 2015). Therefore, students can quickly become familiar with interlocutors' voices with their vocal recognition abilities, especially in extensive input such as audiobooks or TV series. Bimodal input offers knowledge of who is speaking and obviates the need to check the interlocutor's identity on the left margin (or screen in L2 captioned video). Also, bimodal input in audio-visual exposure has been shown to improve L2 speech perception and assist adaptation to unfamiliar accents (Mitterer & McQueen, 2009).

In sum, acoustic variability is a fundamental function in bimodal input. Voice signals are channels that convey not only linguistic information but also the speaker's identity. Speaker variability has been linked with improved L2 vocabulary learning. In addition, it helps the learner skip the speaker's name part in reading (or face in viewing) which makes written texts eminently readable to get more comprehension and acquisition. Clearly again, the argument for bimodal input is premised on the assumption that when used together, reading compensates for the weakness in listening and vice versa.

Prosody Vs. Punctuation.

Bimodal input is indispensable because meaning sometimes can be entirely dependent on non-verbal components of prosody or punctuation. Earlier in this section, I introduced sound-script incongruence, referring to irregularities in the English language. Generally, however, speech conforms to writing at the level of verbal components but not always at the level of non-verbal components. For instance, although a period or comma reflects a pause, intonation provides more variation than punctuation, making prosody and punctuation roughly correlated (Huddleston, 1984).

Firstly, the lack of prosody in reading increases the necessity for bimodal input. Listening provides prosodic, paralinguistic, and extra-linguistic vocal effects,

which play a significant role in conveying meaning (Cruttenden, 1997). Prosodic features are suprasegmental and co-occurrent (e.g., stress, rhythm, loudness, intonation, and pause). Paralinguistic features are interruptive vocalisations that are not words (e.g., whistle, laughter, crying or interjections). Extra-linguistic features are implications that can indirectly give nuances to content meaning (e.g., gender, age, and emotions). For instance, loudness is a prosodic feature that can be used extra-linguistically to express emotions; for example, shouting reveals anger. These aspects of intonational meaning are based on a universal foundation and are characterised by their instant recognition.

Prosodic information is notably powerful. For example, stress often resolves ambiguity that occurs while reading. Considering the example below from Schmitz's work (2008), reading (a) and (b) leads to a degree of ambiguity.

a. John only introduced BILL to Sue.
b. John only introduced Bill to SUE.

However, the prosodic stress that is available to the listener (which I indicated using capitals) helps to recognise that (a) and (b) are answers to the following questions:

a. Who (from a restricted set of people) John introduced Sue?
b. To which woman did John introduced Bill?

This prosodic feature lacking in written texts is essential to understanding the intended meaning and reaching adequate comprehension of texts, which in turn augments opportunities for vocabulary learning. Moreover, the grammatical approach to punctuation often precludes the comma for object/clause distinction. This leads to syntactic ambiguity in meaning that can only be resolved through exploiting prosodic information (Kjelgaard & Speer, 1999) and which bimodal input could offer. Engelhardt, Ferreira, & Patsenko (2010) gave the example of "While the woman cleaned (#) the dog that was big and brown stood in the yard" (p. 640). They explained that unlike in reading, "a boundary tone on *cleaned*…followed by a short pause" represents acoustic features that help the listener effortlessly disambiguate the sentence. In fact, it has been proposed that "readers have fewer cues for parsing than listeners" (Niikuni & Muramoto, 2014, p. 276).

Secondly, the lack of punctuation in listening increases the necessity for bimodal input. Punctuation is a visual cue that helps resolve ambiguity, with a mere

comma having the power to alter meaning based on its position. It groups words meaningfully and leads to greater readability and clarification of the intended meaning of the text. Punctuational distinctions do not always have corresponding intonational cues. Nunberg (1990, p. 13) provided the example below, reflecting the rhetorical approach to punctuation:

a. Order your furniture on Monday, take it home on Tuesday.

b. Order your furniture on Monday; take it home on Tuesday.

Example (a) is conditional; it is the type of advertisement: if you order it on Monday, you can take it on Tuesday. In contrast, (b) is the conjunction of two commands. Hence, the meaning is plain to the eye (the reader) and not the ear (the listener). Moreover, it exceptionally serves to specify possession with nouns by adding an apostrophe which helps avoid the previously noted mondegreens. For instance, the listener may confuse (the girl's room) with plural form (i.e., the girls' room) or mix up the noun "room" for a verb (i.e., the girls room), things that are unlikely to occur to the reader. Additionally, learners are more likely to distinguish adverbs in written than spoken texts because they are often preceded and followed by punctuation marks.

Clearly, punctuation is not a mere decoration in written texts. It is communicatively relevant to the reader. Likewise, prosody patterns are vocal cues that assist rapid processing and comprehension. Prosody provides so much information readily available to the listener: varied syllable duration, stress accentuation, or rise and fall of intonation, all of which are missing when the listener becomes a reader (LeCoultre & Carroll, 1981). Bimodal input raises students' awareness of the expressive meanings that can be realised through both prosody and punctuation, which are vital to words and texts' comprehension. In sum, prosodic and syntactic patterns do not promise a close correspondence. The inconsistency between the two suggests the need to implement bimodal input to comprehend words and texts better.

This section collectively extended insights on the mechanisms underlying the impact of bimodal input on learning. The evidence reviewed suggests a significant role for bimodal input in text comprehension and word decoding and recognition, all of which seem to correlate with incidental vocabulary acquisition. Bimodal input makes meaning intelligible and perspicuous. The review especially shed light on the

circularity in the argument of the bimodal input effect. Reading-only mode entails strengths and weaknesses as much as listening-only mode does; when brought together, each serves to cover up the imperfection of the other.

3.1.3 The Value of Imagery

A picture is worth a thousand words (Bernard, 1921). With the advent of photography, pictures have become an everyday part of classrooms. Pictures can accentuate and punctuate meaningful concepts in ways words could never accomplish. They have an exceptional value in the EFL classroom, as they continue to be the core material to introduce new topics and word meanings and arouse interest in learning. In this section, I will introduce vocabulary research relating to imagery in general, in the format of both static and moving images. I will then review previous research that specifically compared View and Non-View conditions before providing a theoretical understanding of why imagery is valued.

Images have more value for L2 vocabulary learning when presented in animated than static format. It has been shown from a variety of sources that the use of static images (also referred to as still pictures) increases L2 vocabulary intake (Deno, 1968; Goldberg, 1974; Joklová, 2009; Kopstein & Roshal, 1954; Webber, 1978). Nevertheless, other studies provided additional evidence that static images did not provide more effective cues than English translations for recalling the meaning of L2 words, though they inflated learning confidence (e.g., Carpenter & Olson, 2012; Lotto & De Groot, 1998). The results have been attributed to a "metacognitive illusion" (Rhodes & Castel, 2009) of acquisition which pictures create in learners, and which does not necessarily coincide with actual learning (Lenzner, Schnotz, & Müller, 2013), including learning L2 vocabulary (Carpenter & Geller, 2020). For animated images, their use in instructional settings has been found to contribute positively to learning (e.g., Barak & Dori, 2011). Eye-tracking has previously shown more processing in dynamic images in L2 captioned videos than in static images in storybooks (Tragant & Pellicer-Sánchez, 2019). However, this effect was found to be linked to prior knowledge and the target learning outcomes (Ke, Lin Kun Shan, Ching, & Dwyer, 2006). Importantly, they were perceived as helpful modes of vocabulary glossing among L2 learners (Kayaoglu & Akbas, 2011; Ramezanali & Faez, 2019).

Authentic audio-visual input or television programs have been perhaps the most popular format of moving images in language learning and research. In particular, extensive research has been published on the impact of viewing L2 captioned videos on incidental word acquisition (e.g., Ashcroft, Garner, & Hadingham, 2018; Montero Perez, Peters, Clarebout, & Desmet, 2014). Until very recently, viewing research has tended to focus on supplementary input (L1 and L2 captions) rather than the primary imagery input. Studies on the impact of imagery on incidental word learning from this type of multimodal input generally adopt one of three designs, with associated methods: eye-tracking, contiguity, and comparing. Eye-tracking (processing) methods assess whether fixations on imagery predict learning. This method was not considered in this thesis for reasons of infeasibility in the study context. The second method, contiguity, investigates how the co-occurrence of words and their visual referents predicts learning (explored in Chapter 4). The third method is the essence of this chapter. It compares the effects of viewing and non-viewing conditions on learning by obscuring imagery from the screen and keeping only bimodal input as a comparison condition.

Previous Research on Viewing Vs. Non-Viewing

The small number of studies designed to determine the influence of imagery in L2 captioned video on incidental vocabulary learning have produced inconsistent results. The small sample and input size and the number of aspects of word knowledge tested have also limited these studies.

A positive effect of imagery on incidental word learning was found in young learners. A study compared vocabulary gains from 9 but short educational segments in the View and Non-View groups (Neuman & Koskinen, 1992). Segments were in L2 captioned video format and included 90 science target words. Middle school L2 learners (N = 129) were assigned to four groups: (a) uncaptioned video, (b) L2 captioned video (c) reading-while-listening (the same video), (d) textbook only (Control). The Non-View group read captions on scripts rather than the video indicating that reading was inconsistent across the two groups. Results from form recognition and recall and meaning recognition measures showed that the captioned View group significantly outperformed other groups, with learners of high level scoring better than low-level learners. The authors suggested that words' visual

referents in L2 captioned video augment learners' incidental learning of words and does not necessarily overwhelm their attentional capacity.

Studies on adults have provided additional insights. One study raised an intriguing question regarding whether the effect of imagery depends on the presence of L2 captions (Hernandez, 2004). The author investigated the potential of different integrations of modalities, including with and without visual modes of input, on incidental vocabulary learning. She presented four short segments extracted from a film, which included 22 target words, to 115 university students who were intermediate EFL learners. The input was 8 minutes long and presented twice. There were 4 treatment groups: audio + video + text (captioned View, n = 32), audio + text (captioned Non-View, n = 29), which resembles the present study design, in addition to video + audio (uncaptioned View, n = 30), and audio-only (uncaptioned Non-View group, n = 24). Hernandez hypothesised that captioned and uncaptioned View groups would outperform captioned and uncaptioned Non-View groups, respectively, in meaning recognition tests. The findings revealed a significant learning difference between captioned View and captioned Non-View groups, with the former scoring higher. However, there was a parity of results between uncaptioned View and uncaptioned Non-View groups. The results indicated that an advantage of imagery could only be attained in the presence of captions. The researcher, however, commented that the nature of the video led to less visual support for inferring the meaning of words.

A clear-cut effect of imagery was not attained in a recent study either, though its findings suggested that visual input may impede the acquisition of spoken form knowledge (Alshumran, 2019). The researcher implemented a pretest-posttest design and compared incidental L2 vocabulary learning outcomes from four input conditions: video, audio, and caption; video and audio; caption and audio; and audio-only. Input consisted of 4 documentary excerpts of 15 minutes each (i.e., 60 minutes) presented to intermediate EFL learners. Participants were tested on spoken form recognition, meaning recognition, and meaning recall of 36 words. Significant learning effects were marked in all conditions on all vocabulary measures. The View condition without captions contributed to higher meaning outcomes. Participants in the Non-View group with captions (n = 30) scored significantly better

in the spoken form recognition test than the View group with captions (n = 26), indicating that imagery somewhat disrupted auditory perception.

Finally, the most recent study found no difference between viewing and non-viewing. Its purpose was to assess the differential effects of reading (n = 21), listening (n = 15), and viewing (n = 21) documentaries on incidental L2 vocabulary learning (Feng & Webb, 2020). Participants were university EFL learners, 19 of them were assigned to a Control group. Knowledge of 43 target words appearing in a limited input of just 54 minutes and 14 seconds was tested at the level of form recognition using a checklist yes/no test and meaning recognition using multiple-choice-test (i.e., form-meaning connection). Participants demonstrated significant vocabulary gains irrespective of whether they viewed the documentary, read its script, or listened to its audio. The authors attributed the lack of an advantage in the group who had imagery support to unfamiliarity with the viewing mode in the EFL classroom.

The above study outcome corroborates the null comprehension result in Baltova's study (1994). She exposed learners of French to a story either in the format of video + audio (View group), audio-only (Non-View group) or video-only formats. The author reported that visual cues enhanced general comprehension and stimulated more positive attitudes but did not necessarily demonstrate a significant advantage compared to other conditions.

All in all, the focus in the research history of incidental vocabulary learning from audio-visual input (i.e., authentic videos) has chiefly been on the usefulness of L1 and L2 captions. Very little is currently known about the efficacy of imagery itself. Few available studies examined the role of imagery in L2 captioned video by isolating the causal impact of this variable on incidental vocabulary learning using a between-subjects design. The distinctiveness of my study lies in its focus on lengthier exposure. The above studies also were limited by the small sample size and the number of aspects of word knowledge tested. The current study fills these gaps by testing four measures of word knowledge in a larger number of participants based on extensive viewing. Having reviewed studies with the most closely matching design to the present study, I will next provide a theoretical rationale for the significance of imagery.

Theoretical Perspectives on the Value of Imagery

Imagery offers contextual and non-verbal input that compensate for the lack of understanding resulting from having verbal input alone. In this section, I discuss imageability, lip-reading, and motivation effects as three possible mechanisms that account for the strong relationship between imagery and incidental vocabulary learning.

Imageability.

Imageability refers to the ease or difficulty of forming a mental image or arousing a sensory experience to a word (Whaley, 1978, p. 146). The concept is traced back to two studies in which imagery was considered, with concreteness and meaningfulness, as a crucial variable that falls into "the-richness-of-meaning" factor (Paivio, Yuille, & Madigan, 1968; Whaley, 1978). Imageability is a strong determinant of learnability as abstract words are more difficult to learn than concrete words (e.g., Ellis & Beaton, 1993; Paivio & Desrochers, 1979). They are encoded verbally only, and their meanings lack direct imagery representation. Hence, abstract words cause some fatigue and frustration.

One possible mechanism underlying vocabulary learning from L2 captioned video, which has not been addressed previously, is imagery's aid to reduce the fatigue associated with abstractness. This theoretical support relating to imageability is in line with incidental vocabulary learning. On-screen imagery may trigger the incidental unconscious process of linking illustrative images to the meaning of unknown words. The difficulty of encoding abstract words may not always render practice controversial as Paivio (2014) claimed. Pinker (1994) pointed out that we are not born with language; thus, we do not necessarily think merely in words. Educators and professional mnemonics have long been proponents of the view that remembering processes encapsulate the use of images (Paivio, 2014). Paivio proposed that learning abstract words depends on prior learning of concrete words. The former may be imaged after going through a two-steps process," [imaging] requires grounding of the abstract term in a concrete instance, which entails intermediate links through word associations and *illustrative images* [emphasis added]" (p. 46). He gave the example of the abstract word *religion*, which might firstly activate the concrete word *church* as a verbal associate, then as an image of a church; church acts as the illustrative image here. Evidence of the previously noted

fatigue can be deduced from Paivio's two-steps proposal of how abstract words are learnt. The author was of the view that abstractness continues to trouble classroom practices. I make the argument that imagery in videos holds the potential to overcome this learning difficulty by reducing the fatigue that abstract words impose.

Similar to Paivio is my proper example of the word *grief*. It is an abstract word that could only be expressed verbally. It demonstrates the fatigue pertinent to imageability, resulting from the need to link the meaning to concrete words, then to illustrative images of words for retrieval purposes. Thus, if learners encounter the word in a TV series episode, chances are they may encounter concrete images that are pertinent to *grief* and which are needed to build imagery representations to retrieve the meaning of the word. Therefore, the process of word association becomes less frustrating. If we suppose that learners encounter the word grief in a text while viewing a woman sitting and crying, Paivio's two-steps process might no longer be needed or may be effortlessly undergone. This is because the illustrative target images are already available to the learners (e.g., *tears* or the action of *crying*). Thus, I am of the opinion that imagery in videos provides the best intraverbal associative context that is needed to accentuate those word connections. It accelerates the forming of mental images of the target abstract words for retrieving meaning purposes.

The mechanism discussed above also applies to learning concrete words. The non-verbal communication and the environmental context depicted from the visual input in videos assist learners in establishing the meaning of unknown concrete words presented verbally but not visually. To exemplify, the teacher gives the learners the following textual input of two speakers on the phone, in the form of bimodal input:

(a): Hey, where are you?
(b): I went to an *exhibition*

Providing the learners with an L2 captioned video instead will allow them to see where speaker (a) is. This will help them establish the meaning of the unknown word *exhibition* since this word will activate the illustrative image *museum*.

In sum, the impact of imagery on incidental L2 word acquisition should not be centred on the provision of visual referents to unknown words only. Based on

imageability theory, the on-screen visual input supplements the learners with illustrative images. These images are necessary to build mental representations for unknown words (abstract and concrete) that are not depicted through on-screen imagery. These representations are vital for recalling meaning. The following section will show that referents and illustrative images are not all there is to the intrinsic imagery features that are conducive to learning.

Lip-reading: the Visually Perceived Aural Input.

Lip-reading refers to the visual information that is derived from the speaker's mouth movements. The processes that underpin visually perceived spoken word recognition have been an interest to many researchers. Based on their evidence, I argue that the perception of speech from videos that show the speakers' faces involves the cooperation of two sensory modalities (listening and lip-reading). This cooperation forms a coherent representation of input to facilitate spoken form recognition.

Considerable research has highlighted the importance of facial features for speech perception (e.g., tongue, protrusion of lips, shape of mouth). The visibility of the teeth was proposed to play a pivotal role in the distinctiveness of vowels (Montgomery and Jackson, 1983). Empirical evidence showed that subjects were sensitive to teeth visibility, which helped them distinguish vowels (e.g., rounded from unrounded vowels) (McGrath, 1985). In another experiment, subjects were able to identify 78% of vowels correctly in the condition of a natural speaker face. The results were remarkably higher than those attained in the synthetic face condition, in which only lips and teeth were visible. In her study about auditory-visual input interaction in speech perception, Dodd (1980) confirmed that having both the aural and visual (lip-read) input "provided significantly more information than either vision alone, or masked hearing alone" (p. 541).

Researchers explained the significance of lip-reading on speech perception in varied ways. First, lip-reading has been shown to modulate the perception of speech sounds at a prelexical level (Calvert et al., 1997, p. 595). The author found that in audio-visual speech perception, which results from the combination of lip-reading and listening (Summerfield, 1992, p. 71), the linguistic, visual cues can activate the

auditory cortex area at times when the audio is unattainable. This activation can occur even if the movements are meaningless and only speech-like.

Other researchers explained the effect of lip-read input on speech perception in terms of the mental lexicon. Studies on the organisation of the mental lexicon are no longer restricted to the recognition of aurally perceived input but also of visually perceived aural input (lip-read input). Tye-Murray, Sommers, & Spehar (2007) maintained that audio and lip-read speech could help facilitate word recognition. Their underlying argument is that simultaneous activation of competitors in the mental lexicon for both modalities may reduce the possible word candidates.

The majority of studies emphasised the circularity in the argument of the effect of audio and lip-read input. Audio-visual integration in speech can compensate for deficiencies of listening in that the redundancy between input available from listening and lip-reading helps in perceiving speech more accurately. Integrating the speech we see with the speech we hear occurs naturally. Therefore, a "perceptual adjustment" is likely to occur before the speech is recognised when listening while viewing the face (Summerfield, 1992, p. 77). Others explained that auditory input informs about voicing and nasality while visual input informs the place of articulation. Hence, they are complementary, especially if any is degraded (Tye-Murray, Spehar, Myerson, Hale, & Sommers, 2016). Another study also found the interaction between listening and lip-reading as genuinely bi-directional; that is, both modalities affect the perception of the other (Baart & Vroomen, 2010). To explain the functional logic behind the interaction between "seeing and hearing speech", the researchers pointed out two notions as follows:

> "The first is that it is 'ecologically' useful to consult more than one source, primarily because different sense organs provide complementary information about the same external event. For this reason, lipreading is used in understanding speech as it can compensate for interference from external noise and may resolve internal ambiguities of the auditory speech signal. The second reason is that there is internal 'drift' or 'error' within the individual senses that can be adjusted by cross-reference to other modalities." (p. 103).

This type of cross-reference to other modalities was very well known for 100 years, but it has been newly discovered for speech. Interestingly, it coincides with the

fundamental view of the study; that vocabulary learning in L2 captioned video occurs as a result of interchangeable modality effects.

Overall, the visual cues of the mouth of the speaker represent a fundamental variable in the pictorial input in videos. Imagery in videos provides a source for visible articulatory gestures that accompany speech production. Observing this source has been proved to be critical for phonetic adjustments that are needed for word recognition.

Motivation.

A final mechanism to be addressed here which explains the impact of imagery on incidental vocabulary learning is motivation. SLA researchers affirmed that the effectiveness of video lies in its ability to offer attentional and affectional advantages (Baltova, 1994), precisely motivation (Oxford, Park-Oh, Ito, & Sumrall, 1993). L2 learners are incredibly motivated to learn from viewing L2 films or programs (e.g., Rodgers, 2013). In this section, I present two types of motivation, which are believed to be driven by viewing, then explain that motivation promotes strategy use which contributes to vocabulary learning.

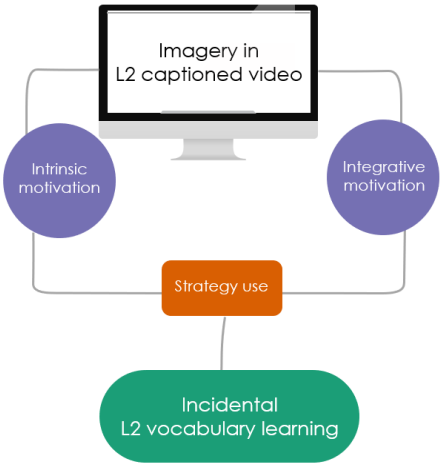
Intrinsic motivation (Crookes & Schmidt, 1991) refers to a learner's desire to know and "the degree of effort a learner makes to learn a second/foreign language as a result of the interest generated by a particular learning activity" (Ellis, 1997, p. 140). One subtype of intrinsic motivation is *stimulation* (Vallerand, 1997) which means that motivation is "...based simply on the sensations stimulated by performing the task" (Noels, Pelletier, Clément, & Vallerand, 2000, p. 61). I argue that images are increasingly motivational and stimulating. They appeal to learners' eyes and raise interest and attention in learning which strongly correlate with greater intake. Television programs are usually designed to impress and captivate viewers. They are the most perceived as a medium of entertainment by learners (Mishan, 2005). Previous findings indicated that the motive behind learners' interest in multimodal input is their visual appeal (e.g., Cutajar, 2017). Learners also tend to approach tasks that have value and relevance to their language orientation (Schmidt, Boraie, & Kassabgy, 1996, p. 9), and TV viewing meets this criterion well. Hence, it could be safe to argue that learners will increase their level of motivation when they can see and hear the target language in use via moving images.

Imagery enhances integrative motivation. The integrative theory of motivation (Gardner, 1985, 2010) is a socio-educational model of SLA which holds that learning a L2 language involves learning cultural values of the target community (Gardener, 2010, p. 2). Integrateviness refers to a student "genuine interest in learning the second language in order to come closer to the other language community" (Gardner, 2001, p. 5). It is the "...willingness to get to know about someone else's culture and to interact with members of that group, coupled with a willingness to learn a language to do so" (Schmidt & Watanabe, 2001, p. 314). It could thus be argued that imagery elicits integrativeness and increases motivation to learn because it provides the cultural component of the target language. Exposure to vivid images of the community and speakers of the target language is a golden opportunity to overcome the cultural strangeness of the foreign language. Videos of speakers of the target language provide a natural context for authentic discourse and permit the picking up of cross-cultural clues, triggering learners' interest in more knowledge. The applicability of integrative motivation in the EFL context has been questioned in the past due to limited interactions with the target language (Schmidt et al., 1996, p. 13). However, with the advent of technology, the target language and speakers of the language are accessible in many forms, especially in narrative TV.

This study holds that audio-visual input fosters intrinsic motivation and integrative motivation (in the case of narrative TV) owing to the vividness of imagery. Motivation research implies that imagery in audio-visual input can motivate the incidental use of vocabulary learning strategies. To explain, high frequency of strategy use has been linked with strong motivation to learn (Oxford, 1990; Oxford & Nyikos, 1989; Pintrich, 1999). Motivation positively correlates with strategy use since students with higher motivation will likely use cognitive and metacognitive strategies (e.g., contextual guessing) to perform tasks (Pintrich & De Groot, 1990). As Figure 3.1 illustrates, imagery could be associated with stimulated strategies. Vocabulary is learnt best when learners value the tasks and the materials. Viewing speakers and communities of the target language via television may provide learners with emotional affinity with the language. The motivation in turn contributes to strategy use that is sometimes needed to incidentally acquire words from audio-visual materials.

Figure 3. 1

Motivated Strategies in Viewing



Taken together, the value of imagery in incidental L2 vocabulary learning from viewing can be explained by the semantic characteristic of imageability (Paivio et al., 1968), visually perceived aural input (Dodd, 1980), and motivated strategies (Oxford & Nyikos, 1989). Worthy of noting is that the materials should also reflect the authenticity of depicted ideas and the truthfulness of the impression they create in the learners' minds (Blanc, 1953, p. 150).

In sum, the overall section identified the limited number of studies comparing viewing and non-viewing condition effects on incidental L2 vocabulary learning. It revealed the inconsistency of research results and the fact that there are currently no comparable data that are substantial in terms of input, sample, and measures of word knowledge, thus, emphasising the need for further research. At last, the section provided a theoretical account of the effect of imagery on incidental L2 word learning before proceeding to unravel the question of how this effect is manifested in the context of L2 captioned video.

3.1.4 L2 Captioned Video

From just 2010 to 2021, more than 150 works examined the relation between captioning/subtitling and vocabulary. The present study builds upon an already robust literature regarding L2 captioned video. Its overall results seem to suggest the effectiveness of the use of L2 captioned video for incidental L2 vocabulary learning

(see the two books Teng, 2021; Vanderplank, 2016, for a comprehensive review). What follows is a discussion of the two conflicting theories prevailing as the arguments upon which positions are made regarding the potential of L2 captioned video.

Dual Coding Theory

An explanatory theory for the positive effects of L2 captioned video on incidental vocabulary learning is the dual coding theory (Paivio, 1971, 1986, 2014). Proponents of the dual-coding theory adhere to the view that learners process verbal (i.e., bimodal) and non-verbal (i.e., imagery) input in L2 captioned video via different cognitive systems. Hence, this type of multimodal input activates both systems and results in increased retention of words.

Results from an overwhelming majority of studies have been consistent with the implication from the information processing theory of dual-coding. Several lines of evidence suggest the positive impact of L2 captions (i.e., same-language subtitling) on L2 vocabulary learning (Bird & Williams, 2002; Borrás & Lafayette, 1994; Garza, 1991; Sydorenko, 2010; Vanderplank, 1988; Zarei, 2009). The most comprehensive review of this literature is Montero Perez et al.'s meta-analysis, based on eighteen empirical studies (2013). The study revealed a large effect size of L2 captions on word knowledge as measured through immediate posttests. Another finding in this literature is that adult learners process both visuals and subtitles irrespective of the language of audio and subtitles (Bisson, Van Heuven, Conklin, & Tunney, 2014b). Mayer expanded upon dual coding theory and introduced the multimedia principle (Mayer, 2001, 2009, 2014), which holds that we are more likely to learn from words and images than from only words.

In addition, many results from eye-tracking studies on viewing/reading behaviour are in congruence with the dual-coding theory. For instance, L2 learners have shown the capability to read captions (e.g., Montero Perez, Peters, & Desmet, 2015). In addition, subtitles appear to change the distribution of visual attention without necessarily increasing cognitive load (e.g., Kruger, Hefer, & Matthew, 2013; Perego, del Missier, Porta, & Mosconi, 2010). Moreover, in a study where the eye movements of 91 persons were examined, learners did not re-read subtitles crossing shot changes but focused on imagery instead (Krejtz, Szarkowska, & Krejtz, 2013).

This behaviour indicated that learners process imagery and words efficiently enough. Evidence of this processing efficiency was also obtained in a study of reading with static images. Findings revealed that audio offers L2 learners the gift of time to skip reading and observe images irrespective of the learners' language proficiency (Pellicer-Sánchez et al., 2018).

Nonetheless, the dual coding theory does not always match results from studies on L2 captioned video. Despite the substantial evidence of the benefits of L2 captioned videos, vocabulary researchers sometimes do not arrive at positive learning findings from this multimodal input (e.g., Birulés & Soto-Faraco, 2016; Bisson et al., 2014b; Peters et al., 2016; Sinyashina, 2019, 2020b). Cognitive overload is a factor that is usually linked to the null result.

Cognitive Load Theory

The cognitive load theory (Sweller, 1988, 1994; Van Merrienboer & Sweller, 2005) of educational psychology and instructional design (Van Gog, Paas, & Sweller, 2010) has been influential. The theory is based on the idea that changes in the amount of information correlate with variations in ease of acquisition due to the limited capacity of working memory (Sweller, 1994). I explain below how cognitive load theory relates to learning from L2 captioned video then describe its underlying limitations.

There are reasons to believe that L2 captioned video may cause cognitive overload in learners. Viewing L2 captioned video entails the learners to attain to three channels: auditory channel, visual non-verbal channel (imagery), and a visual verbal channel (captions). According to cognitive load theory, processing these sources of input simultaneously makes channels in competition for attaining learners' notice, increases cognitive load, and may thus frustrate the learner. Moreover, the integration of bimodal (verbal) input and images consists of redundant information that might be an impediment towards learning (Kalyuga & Sweller, 2014).

However, the theory is principally limited by two factors. First, cognitive load is difficult to measure. Second, it is conditioned upon numerous aspects, including the nature of activities and learners' related variables (e.g., L1, familiarity,

proficiency, hearing status etc.) (Durbahn et al., 2020; Muñoz, 2017; Paas, Renkl, & Sweller, 2003; Taylor, 2005; Winke, Gass, & Sydorenko, 2013). For instance, the cognitive load was found to be linked to proficiency when first students had more difficulty processing audio, imagery, and captions than third-year students (Taylor, 2005). In another multimodal-input study, questions based on both audio and imagery comprehension were more challenging than those based on comprehension of either audio or imagery alone (Durbahn et al., 2020). The authors suggested that participants might have been unfamiliar with questions that require split attention between audio and imagery. In short, this sub-section showed how dual coding (Paivio, 1971) and cognitive-load (Sweller, 1988) theories construct a theoretical framework surrounding the use of L2 captioned video.

Viewing Vs. Non-Viewing in L2 Captioned Video

The evidence from the three strands of research: bimodal input, imagery, and L2 captioned video, suggests that bimodal and multimodal input would likely contribute differently to incidental L2 vocabulary learning. Based on the theoretical perspectives reviewed earlier, there is value for both bimodal and imagery input; each has unique strengths.

Notwithstanding, it is fair to hypothesise that, in line with dual coding theory, extensive viewing in L2 captioned video format (View condition) will result in the increased incidental acquisition of meanings and spoken forms. This is because imagery can offer visual referents, contextual clues, and visual perceptions of the spoken form (lip-reading) to achieve high levels of understanding. It also can motivate the use of strategies. Nevertheless, since knowledge of written form can only be acquired through reading L2 captions, it will be learnt more in the absence of imagery (i.e., bimodal input; Non-View condition), in line with cognitive load theory. In short, imagery is conducive to learning knowledge of meaning and spoken form but unduly load learners with unnecessary information to acquire knowledge of written form. Before I move on to the practical part of this vocabulary research, it is important to explain some analytical decisions.

3.1.5 Word Knowledge as Operationalised in the Present and Previous Studies

If a teacher asks the students whether they know a word, they will likely provide its meaning; however, word knowledge involves more than meaning retrieval. There are three components of word knowledge (Nation, 2001). First, word knowledge involves knowledge of form, which is the orthographic and phonological awareness of the word. Second, there is knowledge of the semantic value of the word and the different meanings associated with it. Third, knowledge of use takes many forms. It involves knowledge of word classes and collocations (i.e., what commonly occurs with a word). It also includes pragmatic constraints (e.g., knowing that to pass away expresses sincere sympathy relative to dying). In addition, a distinction is made between recall and recognition types of vocabulary knowledge. Recognition knowledge refers to the ability to distinguish the correct word form or meaning from other forms or meanings. Recall knowledge refers to the ability to retrieve the proper form or meaning without any assistance.

The results in the present study and subsequent ones are based on measures of meaning recall and recognition and spoken and written form recognition, in response to recent recommendations in viewing research to measure multiple aspects of word knowledge (Feng & Webb, 2020). The opted measures also correspond to those primarily established in the literature. In the early stage of research-decision making, I conducted an informal review of 30 vocabulary studies between 1985 to 2016 that focused on the input modality effect. The results indicated that measuring meaning recognition formed a fundamental element of just above half the studies, followed by meaning recall (47% of the studies). A quarter of studies measured written and spoken form recognition, while 7% tested knowledge of parts of speech and written form recall. Almost no study has considered the measurement of spoken form recall. The present study also assesses word knowledge in isolation. While words can alternatively be tested in context (e.g., Teng, 2016), this method was not used based on the assumption that knowing a word includes reaching a definitional meaning level that permits the successful transfer of words from the familiar to the unfamiliar contexts.

3.1.6 Word-related Variables

Some words are easier to learn than others. Almost every experimental study in vocabulary research references word-related variables and controls for their potential predicting effect. The present research considers six lexical characteristics. First, parts of speech have a potential learning impact; for instance, nouns may be easier to learn than verbs (Ellis & Beaton, 1993). The present research studies learning mainly nouns, adjectives, verbs. The selection depended on the input and the target population's vocabulary and was determined by the norming analysis. Second, length is another factor affecting word learnability (Crystal, 1987), and will be operationalised as the number of syllables and characters. Third, concrete words are easier to learn than abstract words (De Groot, 2006). Concreteness will be controlled using the 5-point rating scale (from abstract to concrete) based on 40 thousand English lemmas (Brysbaert, Warriner, & Kuperman, 2014).

Cognates are another word-related variable that might influence learning outcomes for viewing studies, which lack pseudowords. However, cognates are inevitable in the English language materials. In the present research context, French is an L2 and about one-third of English words originate from French, with 1,700 words being true cognates, while English speakers likely know 15,000 French words without receiving French language instruction (ThoughtCo Team, 2019).While a recent viewing study showed a cognate advantage in learning L2 vocabulary (Peters & Webb, 2018), the cognate language was participants' L1. In contrast, the cognates under the present investigation are participants' L2; thus, they may not be as influential as L1 cognates. An advantage of including a large number of French-English cognates is that cognates are processed faster than non-cognates (Groot & Keijzer, 2000). However, cognates are not always synonymous with easiness (Rogers, Webb, & Nakata, 2015), as Part 1 of the previous chapter clearly showed that students did not recognise many cognates. This is perhaps more relevant when both languages are not L1 as is the case in the context of the present research. In this case, recognition might perhaps be informed more by words' corpus frequency than cognateness.

Some words are massively more common than others. Knowing 3,000 – 4,000 most frequent English language word families allows 95% coverage of TV programs scripts (Webb & Rodgers, 2009a, 2009b), though only 90% coverage may

be needed for adequate viewing comprehension (Durbahn et al., 2020). I will use the logarithmically transformed frequency in corpus SUBTLEX-UK, based on 45,099 BBC broadcasts (201.3 million words) (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014) because it is suitable to the BBC series under investigation. Measures are presented in Zipf-values (low frequency: 1-3; high frequency: 4-7). For instance, 1 refers to a frequency of 1 per 100 million words, and 2 refers to 1 per 10 million words.

Verbal Frequency of Occurrence

The present thesis adopts a minimum total frequency of eight encounters for the target words. Eight encounters were optimal for incidental L2 vocabulary learning in a study on reading (Pellicer-Sánchez, 2016), which is somewhat close to the reading-while-listening mode involved in my study. For viewing research, however, some authors found an advantage of frequency of occurrence for word learning (Peters et al., 2016; Peters & Webb, 2018; Rodgers & Webb, 2019), while others did not (Feng & Webb, 2020). Research suggests that there is no frequency threshold for word acquisition to occur and that the effect of frequency changes according to other word and learner-related covariates (Uchihara, Webb, & Yanagisawa, 2019).

Word Families

It has been common practice to consider word families in vocabulary research, which include the word's base form (e.g., *orbit*), their inflectional forms (*orbiting, orbits*), and derivatives (*orbital*). The present research adopts the flemma as the main word counting unit, which refers to the word's base forms and associated inflections. Inflectional morphology does not change the meaning of the word but only its function (e.g., plural, gender forms, comparative forms). Compounds and derivatives are adopted as a covariate. Worthy of noting is that due to the extremely low number of possible target items, cosmic and cosmos are exceptionally included as two distinct words and so are dense and denser despite not meeting the selection criteria.

Derivatives are formed using affixes and suffixes and have been a point of discussion in vocabulary research. It has been proposed that "once the base word or even a derived word is known, the recognition of other members of the family requires little or no extra effort" (Bauer & Nation, 1993, p. 253). This assumption

has been reflected in numerous vocabulary tests[2] in which derivatives were part of the counting unit (word family) to measure vocabulary size. However, a reasonably safe approach is to exclude derivatives because they encompass knowledge of lemmas and form and meaning entailed by their specific properties. Few studies have broken with tradition and cast doubts on the long-held assumption of counting word families (Gardner, 2007; Kremmel, 2016). For instance, McLean (2018) indicated the inappropriateness of word family as a counting unit for Japanese EFL learners. Also, recent results showed that derivatives do not contribute much to text coverage, but it is still unknown whether knowledge of lemmas extends to that of their derivatives (Laufer & Cobb, 2019).

Given the context of this thesis (i.e., vocabulary acquisition), a good position is to consider derivatives and compounds as a moderator variable of an amplifying effect between learning and verbal frequency. Suppose the learners' morphological awareness is high enough to recognise *sculptor* as a derivative noun for the target word *sculpt.* In that case, frequent encounters of *sculptor* in the input may reduce the learning burden of *sculpt* and, hence, should be considered. Another example is a target derivative (*fusion*) and a verbal derivative (*fuse).* In sum, the present thesis adopts the flemma as an adequate unit of counting while also controls for variation in words' reoccurrences via other related word forms, mainly compounds and derivational forms.

[2] Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001) and Listening Vocabulary Levels Test (McLean, Kramer, & Beglar, 2015)

3.2 The Present Study

This first study assessed the extent to which viewing two full-length seasons of documentary series, in the form of L2 captioned video, promotes incidental learning of knowledge of meaning recall and recognition and spoken and written form recognition. Secondly, it aimed to determine the role of imagery on incidental vocabulary learning from extensive viewing in the format of L2 captioned video. The study implemented a between-participants design. The control group did not receive a treatment, the View group viewed episodes of documentary series extended to eight viewing hours over six weeks at two-week intervals, and the Non-View group had imagery hidden from view and was therefore exposed to the bimodal verbal input only (L2 captions and audio). Participants' word knowledge of 20 target words that were spaced over the documentary series was assessed. Pretests and posttests of forms were administered before and immediately after the treatment while meaning tests were pretested only. As has been pointed out in the above review, this study advances knowledge about incidental word learning from extensive viewing literature by extending exposure to two full-length seasons of documentary series (running for about 8 hours) and to 2-hours viewing sessions. This length is much longer than the average in previous extensive viewing sessions (usually 20 to 30 minutes except Rodgers, which was 42 minutes). It also adds to previous studies that isolated the effect of imagery by increasing sample size and duration of input. On the whole, the study contributes to existing knowledge on the topic by considering four measures of word knowledge.

3.3 Method

3.3.1 Questions and Hypotheses

Study 1 asked the following research questions:

***Research Question 1*:** Does viewing two full-length seasons of L2 captioned documentary series (8 hr) over 2-hour long sessions lead to incidental learning of L2 vocabulary?

> ***Hypothesis*:** It was predicted that extensive viewing would produce gains in knowledge of meaning and form.

***Research Question 2*:** What is the effect of removing imagery and keeping bimodal input?

> ***Hypothesis*:** It was predicted that the View group will outperform in tests of meaning and spoken form, while Non-View group will outperform in the written form test.

3.3.2 Participants

One hundred seventy-three participants took part in Study 1. They were recruited from the population of Algerian EFL learners in their third year of the Linguistics Bachelor programme at the University of Jijel, in the autumn semester in the academic year 2017-2018. Of these, 29 participants were excluded, and data from 144 participants (131 females and 13 males) aged 21-23 years (M = 21.11) were kept for analysis. Participants were excluded if they were absent in any session of the pretests and posttests. Participants in the two experimental groups were also excluded if they missed any treatment session. This was done because I subscribe to the spacing theory, whereby target words need to be encountered in each presentation and missing only one session is believed to create bias in the Study.

Participants were all native Arabic speakers with French as a second language and were targeted because they make a convenient level for authentic materials use. They had studied English for a minimum of 9 years since the age of 12, and are considered as intermediate to upper-intermediate learners of English language. Standardised tests were used to ensure the homogeneity of the sample. They were recruited voluntarily when visiting the study context a few weeks prior to

the study. I distributed invitation cards that requested students to join a Facebook group which helped me communicate and respond to their inquiries instantly.

The study employed a between-subjects design in which participants were divided into three groups: Control (N = 34), View (N = 53), and Non-View (N = 57) using stratified random sampling. Table 3.1 displays the number of participants excluded and included for each class of each group.

Table 3. 1

Number of Excluded and Included Participants in Control, View, and Non-View Groups

		No of participants		
Group	Class	Pre-exclusion	Exclusions	Post-exclusion
Control	3	23	04	19
	8	20	05	15
View	1	24	04	20
	4	24	04	20
	7	17	04	13
Non-View	2	25	03	22
	5	19	02	17
	6	21	03	18
	Total	173	29	144

Note. The population was composed of eight classes at the Linguistics

Sampling

The study adopted a stratified random sampling. The target students were organised into eight groups. They regularly attended the speaking module for two academic years in the language labs with five teachers (a, b, c, d, e). Depending on their teachers' preference, students could have been habituated to different lab activities (audio-based learning, video-based learning, etc.). Familiarity with different presentation modalities might have a decisive effect on performance. In attempt to hold this variable constant, stratified random sampling was preferred ahead of simple random sampling. Students had to be randomly assigned to the control, View, or Non-View group based on their ex-teachers of the speaking module. Due to students' tight university schedules and the availability of the classrooms, it was impracticable to have an assembly of students from many different classes in one

treatment session. Participants were therefore approached as they had been grouped in the department. Classes of similar ex-teachers were spread across the 3 experiment groups; this resulted in a relatively fair distribution, as shown in Table 3.2.

Table 3. 2

The Adopted Stratified Random Sampling

Group	No of classes	ex-lab teacher
Control	2	a, b, c, e
View	3	a, b, d, e
Non-View	3	a, b, c, d

Ethical Considerations

Informed consent was obtained from the head of the English Language Department (Appendix C) and participants (Appendix E). The procedure was identical to that explained in the Norming study (section 3.1.7), except that the actual research aims were not made explicit to preserve the incidental nature of their acquisition. The study was approved by the Education Research Ethics Committee at the University of York. The use of episodes of documentary series for research and educational purposes was covered by *section 34* of the UK's Copyright Designs and Patents Act and by obtained permission from BBC World Wide Learning. The making of highlights for episodes for non-commercial research purposes was covered by *Section 29* and *30* of the UK's Copyright Designs and Patents Act for fair dealing (the length of clips did not exceed 45 seconds). Credits that fully reference the source material were added to the highlight clips or the clips by which open captions were restored after imagery removal (see *Input* in Section 3.3.6 for more details). All clips are destroyed after the examination of the thesis has ended, with the shortest clip being kept available as illustrative material for future researchers and practitioners.

3.3.3 Research Setting

Study 1 (and all the remaining studies in this thesis) took place in the English Language Department at the University of Jijel, Algeria. In particular, it was conducted in the standard classrooms that have a maximum class size of 30 students.

VLC media player and Microsoft PowerPoint on my laptop were used to show the experiment materials. The size of screen displays has been shown to have a psychological impact. Large screens with just about 1.5-meter picture heights have been associated with greater attention to imagery from television and film (Reeves, Lang, Kim, & Tatar, 1999). Therefore, materials were projected into a 2-meter by 1.5-meter screen using my personal projector (Epson EB-X31 Long Throw Office Projector). The projector had 1024 by 768 resolution, 4:3 aspect ratio, and a 2 Watt loud speaker of high quality sound. At the beginning of every sitting, students were requested to sit in the classroom's front row if they had a vision or hearing problem.

In a quest to establish a proper study timetable, I informally discussed with the teaching staff their general overview of the time-of-day effect on the target students' learning. They indicated that students in the specific context tend to be more awake in morning classes. Sessions were therefore scheduled mostly in morning times, with few afternoon sessions being held equally with each of the control, View, and Non-View groups.

3.3.4 Research Schedule

The experiment consisted of 7 sittings which took place over a 10-week period during the autumn semester in the 2017-2018 academic year. The first sitting in Week 1 lasted for about 100 minutes and consisted of an introductory phase and four tasks: consent form, language profile questionnaire, vocabulary pretest, and Oxford Placement Test. The sequence and duration of tasks are provided in Table 3.3.

Table 3. 3

Schedule and Duration for Tasks in First Experiment First Sitting

	Tasks					
	Numerical codes	Consent form	Language profile questionnaire	Vocabulary pretests	Break	Oxford Placement Test
Duration	5	10	10	36		40

Note. Duration is expressed in minutes. Total duration = 101 minutes.

It involved both the control and experimental groups. To guard against the negative effects of pretesting, there was a two-week interval between the first and the second sittings. The four treatment sessions occurred at two-week intervals in Week 3, 5, 7,

and 9. This interval was meant to meet the spacing theory (see Chapter 5). Each session consisted of exposure to two one-hour episodes of the BBC documentary series, which included 28 target words. The fourth treatment session in Week 9 was followed by an immediate vocabulary posttest and a one-week delayed posttest in Week 10. The tests involved both the control and experimental groups and lasted for about 65 minutes each. The 10 weeks culminated with a debriefing survey that aimed at exploring participants' perceptions about different aspects involved in the study. The research schedule is diagrammed in Figure 3.2.

The research schedule was arranged for eight classes in which students were enrolled. The total number of working hours for the in-class procedures for the study was approximately 79 hours. Table 3.4 specifies the duration per each class of the control, View, and Non-View groups.

Table 3. 4

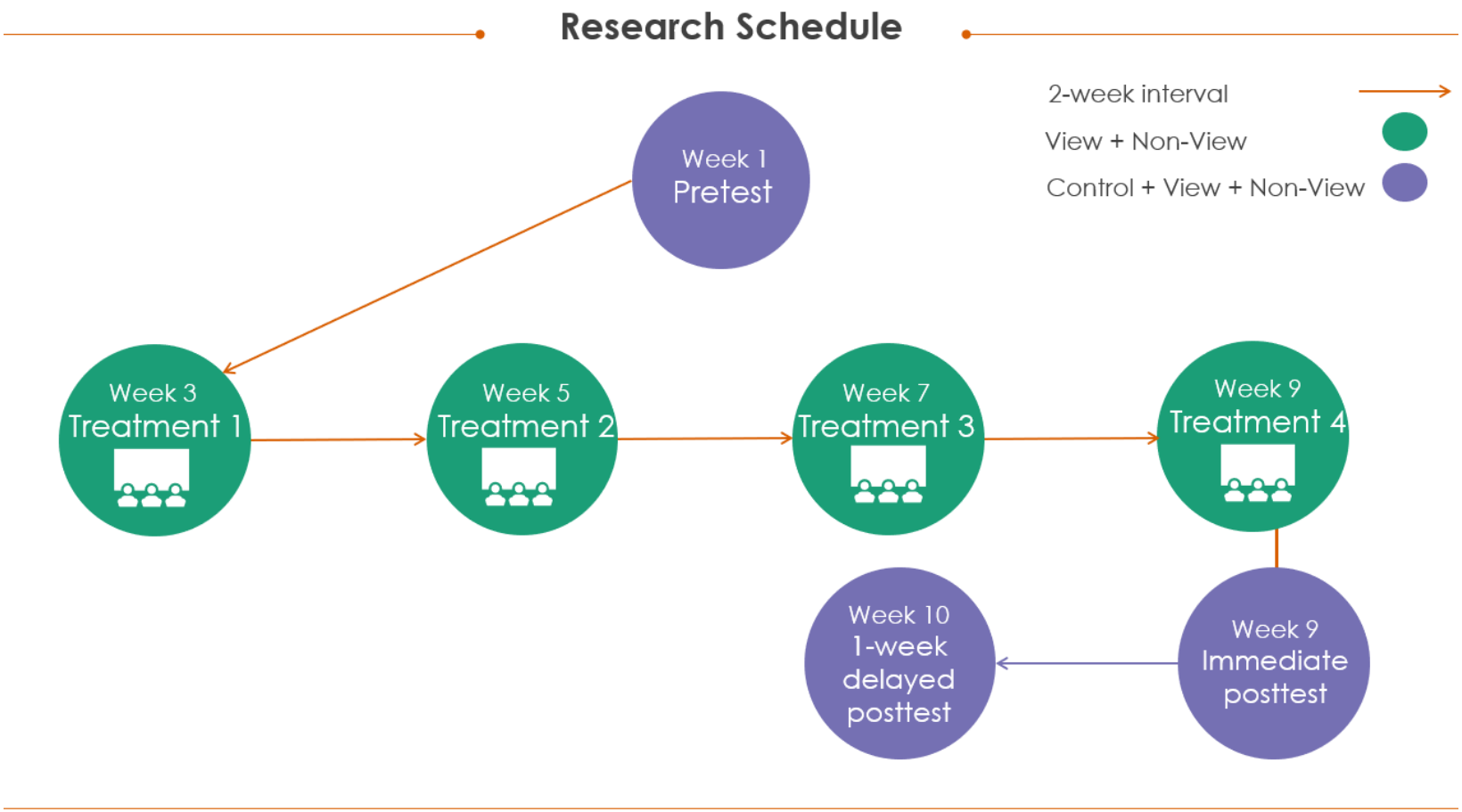
Time Spent per Sitting and Class To Complete Study

		Duration				
Group	Class	First sitting	Treatment sittings	Immediate test sitting	Delayed test sitting	Total
Control	3	100	00	65	65	230
	8	100	00	65	65	230
View	1	100	4 × 120	65	65	710
	4	100	4 × 120	65	65	710
	7	100	4 × 120	65	65	710
Non-View	2	100	4 × 120	65	65	710
	5	100	4 × 120	65	65	710
	6	100	4 × 120	65	65	710
						≈ 79 hr

Note. Duration is expressed in minutes. Treatment phase was composed of 4 sessions

Figure 3. 2

Research Schedule



Note. The experiment was scheduled to be held in 7 sittings over a 10-week period

3.3.5 Pilot Study

A pilot study was carried out at the University of York in the summer term of the 2017 – 2018 academic year, one month before conducting the experimental study in Algeria. Six volunteer participants were divided into two groups: View (N = 3) and Non-View (N = 3). Participants were tertiary students enrolled in different courses. They had relatively similar characteristics to the target population; they were native speakers of Arabic and had studied English as a second language for about nine years. The pilot study was mainly intended to refine the experimental procedures:

- To evaluate and revise tests questions and format of answer sheets.
- To regulate the timings in tests and the length of rest breaks.
- To test the treatment materials on View and Non-View groups.

3.3.6 Materials

This section describes the materials used in this study. The materials include the audio-visual input provided to participants, the vocabulary items targeted, and the instruments developed to collect the data for the study. Instruments are presented in the same order in which they were received.

Input

Based on the results from the norming study, the two full-length seasons of the documentary series *Wonders of the Universe* (Cooter et al., 2011) and *Forces of Nature* (Cooter et al., 2016) were selected as the audio-visual input for the treatment phase in the study. Each series was composed of four episodes that were one-hour long each. Altogether, the total length of the episodes was about eight hours with each session including two episodes. Thus, the amount of the input used well represented the extensive viewing approach intended in this study. Information on the series is detailed in the preceding chapter. In every treatment session, the View group were exposed to the input in the format of L2 captioned video. In contrast, the Non-View group had the imagery removed from the video and were therefore exposed to the L2 audio and L2 captions only.

Captions.

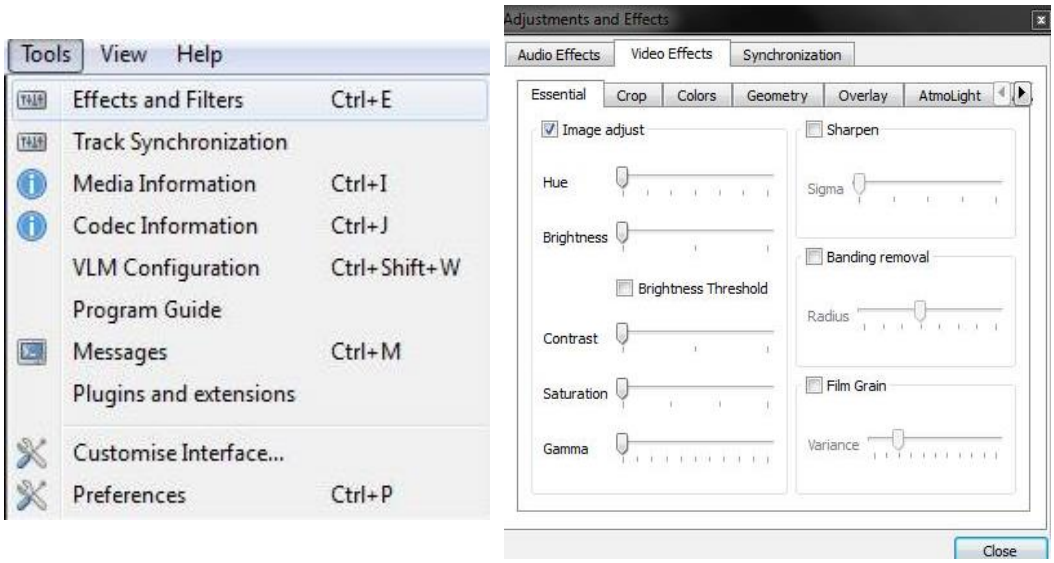
The input was in the form of L2 captioned videos. Also referred to as same-language, intra-lingual, and bimodal subtitles, L2 captions are transcriptions in the same language of the audio (i.e., in English as a second language). The factor that heavily influenced the choice of captions was that these are transcriptions rather than translation. Hence, they involve a pure transformation of aural text into written text which has been found to enrich input and discourage the reliance on and the pursue of L1 text. Research on the usefulness of L2 captions is reviewed in Section 3.1.4. The captions that appeared in the study were in two-line format, which has been shown to be optimal practice concerning learners' enjoyment and cognitive load (Szarkowska & Gerber-Morón, 2019). The readability and accuracy of the captions were checked using the proposed Code of Good Subtitling Practice that was drawn up by Ivarsson and Carroll (1998). Overall, the captions in the study were found to adhere to most of the instructional guidelines in the code.

Hiding Imagery from View.

Imagery was removed from the videos for the Non-View group using VLC media player features. These were accessed via Tools > Effects and Filters> Video Effects, as shown in Figure 3. Under the first tab, Essential, the *Image adjust* option was ticked and all settings that are underneath were dragged to the minimum. This resulted in the screen turning black.

Figure 3. 3

Procedure to Obscure Imagery Using Visual Settings in VLC



The technique helped retain any other variables in the treatment as constant for the View and Non-View groups. These variables include the captions' layout, in particular, and the input presentation, in general. A comparison between the presentation of input for the two experimental groups is illustrated in Figure 3.4.

Figure 3. 4

Screenshots from Input Presentation for View and Non-View Groups

As has been noted in section 2.2.4 of the preceding chapter, few episodes contain a handful of open captions. Removing imagery for these episodes caused open captions to disappear since they appear "… permanently on screen and cannot be switched off by the viewer. The viewer has no control over the style or visibility of the subtitles" (Kashyap, 2011, p 1). The issue was sorted out by capturing the episode (screen + audio) using Screencast-O-Matic software and manually adding the open captions in Windows Movie Maker.

Order of Episodes.

The episodes were not shown in the experiment in their original order found in the DVDs. Instead, episodes 4 of both series were shown first, followed by episodes 1, episodes 3, then episodes 2. This order was followed to fulfil an optimum condition for Study 3 of the spacing effect in learning; a comparison is made between items spaced over episodes and items massed in a single episode. Principally, to minimise the recency effect on immediate posttest results, episodes 2 were selected to be presented last for having the lowest verbal frequency of target words. Both documentaries are not serialised but rather episodic; thus, the followed order exerted no influence on the comprehension of the series.

Target Items

Twenty words that are spread over eight episodes of the two selected documentary series were selected from a broad category as target items. The selection consisted of 8 nouns, 6 verbs, and 6 adjectives and was determined by a norming study (see Chapter 1) that was conducted with a group of 150 participants with similar characteristics to the target population.

All items appeared a minimum total of 8 times (range: 8-40). Other word-related variables in the study included: cognate status, number of characters (range: 4-13), number of syllables (range: 1-4), frequency level (range: 2-9), concreteness (range: 2.36-4.53), and related forms (range: 0-31). Target items for Study 1, along with their aforementioned values, are listed in Table 3.5.

Table 3. 5

Target Words (N = 20) and Related Variables

Item	Verbal Freq	Other forms[a]	Log Freq (Zipf)[b]	Length		Concreteness[c]	Cognate status
				Characters	Syllables		
			Nouns				
supernova	15	0	3.08	9	4	3.78	Yes
constellation	11	0	3.20	13	4	4.31	Yes
sphere	16	31	3.68	6	1	4.44	Yes
spectrum	12	0	3.80	8	2	2.97	Yes
particle	13	0	3.48	8	3	3.78	Yes
temple	09	0	4.03	6	2	4.53	Yes
cosmos	35	12	3.27	6	2	3.19	Yes
tide	12	6	4.25	4	1	4.10	No
			Adjectives				
intricate	8	1	3.60	9	3	2.36	No
dense	24	3	3.74	6	1	3.14	Yes
denser	24	3	3.74	7	2	3.14	Yes
faint	9	1	3.75	5	1	3.74	No
cosmic	10	37	3.35	6	2	2.76	Yes
alien	10	0	4.19	5	2	3.52	No
			Verbs				
stretch	18	0	4.38	7	1	3.62	No
forge	09	0	3.57	5	1	4.04	No
emit	11	0	2.73	4	2	3.22	Yes
sculpt	11	2	2.61	6	1	3.57	Yes
orbit	40	1	3.73	5	2	3.11	Yes
squash	8	0	3.92	6	1	3.04	No

Note. Freq = frequency.

[a] Other forms were derivatives and compounds. [b] Measures were based on the SUBTLEX-UK word frequencies, presented in Zipf-values, a logarithmic scale: 1-3 = low frequency, 4-7 = high frequency (Van Heuven et al., 2014). [c] Measures were based on 40 thousand English lemma words on a 5-point rating scale going from abstract to concrete (Brysbaert et al., 2014).

As can be seen in the previous table, a total of 10 target words occurred in conjunction with other forms of the targets. In this study, target words were defined as the 'flemma' which refers to the baseword and its inflectional forms. Occurrences of compounds and derivational forms made up an additional variable to control for variation in target words' reoccurrences via other forms (Table 3.6).

Table 3. 6

Target Words with Compounds and Derivational Forms (N = 10)

Target word	Other forms	Freq of other forms
orbit	orbital	1
hexagon	hexagonal	6
symmetry	symmetric	2
	symmetrical	6
sulphur	sulphuric	2
fusion	fuse	6
sphere	spherical	5
	hemisphere	5
	atmosphere	21
tide	tidal	6
intricate	intricately	1
faint	faintly	1
dense	condense	2
	density	1
denser	condense	2
	density	1
sculpt	sculptor	2
cosmos	cosmic	10
	cosmology	1
	cosmologist	1
cosmic	cosmos	35
	cosmology	1
	cosmologist	1

Note. "cosmos" and "cosmic" were both a target word and a derivative.

The verb ‘to orbit’ in the study was a little bit problematic. The verb share the exact spelling and pronunciation with the noun ‘orbit’ which also occurred in the input. In total, the form occurred 11 times as a verb, 29 times as a noun, and once as the word form ‘orbital’ (i.e., other forms). Since the occurrence of the noun form contributes to the recognition of the verb’ spoken and written form, it was added to the total count of verbal occurrences for this target word.

Language Profile Questionnaire

A short questionnaire was designed to obtain background information about participants. The survey questions were used to inform the research results.

Procedure.

The questionnaire was administered in a single paper format (see Appendix H). It was composed of an introductory part, in which participants were asked to indicate their gender and the number of years they had been studying English, followed by two main parts. The first part consisted of 12 items and asked participants to self-report their perceived frequency of informal exposure to English language on a six-point Likert Scale. The questionnaire collected information regarding exposure to different genres of authentic videos, different types of captions, and different types of modalities. The out-of-class activities that were addressed are categorised in Table 3.7. The six responses were “everyday,” “several times a week,” “a few times a week,” “a couple of times a month,” “rarely”, and “never”. The second part of the questionnaire consisted of four items that tapped into participants’ perceived difficulty of English language input in four different modality integrations: authentic videos with and without L2 captions and Radio programmes and audiobooks with and without scripts, as shown in the table. Participants responded on a five-point Likert Scale ranging from easy (0) to difficult (4). The questionnaire required a maximum of 10 minutes to fill.

Table 3. 7

Questions' Categories in the Language Profile Questionnaire

	Frequency of exposure					Difficulty	
Gender	English language	Captions	Listening	Listening-while-reading	Genres	Authentic videos	Radio and Audio-book
· Male · Female	· No of years	· English · Arabic · French	· Music · Audiobook · Radio	· Music and lyrics · Audiobook and script	· Films · Documentaries · TV series · TV news · Sports games	· Without captions · With English captions	· Without script · With script

Oxford Placement Test

Participants' English language proficiency level was measured using The Grammar Test of the Oxford Placement Test (OPT) (Allan, 2004). OPT is a valid test used to test the homogeneity of the sample and compare participants with participants in other studies.

Procedure.

The Grammar Test of the OPT is a three-alternative forced-choice (3AFC) test of 100 items. The test used in this study consisted of 75 items presented in two parts. Reducing the number of items was necessary to reduce the time needed to complete the first sitting, which involved four data collection instruments. Participants received the test in a three-page booklet. The test lasted for a maximum of 40 minutes, and the time limit was not imposed. At the scoring level, participants were given one score for each correct response.

Dependent Measures

This study operationalised word knowledge as knowledge of meaning recall, meaning recognition, spoken form recognition, and written form recognition.

Results from multiple aspects are believed to provide a more accurate measure of word knowledge, thus, a more comprehensive picture of participants' incidental vocabulary development. Knowledge of form was measured employing a pretest-posttest design, while knowledge of meaning was measured using posttests only. This was done to prevent exposure to the target words' correct forms before treatment. Moreover, the meaning of the target words were found to be unknown to a sample of participants with similar characteristics to the target population. This result supported the decision of excluding the pretest for meaning measures. The pretests were administered two weeks before the first treatment session, while posttests were administered immediately after the fourth (last) treatment session. A one-week delayed posttest followed the posttests. The following sub-section describes the instrument and procedure implemented for every dependent measure. Instruments are accompanied by audio recordings and are therefore included as supplementary material. Answer sheets for the four instruments are provided in Appendix I in their order of distribution.

Meaning Recall.

Participants were posttested on meaning recall using a meaning translation test.

Procedure.

Target words were presented individually on screen in their written form. Participants were asked to view the word and write the L1 translation or the equivalent English definition/synonym on the answer sheet. Timing for this test was initially set as 45 seconds per item; however, piloting showed that 30 seconds was more than enough to answer each item. The meaning recall test involved 28 target words (including words of Study 3) and lasted for 15 minutes. To reduce the amount of guessing, items were ordered pseudo-randomly in terms of their part of speech.

Meaning Recognition.

Participants were posttested on meaning recognition using a bilingual matching test.

Procedure.

Two blocks were presented on screen, one after another. Each block consisted of 14 target words on the left and 15 L1 (Arabic) translations on the right. One of the L1 translations was a distractor. Participants were asked to match each target word to its equivalent translation. The translations were determined after consulting a lecturer in Arabic language. Based on the piloting results, participants could answer each block within 14 minutes. The test lasted for about 30 minutes.

Spoken Form Recognition.

Knowledge of spoken form recognition was measured in a pretest-posttest design using a 3AFC test (i.e., including the correct spoken form and two distractors).

Procedure.

The three options of items were voice-overed by a British professional sequentially (separately) and preceded by A, B, and C, respectively. Participants were asked to listen to the three options and indicate, on their answer sheet, which option is a correct English language word form. "I don't know" was always added as a final option "D" to minimise guessing. Participants had 10 seconds to answer each word. The spoken form test was in the form of a 3AFC test to preserve the internal validity

of the test, as it was believed that having more than two distractors could involve an extraneous variable which is the participant's memory, thus, skewing the test results.

Fillers were included in the pretest to guard against the negative effects of pretesting. The pretest was composed of 56 items and lasted for about 22 minutes while the posttest was composed of 28 items and lasted for about 12 minutes. Pilot participants found the pretest to be undemanding regarding the sequence of the three items and the time provided. Unlike the pretest which occurred at the beginning of the term, the posttest occurred at the end of it and included tests of meaning. Students were preparing for exams, thus, excluding fillers in the posttest might have helped maintaining the same fatigue level in the two tests.

Filler Items.

Creating filler items was done using Oxford Dictionary, *Compleat Web VP* function from the *Lextutor* (Cobb, n.d) (https://lextutor.ca/vp/comp/), online letter and syllable counters, and word information website (https://yougowords.com). The number of filler items was the total number of words targeted in the studies of this thesis (i.e., 28). Filler items were matched to target items orthographically (i.e., number of characters, number of syllables) and in terms of parts of speech (16 nouns, 6 adjectives, and 6 verbs) (see Tables 3.8 and 3.9) and word difficulty. This was done to reduce the salience of particular items over others by maintaining the same degree of learnability in the two sets of words.

Table 3. 8

Number of Target Words in Relation to the Number of Syllables and Characters

	No. of characters/syllables									
	1	2	3	4	5	6	7	8	9	13
No of target words by No of syllables	11	11	4	2						
No of target words by No of characters				4	5	9	4	3	2	1

Table 3. 9

Fillers in Relation to Number of Characters, Syllables, and Parts of Speech of Target Words

	No. of Characters						
	4	5	6	7	8	9	13
No. of fillers	4 null, fail, file, lush,	5 knife, match, trill, hutch, tense,	9 spleen, regent, mumble, rouble, limpid, wonder, beagle, impale, aviary	4 stealth, reptile, pigtail, devious	3 standard, maintain, emeritus	2 integrate, thesaurus	1 transcription

	No. of Syllables			
	1	2	3	4
No. of fillers	11 null, fail, knife, spleen, match, trill, file, hutch, tense, lush, stealth	11 regent, standard, reptile, mumble, pigtail, rouble, limpid, maintain, wonder, beagle, impale	4 integrate, transcription, devious, aviary,	2 thesaurus, emeritus

	Parts of speech		
	Nouns	Verbs	Adjectives
No. of fillers	16 knife, spleen, file, hutch, tense, stealth, pigtail, rouble, transcription, aviary, thesaurus, trill, regent, reptile, wonder, beagle	6 fail, match, maintain, impale, integrate, mumble	6 null, lush, limpid, devious, emeritus, standard

Distractor Items.

The distractors used in the multiple-choice spoken form test were non-words and were presented for both target and filler items. Distractors were generated by creating phonological neighbours using phoneme deletion, addition, or substitution. An example of the two distractors for the target word intricate (/ɪntrɪkət/) are /ɪntrikeit/ and /ɪntrigət/.

Sequence of Items.

To reduce the amount of guessing in the 3AFC test, items were ordered pseudo-randomly in terms of (1) its function in the pretest (target/filler) (the posttest did not include fillers), and (2) its part of speech in the pre and posttest. Options of each item were also ordered pseudo-randomly in pre- and posttests with respect to the position of the correct form. The sequences were generated using Excel. The first three items of the spoken form recognition pretest are presented in Table 3.10. For the full list of target and filler items as they appeared in the pretest, see Appendix J.

Table 3. 10

First Three 3AFC Items on Spoken Form Recognition Pretest

	Options		
Item	A	B	C
1- match	/ mɒtʃ/	**/matʃ/**	/mɛtʃ/
2- regent	/ˈ**riːdʒ(ə)nt/**	/ˈreiʒ(ə)nt/	/rɒdʒ(ə)nt/
3-consetalltion	**/kɒnstəˈleɪʃ(ə)n/**	/kənstəˈleɪʃ(ə)n/	/kɒntəˈleɪʃ(ə)n/

Note. Options were presented in pseudo-random order in terms of the function of item (target/filler), its part of speech, and the position of the correct spoken form. The item in bold is the correct form.

Written Form Recognition.

Knowledge of written form recognition was measured in a pretest-posttest design using a 4AFC test (i.e., including the correct written form and three distractors). The study used three distractors for being an optimal option in vocabulary testing (e.g., Baghaei & Amrahi, 2011).

Procedure.

The four items' options were projected to screen concurrently and preceded by A, B, C, and D, respectively. Participants were asked to read the four options and indicate, on their answer sheet, which option is a correct English language word form. "I don't know" was always added as a final option, "E", to minimise guessing. Participants had 14 seconds to answer each word which was sufficient time based on the pilot study. Filler items that formed part of the spoken form pretest (see Table 3.9) were included in the written form pretest to guard against the negative effects of pretesting. The pretest was composed of 56 items and lasted for about 14 minutes. The posttest was composed of 28 items and lasted for about 8 minutes.

Distractor Items.

The distractors used in the multiple-choice written form test were non-words and were presented for both target and filler items. Distractors were generated by creating orthographic neighbours using letter deletion, addition, or substitution. An example of the three distractors for the target word intricate are insicrate, enricate, and intrigate.

Sequence of Items.

To reduce the amount of guessing in the 4AFC test, items and options of items were ordered pseudo-randomly each, in the same manner described in spoken form recognition section. The first three items of the written form recognition pretest are presented in Table 3.11. The full list of target and filler items as they appeared in the pretest is available in Appendix J.

Table 3. 11

First Three 4AFC Items on Written Form Recognition Pretest

	Options			
Item	A	B	C	D
1. thesaurus	**thesaurus**	thesaumus	thauresus	thesomus
2. supernova	sperniva	superneve	sperneva	**supernova**
3. alien	**alien**	feillen	feelian	alian

Note. Options were presented in pseudo-random order in terms of the function of item (target/filler), its part of speech, and the position of the correct written form. The item in bold is the correct form.

Scoring.

Responses on the four dependent measures tests, that is, meaning recall, meaning recognition, spoken form recognition, and written form recognition were scored dichotomously (0, 1): "0" for incorrect, missing, and "I don't know" responses, and "1" for correct responses. Scoring was straightforward for all tests except for meaning recall which was in a format that allowed for some subjectively. However, almost all responses were in Arabic definitions and were fairly easy to score as correct or incorrect with only a few partially correct responses. The intended use of dichotomous coding made it impossible to give 0.5 score. Therefore, lenient scoring for accuracy was applied in that response was scored as correct if it revealed

knowledge of at least one semantic feature of the target word. For example, the answer "ظاهرة فلكية" (i.e., astronomical event) for *supernova* was accepted as correct, although a more accurate response is "an explosion or death of a star". These responses that were less easy to score were rare. Thus, an assessment of inter-rater reliability was not necessary; the answers were discussed with a second rater who was a retired teacher of the Arabic language. The data was transferred into an MS Excel spreadsheet. For meaning recognition, spoken form recognition, and written form recognition, the answers were in the form of letters or numbers. Hence, using the conditional format function from MS Excel for these tests allowed an automatic generation of "0" and "1" data for incorrect and correct responses, respectively, depending on the letter/number response opted by the student.

Comprehension Questions

A total of 48 comprehension questions were created to test participants' viewing comprehension every twenty minutes; that is, six questions per episode, 12 questions per session (see Appendix K). Participants were not allowed to read the questions before viewing because this practice may encourage them to find specific information rather than achieve general comprehension. The comprehension questions served three aims: to minimise demotivation that might result from extensive content to convey the impression that the experiment was intended for content comprehension purposes, thus, maintain an incidental context to learning, and explore differences in comprehension between the View and Non-View participants. As such, items consisted of literal questions to ensure construct validity (i.e., the test measures what it purports to measure). Therefore, wrong answers were expected to be the result of either (1) absent-mindedness, (2) poor listening, reading, or comprehension skills, or (3) poor comprehension as a result of an adverse effect of modality. To ensure that this task was inclusive of both groups, the questions were text-based. That is, questions were based solely on what was included in the spoken and written text and were not imagery-based or audio plus imagery-based. It should be noted that a recent study suggested that audio-based questions might require higher lexical demands than audio plus imagery-based questions (Durbahn et al., 2020). The questions consisted of 23 true/false items, 23 3AFC items, and two 2AFC items. Three options were found to be optimal for multiple-choice tests based on 80 years of research (Rodriguez, 2005). Options of each item were ordered

pseudo-randomly with respect to the correct answer. The sequences were generated using Excel. Comprehension items were scored dichotomously using "0" and "1" scores for incorrect and correct responses, respectively.

Debriefing Questionnaire

At the end of the study, participants were given a debriefing form that included a thank you statement for their participation in the study and disclosure of the study's actual purpose. Participants were invited to voice their opinions, observations, or concerns about the study in a short non-directive interview. This was followed by a debriefing questionnaire (N= 63) to participants of the View (N = 34) and Non-View (N = 29) groups. The questionnaire aimed to determine their self-reported perceptions of the treatment regarding multiple aspects. These perceptions were meant to elevate the discussion of the results.

Procedure.

The questionnaire was administered online via Qualtrics (see Appendix L). It consisted of 11 items. The first item asked participants to indicate the group to which they were assigned in the study. The remaining 10 questions were 10-point Likert-scale based items and one 3AFC item. The questions were designed to explore participants' attitudes towards different aspects involved in the treatment. Mainly, the questionnaire contained items assessing the View and Non-View groups on two general areas of interest: information processing and motivation. Both areas pertained to the context of the study and were selected based on their theoretical underpinnings. Information processing-related items included comprehension, input processing, the utility of captions, split attention, and speaker recognition. Motivation-related items included enthusiasm to start learning vocabulary, length of episodes, imagery absence, and intrinsic/integrative motivation as an EFL learner. Overall, the questionnaire items elicited participants' preferences and assessed their self-reported evaluation and satisfaction with the treatment. The questionnaire took about 2-3 minutes to complete.

3.3.7 Experimental Procedure

The experiment adopted a between-participants design. The first sitting consisted of four tasks devoted to the Control, View, and Non-View groups. It began with a 5-minute introductory phase in which numerical codes were distributed; each was unique to each student. Students were requested to use the code throughout the duration of the study to facilitate the accurate tracking of students' data while preserving their anonymity. Students who agreed to participate were given an information sheet about the study's nature and aims, accompanied by a consent form to be signed. Participants then completed the language profile questionnaire and the vocabulary pretests (Spoken form recognition and written form recognition). The vocabulary pretests were administered via PowerPoint slides projected into a 2-meter by 1.5-meter screen and completed by participants in paper formats. None of the target words appeared on the answer sheets and no paper was allowed upon the desk, except the distributed booklets. This was done to prevent participants from reading the correct written forms on their own pace and/or taking notes of them, thus, reducing exposure time to minimise potential pretest effect. The time specified for responding to each item and breaks was set up on PowerPoint using the timing feature on the Transition tab; this allowed a smooth proceeding of the tests since the slides advanced automatically whenever the allotted time had run out. After a 90-minute pause, the participants filled the Oxford Placement Test.

The distribution of answer sheets at the beginning of every new task was thought to cause a disruption. Therefore, participants were given a booklet at the beginning of every sitting. The booklet contained answer sheets (i.e., sub-booklets) for all the tasks due in the sitting. To prevent participants from reading the questions before the task begins, the booklets were placed face down on the participant's table. At the beginning of each task, participants were asked to lift the front answer sheet and write down their given numerical codes in the field designed for them.

A two-week interval was allowed before the start of the treatment sessions. The four treatment sessions were intended for the experimental groups only and took place over a six-week period with two-week interval. In each treatment session, participants in the View group watched two episodes (i.e., 2 hours) of two full-length seasons of documentary series in the form of L2 captioned video. Non-View

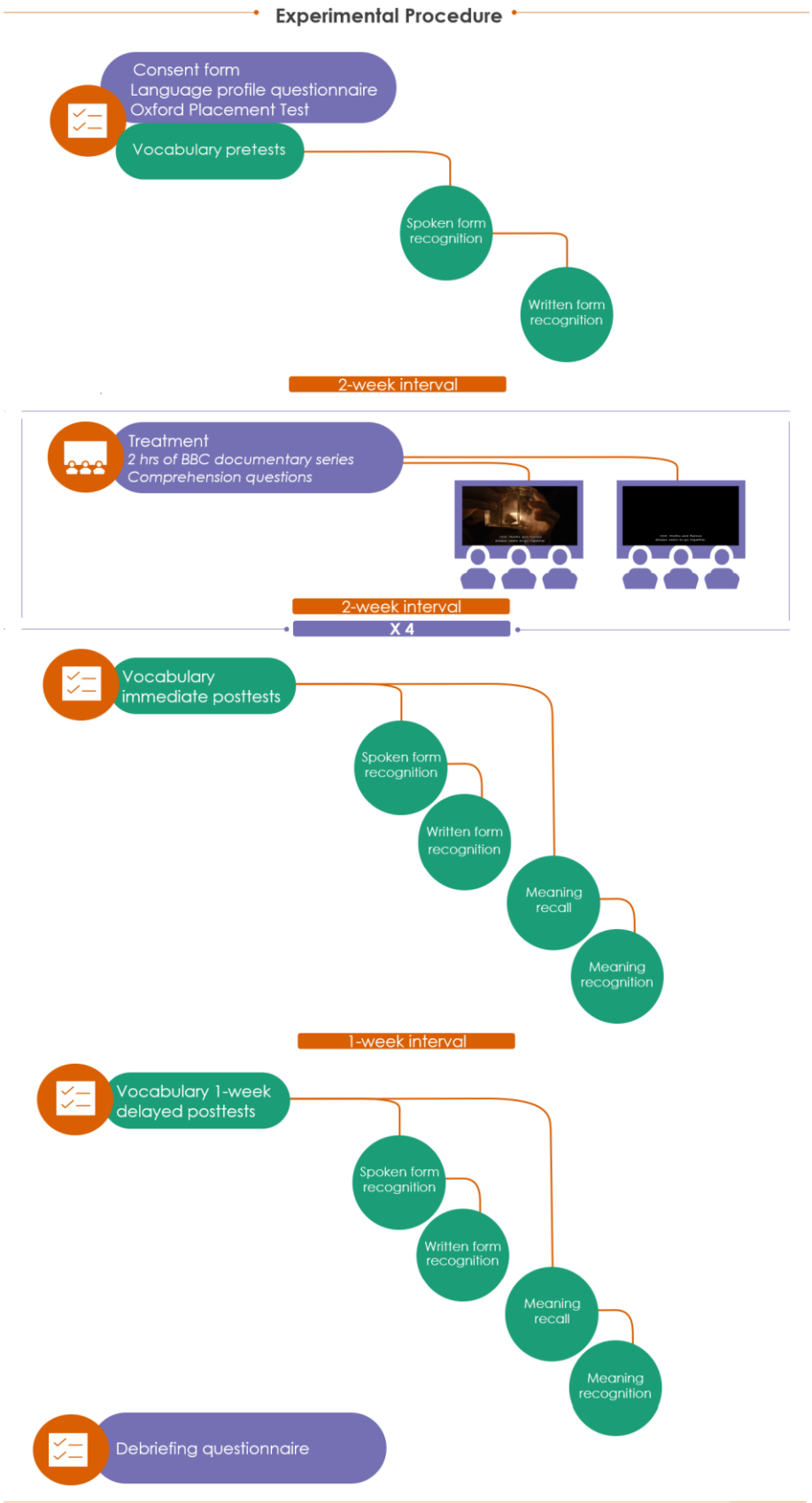
participants were exposed to the same material except that they had imagery removed; thus, they were exposed to audio and L2 captions only. By the end of the experiment, participants in the two groups had a total exposure of 8 episodes (i.e., 8 hours).

In every treatment session, participants were interrupted once every twenty minutes to answer two comprehension questions. The allotted time to answer each question was 30 seconds. Unlike other studies, questions preview was prevented. This was done because the research aim was to measure vocabulary knowledge that results from incidental exposure and the desire to achieve overall comprehension. Question's preview consists of an indirect request to locate specific information in the input to achieve higher scores in an upcoming test. This was thought to divert attention from general comprehension on the one hand and potentially unknown words on the other. Question's preview was avoided by printing the 12 questions of each session on A4 size papers, then cut these into four and place them face down, each of which consisted of two questions intended for the 20 minutes viewing period.

A posttest for the three groups immediately followed the fourth (last) treatment session. The delayed posttest was administered one week later. In posttests, participants were tested at the level of the four dependent measures, unlike the pretests. To prevent transfer of word knowledge between different tests, the tests followed a specific sequence. In the pretest, the spoken form test was administered first, followed by the written form test. In the posttest, form tests preceded meaning tests, and the meaning recall test preceded the meaning recognition test. Also, to minimise frustration, participants took in the pretest a one-minute break between items 30 and 31 of each form test (56 items in each) and a three-minute break between the two form tests. In posttests (i.e., 28 items), participants took a one-minute break between the two form tests and a three-minute break between form and meaning tests and between meaning recall and meaning recognition tests. Finally, participants responded online to the debriefing questionnaire at the end of the study. Figure 3.5. demonstrates the overall experimental procedure.

Figure 3. 5

The Experimental Procedure



3.4 Analyses

In this section, I will outline the procedure followed to analyse data for Study 1. In the section that follows it, I will report on the results of the language profile survey and the OPT test. I will then answer the two research questions before lastly reporting comprehension and debriefing results.

3.4.1 Analysis Procedure

Data were analysed in R (R Core Team, 2018) using Rstudio (version 3.5.1; RStudio Team, 2018). Results were summarised using dplyr package (version 0.8.4; Wickham, François, Henry, & Müller, 2019). They were visualised using ggplot2 package (version 3.2.1; Wickham, 2016) for meaning recognition and recall results and ggpaired function of ggpubr package (version 0.2.4; Kassambara, 2019) for form recognition results (for using a pretest-posttest design). Data were analysed with generalised linear mixed-effects (GLM) logistic regression models using glmer function of the lme4 package (version 1.1-26; Bates, Maechler, Bolker, & Walker, 2015). The procedures implemented to answer the two research questions of Study 1 are explained in what follows.

The first research question examined whether viewing two full-length seasons of documentary series, extending to eight hours, in the form of L2 captioned video over a six-week period of two-week intervals leads to incidental vocabulary learning. The second research question asked whether similar results would be achieved if imagery was hidden from view. A GLM logistic regression analysis was applied to the data set for the View, Non-View, and Control groups to obtain results for both questions. The analyses included all word-related explanatory variables that were theoretically meaningful.

For meaning recall and recognition measures which lacked a pretest, the baseline models specified posttest accuracy as the dependent variable, written form pretest as a control variable, parts of speech, frequency of occurrence, frequency of occurrence of related forms, characters, syllables, concreteness, cognate (cognate = 1, noncognate = 0), and corpus frequency as control covariates, and participants and words as random effects, with random intercepts allowed to vary across participants and words (e.g., random = ~1 | word). The significance of the effect of group was

then assessed using likelihood ratio tests which compared the baseline model to an identical model with group as an additive predictor.

For spoken and written form recognition, data were in a pretest-posttest design. The models specified response accuracy as the dependent variable, time and group as fixed effects, parts of speech, frequency of occurrence, frequency of occurrence of related forms, characters, syllables, concreteness, cognate, and corpus frequency as control covariates, and participants and words as random effects, with random slopes of time for each since the effect of time varies across participants and words (e.g., random = ~Time | word). Significance of the main effects of group and time and time × group interaction were then assessed using likelihood ratio tests which compared the full model with identical models with factor or interaction of interest removed. To further investigate the effect of group, post-hoc pairwise comparisons were performed for the significant interaction between group and time, using Emmeans Package (1.4.4), to compare improvement from pretest to posttest within each group (with Bonferroni adjustment for multiple comparisons, $\alpha = .017$).

The full models were prone to inflated standard errors and were therefore simplified by removing all word-related covariates. If the substantive results differed from the full models, only variables that did not significantly predict response accuracy were removed following a stepwise procedure for model comparisons using the likelihood ratio test. The variable Group was automatically dummy coded by R software as a categorical variable, then releveled so that View group (N = 53) was the reference level. The two factors Parts of speech and Cognates were contrast-coded using "contr.sum" function so that analysis was conducted on the grand mean (intercept) of all levels rather than one specific level.

The study had a large sample size which allowed the implementation of multilevel modelling and inclusion of theoretically meaningful covariates and maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013). This helped control individual variations among participants and across words, thus meeting the independent assumption (in both time points for spoken and written form recognition).

3.5 Results

3.5.1 Language Profile Questionnaire

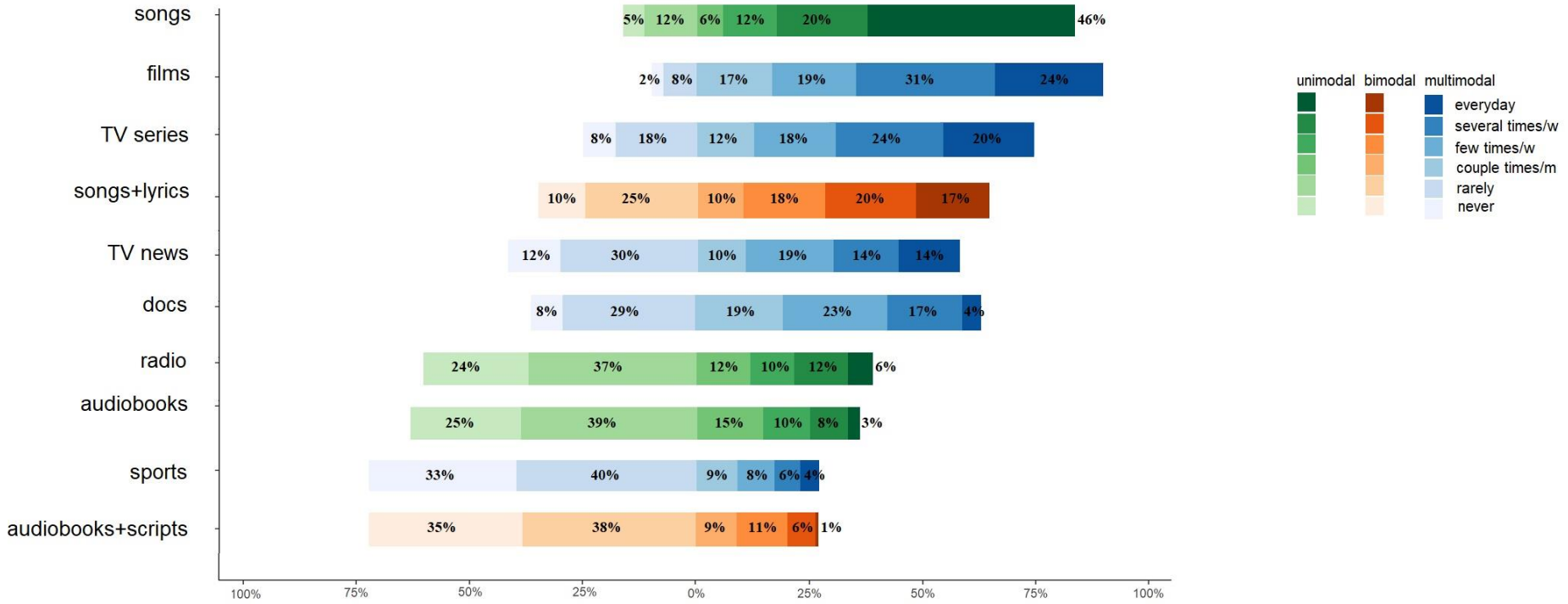
Responses to the language profile questionnaire were analysed and are presented in Figure 3.6. This report explains participants' extensive extra-mural exposure, preferred language of captions, and perceived difficulty of unassisted listening. The results demonstrate that the third-year undergraduate Algerian L2 learners listened to English language songs to a larger extent outside the classroom (78%), with almost 46% of participants indicated that they listened to songs everyday, showing that songs are the most valuable source of input for these learners. The second activity with the highest exposure frequency was watching English language films. About 72% of participants watched films daily (24%), several times a week, or a few times a week, followed by watching TV series (62%). The figure also shows that a high number of learners tended to read lyrics while listening to songs. This indicates that learners (1) often encountered difficulties in listening to English language input, (2) were aware of the benefits of the bimodal input, and (3) were motivated to understand and learn the language. As can be observed, learners were attracted to other genres of audio-visual input, including documentary series. About 44% of participants watched documentaries daily or weekly, while about 20% of participants watched them a couple of times a month.

In contrast to multimodal input, and apart from songs, learners showed a limited exposure to other types of aural input such as radio and audiobooks. The extremely high number of participants who rarely or never watched sports games in the English language is not surprising as 91% of participants were females, and this gender generally does not find sports entertaining as men do (e.g., Deaner & Smith, 2013).

Furthermore, the results showed that participants preferred watching multimodal input with English language captions. About 30% of respondents reported that they rarely or never opted for Arabic and French subtitles, compared to about 16% who ignored English captions. With regards to the perceived difficulty of

Figure 3. 6

Frequency of Out-of-Class Exposure to English Language Input

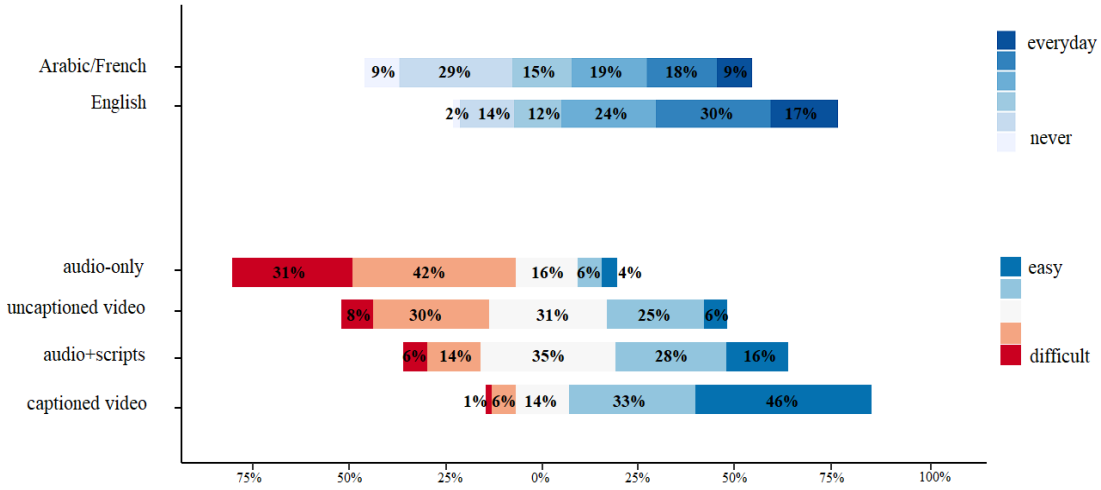


Note. Question = How often do you do the following in ENGLISH language? N= 144; /w = per week; /m = per month.; docs = Documentaries. Percentages may not total 100 due to rounding.

different types of input, audio-only input appears to have been the most difficult English language material to the target participants, with more than 70% checking the two highest difficulty ratings (3+4 points). This result lends credence to the previous result that participants resorted to reading lyrics when they faced listening difficulties. However, the percentage of respondents who perceived uncaptioned video as difficult was 38%. The figure shows that participants perceived scripts (for audio) and captions (for video) as great assists for comprehension, with 78% of respondents showed that they found English language captioned viewing as easy (0+1 points). These results are presented in Figure 3.7.

Figure 3. 7

Preference for Language of Captions and Perceived Difficulty of Unassisted Listening



Note. Questions = How often do you do the following in ENGLISH language? How difficult to understand do you find the following in ENGLISH language? N = 144; audio = radio programmes and audiobooks; video = authentic video. Difficult = 4, easy = 0. Percentages may not total 100 due to rounding

3.5.2 Oxford Placement Test

The descriptive statistics for English language proficiency per each group are provided in Table 3.12.

Table 3. 12

Descriptive Statistics for OPT Scores

Group	*M (SD)*	*Min*	*Max*	*n*	*na.s*
View	43.94 (7.16)	32	63	53	4
Non-View	42.11 (6.97)	18	61	57	3
Ctrl	42.39 (6.02)	31	57	34	1

Note. by(data, data$group, summary). M = mean. Maximum score = 75.

As can be observed above, the Kruskal-Wallis test[3] showed no significant differences in English language proficiency between the three groups, $\chi 2$ (2) = 1.09, p = .58. The one-way non-parametric ANOVA was used because the groups had equal variances (homogeneity)[4], p = 74, but data were not normally distributed[5], p < .05.

[3] kruskal.test(score~group, data = opt)
[4] leveneTest(score~group, data = opt)
[5] aggregate(score~group, data = opt, function(x) shapiro.test(x)$p.value)

3.5.3 Dependent Measures

The mean scores for meaning recall and recognition posttests and written and spoken form recognition pretests and posttests for the View (N = 53), Non-View (57) and Control (N = 34) groups on 20 items were summarised and plotted and are presented in Table 3.13.

Table 3. 13

Descriptive Statistics per Group for all Vocabulary Tests Scores (20 Items)

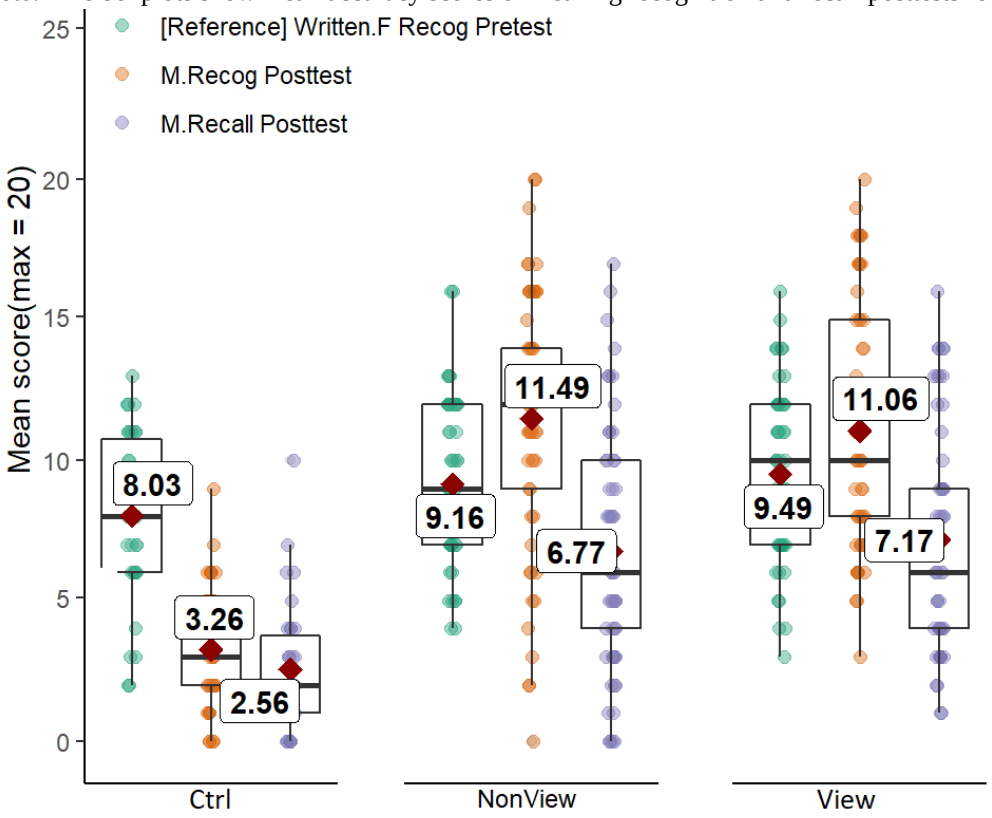
		Mean Scores					
		Pretest			Posttest		
		M (SD)	*Min*	*Max*	*M (SD)*	*Min*	*Max*
Meaning Recall	Control				2.56 (2.26)	0	10
	Non-View				6.77 (4.26)	0	17
	View				7.17 (4.03)	1	16
Meaning Recognition	Control				3.26 (2.29)	0	9
	Non-View				11.49 (4.59)	0	20
	View				11.06 (4.40)	3	20
Spoken Form Recognition	Control	6.44 (2.70)	1	12	6.94 (2.58)	2	12
	Non-View	5.25 (2.70)	2	14	9.91 (3.05)	3	17
	View	8.09 (2.96)	3	15	10.79 (3.46)	3	17
Written Form Recognition	Control	8.03 (2.98)	2	13	7.21 (2.95)	2	13
	Non-View	9.16 (2.75)	4	16	12.05 (2.88)	4	17
	View	9.49 (3.06)	3	16	12.57 (3.17)	6	19

Note. data %>% group_by(group, Time) %>% summarise (mean = mean(Response), sd = sd(Response), max = max(Response), min = min(Response)).
M = mean. SD = standard deviation. Maximum score = 20.

Meaning accuracy data (Figure 3.8) show that, regardless of viewing method, participants who were exposed to eight episodes of L2 captioned documentary series scored substantially higher than the Control group at both levels of tests (i.e., recall and recognition). Moreover, both View and Non-View groups appear to have scored equally well, with the score being greater at recognition test than at recall test.

Figure 3. 8

Mean Accuracy in Meaning Recall and Recognition

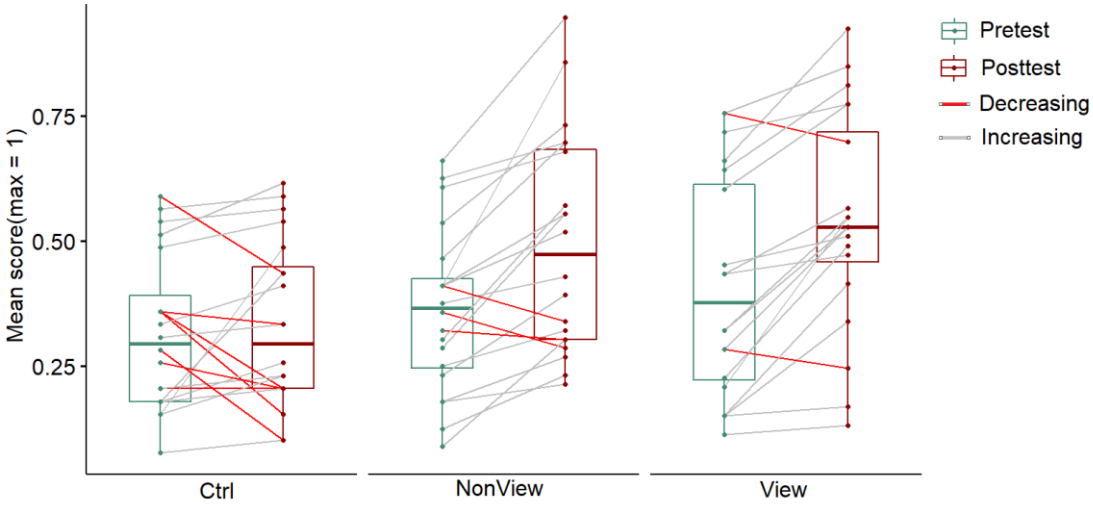


Note. The boxplots show mean accuracy scores of meaning recognition and recall posttests for 20 words by subject and across Control (N =34), Non-View (N = 57), and View (N = 53) groups. Meaning was not pretested to prevent prior exposure bias, written form recognition pretest was used as a baseline reference. Means are represented in the figure by the red points.

From the paired boxplots shown in Figure 3.9, it is apparent that accurate recognition of spoken form increased considerably from pretest to posttest for both View and Non-View groups. In contrast, the Control group showed minimal gains with a commensurate number of decreases in performance compared to experimental groups. Both View's and Non-View's recognition were fairly comparable.

Figure 3. 9

Mean Accuracy in Spoken Form Recognition

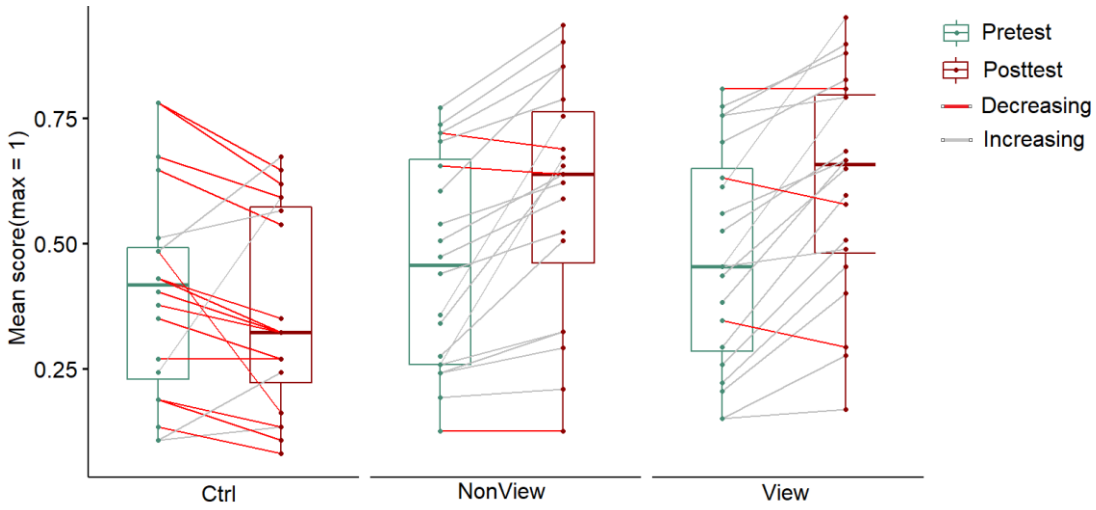


Note. The paired boxplots (by word) show mean accuracy scores of spoken form recognition across Control (N =34), Non-View (N = 57), and View (N = 53) groups for 20 words. Grey and red lines match mean scores from pretest to posttest.

A similar trend was marked for mean accuracy data for written form recognition, in which decreases in scores for Control participants were increasingly higher (Figure 3.10).

Figure 3. 10

Mean Accuracy in Written Form Recognition



Note. The paired boxplots (by word) show mean accuracy scores of written form recognition across Control (N =34), Non-View (N = 57), and View (N = 53) groups for 20 words. Grey and red lines match mean scores from pretest to posttest.

Preview

To preview the results, a significant effect of viewing two full length seasons of documentary series (amounting to 8 hours) on incidental vocabulary learning was revealed in all aspects of word knowledge tested. There was, however, no significant difference between the group who was exposed to the multimodal input with imagery included and the group who had imagery removed and was exposed to bimodal input only on all dependent measures tested. A detailed review of the results is provided next.

Statistical tests will next be performed to determine whether differences between groups are statistical. Reported in tables are the coefficients of the final model fixed effects and random effects on response accuracy by participants. The first column provides the change in the log odds of response accuracy associated with a change in group conditions. A positive coefficient indicates an increase in accuracy, while a negative coefficient indicates a reduction in accuracy relative to the baseline category. Odds ratio can act as a useful effect size statistic. For all the four dependent measures, the word-related covariates were pruned from the model without an effect on the significance of the variable of interest and overall substantive results to enable a more parsimonious model without inflated standard errors. Worthy of noting is that concreteness, cognate, and corpus frequency emerged as significant predictors of spoken form accuracy in the full model. Parts of speech and frequency of related forms were predictors of written form accuracy, and characters predicted both forms. For every dependent measure, the results of both research questions are obtained from the same model; hence, results are reported by dependent measures. The results for the four dependent measures in this and the subsequent studies are reported separately and arranged in the following order: meaning recall and recognition and spoken and written form recognition. This hierarchy by which recall and meaning tests are prioritised over recognition and form tests, respectively, reflects the limitation arising from the use of pretest in form recognition measures and the multiple-choice format in recognition measures. The latter has recently been known for its overestimation issue (Gyllstad, Vilkaitė, & Schmitt, 2015; Schmitt, Nation, & Kremmel, 2020). Spoken form results precede written form results in this report because it represents a stern test since its multiple-choice items were voiced-over instead of projected into the screen.

Meaning Recall

The results showed a significant group effect in meaning recall accuracy, $\chi^2(2) = 38.61$, $p < .0001$. The odds of a correct response in the View group were more than five times as high compared to the Control group (*OR* = 1/Exp(B) = 1/.18 = 5.56, 95% CI [0.10, 0.31]). There was no significant difference between the View and Non-View groups ($p = .522$). Pairwise comparisons with Bonferroni adjustment showed that the Non-View group performed significantly better than the Control group ($p < .0001$). The coefficients estimates for meaning recall responses are reported in Table 3.14. Taken together, these results demonstrate that meaning recall accuracy depended on whether participants were exposed to the extensive documentary input, but not on whether they viewed imagery or not.

Meaning Recognition

The results revealed a significant effect of group in meaning recognition accuracy, $\chi^2(2) = 81.24$, $p < .0001$ with accurate scores under View condition being significantly higher than Control condition. The odds of accurate recognition of meaning in the View group were 11 times higher in the View group compared to the Control group (*OR* = 1/Exp(B) = 1/.09 = 11.11, 95% CI [0.05, 0.15]). No significant difference in accuracy was found between the View and Non-View groups ($p = .684$). Pairwise comparisons with Bonferroni correction showed that scores in the Non-view group were significantly higher than scores in the Control group ($p < .0001$). The coefficients estimates for response accuracy in meaning recognition are reported in Table 3.14. Overall, the results indicate that meaning recognition accuracy was heavily affected by exposure to the two full-length seasons of the documentary series, irrespective of imagery presence.

Table 3. 14

GLM Logistic Regression Predicting Meaning Accuracy

	Meaning recall					Meaning recognition				
Parameters	*B*	*SE*	*z*	*p*	*OR*	*B*	*SE*	*z*	*p*	*OR*
Fixed effects										
Intercept	−0.99	0.27	−3.71	***	0.37	0.14	0.26	0.56	.575	1.15
Group = Ctrl	−1.71	0.28	−6.13	***	0.18	−2.44	0.28	−8.87	***	0.09
Group = Non-View	−0.15	0.23	−0.64	.052	0.86	0.09	0.22	0.41	.684	1.10
Group = View										
Written. F	0.44	0.11	3.96	***	1.54	0.42	0.11	3.99	***	1.52
Random effects				Variance	*SD*				Variance	*SD*
Participant (intercept)				1.11	1.05				1.10	1.05
Item (intercept)				0.85	0.92				0.75	0.87

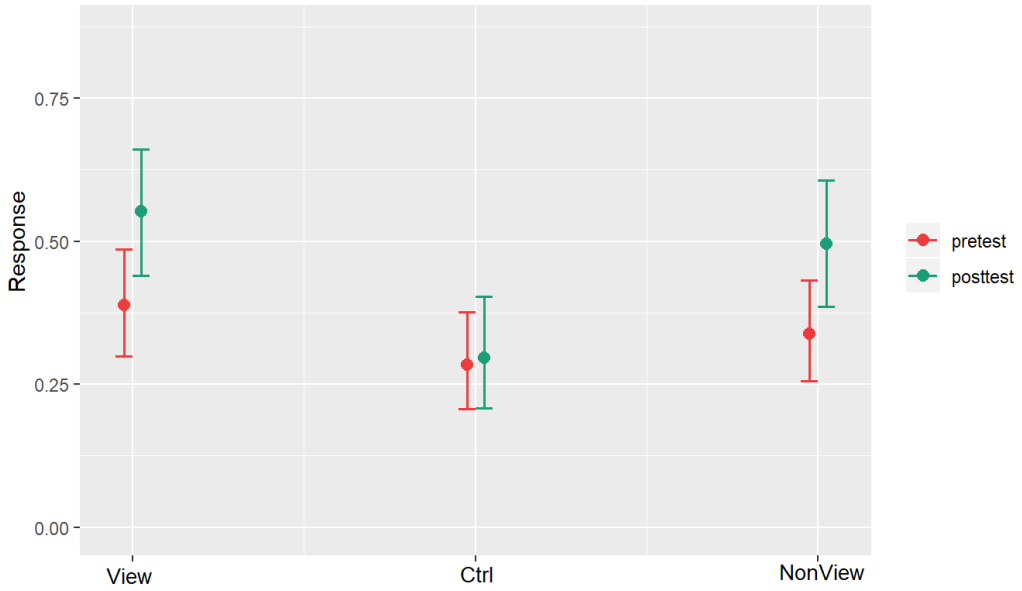
Note. Posttest ~ group + written form pretest + (1|participant) + (1|item). Baseline category = View group. Model fitted to 2880 observations across 20 words. N = 144.
****p* <.001

Spoken Form Recognition

The analysis showed a significant main effect of both group, $\chi^2(2) = 18.77$, $p < .0001$, and time $\chi^2(1) = 84.68$, $p < .0001$. Random slopes of time were significant for items, $\chi^2(2) = 18.55$, $p < .0001$, but not for participants, $\chi^2(2) = 3.47$, $p = .176$, but were retained both in the model to allow a maximal random effects structure. The two-way interaction between group and time was significant, $\chi^2(2) = 13.52$, $p < .01$, indicating that gains from pretest to posttest were different between groups. A negative estimate for group (view) × time (posttest) interaction indicated that learning gains were significantly stronger in the View group compared to the Control group, with no significant difference in gains between the two experimental groups ($p = .929$) (see Figure 3.11).

Figure 3. 11

Condition Effects in Spoken Form Recognition



Note. The predicted probabilities plot shows the probability values for response accuracy on spoken form recognition of 20 words by View (N = 53), Non-View (N = 57), and Control (N = 34) groups, calculated from GLM logistic regression analysis.

To follow up on this significant interaction, pairwise comparisons were conducted to determine the significance of time within each of the three groups separately. There were significant gains from pretest to posttest in the View group, (B = -0.67, SE = 0.13, z = -5.06 , p < .0001) and the Non-View group (B = -0.66, SE = 0.13, z = 5.05, p < .0001) but not in the Control group (B = -0.10, SE = 0.16, z = -0.64 , p = .525). The coefficients estimates for response accuracy in spoken form recognition are reported in Table 3.15. In sum, the findings indicate that exposure to the two full-length seasons of the documentary series produced significant gains in knowledge of spoken form recognition, regardless of whether participants viewed imagery or not.

Written Form Recognition

The results revealed a significant main effect of group, $\chi^2(2) = 21.98$, $p < .0001$, and time $\chi^2(1) = 83.49$, $p < .0001$. Random slopes were significant for items, $\chi^2(2) = 28.19$, $p < .0001$, and participants, $\chi^2(2) = 13.19$, $p < .01$ and were retained both in the model. The two-way interaction between group and time was significant, $\chi^2(2) = 42.21$, $p < .01$, indicating that gains from pretest to posttest were different between groups. A negative estimate for group (view) × time (posttest) interaction indicated that learning gains were significantly stronger in the View group compared to the Control group, with no significant difference in gains between the two experimental groups (p = .621) (see Figure 3.12).

To follow up on the significant interaction, pairwise comparisons were carried out to determine the significance of time within each group. Significant gains in knowledge of written form were achieved in the View group, (B = -0.85, SE = 0.15, z = -5.59 , p < .0001) and the Non-View group (B = -0.78, SE = 0.15, z = 5.22, p < .0001) but not in the Control group (B = 0.29, SE = 0.17, z = 1.64 , p = .100). The coefficients estimates for response accuracy in written form recognition are reported in Table 3.15. In sum, the findings indicate that exposure to the two full-length seasons of the documentary series produced significant gains in knowledge of written form recognition, regardless of whether participants viewed imagery or not.

Table 3. 15

GLM Logistic Regression Predicting Form Accuracy

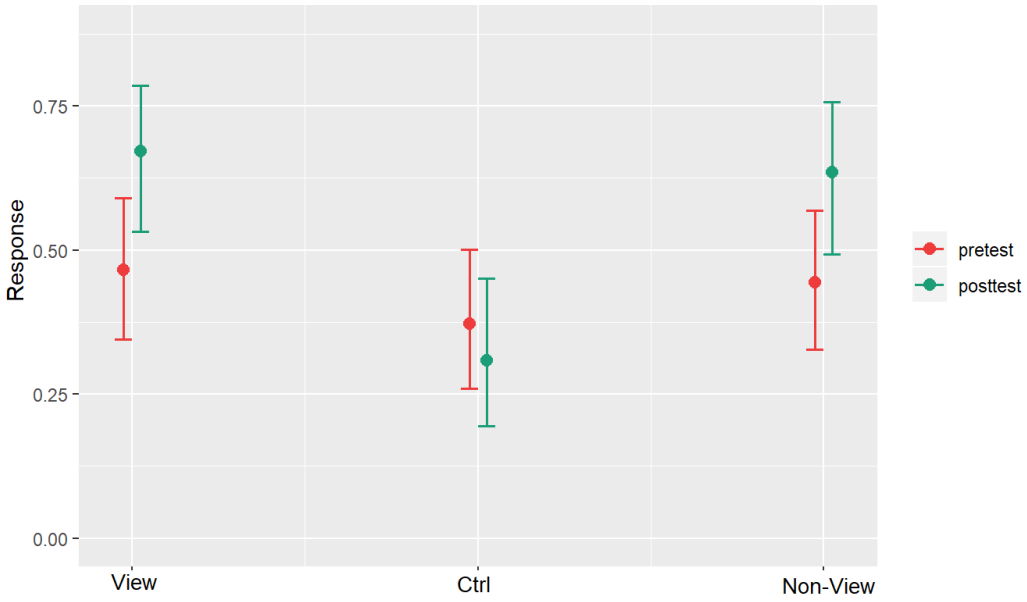
	Spoken form recognition					Written form recognition				
Parameters	*B*	*SE*	*z*	*p*	*OR*	*B*	*SE*	*z*	*p*	*OR*
Fixed effects										
Intercept	−0.46	0.20	−2.25	*	0.63	−0.14	0.26	−0.53	.595	0.87
Group = Ctrl	−0.43	0.16	−2.74	**	0.65	−0.39	0.17	−2.33	*	0.68
Group = Non-View	−0.21	0.13	−1.60	.112	0.81	−0.09	0.14	−0.60	.547	0.92
Group = View										
Time = Posttest	0.67	0.13	5.10	***	1.95	0.85	0.15	5.59	***	2.34
Time = Pretest										
Group (Ctrl) × Time (Posttest)	−0.57	0.17	−3.42	***	0.57	−1.13	0.18	−6.40	***	0.32
Group (Non-View) × Time (Posttest)	−0.01	0.14	−0.10	.929	0.99	−0.07	0.15	−0.50	.621	0.93
Group (View) × Time (Posttest)										
Random effects				Variance	*SD*				Variance	*SD*
Participant = intercept				0.24	0.49				0.31	0.56
Participant = Posttest				0.04	0.21				0.07	0.27
Item = Intercept				0.64	0.80				1.10	1.05
Item = Posttest				0.14	0.38				0.22	0.47

Note. Response ~ group × time + (Time|participant) + (Time|item). Baseline category = View group. Model fitted to 5760 observations across 20 nouns (N = 144).

*p < .05. **p < .01. ***p <.001

Figure 3. 12

Condition Effects in Written Form Recognition



Note. The predicted probabilities plot shows the probability values for response accuracy on written form recognition of 20 words by View (N = 53), Non-View (N = 57), and Control (N = 34) groups, calculated from GLM logistic regression analysis.

Summary of Findings

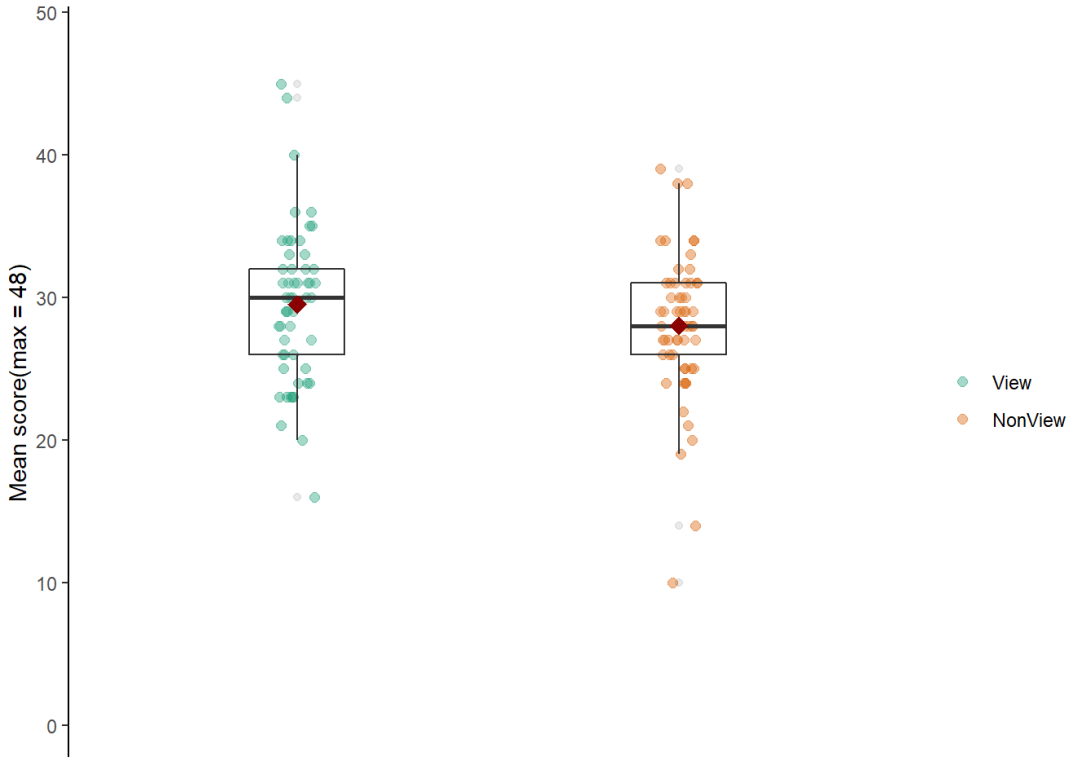
To summarise, the first research question results indicate that extensive viewing of two full-length seasons of documentary series in the form of L2 captioned video, and over four 2-hour long sessions significantly increases incidental acquisition of word meanings and forms. The second research question finding showed that participants in the View and Non-View groups acquired vocabulary equally well, suggesting a strong learning effect on all four levels of measurement, even in the absence of imagery.

3.5.4 Comprehension Questions

As it is shown in Figure 3.13, there was no significant difference between the View group (M = 29.51, *SD* = 5.49) and the Non-View group (M = 28.02, *SD* = 5.14) on comprehension scores over the treatment period ($p > .062$)[6]. On average, participants scored accurately about 60 % of comprehension questions in both groups, with 45 (94%) and 10 (20%) as the highest and lowest scores, respectively.

Figure 3. 13

Mean Comprehension Scores in View and Non-View Groups



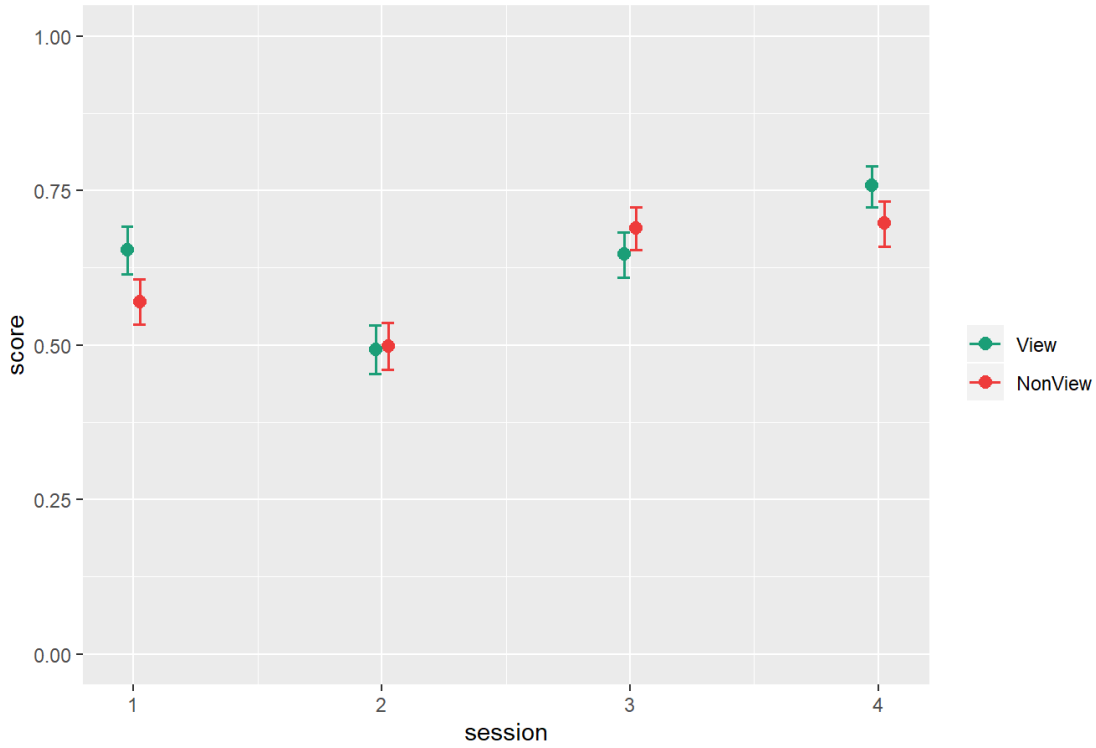
Note. The boxplots show mean accuracy scores of comprehension based on 48 questions (12 per session) by subject and across Non-View (N = 57) and View (N = 53) groups.

[6] glm(score~group, data = comprehension).
Question 8 was missing in 33% of Non-View data, a random equivalent of answers was then filtered out from View data to maintain balanced data between groups.

There was, however, a significant main effect of session ($p < .001$) and a significant interaction between group and session ($p < .01$). Comparisons on the interaction were run[7] , and results revealed that the View group performed significantly better than the Non-View group in the first ($p = .002$) and the last session ($p = .016$). Further comparisons at the level of each group (with Bonferroni adjustment for multiple comparisons, $\alpha = .008$) demonstrated a significant decrease in comprehension scores in Session 2 by View participants ($p < .0001$). However, scores increased steadily afterwards from one session to another (all ps < .001). The Non-View group showed approximately a similar pattern as shown in Figure 3.14. The low scores in Session 2 indicate that this session's topic or designed questions might have been more challenging than those in other sessions.

Figure 3. 14

Session Effects on Comprehension Scores



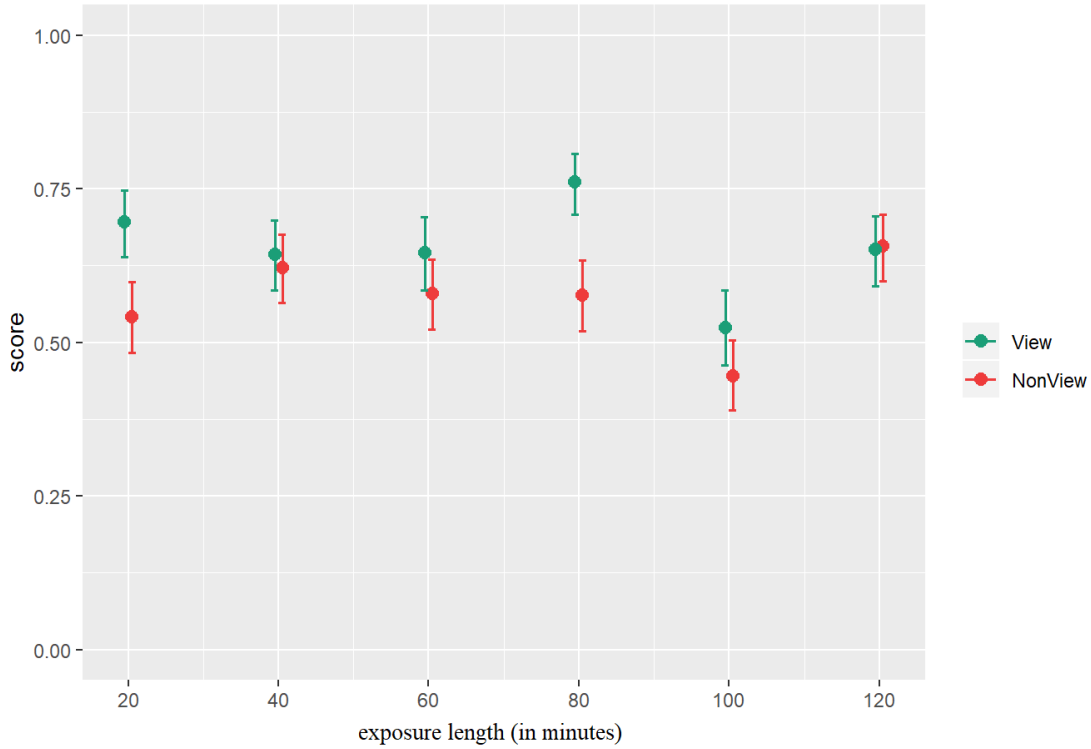
Note. The predicted probabilities plot shows the probability values for score accuracy on comprehension by View (N = 53) and Non-View (N = 57) groups in four sessions, calculated from GLM logistic regression analysis. Observations = 5156.

[7] Using emmeans function in emmeans package

There was also a significant main effect of length and a significant interaction between group and length (both ps < .001). Post-hoc comparisons revealed that the View group scored significantly better than the Non-View group at 20 minutes (p = .004) and 80 minutes (p = .0001) of viewing. Further within-group analyses (with Bonferroni adjustment for multiple comparisons, α = .003) showed a significant drop in scores at 100 minutes exposure length for both the View (p < .0001) and the Non-View groups (p = .002) followed by a significant increase in scores at 120 minutes in the View (α = .003) and Non-View groups (p < .0001). These results can be observed in Figure 3.15.

Figure 3. 15

Exposure Length Effects on Comprehension Scores



Note. The predicted probabilities plot shows the probability values for score accuracy on comprehension by View (N = 53) and Non-View (N = 57) groups at 6 intervals within session, calculated from GLM logistic regression analysis. Observations = 5156.

Summary of Findings

The above results revealed a parity in the comprehension of episodes between the View and Non-View groups. Overall, both groups were able to maintain focus throughout the 2 hr exposure period. However, comprehension scores of the View participants significantly increased after 80 min of viewing, decreased at 100 min, then increased at 120 min. The findings also showed that the View group outperformed the Non-View group in sessions 1 and 4 and after 20 min viewing in overall sessions.

3.5.5 Debriefing Questionnaire

Results of the debriefing survey are arranged into two categories: information processing items and motivation related items. The graphical method for displaying the 10-point Likert scale results is the divergent stacked bar to provide a comparative visual presentation of scaled responses in the View and Non-View groups.

Information Processing Items

There were five questions about information processing. The 3AFC question tested participants' recognition of the series' speaker voice among two other voices. The results showed that almost half of participants recognised the series' speaker voice (47% in the View group and 45% in the Non-View group) despite a somewhat long interval between the last exposure and the debriefing survey. The remaining items were 10-point Likert scale based questions on comprehension, input processing, L2 captions, and split attention. Except for the latter, the data showed generally positive results in both groups.

Comprehension questions requested students to indicate the extent to which their overall comprehension of the episodes of the documentary series was good or bad. The percentage of View participants who gave a positive rating (i.e., ≥ 5) was 94.1% (with the majority indicating moderate comprehension) compared to 79.3% by Non-View participants (with the majority opting for a 7-point rating). Around 14.7% provided a score of 10 (extremely good) in the View group compared to none in the Non-View group. The second question requested students to indicate the extent to which their processing of information in episodes was easy or difficult. Only 17.6% and 13.8% indicated that input processing was difficult in the View and Non-View groups, respectively. The majority provided neutral responses in both

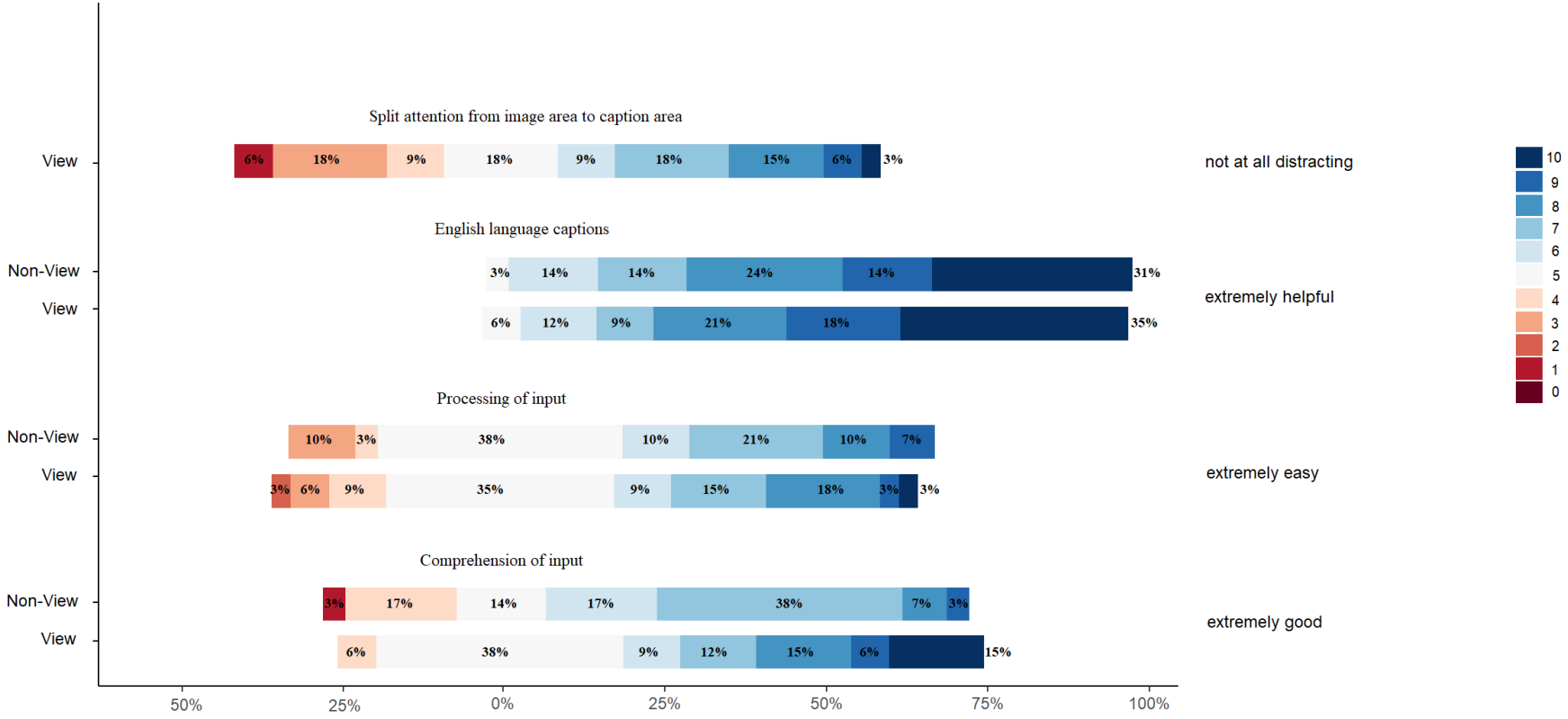
groups. The third question asked students to indicate the extent to which they found L2 captions helpful or unhelpful. In both groups, the majority provided a 10-point rating, indicating that they found captions as extremely helpful, while none of the participants perceived captions as unhelpful. GLM logistic regression analysis[8] showed that the View and Non-View groups did not vary significantly in comprehension scores and processing and caption items (all ps > .130).

Finally, the last question in this category was addressed to View group only. It aimed to determine the extent to which splitting attention from the image area to the caption area was perceived as distracting/not distracting to participants. Results revealed that 32.4% found splitting attention from image to caption area distracting (1, 3, or 4 scores). Though 67.6% gave a neutral or positive response (i.e., 5 ≥), only 8.8% provided 10- and 9-points ratings. Results for information processing items in the View and Non-View groups can be observed in Figure 3.16.

[8] Mod <- lm(score~ group, data = df)

Figure 3. 16

Perceptions on Documentary Input Processing



Note. View = 34; Non-View = 57. Percentages may not total 100 due to rounding.

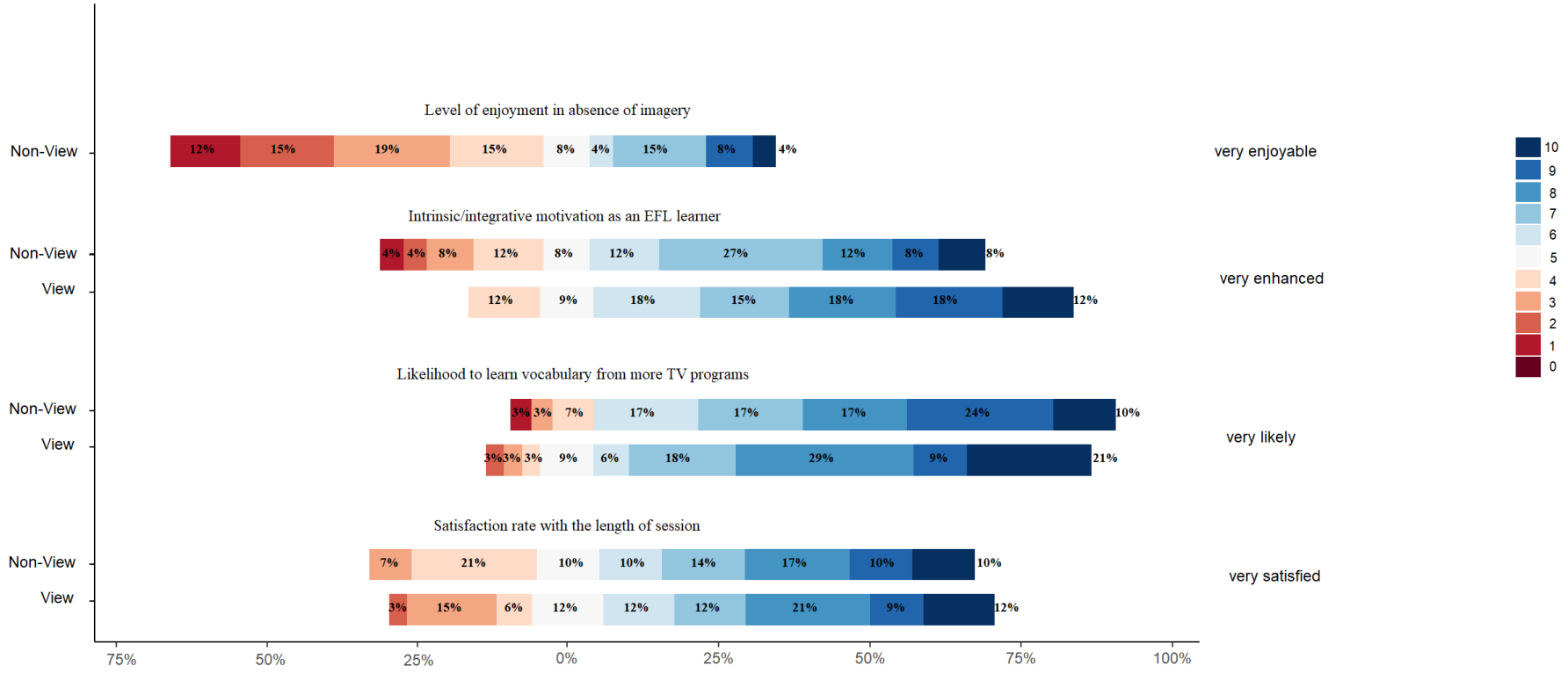
Motivation-Related Items

The debriefing questionnaire included four items with a motivation focus. The first item requested participants to provide their satisfaction rating regarding the length of episodes. The percentage of participants whose satisfaction rating was either neutral or positive is 76.5% and 72.4% for the View and Non-View groups, respectively. More than 10% in both groups indicated that they were very satisfied (10-point rating). This result correlates well with findings from comprehension scores analysis which showed that students could maintain focus throughout the session period. The second item questioned participants' likelihood to attempt to learn vocabulary through viewing more television programs. Around 82.4% among View participants were likely to, compared to 86.2% in the Non-view group. The third question requested students to indicate the extent to which the documentary series affected their overall intrinsic/integrative motivation as EFL learners. Around 79.4% of participants in the View group indicated that the documentary series enhanced their motivation, that is a 10% gap than the Non-View group (69.2%). GLM logistic regression analysis did not reveal a significant difference between the View and Non-View groups on scores of motivation-related items (all ps > .069).

The last item was directed to Non-View participants only and explored the impact of the absence of imagery on their level of enjoyment. About 61.5% of participants reported that following the documentary series without imagery was not enjoyable. The results for items in the motivation category for the two experimental groups are presented in Figure 3.17.

Figure 3. 17

Perceived Motivation During and After Treatment



Note. View = 34; Non-View = 57. Percentages may not total 100 due to rounding.

3.6 Discussion

In this study, I empirically assessed the impact of extensive TV viewing in the form of L2 captioned video on incidental vocabulary learning and the role of imagery in producing this effect. The study augments current developments in L2 research on word learning from viewing in three ways. First, by addressing out-of-class viewing through extending exposure length to eight hours and single session length to two hours using two full-length seasons of documentary series. Second, by assessing the significance of imagery through comparing a View condition to a Non-View condition. Lastly, by measuring vocabulary learning at the level of meaning recall, meaning recognition, spoken form recognition, and written form recognition.

Does viewing two full-length seasons of L2 captioned documentary series (8 hr) over 2-hour long sessions lead to incidental learning of L2 vocabulary?

I tested the hypothesis that viewing L2 captioned documentary series over four extensive sessions of 2-hour length each at two-week inter-session intervals would result in significant L2 vocabulary gains, compared to the Control group. The results strongly supported this hypothesis. GLM logistic regression analyses demonstrated that extensive viewing of L2 captioned documentary series (8 hr) over 2-hour long sessions promotes incidental learning of word meanings and forms. Participants in the View group were five times and 11 times more likely to recall and recognise the meaning of words, respectively, compared to the Control group. Similarly, accuracy in spoken and written form was significantly more robust in the View group compared to the Control group. These positive results are discussed next in light of previous findings and possible causal mechanisms.

Research Question 1: Previous Studies

The present findings align with the positive learning outcomes observed in the study with the most closely matching design (Rodgers & Webb, 2019). The study was marked by +7 hr viewing, conducted via L2 captioned video over sessions that extended to about 42 minutes and at a short inter-session interval of one week. The majority of studies reviewed in Section 3.1.1. (*Length of Exposure*) differ from the present study in design (e.g., exposure length, pre-teaching, home viewing). Comparisons may therefore present several interpretation problems. In general, the

results support previous findings that viewing programs in the form of excerpts (e.g., Peters, 2019), one-hour episode (Peters & Webb, 2018), or multiple 25-minutes episodes (+5 hr viewing) (Frumuselu et al., 2015) leads to incidental enhancement of word knowledge. The present outcomes reflect those of +3 hr out-of-class viewing studies, which did not control the duration and the timing of exposure sessions (Sinyashina, 2020a; Zarei, 2009) but are in contrast to those of +5 hr home viewing (Sinyashina, 2020b).

Multimodal Input Processing.

The finding that learners could acquire knowledge of words present in the multimodal input accord with the dual coding theory (Paivio, 1971). Its concept is that learners can process verbal and visual input via two independent systems and make referential connections between them to retrieve information. Watching documentary episodes in L2 captioned video necessitates a split of attention between captions and imagery. This divided attention could push learners into cognitive overload. In the debriefing questionnaire, approximately 33% of participants perceived split attention as distracting. Almost 68% of participants either gave a neutral response or reported that they did not perceive divided attention as distracting, with about 8.8% indicating that it was not distracting at all. Moreover, only less than 18% of participants perceived input processing as difficult. These results from third-year university students reflect those of Taylor (2005), who suggested that cognitive overload depends on prior knowledge. Furthermore, the finding that viewing complete seasons of documentary series leads to a growth in word knowledge is in line with the narrow viewing principle (Chang & Renandya, 2019; Rodgers & Webb, 2011). Episodes were of a similar genre and presented by the same speaker. What was interesting is the ability of 47% of participants to distinguish the series presenter's voice among other voices[9] sometime after the end of the study.

[9] speaking for few seconds about the same topic

Motivation.

Watching videos stimulates learners' motivation to learn (Oxford et al., 1993). The observed increase in knowledge of words post documentary viewing is likely to be related to increased motivation stemming from viewing. The debriefing questionnaire results confirmed this relationship. More than half of the Non-View group participants reported that the absence of imagery detracted from the enjoyment of the series. Essentially, almost 80% of participants in the View group reported that viewing the episodes improved their intrinsic/integrative motivation. More than this percentage indicated that they were more likely to attempt to learn vocabulary through watching more television programs in the future.

Dependent Measures.

Meaning recognition results were noticeably more substantial than meaning recall results. This outcome is in line with the consensus among vocabulary researchers that recognition knowledge precedes recall knowledge in acquisition. Consistent with Peters & Webb's (2018) findings, the present results also indicate that vocabulary gains were marked more at the level of meaning tests than form tests. Nevertheless, this conclusion cannot be relied upon uncritically due to differences in the tests' designs of meaning (posttest-only) and form (pretest-posttest).

Imageability.

The stronger meaning results are consistent with "the-richness-of-meaning" concept of imagery (Paivio et al., 1968). The finding may further support my suggestion in the first section of this chapter. I put forward the view that the significance of vocabulary learning from viewing lies not only in the availability of visual referents that assist meaning recognition of unknown words. It also lies in the fact that these visual referents can act as "illustrative images" (Paivio, 2014) to form mental representations for the co-occurring unknown words that lack a visual referent. These mental images are essential for the retrieval of meaning.

Comprehension.

Students' ability to incidentally acquire words can also be the result of adequate content comprehension, as was evident from the comprehension scores and the debriefing survey. Mean comprehension scores showed that participants generally scored correctly on more than half the questions (29/48). Participants' perception in

the debriefing questionnaire matched these results as the majority self-rated their comprehension as moderate, and 95% of participants gave a positive rating). The supposition that vocabulary learning could have resulted from good comprehension comes from previous studies that showed a positive correlation between general comprehension of L2 multimodal input and vocabulary gains, especially in the presence of L2 captions (e.g., Pujadas, 2019).

Out-of-class Viewing.

There are other possible explanations for the positive vocabulary learning results following the extensive viewing sessions. Students' familiarity with multimodal input outside the classroom, as was detected from the language profile survey, might have facilitated acquisition. Students were habituated to English language TV viewing and intra-lingual captions and showed a high exposure frequency to TV programs, including documentary series. In fact, extramural L2 exposure has been positively linked with vocabulary knowledge (Puimège & Peters, 2019).

Session Length.

Previous studies on the impact of extensive TV viewing on incidental vocabulary learning were perhaps limited by the relatively low ecological validity of the findings due to the implementation of short viewing sessions. The result that incidental vocabulary learning can occur following multiple long TV viewing sessions (of +1 hr long each) has not previously been reported in controlled studies where viewing-related variables (e.g., inter-session interval, session length) are kept constant. This study supports the 2 hr session length I proposed for extensive viewing research and weakens the consensus that long viewing sessions cause fatigue. Results from the debriefing survey support my proposal of +1 hr sessions. Approximately 77% of participants provided a neutral or positive rating regarding the session length in the experiment. Results from the comprehension instrument further substantiate my suggestion. Scores indicated that participants maintained focus throughout the session except at 100 minutes. Participants soon regained focus since they significantly improved in scores from 100 minutes to 120 minutes. Importantly, scores at this last interval did not significantly vary from scores at 20, 40, 60, and 80 minutes.

What is the effect of removing imagery and keeping bimodal input?

I further predicted that participants in the View group, who watched episodes of the documentary series in L2 captioned video format, would produce more vocabulary gains in meaning and spoken form tests than participants in the Non-View group, who were exposed to bimodal verbal input only. On the other hand, I hypothesised that the Non-View group would outperform the View group in written form tests. The results did not offer support for these two hypotheses. GLM logistic regression analyses showed that learners scored significantly well on all dependent measures tested independently of the experimental condition.

Research Question 2: Previous Studies

The increased vocabulary gains in Non-View participants corroborates studies that found an advantage of bimodal input in incidental L2 word development (e.g., Tangkakarn & Gampper, 2020; Teng, 2016; Webb & Chang 2012). This result may be explained by the bundle of arguments for bimodal input discussed in Section 3.1.2 such as the increase in speed of lexical access or the availability of segmented text.

The parity of L2 word gains between the group without imagery support (reading-while-listening) and the group with imagery support (L2 captioned video) was recently reported (Feng & Webb, 2020). However, the study differs from mine in research design. The authors implemented a limited input and compared a viewing condition to reading-only and listening-only conditions.

The present study did not demonstrate an imagery effect in L2 captioned video. This finding contrasts with previous studies that isolated the effect of imagery using a between-subjects design and found more vocabulary gains in the presence of imagery. Nevertheless, all previous findings were based on minimal input and small sample size compared to the current study. Neuman & Koskinen (1992) implemented short segments of few minutes, Hernandez (2004) used an 8-minutes video twice, and Alshumrani (2019) used four sessions of 15 minutes each. This makes direct comparisons with a study with four 2 hr sessions (i.e., 8 hr) problematic.

Finally, contrary to expectations, participants in the View and Non-View groups scored equally well on written form tests. This result is in line with Vanderplank's (1988) assertion that the presence of captions stimulates conscious

attention to words' written forms. The finding also aligns with the fact that students were highly aware of the value of L2 captions regardless of whether they were in the multimodal or bimodal condition. The majority assigned the highest satisfaction rating of 10 for captions, and none of the participants regarded them as unhelpful. Importantly, similar results in the two groups at the level of written form suggest that the presence of imagery does not distract students from noticing written forms in captions.

Differences in Response to Stimuli.

The above results lead to two hypotheses. The first posits that learners in the View and Non-View groups learnt words from L2 captioned documentary series similarly because imagery did not influence incidental L2 word learning. The second hypothesis argues for a pivotal role of imagery in promoting knowledge of L2 words and suggests that the results may be the consequence of differences in behaviour in the two groups. In the section below, I will review evidence for both hypotheses, concluding that the parity of learning gains between groups is more likely attributable to differences in learning behaviour than to a null effect of imagery.

Imageability.

The first hypothesis positing that imagery did not influence learning suggests that target words were not adequately represented with visual referents within the documentary series. It assumes that visual referents of target words might have either been lacking throughout the episodes or existing with a low frequency that did not assist vocabulary learning. Nevertheless, the materials selected for the present study do not support this hypothesis (as evidenced in Chapter 4). In line with the second hypothesis, I postulate that on-screen imagery consisted of visual referents that were effective in facilitating learning in the View group. In contrast, learners in the Non-View group were able to build up their own representations of newly recognised words due to their high average imageability rating (i.e., 3.5). That is, the exclusion of imagery pushed learners into a state of constructing and imaging word meanings, which assisted later retrieval.

Noticing hypothesis.

A possible explanation for the positive finding in the Non-View group might be that the lack of imagery compelled learners to notice unknown words. Schmidt (1990)

stated that learning results from conscious noticing of input. In line with the noticing hypothesis and the study, it could be suggested that learners are more likely to notice unknown words when there is less distraction. For instance, Non-View participants might have noticed spoken forms for being more salient in the absence of imagery.

The result that there was no significant difference between groups in spoken form gains does not necessarily indicate that View participants did not depend on lip-read input as previously hypothesised. Participants of different groups might have relied on different strategies to acquire spoken forms. An analysis of whether eye-tracking data of fixations on lip-read input predicts spoken form gains might better demonstrate the importance of lip-reading in viewing.

Motivated strategies.

Moreover, it is possible to hypothesise that noticing unknown words triggered essential vocabulary learning strategies. Guessing the meaning from the linguistic context, for instance, might have fostered the observed vocabulary gains in Non-View participants, including in minimal users who favour enjoyment over interruption. Participants were intermediate to upper-intermediate learners. They had a vocabulary size that warrants adequate contextual guessing (Laufer, 1997a).

I proposed in Section 3.1.3 that the value of imagery lies in its ability to motivate learning and stimulate strategy use. However, improved L2 vocabulary knowledge in Non-View participants suggests that bimodal input also has its motivating strength. Findings from the debriefing questionnaire may substantiate this conclusion. Although more than half of participants indicated that they were not satisfied with the exclusion of imagery, about 70% indicated that the treatment promoted their motivation as EFL learners, and 87% showed an eagerness to learn vocabulary through more television programs in the future (i.e., maximal users). Moreover, more than 70% of participants were either neutral or satisfied with the long session length. Other results from the language profile survey (perceived difficulty) revealed learners' consciousness of the benefits of bimodal input in text comprehension before treatment. All in all, it seems that motivational factors probably prompted strategy use and promoted word acquisition.

The words learned.

One possibility that is worth examining is that participants from the two different groups learned different words but in equal amounts. Table 3.16 shows the descriptive statistics for meaning recognition posttest by words, with the green cells representing the top five most recognised words in each group while red cells represent the five least recognised words. Although the table shows that students from the two groups performed comparably on many words, there are some noticeable differences. This result supports the previously considered hypothesis that students in the two conditions responded differently to the stimuli. For instance, it could be that Non-View participants learnt many abstract words that do not require visual representation. This hypothesis supports a role of imagery in L2 captioned video in increasing vocabulary learning.

Table 3. 16

Meaning Recognition Mean Scores by Word (20 Words)

	Mean Scores	
Words	Non-View	View
temple	0.96	0.87
cosmic	0.81	0.92
particle	0.74	0.55
cosmos	0.72	0.94
sculpt	0.67	0.51
emit	0.65	0.45
tide	0.65	0.49
supernova	0.61	0.57
faint	0.60	0.66
alien	0.58	0.58
stretch	0.58	0.64
squash	0.56	0.45
denser	0.51	0.51
spectrum	0.47	0.34
orbit	0.44	0.47
sphere	0.44	0.34
dense	0.44	0.55
constellation	0.42	0.53
intricate	0.42	0.32
forge	0.23	0.36

Note. Recog %>% group_by(word, group) %>% summarise(mean = mean(Posttest)) -> Recog.mean
Green cells = 5 most recognised words. Red cells = 5 least recognised words.

Comprehension.

The lack of a significant difference between the View and Non-View groups in incidental L2 vocabulary learning is also consistent with comprehension test results. The two groups did not differ overall in comprehension scores and showed a similar score pattern. Nonetheless, a few points should be noted. Firstly, the low scores obtained in the Non-View group, compared to the View group, for questions of the first 20 minutes indicate that the presence of imagery may assist learners to engage with the content more rapidly than when there is a lack of visual input. Secondly, it was found that the View group outperformed the Non-View group in the first and the last sessions. It is difficult to explain this result, but it might be related to the characteristics of the visual input within episodes of these two sessions. For instance, these sessions probably consisted of visual referents that were pivotal to achieve content comprehension, and referents were perhaps great in frequency and strength.

Out-of-class reading-while-listening.

The finding that learners acquired knowledge of words irrespective of the presence of imagery may also be due to learners' familiarity with bimodal input outside the classroom. In the language profile questionnaire, about 65% of participants indicated that they were used to reading lyrics as they listened to songs, with 55% indicating that they were doing this daily or weekly. Out-of-class exposure to L2 input may facilitate learners' input processing over time. Only about 14% of the sample perceived the bimodal input in the study as difficult to process. Effective processing of input may assist learners in reaching the necessary mastery of the language to achieve content comprehension and high levels of word acquisition.

3.7 Conclusion

Several studies have confirmed the effectiveness of exposure to multimodal input and L2 captioned video, in particular, on developing incidental L2 vocabulary learning. The substantial evidence has heightened the need to investigate incidental L2 vocabulary learning from extensive TV viewing that resembles real-life conditions. A search of the literature revealed only a few studies in this area of research. Moreover, the studies had relatively low ecological validity as researchers have commonly employed 20 to 30 minutes as the optimal session length for

extensive TV viewing. This duration does not correspond to actual out-of-class viewing. In addition, an understanding of how imagery contributes to incidental L2 vocabulary learning from extensive viewing is still lacking.

The present study is, to my knowledge, the first experiment with high ecological validity to investigate incidental L2 vocabulary learning outcomes from extensive TV viewing, specifically, two full-length seasons of L2 captioned documentary series. The study maximised ecological validity by using 2 hr length sessions totalling 8 hr viewing over a six-week period of two-week intervals. Furthermore, this is the first study of substantial duration and large sample size to isolate imagery variable in L2 captioned video to assess the effect of imagery on the incidental acquisition of L2 words. The study is also characterised by specificity in word knowledge by testing meaning recall, meaning recognition, spoken form recognition, and written form recognition. The study results were discussed in light of dual-coding theory, imageability, comprehension, out-of-class viewing, motivation, and motivated strategies.

These findings contribute in several ways to our understanding of the effect of extensive TV viewing in general, and imagery in L2 captioned video, in particular, on incidental L2 vocabulary learning. The study demonstrates that viewing two full-length seasons of L2 captioned documentary series, amounting to 8 hr, over six weeks at two-week intervals improves incidental L2 vocabulary learning in L2 learners who are third-year university students. The improvement can be seen at the level of meaning recognition, meaning recall, spoken form recognition, and written form recognition. The 2 hr session length is clearly supported by the current findings that show that intermediate to upper-intermediate L2 learners can comprehend the content and maintain focus throughout this period. The second significant finding is that the learners are able to incidentally acquire words from viewing L2 captioned documentary series at the same level as if they had imagery removed from the episodes and were exposed to bimodal input only. In terms of comprehension, learners can achieve good comprehension of episodes of L2 captioned documentary series irrespective of imagery, though the latter may seem to accelerate engagement with the content in initial viewing. The language profile questionnaire suggests that regular out-of-class exposure to L2 input may facilitate

L2 content comprehension and vocabulary acquisition. Moreover, the language profile and the debriefing questionnaires suggest a role for L2 captioned documentary series in promoting EFL learners' motivation to learn in both multimodal and bimodal modes.

Notably, the above findings raise important questions regarding the effect of visual input. Are the similar results attributable to a null effect of imagery or to the greater noticing of words and strategy use, as stimulated by the absence of imagery? In this study, I favour the second hypothesis. More work needs to be done to address the possible alternative explanation for the results to establish whether L2 learners benefited at all from imagery input. A natural progression of this study is to investigate the mechanisms that may underlie imagery effects. In the following chapter, I will extend the present study by investigating the effects of imagery in L2 captioned video by analysing the effects of contiguity between word forms and their visual referents.

Chapter 4

Study 2. Contiguity Effects in Learning

In Study 1, no significant difference in incidental vocabulary learning was evident between the group who watched full-length seasons of L2 captioned documentary series and the group who was exposed to the same material with imagery being removed from the episodes. In the discussion section of the chapter, I speculated the possibility that a clear benefit of imagery might not have been identified in the analysis due to an increased reliance in the Non-View group on contextual clues provided from the bimodal verbal input. The elimination of the treatment (imagery effect) might have stimulated other vocabulary learning strategies and brought about equal effects to that of viewing. The prospect of being able to further my research and take a new look at a clear-cut effect of imagery in L2 captioned video served as an incentive for Study 2, by calling into question the effect of contiguity between words' verbal forms and their visual referents on incidental vocabulary learning.

This second study builds on recent second language research (Peters, 2019; Rodgers, 2018) exploring the effect of imagery in authentic video on vocabulary acquisition by focusing on the role of the synchronous occurrence of words and their visual referents. Contiguity is broadly defined as the state of having two things close to each other. According to a psychologically-based definition, however, contiguity is the principle that constant perception of two stimuli together leads to a stronger association between these stimuli in mind (Ellis, 2003; Hebb, 1949). Although contiguity has long been investigated (Froeberg, 1918; Guthrie, 1933; Nodine, 1969), it was not until recently that the concept has gained a footing in the domain of education and pedagogy. Specifically, Mayer and Anderson (1992) were the first to incorporate contiguity into the area of instructional design when they introduced the concept as a cognitive principle of multimedia learning. The term verbal-visual contiguity is used throughout this chapter to refer to the synchronous occurrence of words and their visual referents.

A focus on contiguity is warranted for several reasons. Firstly, research on contiguity in multimedia learning has exclusively been limited to explicit teaching and non-authentic materials due to its orientation towards classroom materials development. The significant positive effect of simultaneous processing of the verbal narration and the visual representations from a video material was first demonstrated experimentally on form recall via static videos (Baggett, 1984; Baggett & Ehrenfeucht, 1983) and problem-solving tasks via animated videos (Mayer & Anderson, 1991). This was achieved by moving the narration forward and backwards to examine the outcome effects compared to presenting the soundtrack simultaneously with the visuals. However, whether contiguity learning effects between vocabulary items and visual referents exist in authentic videos in incidental learning contexts remains unclear but is an intriguing question to ask. Secondly, one cannot undertake vocabulary research without accounting for word properties (e.g.,Hulme et al., 2019; Peters & Webb, 2018). It is now well established from a variety of studies that vocabulary learning may vary as a function of cognateness (Granger, 1993; Puimège & Peters, 2019), length (Crystal, 1987), part of speech (Laufer, 1997; Rodgers, 1969), concreteness (Brysbaert et al., 2014; De Groot, 2006), verbal frequency (Pellicer-Sánchez, 2016; Saragi, 1978; Schmitt, 2010) and other factors. This thesis puts the view that verbal-visual contiguity could also function as a predictor of word learning from videos.

No previous study has investigated the effect of verbal-visual contiguity on incidental vocabulary learning from TV viewing based on extensive exposure (e.g., 8 hr) and different parts of speech. Despite extensive research into vocabulary learning through videos, existing studies have not treated imagery in much detail (See Chapter 3 for a review). For instance, in 2016, Peters et al. only acknowledged imagery's critical role. Quantifying the visual referents that co-occur with words was beyond researchers' scope until 2018, when Rodgers first conducted a descriptive demonstration of this contiguity. This study brought into the vocabulary literature a valuable word-related variable that merits careful consideration in further studies. Rodgers logged visual referents of target words occurring in authentic videos and measured their frequency in his corpus study. He found that verbal-visual contiguity was more frequent in documentary series than in narrative

television. Rodgers looked at verbal-visual contiguity as a unidimensional construct operationalised as frequency. The present study makes an original contribution in this area of research by introducing contiguity duration (*contigduration*), contiguity frequency (*contigfrequency*), and contiguity ratio (*contigratio*) as three conceptual elements that make up the construct of verbal-visual contiguity. Rodgers found that over 65% of visual referents occurred concurrently with spoken form, while 70% of referents occurred within 5 seconds of the verbal occurrence. In light of the support of imagery shown in these results, the question that naturally arises is to what extent does such a word factor potentiate incidental vocabulary learning from videos?

The question described above was addressed a year later in an intervention study by Peters (2019). By exposing students to a 12-minute documentary excerpt, she found that visual referents within the timespan of 5 seconds before and after the verbal occurrence of words promote acquisition at the level of form recognition and meaning recall. As highlighted by Peters herself, however, the extent of this effect remains ambiguous until an empirical longitudinal study of extensive viewing is conducted. Within the same year, Pujadas Jorba (2019) followed Peters' approach for coding imagery and looked at the effect of this binary variable on incidental learning of nouns from viewing 2 hours and 55 minutes of TV series. She found that meaning of words occurring along their visual referents were 2.33 times more likely to be learnt than words occurring without visual referents.

Finally, Ahrabi Fakhr et al., (2021) later examined the effect of different factors on incidental vocabulary learning from viewing an episode of a captioned English language TV program. Among the item-related variables was what they termed visual imagery. Using Peters' methodology (2019), the authors adopted a 3-levels scale (partly available image, available image, no available image). They found that words with imagery were at least 2.5 times more likely to be learnt. While the authors mentioned that they "… took into consideration the degree to which an image co-occurred its aural form" (p. 7), the procedure to achieve this and the overall aspects of imagery coding, including the adopted timespans, were not clarified.

It is worth noting that Peters' approach has possibly been first used by Neuman and Koskinen (1992) in their study on incidental vocabulary learning from

short children educational videos (see Chapter 3, Section 3.1.3, for a summary of the study). The authors formed a contextual support measure of both verbal and visual clues. Visual support from the video was rated using a 4-point scale: (a) word shown (b) word described (c) word not shown (d) word shown with contrasting video. Results indicated that learnt words were those which had both verbal and visual support.

A source of concern in Peters' methodology is the operationalization of imagery. To make this clearer, Peters and followers treated this predictor as a categorical variable, assigning binary codes "image" and "no image". They did not attempt to count how many verbal-visual occurrences a single word had (i.e., frequency, Rodgers, 2018). An approach of this kind carries a well-known limitation since binary coding leads to a potential loss of information. For instance, '*calf*' and '*steep'* had verbal frequencies of '4' and '1', respectively, implying that '*calf*' could have had more visual referents. The assumption that the number of visual referents could be attributed to concreteness is eliminated as the words have only a 0.72 concreteness difference. Therefore, since imagery was treated as a categorical predictor, Peters' analysis assumes that '*calf*' and '*steep*' are equivalent in terms of verbal-visual contiguity. This operationalisation could limit the experiment's internal validity through the introduction of measurement error since the data may not precisely represent the construct they are intended to.

Moreover, knowledge of this effect was based mainly on limited timespans (5 seconds). Further work needs to be done to determine whether the verbal-visual contiguity effect could still be observed if longer timespans are considered. Verbal-visual contiguity effects in authentic videos on learning verbs and adjectives in addition to nouns, and based on extensive exposure are not clear yet. The present study seeks to obtain more data that will help to fill these gaps.

It is clear from the foregoing discussion that what we know about contiguity in multimedia learning is primarily based upon studies investigating synchronicity between an instructional video and its narration for explicit teaching/learning purposes. Therefore, it is still unknown whether contiguity learning effects arise in incidental learning contexts, specifically between vocabulary items and visual referents in authentic videos. Remarkably, research on incidental vocabulary

learning from videos has gained significant popularity in recent years. Despite the importance of verbal-visual contiguity in videos, there is a paucity of evidence of the differential effects of such a word factor on word learning. Notably, a systematic understanding and assessment of whether and how verbal-visual contiguity contributes to incidental vocabulary learning from extensive viewing are still lacking.

The present study explores the construct of verbal-visual contiguity in authentic videos, measures its dimensions, and statistically determines its importance in predicting incidental vocabulary learning from extensive viewing. This chapter reports on the results of an experimental study that was designed to assess the effect of verbal-visual contiguity in L2 captioned videos on incidental vocabulary learning from viewing two full-length documentary series (a total of 8 episodes, one hour each). The study extends current knowledge of contiguity in videos by operationalising and investigating three conceptual dimensions: contigduration, contigfrequency, and contigratio. The dimensions were carefully measured within two timespans: within $\mp$7 seconds and $\mp$25 seconds of the verbal referent. These spans were chosen based on a review of the role of phonological and visual short-term memory and are discussed below. It is hoped that the present study will contribute to a deeper understanding of the effect of contiguity, a multidimensional measurable construct, on incidental vocabulary learning from L2 captioned videos. The study will also delineate how long the verbal-visual contiguity timeframe could be.

Following this introduction of the gaps in contiguity literature, the next section will give a review of verbal-visual contiguity research. It will first present the theoretical foundation of the contiguity principle in education, then address verbal-visual contiguity in the context of L2 captioned video. The following three subsections will represent the three contiguity measures established in the present study. The last two subsections will explain two aspects related to contiguity, these are timeframe and sequence.

4.1 Verbal-visual Contiguity Effect in Incidental L2 Vocabulary Learning from Viewing L2 Captioned Video

The contiguity principle is inspired by the dual coding theory. The theory was firstly proposed by Paivio (1971) and is one of the most influential theories of cognition underpinning the growth in interest in imagery research. Paivio put forward the view that verbal input (words) and visual input (e.g., imagery) are stored via two separate but interacting codes in the human mind. These codes have qualitatively different subsets of mental representations (often referred to as the verbal and imagery codes). When the input is stored in two systems instead of one, verbal input and its visual referent in verbal and imagery codes, respectively, it allows between-codes referential connections that are likely to augment chances of input retrieval.

Working memory has a severely limited capacity and duration, which often causes a heavy cognitive load in learning (Sweller, Chandler, Tierney, & Cooper, 1990). Therefore, it was postulated that referential connections between verbal and imagery codes could become easily constructed if words and their visual referents occur contiguously in time or space. This supposition gave rise to the principle of contiguity, which states that "the effectiveness of multimedia instruction increases when words and pictures are presented contiguously (rather than isolated from one another) in time or space" (Mayer & Anderson, 1992, p. 444). As mentioned in the introduction to this chapter, the contiguity principle has been well-documented in the area of intentional learning from non-authentic materials. Nonetheless, there is still considerable ambiguity concerning the association between verbal-visual contiguity and incidental vocabulary learning from authentic materials, namely L2 captioned authentic videos as the focus of this study.

In addition, it could be technically challenging to determine the impact of verbal frequency of occurrence on word learning when verbal-visual occurrences are not controlled. Few researchers have sounded a note of caution concerning such findings. Namely, Peters et al. (2016) revealed that certain words with only a single verbal occurrence were among the best learnt items even though their acquisition likelihood was much lower compared to words with higher occurrences. English captions were also found to facilitate learning regardless of proficiency level. The authors, therefore, explained that a semantic match between the verbal form of the

target words and their corresponding imagery might have assisted in forming initial associations between form and meaning and thus played an important role in bringing about these results. To remind the reader of the significance of this word-related factor, the authors commented:

> "Another issue that needs to be addressed here is the potentially mediating role of imagery. Although the relationship between imagery and the aural presentation of the lexical items was not the focus in our study, it is not unlikely that such visual clues may have helped the learning of some items… as there was more *visual support* [emphasis added] in The Simpsons episode than in the documentary"(p. 145).

More authors have been addressing the issue recently in the limitations of their studies by calling for research on verbal-visual contiguity: "It would therefore be interesting to analyse the extent to which the visual and verbal representations of the forty TWs co-occurred, and to investigate whether this had any association with participants' learning" (Suárez & Gesa, 2019, p. 511).

What should be considered as visual support to vocabulary learning, how to measure it, and what is the extent of its effect? These are the three fundamental questions addressed in this chapter to unravel some of the mysteries surrounding the verbal-visual contiguity principle in L2 captioned video and its association with incidental vocabulary learning.

4.1.1 Contiguity Principle in L2 Captioned Video

The act of paring words and potential visual referents across many situations has a facilitative effect on learning. This phenomenon is known as "cross-situational learning", and awareness of it is not recent (Gleitman, 1990; Pinker, 1984). The research on verbal-visual contiguity in L2 vocabulary learning from authentic video thus far has not been drawn on but could potentially benefit from work investigating underlying learning mechanisms of cross-situational learning in adults (e.g., Berens, Horst, & Bird, 2018; Yu & Smith, 2007, Kachergis, Yu, & Shiffrin, 2012).

Studies originally arose to investigate how children learn to map meaning to word forms when mappings are probabilistic (e.g., Akhtar, 2002; Akhtar &

Montague, 1999; Gleitman, 1990; Siskind, 1996; Smith & Yu, 2008; Vogt & Smith, 2005). Studies focused on the mechanisms that allow indeterminacy resolution and how the co-occurrence of a word and visual referent candidates over multiple moments leads to correct pairings, allowing the meaning of words to be acquired cross-situationally. In Smith and Yu's words, the learner must "store possible word-referent pairings across trials, evaluate the statistical evidence, and ultimately map individual words to the right referents through this cross-trial evidence" (p. 414). As they emphasised, the essence of these studies is in unravelling the processes underlying learning in ambiguous real-world situations.

Within the context of the study at hand, the above complexity could well be illustrated in incidental L2 vocabulary learning from authentic videos in which words occur with multiple possible referents, provided by imagery input. The more word-referent pairs appear to the second language learner, the more opportunities they have to segregate repeated pairs from unrepeated pairs incidentally. Segregation was found to potentiate rapid learning for adult learners even under high referential uncertainty conditions (Kachergis, Yu, & Shiffrin, 2009).

Before proceeding to operationalise the concept of verbal-visual contiguity in L2 captioned video, I will first explain two sub-categories of the contiguity principle: spatial contiguity, or the coordination of imagery with text (written) in space, and temporal contiguity, or the coordination of imagery with narration (spoken) in time. As will be noted later, the second of these is the most relevant for the present study and has been chosen to operationalise the construct of contiguity.

Spatial Contiguity

Spatial contiguity is the state of having pictures or illustrations appearing closely together with written text on a book page, whiteboard, PowerPoint slide or, more pertinently, a computer screen. L2 captioned video is a type of multimodal input that fulfils the principle of spatial contiguity between moving images and written words since captions are integrated slightly above the bottom of the screen. A large amount of research has found that this state has a positive influence on teaching and learning. In 1989, Mayer found students to be more efficient at solving transfer problems from multiframe illustrations when presented with verbal descriptions within the frame than students who had them at a distant location on a page. Since then, a

considerable quantity of research evidence has borne out the assumption that spatial contiguity enhances learning (Bodemer, Ploetzner, Feuerlein, & Spada, 2004; Chandler & Sweller, 1991, 1992; Florax & Ploetzner, 2010; Holsanova, Holmberg, & Holmqvist, 2009; Johnson & Mayer, 2012; Kester, Kirschner, & Van Merriënboer, 2005; Mayer & Gallini, 1990; Mayer, Steinhoff, Bower, & Mars, 1995; Moreno & Mayer, 1999; Owens & Sweller, 2008; Pociask & Morrison, 2008; Sweller et al., 1990; Tindall-Ford, Chandler, & Sweller, 1997).

Temporal Contiguity

Temporal contiguity is the state of hearing a narration of an event and seeing what depicts it at the same time (Mayer, 2009). Like spatial contiguity, temporal contiguity exists in L2 captioned video whenever spoken words and their visual referents co-occur. Several studies have examined this principle (Baggett, 1984; Baggett & Ehrenfeucht, 1983; Mayer & Anderson, 1992; Mayer, Moreno, Boire, & Vagge, 1999; Mayer & Sims, 1994; Michas & Berry, 2000; Owens & Sweller, 2008). The studies have consistently shown that students learn better when words and pictures are contiguous in time.

Spoken Form as a Reference of Measurement.

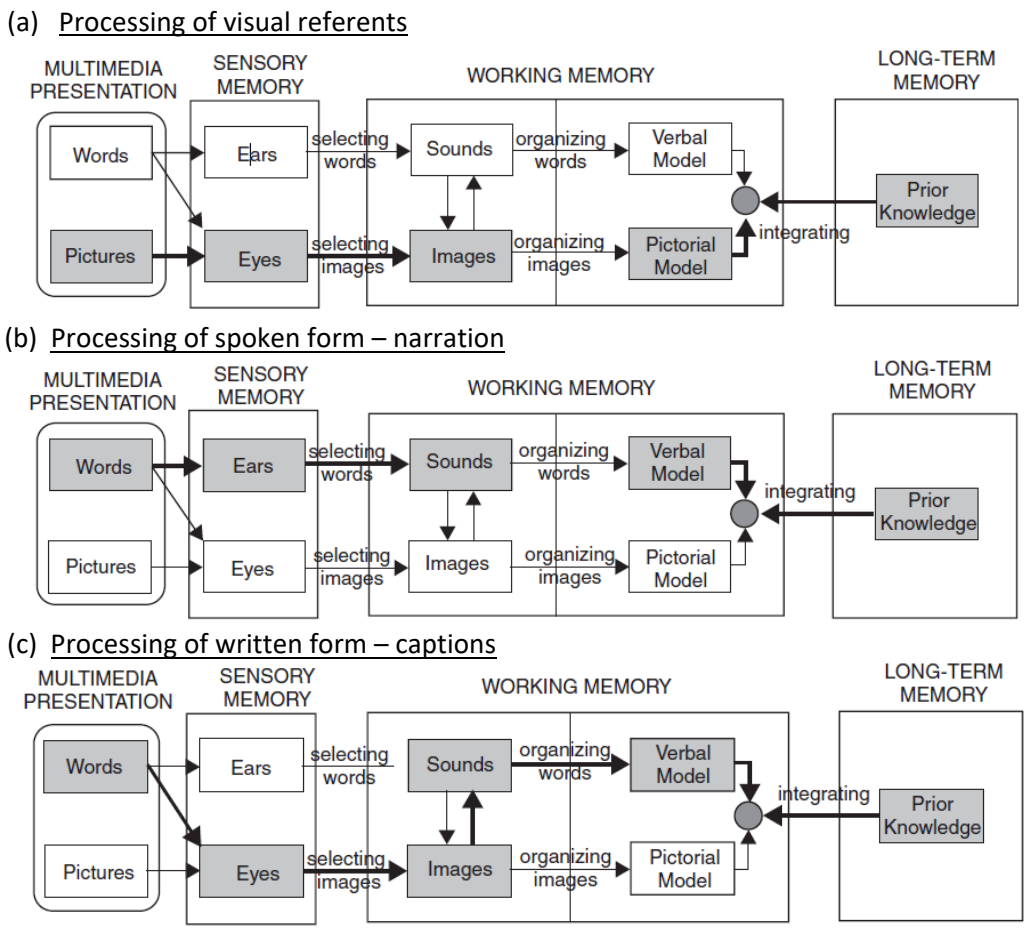
Verbal-visual contiguity in video is measured by quantifying associations between words' visual referents and verbal forms. In his seminal work, Rodgers (2018) measured verbal-visual contiguity in video by logging time for visual referents in relation to spoken form occurrence. Difficulties arise, however, when attempts are made to measure contiguity in L2 captioned video in which two verbal (spoken and written) forms are involved. Figure 4.1 illustrates the multimodal processing that is expected to take place during incidental word learning from L2 captioned video. Captions appear on screen for a maximum of seven seconds (Ivarsson & Carrol, 1998); thus, contiguity relating the image to the occurrence of the written form may be a more potent variant relative to contiguity relating the image to the spoken form. Nonetheless, after careful consideration, I chose to measure contiguity between image and spoken form for the measurement reasons detailed next.

Several considerations suggest that it might be preferable to measure contiguity in relation to written form: captions appear together with imagery in both time and space, reflecting two contiguity principles of multimedia learning.

Choosing spoken form at the expense of written form also implies that long periods of contiguity through captions and images are neglected. In addition, features of authentic videos such as speaker accent and noise may interfere in speech segmentation, making L2 listening difficult in contrast to reading (Alderson et al., 2006; Brindley & Slatyer, 2002; Goh, 2000; Vandergrift, 2004; Wagner, 2010), and suggesting that spoken word units could at times be hard to recognise.

Figure 4. 1

Processing of Visual Referents and Their Verbal Forms in Incidental Vocabulary Learning from L2 Captioned Videos



Note. Adapted from "Multimedia Learning," by R. E. Mayer, 2009, p. 77.

Notwithstanding the previous points, there remain several aspects of exposure to captions about which questions may be raised. Contiguity occurs in the instant in which imagery and written form intersect. This may be problematic if one is viewed at the expense of the others. The research on potential trade-offs to date is

not conclusive. A similar cognitive load was reported when looking at the captions or screen area (Kruger, Hefer, & Matthew, 2013). However, the fact that the video was an academic lecture makes it difficult to make inferences to authentic video studies in which complex visuals are involved. Eye-tracking research revealed that adults do not skip captions as often as they skip L2 subtitles (Muñoz, 2017). In the study by Winke et al., (2013), fixations were found to be L1 specific, and Arabic language learners fixated on the captions area 75% of the time the captions were on screen. The authors also found that Chinese language learners employed a strategy of reading captions at the expense of images whenever comprehension was obstructed. This finding is in line with Montero Perez's results (2019), who found that learners spend some time on unknown words in the captions. These results are promising, as they indicate processing, thus potentially learning; however, they also indicate a source of uncontrolled variation in the measurement of contiguity between images and captions.

Another argument for choosing spoken over written form is the uncontrolled differences that emerge from processing captions. Firstly, target words may be read immediately or seconds later, depending on whether they are situated at the beginning or the end of the caption text. Therefore, selecting the moment in which a caption appears and ignoring the position of the target word may introduce further measurement error. Unfortunately, considering the position of the word in the caption is also practically demanding and beyond the scope of this study. In addition, while L2 learners hear the target words all at the same instant relative to the imagery, they differ from each other regarding the time spent reading captions (Specker, 2008). In fact, patterns of shifting between scenes and captions differ not only between but also within learners. To illustrate, it has been shown that viewers altered their reading patterns as they moved through video material and rhythmic captions (Perego et al., 2010; Specker, 2008). This trend, in turn, entails possible changes in attentional processing of written forms, depending on whether the target word occurs at the beginning or the end of the video.

There are other factors to bear in mind when discussing fixations on captions. In the first place, there is an interaction between attention to captions and the video content. Emotional scenes and striking imagery are strong predictors of attention

(Bradley, Greenwald, Petry, & Lang, 1992; Lang, Dhillon, & Dong, 1995; Lang, Newhagen, & Reeves, 1996). Therefore, it could be postulated that captions of video segments with highly arousing imagery are likely to receive less attention than captions of segments with less exciting imagery. Large screen sizes have also been found to produce greater attention to imagery in participants watching a film (Reeves et al., 1999), making the extent to which highly stimulating scenes divert attention from captions an interesting question to address in eye-tracking studies.

Furthermore, auditory perception is faster than visual perception, with only 0.05 seconds needed for the brain to recognise a sound wave, which is 10 times faster than the blink of an eye (MED-EL, 2020). This speed has also been observed in athletes' reaction time to auditory versus visual stimuli (Pain & Hibbs, 2007; Schaffert, Janzen, Mattes, & Thaut, 2019; Shelton & Kumar, 2010).

Previous contiguity measurements in L2 captioned video were not explicit about their choice of the verbal reference of measurement. In view of all that has been mentioned in this section, it could be contended that, in the context of captioned video where a split-attention effect is involved, we cannot guarantee that every written form is read. However, we could strongly assume that, relatively, spoken words are heard. On this basis, I chose to use spoken form instead of written form to measure verbal-visual contiguity in L2 captioned video in this study.

4.1.2 Contigfrequency

This study identified three measurable elements subsumed under the term verbal-visual contiguity. The first element is the subset of the number of verbal occurrences (verbal frequency) of words that have visual occurrences in video material. Though a precise term for its concept has been elusive, the construct was proposed by Rodgers (2018) to describe verbal-visual contiguity in videos. In second language vocabulary research, the term frequency is generally understood to mean verbal occurrences of words. This shows a need to use a specific term to refer to the number of verbal occurrences contiguous with visual referents. In this study, the term that will be used to describe this meaning is contigfrequency.

No previous study has been found that investigated the effect of contigfrequency in videos on incidental vocabulary learning. As indicated in the

introduction to this chapter, the available experimental data have been limited to categorical variables. Counting the number of occurrences of a target word to study the effect of its verbal presence, instead of categorising the target word as either present or not, has been the norm among vocabulary researchers. Likewise, visual presence should be treated similarly in that the total number of visual occurrences of each word is included.

For reasons previously outlined (in Chapter 3, Section 3.1.6), researchers investigating the effect of verbal-visual contiguity might need to closely examine the link between incidental learning of target words and visual frequency of their related forms. Very little was found in the literature on whether the relationship between occurrence frequency and incidental vocabulary learning is moderated by the account of related forms of the target words. In particular, whether the word family was included as a unit of counting has not been made clear in most studies of incidental learning as a function of verbal occurrence.

4.1.3 Contigduration

Peters (2019) found that words with visual referents "are almost three times more likely to be picked up incidentally than words without imagery" (p. 16). However, she used a sole ~~measure~~ of contiguity. A potentially more systematic approach would identify how this measure interacts with other variables that capture alternate dimensions of the contiguity learning effect, especially in studies where captions, and thus split attention effects, are involved. For instance, a visual referent might appear both infrequently but for longer durations or frequently but for brief durations; it is unknown whether either of the two scenarios is more conducive to learning.

The current study seeks to examine contigduration, which is the amount of time a visual referent is displayed on the screen, an alternate dimension of verbal-visual contiguity. Duration as a metric of verbal-visual contiguity has been overlooked in previously published studies. In addition to contigfrequency, it is vital to examine how much time is available to the learner to have his eyes fixed on the specific visual stimulus.

Verbal-visual contigduration in the present study is firmly grounded in the learning theory of contiguity. Another study was encountered at the time of writing, which examined the effect of visual referents' duration on incidental vocabulary learning irrespective of whether they were in temporal contiguity with the spoken/written form (Pujadas Jorba, 2019). The author measured the total seconds an image is present on-screen, what they termed the image time of screen (ITOS), throughout the audio-visual input that amounted to 2 hr and 55 min. That is, the duration of all visual referents present in the episodes were included. Results showed that for every additional minute in ITOS, the odds of a correct response increased by 18% and 24% for word form and word meaning, respectively. While these results appear promising, the decision to include whatever visual referent in the video without accounting for the position and the frequency of the word form was not well justified in the text. In fact, there remains uncertainty as to whether results were input-dependent. In other words, since the position of visual referents in relation to word forms was not examined, it is not clear whether this effect would still be observed in materials with distribution patterns of verbal and visual occurrences that are entirely distinct from those examined in Pujadas Jorba's study. The researchers also considered knowledge of nouns only. The present study seeks to provide more definitive evidence to contiguity effects in learning from extensive TV viewing (5 hr longer than Pujadas's exposure). This could be done by including verbs and adjectives, careful selection of contiguity timespans, and detailed examination of visual referents.

4.1.4 Contigratio

A final potentially significant aspect of verbal-visual contiguity is contigratio. Counting how often verbal occurrences have a corresponding image, by itself, might not be enough to fully measure the potential effects of verbal-visual contiguity. It may be important to ask as well how high or low this count is relative to existing verbal occurrences. For example, two words, '*squirrel*' and '*vulture*', have 6 and 8 contigfrequency, but 6 and 15 verbal frequency, respectively. Every time the word '*squirrel*' appears there is a visual referent for it (i.e., ratio = 6/6 = 1), unlike '*vulture*', which appears seven times without any corresponding image; ratio = 8/15 = 0.53. Given the lack of research in this area, evidence for the potential influence of high or low contigratio could not be traced in the literature. Thus, one possibility

is that contigratio exerts an additional influence on learning, independent of contigfrequency effect

In summary, using frequency as an exclusive measure of the degree of verbal-visual contiguity involves neglecting other potentially influential dimensions of contiguity. For instance, longer visual durations can reasonably be expected to generate longer fixations on the stimulus, and hence a potential increase in the probability of encoding and later successful retrieval. Study 2 attempts a more comprehensive approach in that it considers contigfrequency, contigduration, and contigratio as potential measures contributing unique variance to incidental learning of words.

4.1.5 Timeframe

In this study, timeframe or timespan refers to the duration in seconds between the occurrence of a target word's verbal form and the occurrence of its visual referent. The researcher needs to set out one specific timespan to abide by while quantifying contiguity (i.e., frequency, ratio, and duration). As indicated previously, Rodger's work (2018) was used as a jumping-off point for a well-thought-out and accurate measurement. Rodgers opted for shorter timespans: ∓2 seconds and ∓5 seconds. These choices were based on the six seconds subtitling rule adhered to by television stations worldwide and grounded in the work of d'Ydewalle, Van Rensbergen, & Pollet (1987). Choosing five seconds instead of six ensured that the image occurred within the processing time rather than its end. Subsequent studies implemented 5-second timeframe (Peters, 2018; Pujadas Jorba, 2019). This study extends the timeframes used in previous studies to ∓7 seconds and ∓25 seconds based on memory research findings. It makes use of two timeframes for two reasons: first, to prevent selection bias from being a potential concern, especially given that memory studies are specific to explicit rather than incidental context and also have shown inconsistent results; second, to address the need for another study to approximate the maximum length of time verbal forms and visual referents can be separated before contiguity learning effect is no longer observed, as recommended by Rodgers.

One question that is important to ask is whether temporally distant visual referents have a contiguity learning effect. Visual referents were logged by observing referents before and after a verbal occurrence (this will be discussed

further in the upcoming section). Therefore, it is fair to suggest that a typical contiguity timespan is rooted in visual and verbal short term memory (STM) research. Two scenarios exist: ' – before' verbal occurrence; representing the ability to remember a visual referent until you hear its verbal form (i.e., visual STM); and ' + after' verbal occurrence; representing the ability to remember a verbal form until you see its visual referent (i.e., verbal, phonological STM; spoken form is the reference here).

On the one hand, there is reason to believe that the verbal-visual contiguity learning effect might extend to up to 7 seconds. Perhaps the most relevant research to the subject of this review is studies of memory recall. To begin with, it is unlikely that we cease to process a stimulus beyond 5 seconds. According to informational and visual persistence (Coltheart, 1980), a brief visual referent of the target word will continue to be visible and processed by students after it disappears from a video, until another shot interrupts the process.

Moreover, memory span is commonly measured as the longest series of pictures (or numbers; digit span) a person can hold in their memory after seeing or hearing them at the rate of one per second. According to Miller's law (1956), on average, this span is about seven items in length, referred to as the "magical number". So, what could this tell us? In contiguity terms, when a learner sees a visual referent while watching a video, they might be able to recall up to the last seven word forms they have just heard (or read in captions) and possibly retrieve the corresponding word form. As previously touched upon in connection with cross-situational learning, the more the encounters, the more the inferring process is successful.

In a similar vein, at the time a learner hears (or read) an unknown word, they might be able to recall up to the last seven visual candidates they have just seen, if referents are clearly distinguishable from the scene, and hence, possibly arrive at the correct visual referent. According to an American film theorist, films usually average no less than a minute per scene (Bordwell, 2006, p. 57). This average scene length was also demonstrated based on 20 top return-on-investment films (Velikovsky, 2012). Though there could be several potential visual referents displayed in one scene, "this maybe less likely in a documentary television

programme where the narration is designed to explicitly describe what is on-screen and the significance of what is being seen" (Rodgers, 2018, p. 205). Overall, this feature in documentaries generally mitigates against there being more than seven distinct items in a seven-second timeframe. This increases the likelihood of having the visual referents retained in the memory when encountering its equivalent verbal form.

In addition, a relevant contiguity study to our research dates to 37 years ago by Baggett (1984). While my study investigates word-image contiguity in authentic video, Baggett examined narration and video contiguity in educational video by manipulating the start of narration in relation to the video. Seven groups watched a 30-minute instructional film of an assembly kit either in synchrony with narration, 21, 14, 7 seconds before narration, 21, 14, or 7 seconds after the narration, followed by an immediate or seven-day recall test. In both tests, the contiguity and the (-) 7 seconds group (i.e., before narration) performed substantially better than the remaining groups. In another study, 6 seconds was found to be the minimum duration before any visual STM information was lost (Murdock, 1971). The studies presented thus far indicate that an optimal timespan for contiguity measurement could reach 7 seconds.

On the other hand, support for choosing longer timespans, up to 25 seconds, can also be found in several studies. Peterson and Peterson (1959) conducted what is perhaps the most replicated study in verbal memory research, for its Brown-Peterson distractor technique. Students were presented with verbal items, then were asked to count backwards three-digit numbers in the hope of minimising their rehearsal of target items prior to recall. Results indicated that STM duration was less than 18 seconds; however, rehearsal, which is believed to improve recall, was prevented in this study, suggesting duration could be much longer. Another indicator of longer retention lengths is Baddeley and Levy's use (1971) of a distractor task that was 20 seconds long in an attempt to block STM.

Further evidence could be provided by another study in which 10 participants were given one male-voiced letter, either aurally or visually, to remember while repeating aloud female-voiced letters (Kroll, Parks, Parkinson, Bieber, & Johnson, 1970). Participants recalled more visual than aural letters, with 100% correct

responses at 1 second, 92% at 10 seconds, and 88% at 25 seconds. These results indicate that "humans…have some ability to hold a visual image for at least 25 seconds" (p. 223).

Though recall studies have closer ties with our focus here, studies of recognition memory also usefully contribute to our understanding. Participants correctly recognised 90% of 2560 photographic stimuli presented for 10 seconds each, even after three-days interval (Standing, Conezio, and Haber, 1970). Most importantly, presentation duration could be reduced to one second per image without affecting results. This vast capacity for remembering images was also found in other studies (e.g., Nickerson, 1965, 1968; Shepard, 1967). In an experiment with the famous patient H.M who suffered long-term memory loss, Prisko (1963) showed him two images (shapes), one after the other, then asked him to indicate whether the images were identical. The interval between the pair ranged from 0 to 30 seconds. The result showed that H.M was able to keep the first image in his memory for about 15 to 30 seconds. The claim that forgetting from STM is complete within 30 seconds can also be found in Shiffrin and Atkinson's work (1969). Together, the studies reviewed here support a view that contiguity learning effects might extend to longer durations and may well exceed 7 seconds to up to 25 seconds.

Although each has a different aim and focus, the studies reviewed in this section highlight the need to extend the currently used contiguity measurement timeframe; from 2 and 5 seconds to notably 7 and 25 seconds. Admittedly, my argument relies heavily on evidence from studies of explicit learning in controlled experimental settings, and there could be a degree of uncertainty as to how much the results could inform my decisions. In short, support for my claim in vocabulary learning through videos studies is difficult to find; however, choosing two timespans in this study might help to explore opportune contiguity timespans for incidental vocabulary learning from the study materials.

4.1.6 Sequence

In addition to the length of the timeframe within which verbal-visual contiguity is measured, another operationalisation of verbal-visual contiguity is whether to observe visual referents occurring before or after verbal forms. Similar to previous

studies, this study adopts both sequences without attempting to control for this variable.

The sequence could exert an influence on the contiguity learning effect. In Baggett (1984) study, having the narration preceding the video was detrimental; this could suggest that we look at visual referents that occur “before” verbal occurrences. Conversely, spoken letters may be recalled more than visual letter (Kroll et al., 1970), suggesting that it might be useful to look at referents “after” verbal occurrences. Drawing on an extensive range of sources (42 studies), Eitel and Scheiter (2015) concluded that it all boils down to the complexity of text and images.

Although the sequence factor may play a crucial role, authentic videos in which target words are often ‘bunched’ (i.e., verbal occurrences are clustered or repeated) present an additional difficulty. A visual occurrence may both precede one verbal occurrence and follow another within overlapping timespans. As such, a combination of both was used and it was outside the scope of the thesis to study sequence effects, for reasons of space, and the topic is deferred to future work.

4.2 The Present Study

This second study investigated the effect of verbal-visual contiguity in L2 captioned video on incidental acquisition of knowledge of meaning recall and recognition and spoken and written form recognition, from extensive viewing of two full-length seasons of documentary series. Few accounts exist in the literature: a descriptive study (Rodgers, 2018) quantified contiguity (contigfrequency), and three experimental studies (Ahrabi Fakhr et al., 2021; Peters, 2019; Pujadas Jorba, 2019) categorised contiguity. The present study distinguished three conceptual elements that are believed to make up the contiguity construct: contigfrequency, contigratio, and contigduration. These were measured using longer timespans (∓7 seconds, ∓25 seconds) relative to what has been previously observed (∓2 seconds, ∓5 seconds). The study adopted a within-participants design to the viewing group in Study 1, who watched two full-length documentary series extending to eight viewing hours in the form of L2 captioned video, over six weeks at two-week intervals. Participants' word knowledge of 28 target words present in the input was assessed at the level of meaning recall, meaning recognition, spoken form recognition, and written form recognition. Pretests and posttests were administered before and immediately after the treatment phase, except for meaning tests that were pretested only.

4.3 Method

4.3.1 Questions and hypotheses

Study 2 formulated the following research questions:

***Research Question 1*:** *(model building)* Is the effect of verbal-visual contiguity on incidental word learning from extensive viewing of L2 captioned documentary series moderated by:

(a) The length of the timespan within which contiguity was measured

(b) The inclusion/exclusion of weak visual referents and related word forms?

***Research Question 2*:** *(main)* What is the effect of three verbal-visual contiguity measures: contigduration, contigfrequency, and contigratio on incidental word learning of different parts of speech from extensive viewing of L2 captioned documentary series?

> ***Hypothesis*:** It was predicted that contigduration, contigfrequency, and contigratio will influence incidental vocabulary learning. Higher contiguity measures will contribute to higher accuracy scores on both meaning and form measures of word knowledge.

Research Question 3: *(exploring)* What are the relative strengths of the three predictors; contigduration, contigfrequency, and contigratio of incidental vocabulary learning from extensive viewing of L2 captioned documentary series?

4.3.2 Participants

Sixty-five Algerian EFL learners in their third year of the Linguistics Bachelor programme at the University of Jijel, in the autumn semester in the 2017-2018 academic year, took part in Study 2. Of these, 12 participants were excluded: if they were absent in any session of the pretests and posttests and if they missed any session of the treatment phase. Data from 53 participants (47 females and 6 males) aged 21-23 years (M = 21.11) were kept for analysis. Participants had an intermediae to upper-intermediate English language proficiency. They were all native Arabic speakers with French as a second language. The study was approved and consent was obtained (See Chapter 3, Section 3.3.2 for recruitment, full description of participants, and ethical considerations).

4.3.3 Materials

The materials were the two full-length seasons of the documentary series previously used in Study 1. Information on the series is detailed in Chapter 2, section 2.2.1.

4.3.4 Target Items

Twenty-eight words appearing in the documentary materials made up the target items for Study 2 (Table 4.1). Twenty of these were the target spaced words used in Study 1, and the remaining eight words were the massed nouns previously seen in the Norming study (Chapter 2), which make part of the target word pairs of Study 3.

Table 4. 1

Target Words (N = 28) and Related Variables

Item	Verbal Freq	Other forms[a]	Log Freq (Zipf)[b]	Length		Concreteness[c]	Cognate status
				Characters	Syllables		
			Nouns				
supernova	15	0	3.08	9	4	3.78	Yes
constellation	11	0	3.20	13	4	4.31	Yes
sphere	16	31	3.68	6	1	4.44	Yes
spectrum	12	0	3.80	8	2	2.97	Yes
particle	13	0	3.48	8	3	3.78	Yes
temple	09	0	4.03	6	2	4.53	Yes
cosmos	35	12	3.27	6	2	3.19	Yes
tide	12	6	4.25	4	1	4.10	No
hexagon	10	6	2.63	7	3	4.52	Yes
fusion	10	6	3.60	6	2	3.30	Yes
pile	8	0	4.23	4	1	4.56	No
photon	32	0	2.45	6	2	3.38	Yes
moth	11	0	3.64	4	1	4.69	No
symmetry	11	8	3.25	8	3	2.79	Yes
sulphur	16	2	3.35	7	2	4.23	Yes
manatee	14	0	2.08	7	3	4.66	No
			Adjectives				
intricate	8	1	3.60	9	3	2.36	No
dense	24	3	3.74	6	1	3.14	Yes
denser	24	3	3.74	7	2	3.14	Yes
faint	9	1	3.75	5	1	3.74	No
cosmic	10	37	3.35	6	2	2.76	Yes
alien	10	0	4.19	5	2	3.52	No
			Verbs				
stretch	18	0	4.38	7	1	3.62	No
forge	09	0	3.57	5	1	4.04	No
emit	11	0	2.73	4	2	3.22	Yes
sculpt	11	2	2.61	6	1	3.57	Yes
orbit	40	1	3.73	5	2	3.11	Yes
squash	8	0	3.92	6	1	3.04	No

Note. Freq = frequency; The last 8 nouns were massed words.

[a] Other forms were derivatives and compounds. [b] Measures were based on the SUBTLEX-UK word frequencies, presented in Zipf-values, a logarithmic scale: 1-3 = low frequency, 4-7 = high frequency (Van Heuven et al., 2014). [c] Measures were based on 40 thousand English lemma words on a 5-point rating scale going from abstract to concrete (Brysbaert et al., 2014).

4.3.5 Procedure

This study adopted a within-participants design. Participants watched eight episodes (i.e., 8 hr) of two full-length seasons of documentary series in the form of L2 captioned video, over a period of six weeks. The research schedule and vocabulary tests are the same as that of the viewing group in the previous study (See Sections 3.3.4 and 3.3.6 [*Dependent Measures*] of Chapter 3 for a full description).

4.3.6 Scoring

Responses of tests of meaning recognition and recall and spoken and written form recognition were scored in the same way as for Study 1: "0" for incorrect, missing, and "I don't know" responses, and "1" for correct responses (see Chapter 3, Section 3.3.6).

4.4 Analyses

4.4.1 Measuring Contiguity

What we know about how to measure contiguity is largely based on Rodgers' study (2018). The current study somewhat deviated from his approach. Rodgers marked four categories of occurrence: no occurrence, concurrent, $\overline{+}2$ seconds, and $\overline{+}5$ seconds. While his study was descriptive and required him to count the frequency of concurrent and non-occurrences per every item, my study focused on capturing all instances of contiguity within two timespans (i.e., categories). I opted for longer timespans: $\overline{+}7$ seconds and $\overline{+}25$ seconds and the justification for this selection can be found in Section 4.1.5. Lastly, contigduration is one of the distinctive features of the present study. To calculate it, the start and endpoints of every visual referent segment were logged.

Timespans were stored in two database tables: $\overline{+}7$ seconds and $\overline{+}25$ seconds. As illustrated in Figure 4.2, each data table had the following as vectors: word; the target word, word type; whether the word was exactly the target word or a related form (see *Target Items* in Chapter 3, Section 3.3.6 for details on related forms), image quality; the quality of the visual referent coded as "strong" or "weak" (this will be detailed later), episode ID; the episode in which the word occurred, position (-); the timespans for visual referents appearing before verbal occurrence; verbal

Figure 4. 2

Tabular Presentation of Logged Contiguity Timespans

	A	B	C	D	E	F	G	H	I
1	word	word type	image quality	episode ID	position (-)		verbal occurrence	position (+)	
2					start-point	end-point		start-point	end-point
3	Hexagon	exact	strong	F1	02:56		2:58		
4		exact	strong				16:58		17:05
5		exact	strong		17:21		17:28		17:30
6		exact	strong				22:22		22:24
7		related	strong		22:22	22:24	22:46		
8		exact	strong		23:26		23:29		23:31
9		exact	strong		23:55	23:57	24:02		
10		exact	strong		24:21		24:22		
11		related	strong		25:30		25:35		25:42
12		related	strong		51:09		51:10		51:17
13		related	strong		54:03	54:05	54:10		
14		exact	strong		54:09				54:12
15	tide	exact	weak	U1	10:08		10:15		10:22
16		exact	strong	F2	20:40		20:47		20:52
17		exact	weak	F2				20:52	20:54
18		related	weak	F2	22:03		22:10		
19		exact	strong	F2			22:32		22:39

25 seconds | 7 seconds

Note. Excel was used to store the timespans, in minutes and seconds. U = Wonders of the Universe; F = Forces of Nature.

occurrence; the instant in which the word occurred, and position (+); the timespans for visual referents appearing after the verbal occurrence.

Using time-stamped scripts of the eight episodes and the Ctrl+T function in VLC media player, the instant in which each spoken form of a target word occurred was logged into the 25 seconds sheet in min:s format. The justification for choosing spoken over written forms (L2 captions) as a reference to measure verbal-visual contiguity is detailed in Section 4.1.1. The scene was then examined for the occurrence of visual referents (simultaneous occurrences) and were logged into the sheet by highlighting the cell.

Secondly, the video was jumped 25 seconds backwards to identify visual referents starting 25 seconds before verbal occurrence (-). The start and the endpoints of every segment that includes a visual referent were logged in separate columns. Fleeting presentations of visual referents (< 1 second) that occur fleetingly were also logged; this was done because neuroscientists suggest that the brain can detect images that are as little as 13 milliseconds (Trafton, 2014). However, if a visual referent prolonged beyond 25 seconds on screen, the extra seconds were not considered. Thirdly, the video was jumped ahead to identify visual referents appearing 25 seconds after verbal occurrence (+) following the previous approach.

I repeated manually this procedure of observing visual referents that precede and follow verbal forms 536 times (i.e., verbal occurrences including related forms). It resulted in redundant timespans for close occurrences of one verbal form and were then removed, bringing the total length of analysed material to 26800[10] seconds. Given the nature of documentary series, visual referents in a 25-second scene were frequently interrupted by scene changes, resulting in multiple timespans per 25-second scene. The analysis revealed 25 words with visual referents and three words lacked imagery: *alien*, *sculpt*, and *intricate*. The latter has the lowest concreteness rating ‘2.36’, thus, it was unsurprising not to be visually represented.

[10] (520 × 25 seconds × 2).

Analysis of scenes produced 356 and 359 timespans of visual referents before and after verbal forms, respectively. The two categories were then combined, and 167 redundant timespans were manually removed. Hence, there were 548 visual referents' timespans for 25 verbal forms falling under the 25-second category to be sampled for inter-coder agreement. The words *cosmos* and *cosmic* as well as *dense* and *denser* were treated as one item, for sharing similar meanings and exact timespans. This reduced the timespans to sample from to 470 timespans for 23 words. Another independent researcher inter-coded a sample of 97 out of 470 timespans (20%), which will be detailed in the upcoming section.

Following the inter-coder agreement, the adjusted coded data were used to calculate the three variables of interest. Contigduration from 527 ultimate timespans for the 25-second category was calculated (in seconds) using lubridate package (Version 1.7.4; Grolemund & Wickham, 2011) in R (R Core Team, 2018), RStudio (version 1.2; RStudio Team, 2018) by subtracting the segment start point from the endpoint. Contigfrequency was calculated by summing the number of occurrences that had visual referents (priority was given to strong over weak visual referents). Contigratio was calculated by dividing contigfrequency by verbal frequency.

Once the data for ∓25 seconds sheet was finalised, it was copied into the ∓7 seconds sheet. Data for ∓7 seconds was created by deducting (backward) and adding (forward) 7 seconds to the instant of verbal occurrence to mark the two ending points for this category segments. Extraneous timespans were manually removed, and contiguity was calculated in the same manner as detailed above.

4.4.2 Imagery Coding

While some visual referents may be very straightforward to code, others are not and may involve some guessing. The coded data are the baseline of Study 2 analysis and were coded systematically to ensure their reproducibility and generalisability (Bolibaugh, Vanek, & Marsden, 2021). This was achieved by having a second coder, so the researcher could "safely rely on per cent agreement to determine interrater reliability" (McHugh, 2012, p. 282), and by establishing a set of criteria for what constitutes a visual referent and what distinguishes strong from weak referents. It could be noted that, prior to assessing inter-coding reliability, I reconducted a full

second coding; however, the intra-coding agreement test was not checked at that point.

Inter-coding Agreement

Another external researcher (IELTS = 6.5) coded 97 (20%) out of 470 segments of 25 seconds timespan category for 23 target items. A small number of timespans with no visual referents were added to the sample to increase the validity of inter-coding agreement test. The number of timespans for every word ranged from 2 to 68; this variation made it impossible to sample evenly across words. Stratified sampling into the number of timespans was then built into five different categories as shown in Table 4.2. Hierarchically, every category was assigned the number of timespans to be inter-coded.

The selected timespans for every target word were inclusive of all possible visual referents. To put it otherwise, one word could have different kinds of referents, and a referent could appear multiple times. The number of timespans the sample required was reached by prioritising disparate referents before adding repetitive ones. In a similar vein, spaced words occur in multiple episodes, variety in episodes was hence maintained for every spaced word as much as possible.

A joint meeting between the two coders was held. I first obtained consent from the second coder (Appendix F), handed him the coding protocol (Appendix M), and explained the adopted sequential coding procedure, word by word, and segment by segment, along with the criteria that I followed for coding. Afterwards, I started to show these, one after another. The coder had to indicate on the protocol whether a supporting image for the target word existed along with its quality (weak or strong). The researcher showed the segments to the second coder because asking him to move to specific scenes using what have probably been unfamiliar VLC features was thought to be impractical.

Inter-coding agreement for categorisation of visual referents was assessed by comparing the two data. The agreement obtained for visual referents was high: 0.96, while the agreement score for referents quality was lower: 0.87. A consensus

Table 4. 2

Stratified Sampling for Inter-coder Reliability

word	Total timespans	Timespan category	Timespans to be inter-coded
particle	2	<10	2
supernova	6		2
pile	6		2
emit	11	10 < N < 15	3
forge	12		3
temple	14		3
squash	14		3
spectrum	15	15 < N < 20	4
stretch	15		4
faint	16		4
sulphur	16		4
dense/denser	17		4
fusion	17		4
photon	17		4
manatee	17		4
tide	20		4
moth	20		4
hexagon	21	20 < N < 30	5
constellation	25		5
symmetry	39	35 < N < 45	7
orbit	40		7
sphere	42		7
cosmos/cosmic	68	> 50	8
Total	470		97

Note. dense/denser and cosmos/cosmic share similar visual referents' characteristics.

approach was adopted; discrepancies were discussed, and controversial referents were solved by using a third coder. Adjustments were made to the rest of the coding data. Analysis of scenes resulted in 320 and 331 timespans of visual referents before and after verbal forms, respectively. The two categories were then combined, and this led to the manual removal of 124 redundant timespans.

Coding Criteria: What Makes a Visual a Referent?

To judge whether a frame depicts a visual referent for a co-occurred word, one should first ask: what makes a visual a referent? A source of uncertainty in preceding investigations lies in the obscurity of the distinctions and the criteria used during the coding process, especially when different parts of speech and concreteness ratings are involved, as in Peters (2019). A more systematic approach would clarify and illustrate (i.e., with pictures) how different visuals were believed to have depicted words with disparate characteristics. Specifically, it is crucial to show how coders identified a displayed image "… as being of a given type or as having given properties. The latter, the depicted properties, are in general the most problematic" (Brown, 2010, p. 208). Ziska (2018) put it in this way: "pictures possess both the capacity to approximate to the appearance of things, but also to do so in a misleading way" (p. 232). Hence, if we consider perceptual ambiguities associated with seeing pictures (in visual arts), it becomes essential to devise a set of coding criteria for referents analysis to minimise subjectivity.

In this study, referents were coded into two separate categories: strong and weak. The justification for coding an image as depicting of a strong or weak referent depended on few aspects. Large referents were considered as strong, since they are more potent than small ones and give a "well-focused eyeful", as highlighted by Brown in "Seeing Things in Pictures" (2010), as shown in Figure 4.3 (see Appendix A for examples of visual referents for each word).

Figure 4. 3

Examples of Strong Visual Referents for Large Size

spectrum supernova hexagon

Nonetheless, to ensure that pictures were not arbitrarily ignored, small distant pictures were coded as referents (strong category) in some instances. First, if they had support from the non-verbal signs, as illustrated in Figure 4.4.

Figure 4. 4

Examples of Strong Visual Referents for Non-verbal Signs

moth

supernova

Second, if the image was outstanding relative to the ambient space. For instance, when there were no distracting objects in the frame (as *sphere* in Figure 4.5), or when the image was heightened with chiaroscuro (Brown, 2010), that is, the involvement of low and high-contrast lighting which creates areas of both light and darkness (as *faint*).

Figure 4. 5

Examples of Strong Visual Referents Due to Chiaroscuro Effect and Absence of Distracting Images

faint

sphere

Furthermore, a word whose meaning was not plainly visible but rather embedded in another image was coded as a visual referent, especially if it had narrative saliency. This criterion was inspired by Wollheim's influential theory of depiction and pictorial representation "Seeing-in" (1980, 2001, 2003). Accordingly, we are likely to see more than one thing in one image (Wollheim, 2001, p. 26). This could be best exemplified in Figure 4.6. However, these referents were included in the weak category. It was unknown whether the target viewers had shared similar experiences that would allow them to conform to the same conventions regarding the image being observed and processed.

Figure 4. 6

Examples of Weak Visual Referents for Low Visibility

photon in aurora

emit in lense flare

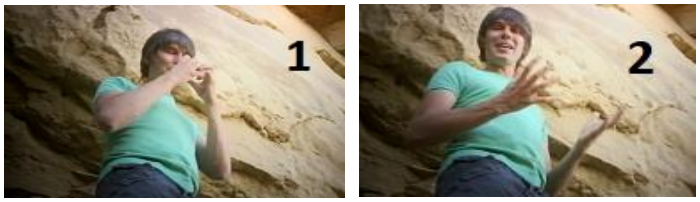
Moreover, the presenter in the analysed documentary series, Brian Cox, often gesticulated to depict the meaning of the co-occurring target word. Known as iconic gestures (Feyereisen & De Lannoy, 1991), these tend to " … represent through some form of depiction or enactment something relevant to the referential content of what is being said" (Kendon, 2004, p. 106). Thus, these gestures were considered in the study as visual referents (see Figure 4.7). Unlike pictorial referents, gestural referents can rarely be interpreted by their mere shapes. Instead, they require the viewer to understand the co-occurred verbal speech since they depend on ideational equivalence instead of social conventions (Hadar & Butterworth, 1997, p. 148). Semantic specificity refers to the unambiguity with which gestures depict the meaning referred to (Hadar & Pinchas-Zamir, 2004, p. 204). As a result, it was decided that referents with high and low "semantic specificity" were included in the strong and weak categories, respectively.

Figure 4. 7

Examples of Gestural Visual Referents

hexagon (n)

stretch (v)



supernova (n)

faint (adj)

In summary, due to the indeterminacy of pictures, the materials were scrutinised for evidence of visual referents for the target words by establishing a set of criteria related to visual and narrative saliency of referents. Gestural referents were coded as strong or weak depending on their semantic specificity. Overall, an image, be it still (e.g., object) or dynamic (e.g., action, event), was coded as a *strong* referent if it met the following:

› It was large enough to be noticeable.
› It conveyed the meaning accurately, especially in the case of a dynamic image.
› It was immediately perceived as a visual referent for the target word due to its overall observable properties.
› It was small/distant but supported with non-verbal signs or high visual saliency.

Images were coded as *weak* referents if they had some degree of narrative saliency and met the following:

› It was small and distant with distracting images.
› It might not be straightforward to discern if the image possessed the required properties to be a visual referent.

4.4.3 Analysis Procedure

Data were analysed in R (R Core Team, 2018) using Rstudio (version 1.2; RStudio Team, 2018). Results were summarised using dplyr package (version 0.8.3; Wickham, François, Henry, & Müller, 2019). They were visualised using ggplot2 package (version 3.2.1; Wickham, 2016) for meaning recognition and recall results and ggpaired function of ggpubr package (version 0.2.4; Kassambara, 2019) for form recognition results (for having a pretest-posttest structure). Finally, the binary data were analysed with generalised linear mixed-effects (GLM) logistic regression models using glmer function of lme4 package (version 1.1-21; Bates, Maechler, Bolker, & Walker, 2015). The remaining part of the section explains a sequence of procedures adopted to answer the three research questions for Study 2.

Research Question 1

The first research question explored whether effects of verbal-visual contiguity on incidental word learning are moderated by the length of the timeframe within which contiguity was measured and the inclusion/exclusion of weak visual referents and related word forms. These potential moderator effects were explored within the contigduration measure because it takes on a sufficiently large number of different values; thus, it is a more continuous measure of contiguity. A model comparison approach was adopted. Eight models were created which systematically varied their inclusion of weak visual referents and related verbal forms within the two

timeframes of ∓ 7 seconds, and ∓ 25 seconds. The data used were those of the viewing group (see Study 1).

For meaning recall and recognition measures that lacked a pretest, the model specified 'posttest' as a dependent variable, contigduration as a fixed effect, spacing (spaced = 1, massed = 0) as a control variable, and participants and words as random effects, with random intercepts allowed to vary across participants and words (e.g., random = ~1 | participant). Spacing was automatically dummy coded by R software as a categorical variable; then, it was releveled to set spaced words (N = 20) as the reference level. Contigduration was centred to avoid convergence issues.

For spoken and written form recognition measures, data were in a repeated-measures design (since participants had sat a pretest for these measures). The models had specified response accuracy as the dependent variable, time × contigduration as an interaction term (since contigduration is assumed to be a treatment effect), spacing as a control variable, and participants and words as random effects, with random slopes of time for each to denote that the effect of time varies across participants and words (e.g., random = ~Time | Word). Adding an interaction between time and contigduration was assessed for all the candidate model sets. Evidence for its significance was found for six models out of 8, hence, it was retained.

The restriction on the number of covariates was derived from the study sample size, which was 53 participants, permitting for only five parameters (including the intercept and random effects) in each model (about 10 to 15 participants per variable as recommended by Nunnally, 1978; Kass & Tinsley, 1979). This was contrary to Study 1, in which most of the theoretically meaningful variables were able to be included due to the larger sample size. Multilevel modelling and inclusion of maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013) helped meeting the independence assumption by controlling for individual variations among participants and across words (in both points of time for form recognition), and these were justified using the likelihood ratio test (Pinheiro & Bates, 2000).

The models were then compared based on the Akaike Information Criterion (AIC) measure corrected for small sample size[11] (AIC_c; Burnham & Anderson, 2002) using aictab function of the AICcmodavg package (Version 2.2-2, Mazerolle, 2019). AIC penalizes models for having complex fits. It was used instead of the traditional likelihood ratio test because the latter is not valid for non-nested models (i.e., where one model is not a special case of another model). Higher AIC indicates that the fit is worse, while lower AIC indicates a better fit (i.e., a more parsimonious model). For each candidate model, AIC weight (*w)* and AIC difference *(*Δ*)* were calculated ($\Delta_i = AIC_i - AIC_{min}$). A model has the best support from data when it has the highest AIC *(w)*. It has substantial support when $\Delta < 2$, less support when $2 < \Delta < 4$, and essentially no support when $\Delta > 10$ (Burnham & Anderson, 2004, p. 70). Models fitting and comparison were carried out for all the four dependent variables (Meaning recognition and recall, spoken and written form recognition).

Research Question 2

The second research question assessed the effects of three verbal-visual contiguity measures: contigduration, contigfrequency, and contigratio, on incidental vocabulary learning. The model comparison results from Research question 1 were used to select the optimal length of timespan (7 vs. 25 seconds) and inclusion and exclusion of weak referents and related word forms to calculate contigfrequency and contigratio. Afterwards, the three contiguity variables were centred around their means (to assist model convergence) and entered in a model as three predictors of accuracy for each dependent variable.

A key concern when conducting multiple regression is multicollinearity. The issue exists when there is a strong correlation between predictors. There is currently no consensus on the best approach to identify multicollinearity. Nevertheless, simple examination of correlation matrices has been claimed to be less optimal relative to an examination of condition indexes and the variance inflation factor (VIF) (Belsley, D., Kuh, E., & Welsh, 1980; Flom, 1999; James, Witten, Hastie, & Tibshirani, 2013). This study tested independent variables for potential

[11] AIC_c was used instead of AIC because the sample size was small (N = 53) compared to the number of estimated parameters. The rule of thumb is $n/k < 40$ as advocated by Burnham & Anderson (2002).

multicollinearity using VIF, the frequently used diagnostic in empirical studies and advanced statistical books. VIF values greater than 5 and 10 are the most common points of concern (Marquaridt, 1970; Myers, 1990; Imdadullah, Aslam, & Altaf, 2016; Menard, 2002), and tolerance statistics (1/VIF) lower than 0.1 and 0.2 are considered problematic (Hair, Anderson, Babin, & Black, 2010; Menard, 2002). Using imcdiag function of mctest package[12] (version 1.2.5; Imdad & Aslam, 2018), VIF diagnostics showed no multicollinearity issues (VIFs < 4; 1/VIF > 0.20). Therefore, I performed a GLMregression that included the three contiguity measures (contigduration, contigfrequency, contigratio) as predictors in one model of the previously used specifications. Statistical significance was determined using Wald's Z values.

In addition to this approach in which the three predictors were included in single models, I felt it was prudent to estimate coefficients for three separate models each consisting of an individual contiguity predictor (an approach that is not uncommon, e.g., Pawlicz & Napierala, 2017) to unmask the unique relationship of every contiguity measure with response accuracy[13] (Marquardt & Snee, 1975; Lavery, Acharya, Sivo, & Xu, 2019). This additional approach is supported by a strong Pearson's correlation (r = 0.84) between contigduration and contigfrequency, even though this measure has not been advocated as a sole diagnostic (Belsley, Kuh, & Welsch, 1980; Weissfeld & Sereika, 1991). In fact, this step was thought to help find out whether future contiguity observations would yield similar results to this study despite differences in the consistency between predictors.

Therefore, two approaches were adopted for meaning recognition and recall data (full model with three predictors, three models for each predictor). For spoken and written form recognition, predictors were entered in separate models for having a repeated measures structure. There was an inevitable two-way interaction with time for three predictors, which causes overfitting issues due to a larger number of parameters relative to sample size.

---

[12] mctest package integrates the mostly used multicollinearity diagnostic measures into one computation.

[13] An alternative to few researchers has been standardising the regressors (e.g., Blything & Cain, 2016). Critics, however, have considered this as a big misconception in the field (e.g., Assaf, Tsionas, & Tasiopoulos, 2019).

Research Question 3

The last research question explored the relative strengths of contiguity predictors on incidental vocabulary learning. It was addressed using two alternatives for comparing non-nested models: the goodness of fit AIC_c and the predictive power R^2.

Firstly, the goodness of fit was compared across the three models using AIC_c criteria following the procedure in Research Question 1. The Akaike weight (*w*) and difference (*Δ*) can be used to rank the relative importance of predictors (Burnham & Anderson, 2002). Secondly, marginal r-squared change (ΔR^2m) (Vonesh, Chinchilli, & Pu, 1996) with and without the contiguity predictor was computed using the r.squaredGLMM() function of MuMIn package (version 1.43.15, Barton, 2019). This change denotes the incremental increase in the model predictive power resulting from adding a contiguity predictor of a specific explained variation. That is, values equal to zero indicate that the variable does not contribute to the model R^2. The function computes the marginal (fixed effects only) and conditional (fixed + random effects) delta R^2 (Vonesh, Chinchilli & Pu, 1996) and with the theoretical distribution-specific variance for binomial distributions. These values are based on Nagelkerke's pseudo r-squared values are less prone to common problems than other measurements of R^2 (Nakagawa & Schielzeth, 2013).

Both AIC_c and R^2 are typical indices that uniquely contribute to a model's overall strength; thus, need to be jointly considered as points of comparison. In fact, it has been recommended that R^2 should be supplemented with other methods such as AIC in nonlinear modelling (Spiess & Neumeyer, 2010). Fortunately, the performance package (version 0.4.4, Lüdecke, Makowski and Waggoner, 2020) allows for such a comparison. Indices of quality and goodness of fit were compared across models using the compare-performance() function. This computation returns a performance score, an exploratory index based on a mean value of normalised 10 statistical criteria for each model and ranges from 0% to 100%, including the Bayes factor (BF) for models against the contigduration model (reference model). Only AIC and R^2 metrics were selected for comparison; however, full details of the remaining indices and a visualisation of the produced results are provided in Appendix N. The results for the three separate models of contigduration,

contigfrequency, and contigratio were compared to arrive at a hierarchy of strength for meaning recall and recognition and spoken and written form recognition.

4.5 Results

Before proceeding to answer the set of research questions, verbal-visual contigduration for the 28 items measured within 25 seconds timeframe (inclusive of weak referents and related forms) produced a total duration of one hour and 27 minutes visual referents (i.e., 5236 seconds). Following the exclusion of redundant timespans for close occurrences of one verbal form, referents were found to be located along a total of 527 timespans. The results for the four dependent measures are reported separately and arranged in the following order: meaning recall and recognition and spoken and written form recognition. The reason for this arrangement is outlined in Chapter 3, Section 3.5.3.

The mean scores for meaning and recall recognition posttests and written and spoken form recognition pretests and posttests for the View (N = 53) and Control (N = 34) groups on 28 items were summarised in Table 4.3. They were also plotted and are presented in figures.

Table 4. 3

Descriptive Statistics per Group for all Vocabulary Tests Scores (28 Items)

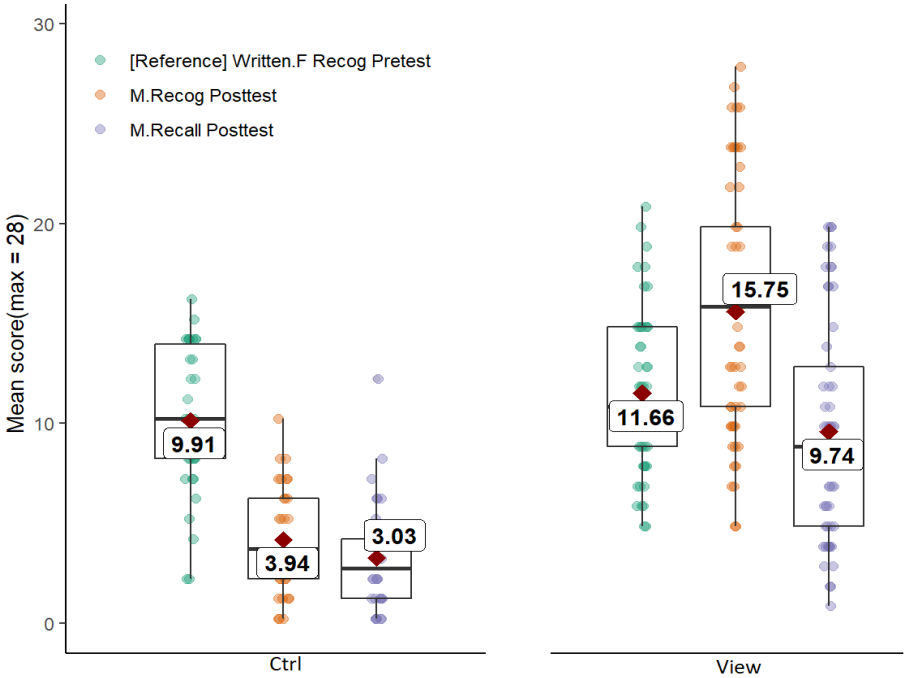
		Mean Scores					
		Pretest			Posttest		
		M (SD)	*Min*	*Max*	*M (SD)*	*Min*	*Max*
Meaning Recall	Control				3.03 (2.66)	0	12
	View				9.74 (5.53)	1	20
Meaning Recognition	Control				3.94 (2.62)	0	10
	View				15.75 (6.30)	5	28
Spoken Form Recognition	Control	8.67 (3.30)	3	16	8.69 (3.15)	3	16
	View	11.06 (3.76)	4	19	14.51 (4.34)	5	24
Written Form Recognition	Control	9.80 (3.71)	2	16	8.63 (3.44)	3	17
	View	11.66 (3.95)	5	21	15.62 (4.59)	6	26

Note. data %>% group_by(group, Time) %>% summarise (mean = mean(Response), sd = sd(Response), max = max(Response), min = min(Response)).
M = mean. SD = standard Deviation. Maximum score = 28.

The meaning accuracy data (Figure 4.8) show major differences between control and view group performance following the treatment. Compared to the Control group, participants who viewed eight episodes of L2 captioned documentary series were able to recognise and recall the meaning of a substantial number of words.

Figure 4. 8

Mean Accuracy in Meaning Recall and Recognition

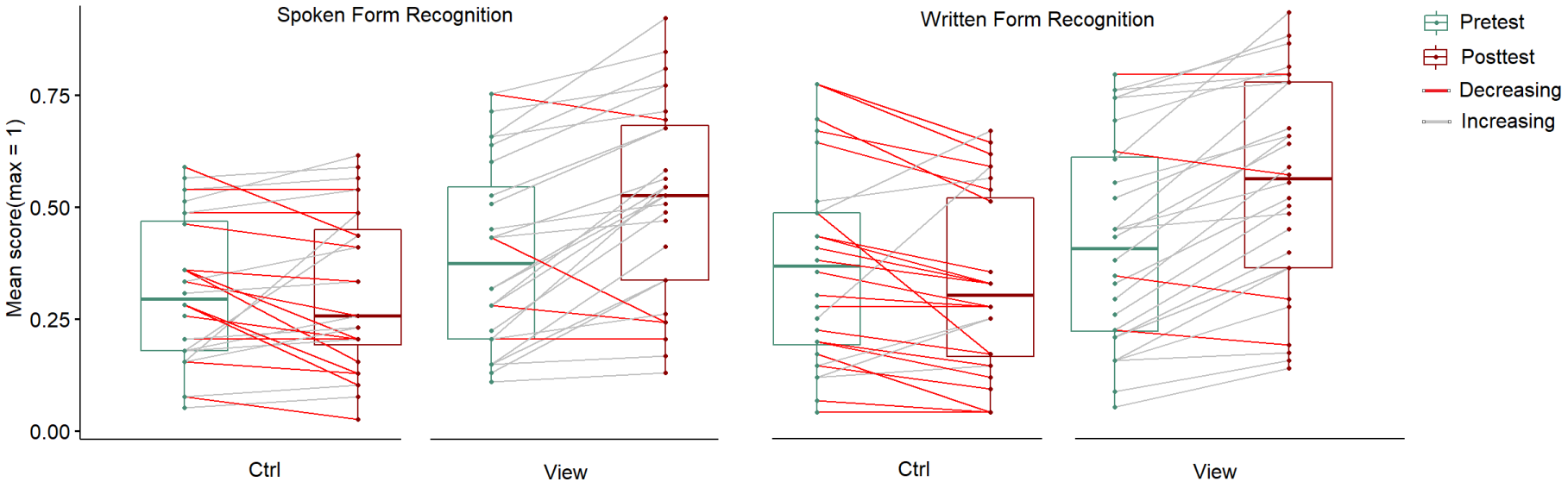


Note. Boxplots showing mean accuracy scores of meaning recognition and recall posttests for 28 words by subject and across View (N = 53) and Ctrl (N = 34) groups. Meaning was not pretested to prevent prior exposure bias, written form recognition pretest was used as a baseline reference.

The paired plots of individual participants scores on spoken and written form tests are shown in Figure 4.9. It is apparent that, again, in contrast to the Control group, the View group participants recognised in the posttest (at both levels of measurement) many word forms that were unknown in the pretest. This reflects a considerable amount of vocabulary learning gains at the level of spoken and written form recognition.

Figure 4. 9

Mean Accuracy in Form Recognition



Note. Paired boxplots (by word) showing mean accuracy scores of spoken and written form recognition on 28 words (dots) across View (N = 53) and Control (N = 34) groups. Grey and red lines match mean scores from pretest to posttest.

4.5.1 Research Question 1: Model Building

Is the effect of verbal-visual contiguity on incidental word learning from extensive viewing of L2 captioned documentary series moderated by the length of timespan and inclusion/exclusion of weak visual referents and related word forms?

Table 4. 4

Contigduration Measures Within ~~F~~7 and ~~F~~25 Seconds Timeframes, With and Without Related Word Forms and Weak Visual Referents

Timespans	7 sec				25 sec			
related forms	Without		With		Without		With	
weak referents	Without	With	Without	With	Without	With	Without	With
supernova	65	65	65	65	125	125	125	125
alien	0	0	0	0	0	0	0	0
constellation	56	75	56	75	122	182	122	182
sphere	82	103	168	193	171	235	366	462
intricate	0	0	0	0	0	0	0	0
spectrum	71	74	71	74	151	165	151	165
dense	42	56	43	57	123	140	124	141
stretch	30	41	30	41	50	70	50	70
particle	35	35	35	35	35	35	35	35
denser	42	56	43	57	123	140	124	141
temple	115	115	115	115	330	330	330	330
forge	43	43	43	43	87	87	87	87
hexagon	40	40	53	53	75	75	101	101
fusion	11	18	17	24	32	66	61	95
emit	43	43	43	43	74	86	74	86
sculpt	0	0	0	0	0	0	0	0
pile	30	41	30	41	58	69	58	69
orbit	154	161	164	171	311	346	326	361
cosmos	86	140	161	215	191	399	443	652
faint	77	77	82	82	177	177	182	182
squash	19	19	19	19	36	36	36	36
photon	52	116	52	116	66	187	66	187
tide	12	62	26	83	18	140	65	212
moth	125	125	125	125	147	147	147	147
cosmic	62	62	161	207	282	286	665	828
symmetry	41	41	77	77	117	117	164	164
sulphur	77	77	77	77	150	150	150	150
manatee	143	143	143	143	228	228	228	228
Total	1531	1771	1924	2225	3179	3879	4288	5236
Median	43	59	52.5	69.5	119.5	140	123	144
sd	41.26	44.28	52.95	60.82	91.15	104.75	151.23	190.74
mad	38.55	27.43	37.81	46.70	88.21	100.08	89.70	93.40

Note. strg = strong; w = weak ; sd = standard deviation; mad = median absolute deviation. Values are in seconds.

Table 4.4 presents the descriptive statistics of contigduration for the target items. Data with and without weak visual referents and with and without related word forms are provided in 7 seconds and 25 seconds timespans. The medians for each measure are provided, with standard deviations given in parentheses and median absolute deviation underneath.

Meaning Recall

Eight GLM logistic regression models were built for meaning recall data set from 53 participants who watched two L2 captioned documentary series. The models sought to predict response accuracy from eight contigduration measures to consider the influence on contigduration estimates and improvement in fit from (1) the length of observed timespan, (2) the weak visual referents and (3) the related word forms (Table 4.5). The model in bold (Model 7) of contigduration predictor measured within 25 seconds timespans and inclusive of related forms and weak visual referents had the lowest AIC difference value ($\Delta_7 = 0.00$), followed by a similar model (Model 6) excluding weak referents ($\Delta_6 = 0.61$). All other candidate models had less support $6 \geq \Delta_i \geq 2$).

Table 4. 5

Summary of Model Results for Eight Measures of Contigduration as a Predictor of Meaning Recall

	Parameters				
Model	B(contigduration)	AIC_c	Δ_i	w_i	D
7 s					
1 $^{+ \text{weak}}$	0.22(0.20)	1634.32	5.48	.024	1403.31
2 $^{+ \text{related}}$	0.29(0.20)	1633.36	4.51	.039	1403.61
3 $^{+ \text{weak} + \text{related}}$	0.35(0.19)	1632.42	3.58	.063	1403.75
4	0.17(0.20)	1634.73	5.89	.020	1403.26
25 s					
5 $^{+ \text{weak}}$	0.42(0.19)*	1630.88	2.04	.136	1403.79
6 $^{+ \text{related}}$	0.48(0.18)**	1629.45	0.61	.278	1404.22
7 $^{+ \text{weak} + \text{related}}$	**0.50 (0.18)****	**1628.84**	**0.00**	**.377**	**1404.27**
8	0.34 (0.19)	1632.45	3.61	.062	1403.60

Note. Posttest ~ contigduration + spacing + (1|participant) + (1|item). Model fitted to 1484 observations across 28 words. N = 53. β = coefficient (standard errors are in parentheses); AIC_c = akaike information criterion corrected; Δ_i = change in AIC [AICi – AIC_{min}]; W_i = AIC weight; D = residual deviance (degrees of freedom: df = 1477). 7 and 25 s refer to the length of timespan within which contigduration was measured. $^{+ \text{weak}}$ = including weak referents; $^{+ \text{related}}$ = including related forms.

· $p < 0.10$. **$p < 0.01$.

The two models had a higher explained deviance and their parameter estimates revealed a significant positive increase contigduration effect on meaning recall accuracy. Based on its AIC weight, Model 7 was 1.4 times more likely to fit the data than Model 6. It can also be seen that Models 1 to 4, which included contigduration measured within only 7 seconds timespans, received no support. Also there was still support when weak referents were excluded (Model 6) or related forms were excluded (Model 5), but not when both were excluded (Model 8). Instead, a combination of both resulted in a considerable improvement in fit.

Meaning Recognition

The results list of the eight candidate contigduration models for meaning recognition accuracy according to differences in AIC is provided in Table 4.6. Despite strong positive effects for all contigduration predictors measured within 25 seconds, only those inclusive of related forms received considerable support. Model 7, inclusive of weak referents, fitted the data better than any other model and 33 times better than a similar model with contiguity measured within seven seconds timeframe.

Table 4. 6

Summary of Model Results for Eight Measures of Contigduration as a Predictor of Meaning Recognition

	Parameters				
Model	B(contigduration)	AIC_c	Δ_i	w_i	D
7 s					
1 $^{+ \text{weak}}$	0.27 (0.18)	1737.89	11.10	.002	1495.08
2 $^{+ \text{related}}$	0.39 (0.18)*	1735.25	8.45	.008	1495.12
3 $^{+ \text{weak} \ + \text{related}}$	0.45 (0.17)**	1733.80	7.01	.016	1495.08
4	0.23 (0.18)	1738.46	11.67	.002	1495.13
25 s					
5 $^{+ \text{weak}}$	0.52(0.17)**	1731.34	4.54	.055	1495.36
6 $^{+ \text{related}}$	0.61 (0.16)***	1727.56	0.75	.366	1495.69
7 $^{+ \text{weak} \ + \text{related}}$	**0.64(0.16)*** **	**1726.80**	**0.00**	**.534**	**1495.60**
8	0.45(0.17)*	1733.71	6.91	.017	1495.50

Note. Posttest ~ contigduration + spacing + (1|participant) + (1|item). Model fitted to 1484 observations across 28 words. N = 53. β = coefficient (standard errors are in parentheses); AIC_c = akaike information criterion corrected; Δ_i = change in AIC [AICi – AIC_{min}]; W_i = AIC weight; D = residual deviance (degrees of freedom: df = 1477). 7 and 25 s refer to the length of timespan within which contigduration was measured. $^{+ \text{weak}}$ = including weak referents; $^{+ \text{related}}$ = including related forms.

· $p < 0.10$. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Model 7 was followed by Model 6 ($\Delta_6 = 0.75$, without weak referents), which was 46 times better than a similar 7-second timeframe model. Both models showed a higher accuracy probability with increased contigduration.

Based on AIC weight, the model fit the data 1.5 times better if weak referents are included (i.e., Model 7 vs. Model 6) and 9.7 times better if related forms are included (i.e., Model 7 vs. Model 5). Remarkably, meaning recognition results revealed strong positive effects for contigduration predictor measured within 7 seconds timespans and inclusive of related forms. However, these models, along with all remaining candidates, had a Δ_i larger than 4.5.

Spoken Form Recognition

What stands out in the model selection results for spoken form recognition data which has a repeated-measure structure (Table 4.7), is the lower AIC_c estimates ($\Delta_i < 4$). Importantly, model 7 again fit the data better than the evaluated models, followed by Model 6 (without weak referents) with a difference just under 0.5.

Table 4. 7

Summary of Model Results for Eight Measures of Contigduration as a Predictor of Spoken Form Recognition

	Parameters				
Model	B (posttest:contigduration)	AIC_c	Δ_i	w_i	D
7 s					
1 $^{+ \text{weak}}$	0.16 (0.12)	3552.26	3.35	.048	3235.08
2 $^{+ \text{related}}$	0.21 (0.11)	3551.32	2.41	.077	3236.02
3 $^{+ \text{weak} + \text{related}}$	0.24 (0.11)*	3550.26	1.35	.130	3236.61
4	0.13 (0.12)	3552.71	3.80	.038	3234.78
25 s					
5 $^{+ \text{weak}}$	0.25 (0.11)*	3550.02	1.11	.147	3236.66
6 $^{+ \text{related}}$	0.26 (0.11)*	3549.27	0.37	.213	3237.32
7 $^{+ \text{weak} + \text{related}}$	**0.27 (0.11)***	**3548.91**	**0.00**	**.256**	**3237.45**
8	0.22 (0.11)*	3550.96	2.05	.038	3236.16

Note. Response ~ contigduration × time + spacing + (Time|participant) + (Time|item). Model fitted to 2968 observations across 28 words. N = 53. β = coefficient (standard errors are in parentheses); AIC_c = akaike information criterion corrected; Δ_i = change in AIC [AICi – AIC_{min}]; W_i = AIC weight; (D = residual deviance (degrees of freedom: df = 2956). 7 and 25 s refer to the length of timespan within which contigduration was measured. $^{+ \text{weak}}$ = including weak referents; $^{+ \text{related}}$ = including related forms.

· $p < 0.10$. *$p < 0.05$.

Support was still preserved following the exclusion of related forms from 25 seconds model (Model 5). For 7 seconds timespans models, the best fit was observed when both related forms and weak referents were maintained (Model 3), which also showed a significant increase in the predictor effect on spoken form recognition accuracy.

Written Form Recognition

The table below presents written form recognition data. In contrast to earlier selection results, the most parsimonious model was Model 8 that contained contigduration predictor measured within 25 seconds timespans without weak referents and related forms. To explain, adding 2061 seconds of weak visual referents and related forms to contigduration increased the AIC by $\Delta_i = 4.79$. That is, they were unimportant measures for the model to explain written form recognition data.

Table 4. 8

Summary of Model Results for Eight Measures of Contigduration as a Predictor of Written Form Recognition

	Parameters				
Model	B (posttest:contigduration)	AIC_c	Δ_i	w_i	D
7 s					
1 + weak	0.31 (0.13)*	3197.94	6.87	.018	2875.75
2 + related	0.30 (0.13)*	3197.69	6.62	.020	2875.73
3 + weak + related	0.27 (0.13)*	3197.76	6.68	.020	2875.49
4	0.32 (0.13)*	3196.86	5.79	.031	2876.19
25 s					
5 + weak	0.40 (0.12)***	3192.80	1.73	.236	2877.50
6 + related	0.31 (0.13)*	3195.41	4.34	.064	2876.38
7 + weak + related	0.26 (0.13)*	3195.86	4.79	.051	2875.76
8	**0.43 (0.12)*** **	**3191.70**	**0.00**	**.560**	**2878.14**

Note. Response ~ contigduration × time + spacing + (Time|participant) + (Time|item). Model fitted to 2968 observations across 28 words. N = 53. β = coefficient (standard errors are in parentheses); AIC_c = akaike information criterion corrected; Δ_i = change in AIC [AICi – AIC_{min}]; W_i = AIC weight; (D = residual deviance (degrees of freedom: df = 2956). 7 and 25 s refer to the length of timespan within which contigduration was measured. + weak = including weak referents; + related = including related forms.

· $p < 0.10$. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Research Question 1 Summary

Model selection was performed for each level of vocabulary measurement to find out which among eight contigduration measures provides a better-fitting model (within 7 seconds vs. 25 seconds timespans; inclusive vs. exclusive of weak referents; inclusive vs. exclusive of related forms). Contigduration measured within 25 seconds timespans came out as the most potent predictor in all the models, indicating that 25 seconds timespan was most opportune for incidental learning. The model that included contigduration measured within 25 seconds timespans and inclusive of weak visual referents and related forms was selected for upcoming predictive modelling. It produced the best fit (lower AIC and higher weight and deviance) for three levels of word knowledge measurement: meaning recall, meaning recognition, and spoken form recognition. The exception was the written form recognition results; the model that included contigduration within 25 seconds timespans without weak referents nor related forms was the top-ranked model; thus, it was selected for later written form recognition data analysis.

4.5.2 Research Question 2: Main

What is the effect of three verbal-visual contiguity measures: contigduration, contigfrequency, and contigratio on incidental vocabulary learning from extensive viewing of L2 captioned documentary series?

The previous section provided information that was necessary for upcoming analyses in this chapter. It answered the question of whether to choose verbal-visual contiguity measured within 7 seconds versus 25 seconds timespans, inclusive versus exclusive of weak referents, and inclusive versus exclusive of related forms. The results specified that contiguity measures within 25 seconds timespans provided a substantially better fit for the data. Inclusion of weak referents and related forms yielded the best fit for meaning recall, meaning recognition, and spoken form data, while excluding them yielded the best fit for written form recognition data.
Built on these findings, contigfrequency and contigratio were measured and presented in Table 4.9 along with contigduration values. To answer the second research question, the results for the four dependent measures are reported separately.

Table 4. 9

Descriptive Statistics for Contigduration, Contigfrequency, and Contigratio Variables for 28 Target words

	Meaning and spoken form			Written form		
Word	Contig-duration	Contig-frequency	Contig-ratio	Contig-duration	Contig-frequency	Contig-ratio
supernova	125	7	.500	125	7	.500
alien	0	0	.000	0	0	.000
constellation	182	8	.727	122	8	.727
sphere	462	32	.695	171	9	.600
intricate	0	0	.000	0	0	.000
spectrum	165	11	.916	151	10	.833
dense	141	14	.518	123	11	.458
stretch	70	10	.588	50	9	.529
particle	35	6	.500	35	6	.500
denser	141	14	.518	123	11	.458
temple	330	10	.909	330	10	.909
forge	87	5	.555	87	5	.555
hexagon	101	15	.937	75	9	.900
fusion	95	7	.437	32	1	.010
emit	86	6	.545	74	6	.545
sculpt	0	0	.000	0	0	.000
pile	69	7	.875	58	6	.750
orbit	361	29	.707	311	26	.650
cosmos	652	29	.617	191	15	.428
faint	182	8	.800	177	7	.777
squash	36	6	.750	36	6	.750
photon	187	14	.437	66	7	.219
tide	212	9	.529	18	1	.090
moth	147	11	1.00	147	11	1.00
cosmic	828	29	.617	282	9	.900
symmetry	164	16	.842	117	8	.727
sulphur	150	17	.944	150	15	.937
manatee	228	14	.999	228	14	1.00
Total	5236	334	17.46	3279	227	15.75
Median	144	10	.62	119.5	8	.58
sd	190.74	8.71	.28	91.15	5.49	.32
mad	93.40	5.93	.23	88.21	2.97	.28

Note. N = 28 words.

Reported in Table 4.10 are the coefficients of the fixed effects and random effects on response accuracy by 53 participants for meaning recall and recognition outcomes from a full model. The first column provides the change in the log odds of response accuracy associated with a unit change in contiguity predictors. A positive coefficient indicates that the predictor will improve accuracy, while a negative coefficient indicates that the predictor will impede it. In logistic regression models, the odds ratio can be used as an effect size statistic.

Meaning Recall

The full model estimating results for meaning recall showed that contigduration was the only significant predictor of a correct response (β = 0.99, SE = 0.34, Z = 2.94, p = .003) with an odds ratio of 2.68, indicating that the odds of recalling the meaning of one word was 2.68 times higher when the duration of the visual referent increased by one unit. There was no main effect for contigfrequency and contigratio. To follow up on the full model, three simple effects analyses were carried out to determine the significance of each contiguity predictor separately while maintaining the same structure as the full model. A main effect of contigduration was found, x^2 $(1) = 6.60$, p = .01, but neither contigfrequency, x^2 $(1) = 1.82$, p = .18, nor contigratio, x^2 $(1) = 0.47$, p = .49, predicted accurate recall, thus, were not examined further. These results are in line with the findings reported from the former model incorporating the three predictors. The odds ratio for contigduration (β = 0.49, SE = 0.18, Z = 2.75, p = .006) in favour of a correct recall was 63% higher (OR = 1.63) when duration was one unit longer, which is lower than the full model. Of particular concern is the sharp drop in the standard error (from 0.34 in the full model to 0.18 in the separate model), which implies that there could be indeed a violation of independence assumption in the full model that inappropriately inflated the estimates, making separate models more likely to be consistent.

Meaning Recognition

The full model estimating results for meaning recognition showed that both contigduration (β = 1.39, SE = 0.27, Z = 5.19, p < .001) and contigfrequency (β = −0.85, SE = 0.26, Z = −3.34, p = < .001) significantly predicted accuracy in meaning recognition test, while contigratio did not. The odds of recognising the meaning of a word (OR = 4) was 4 times higher with every one-unit increase in contigduration,

Table 4. 10

GLM Logistic Regression Predicting Meaning Accuracy from Contigduration, Contigfrequency, and Contigratio

	Meaning recall					Meaning recognition				
Parameters	*b*	*SE*	*z*	*p*	*OR*	*b*	*SE*	*z*	*p*	*OR*
Fixed effects										
Intercept	−0.89	0.26	−3.49	***	0.41	0.24	0.24	1.07	.285	1.27
Contigduration	0.99	0.34	2.94	**	2.68	1.39	0.27	5.19	***	4.00
Contigfrequency	−0.59	0.34	−1.71	.088	0.56	−0.85	0.26	−3.34	***	0.43
Contigratio	0.04	0.21	0.20	.843	1.04	−0.00	0.15	−0.02	.983	1.00
Spacing = massed	0.17	0.44	−0.39	.696	1.19	0.66	0.32	2.07	**	1.94
Spacing = spaced										
Random effects				Variance	*SD*				Variance	*SD*
Participant (intercept)				1.12	1.06				1.40	1.18
Item (intercept)				0.67	0.82				0.31	0.56

Note. Posttest ~ contigduration + contigfrequency + contigratio + spacing + (1|participant) + (1|item). Model fitted to 1484 observations across 28 words. N = 53.
***p* < .01. ****p* <.001.

whereas for a one-unit increase in contigfrequency, a 43% (OR = 0.43) decrease in the odds of a correct response was expected. The results contrast, however, with the findings from separate models' analyses: contigduration was the only significant predictor x^2 (1) = 13.19, p < .001, while there was no indication that contigfrequency, x^2 (1) = 2.78, p = .1, influenced recognition. Contigduration predicted recognition accuracy (β = 0.62 , SE = 0.16, Z = 3.92, p < .001, with an odds ratio that was 87% (OR = 1.87) higher, which greater than that of meaning recall (63%). It could be noted that contigfrequency had a marginal but non-significant positive effect (β = 0.30 , SE = 0.18, Z = 1.69, p = .09, OR = 1.35). No evidence for a contigratio effect was found in neither of the two analyses (i.e., via full and separate models) (ps > .95).

Spoken Form Recognition

Table 4.11 presents coefficients estimates on response accuracy from 3 separate models, one for each contiguity predictor, for spoken and written form recognition outcomes. The main effect of time on response accuracy was significant, x^2 (1) = 17.18, p < .001 but none of the main effects of contiguity predictors were (all $ps >$ .85). However, the addition of the two-way interaction between time and contiguity predictors contributed significantly to contigduration model, x^2 (1) = 5.72, p = .02, slightly but not significantly to contigfrequency model, x^2 (1) = 2.98, p = .08, and not at all to contigratio model, x^2 (1) = 0.12, p = .73; therefore, only contigduration data were further assessed. As hypothesised, the odds of recognising one spoken form after watching full-length documentary series were 30% (OR = 1.30) higher when the visual referent was one unit longer (β = 0.27, SE = 0.11, Z = 2.52, p < .01). This finding indicates that the influence of contigduration on spoken form recognition is less than its influence on meaning recall and recognition.

Written Form Recognition

Similar to spoken form recognition, the main effect of time on response accuracy was significant, x^2 (1) = 24.56, p < .001 but that of contiguity predictors was not (all $ps > .25$). The main effects were qualified by a significant interaction between time and contiguity predictor for only contigduration model, x^2 (1) = 12.31, p = .001, but not for contigfrequency, x^2 (1) = 2.78, p = .10 or contigratio, x^2 (1) = 0.67, p < .41, models. Looking at contigduration model estimates (β = 0.43, SE =

0.11, Z = 3.79, $p <$ .001), for each unit increase in the duration of the visual referent, participants were 53% (OR = 1.53) more likely to recognise the written form of the target word following their exposure to full-length documentary series, which is larger than the spoken form recognition (30%).

Table 4. 11

GLM Logistic Regression Predicting Form Recognition Accuracy from Contigduration, Contigfrequency, and Contigratio

Models	Contigduration					Contigfrequency					Contigratio				
Fixed effects	*b*	*SE*	*z*	*p*	*OR*	*b*	*SE*	*z*	*p*	*OR*	*b*	*SE*	*z*	*p*	*OR*
Spoken form															
Intercept	−0.46	0.24	−1.96	*	0.63	−0.46	0.24	−1.95	.051	0.63	−0.46	0.24	−1.88	.060	0.63
Contig predictor	−0.11	0.19	−0.56	.579	0.90	−0.03	0.19	−0.17	.868	0.97	0.01	0.21	0.05	.964	1.01
Spacing = massed	−0.23	0.43	−0.55	.580	0.79	−0.23	0.42	−0.55	.579	0.79	−0.25	0.47	−0.53	.596	0.78
Spacing = spaced															
Time = Posttest	0.63	0.12	5.08	***	1.88	0.63	0.13	4.89	***	1.88	0.63	0.13	4.72	***	1.88
Time = Pretest															
Time (Posttest) × contig predictor	0.27	0.11	2.52	*	1.30	0.20	0.11	1.78	.075	1.22	0.04	0.12	0.35	.730	1.04
Random effects				Variance	*SD*				Variance	*SD*				Variance	*SD*
Participant = Intercept				0.30	0.54				0.29	0.54				0.30	0.54
Participant = Posttest				0.21	0.46				0.21	0.46				0.21	0.46
Item = Intercept				0.90	0.95				0.91	0.95				0.92	0.96
Item = Posttest				0.12	0.34				0.15	0.39				0.19	0.44
Written form															
Intercept	−0.13	0.30	−0.43	.667	0.88	−0.15	0.29	−0.52	.605	0.86	−0.16	0.30	−0.55	.584	0.85
Contig predictor	−0.13	0.24	−0.53	.593	0.88	−0.32	0.24	−1.36	.173	0.73	−0.23	0.25	−0.91	.362	0.80
Spacing = massed	−0.34	0.54	−2.49	*	0.26	−1.24	0.53	−2.35	*	0.29	−1.19	0.55	−2.15	*	0.30
Spacing = spaced															
Time = Posttest	0.88	0.12	7.43	***	2.41	0.86	0.13	6.41	***	2.36	0.86	0.14	6.14	***	2.35
Time = Pretest															
Time (Posttest) × contig predictor	0.43	0.11	3.79	***	1.53	0.21	0.13	1.71	.087	1.24	0.08	0.13	0.63	.530	1.09

(Continued)

Random effects	Variance	*SD*	Variance	*SD*	Variance	*SD*
Participant = Intercept	0.46	0.68	0.46	0.68	0.46	0.68
Participant = Posttest	0.12	0.34	0.12	0.34	0.12	0.34
Item = Intercept	1.52	1.23	0.41	1.19	1.46	1.21
Item = Posttest	0.08	0.29	0.20	0.45	0.24	0.49

Note. Response ~ contiguity predictor × Time + spacing + (Time|participant) + (Time|item). The model fitted to 2968 observations across 28 words, N = 53.
*p < .05. ***p <.001.

4.5.3 Research Question 3: Exploring

What are the relative strengths of the three predictors: contigduration, contigfrequency, and contigratio of incidental vocabulary learning from extensive viewing of L2 captioned documentary series?

The final research explored the relative predictive strengths of the three predictors. This was addressed by comparing across the three models:
(1) the model goodness of fit, (2) the variation explained by contiguity predictor, and (3) the overall performance of the model based on 10 typical indices, including the already compared AIC_c and R^2 metrics. Important indices are shown in Table 4.12.

Table 4. 12

Summary of Model Results for Three Measures of Contiguity Predicting Accuracy of Meaning Recall and Recognition and Spoken and Written Form Recognition

Model	Δ_{aicc}	W_{aicc}	$R^2_{marginal}$	$\Delta R^2_{marginal}$	$R^2_{conditionall}$	ICC	BF	Score %
				Meaning recall				
C_1. Contigduration	0.00	.879	0.045	0.044	0.390	0.36	1.00	75.00
C_2. Contigfreq	4.78	.080	0.015	0.014	0.391	0.38	0.05	25.91
C_3. Contigratio	6.13	.041	0.005	0.004	0.392	0.39	0.09	25.00
				Meaning recognition				
C_1. Contigduration	0.00	.993	0.069	0.068	0.410	0.37	1.00	100.00
C_2. Contigfreq	10.40	.005	0.017	0.016	0.407	0.40	0.01	17.69
C_3. Contigratio	12.71	.002	0.003	0.002	0.406	0.40	0.00	00.00
				Spoken form recognition				
C_1. Contigduration	0.00	.761	0.027	0.004	0.310	0.29	1.00	79.15
C_2. Contigfreq	2.74	.193	0.026	0.003	0.310	0.29	0.06	34.92
C_3. Contigratio	5.61	.046	0.023	0.000	0.311	0.29	0.25	25.00
				Written form recognition				
C_1. Contigduration	0.00	.98	0.100	0.013	0.464	0.40	1.00	100.00
C_2. Contigfreq	8.18	.016	0.095	0.008	0.459	0.40	0.02	17.25
C_3. Contigratio	11.25	.004	0.092	0.005	0.459	0.40	0.00	0.78

Note. The model fitted to 1484 and 2968 observations for meaning and form data, respectively. N = 53. Contigfreq = contigfrequency. Δ_{aicc} = change in AIC [AICi – AIC_{min}]; W_{aicc} = AIC weight. $R^2_{marginal}$ = fixed effects R^2; $\Delta R^2_{marginal}$ = change in $R^2_{marginal}$ with and without predictor; $R^2_{conditionall}$ = fixed and random effects R^2. ICC = intraclass correlations coefficients; BF = Bayes factor. Score % = ranges from 0% to 100% and based on mean value of normalised AIC and R^2.

Model Performance Comparison Based on AIC_c

AICs measures were used to rank the contiguity predictors in terms of the relative strengths of their models. For each of the four vocabulary measurements, the model with the minimum AIC value was always contigduration (c_1), followed by contigfrequency (c_2), then contigratio (c_3).

For meaning recall responses, there was an 87.9 % (w_1 = .879) chance that contigduration was the best-approximating model describing accuracy, with the next highest weight being only 8 % for the contigfrequency model (w_2= .080). That is, model c_1 was about 10 (w_1 / w_2 = 10.99) and 20 (w_1 / w_3 = 21.44) times more likely to be the best-approximating model than model c_2 and model c_3, respectively, (this is known as the evidence ratio or the relative likelihood of a model versus another model). The evidence ratio for model c_2 versus model c_3 was only about 2 (w_2 / w_3 = 1.95), which is relatively weak support. The contigduration model fitted meaning recognition data 99.3 % (w_1 = .993) better than the other evaluated models with the next highest weight being only 0.5 % for contigfrequency model (w_2= .005), followed by 0.2 % for contigratio model (w_3= .002). In other words, model c_1 was almost 200 times (i.e., w_1 / w_2 = 198.6) more likely to fit the data than model c_2, and approximately 500 times (w_1 / w_3 = 496.5) better supported by meaning recognition data than model c_3. More precisely, models c_2 and c_3 had $\Delta_{aicc} > 10$, thus, were extremely implausible relative to model c_1.

For spoken form recognition data, model c_1 was only 2.74 AICc units from the second-best model c_2. This resulted in an evidence ratio for model c_1 versus model c_2 of only 4, which is similar to that for model c_2 versus model c_3 (w_1 / w_2 = 3.94, w_2 / w_3 = 4.20), and an evidence ratio versus model c_3 of about 17 (w_1 / w_3 = 16.54). Slightly similar to meaning recognition results, written form recognition data clearly supported contigduration model with an Akaike weight value of w_1 = .98, which is considerably higher (98 %) than contigfrequency model (w_2 = .016). That is, model c_1 was 61 times (w_1 / w_2 = 61.25) more likely to be the best-approximating model than model c_2. As for the contigratio predictor, model c_3 was implausible for written form recognition data for which Δ_3 value was 11.25.

Model Performance Comparison Based on R_2

In meaning recall results, the three contiguity models revealed a coefficient of determination of $R^2_{conditional}$= .39. That is, 39 % of the variance in response accuracy was explained by the overall model. Marginal delta R-squared value for contigduration predictor was only 4.4 % ($\Delta R^2_{marginal}$ = .044) but the highest relative to contigfrequency (1.4%, $\Delta R^2_{marginal}$ = .014) and contigratio (0.4 %, $\Delta R^2_{marginal}$ = .004). In addition, the R-squared value for meaning recognition models was $R^2_{conditional}$= .41, indicating that the models explained about 41% of the variation. A high variance of 6.8 % in meaning recognition responses was accounted for by contigduration ($\Delta R^2_{marginal}$ = .068), while the R^2 associated with contigfrequency (1.6%) and contigratio (0.2 %) were markedly lower.

For spoken form recognition models, the predictive power of 31 % ($R^2_{conditional}$= .31) was obtained. Results revealed that, respectively, a mere 0.4 %, 0.3 %, and 0 % variance in accuracy scores were attributed to contigduration, contigfrequency, and contigratio. The amount of variance in written form recognition responses that can be ascribed to contigduration was just 1.3 %, a little bit above contigfrequency variance (0.8 %). In comparison, contigratio accounted for only 0.5 % of the variation.

Intraclass correlation (ICC), a statistic that is commonly reported in multilevel analysis, is the fraction of the total variation in scores that is accounted for by the between-participants and the between-words variation. The results yielded slightly different values for meaning recall models (c_1 ICC = .36 , c_2 ICC = .38 , c_3 ICC = .39) and meaning recognition models (c_1 ICC = .37, c_2 ICC = .40, c_3 ICC = .40). However, spoken and written form models captured, respectively, 29 % (ICC = .29) and 40 % (ICC= .40) of diversity that is attributable to participants and words.

Model Performance Comparison Based on AIC_C and R^2

Taken together, model c_1 fitted to meaning recall data scored best (75 %, Table 4.12, based on AIC_c and R^2 estimates, while model c_2 and model c_3 scored a quarter (about the same, c_2 = 26, c_{23} = 25). Model c_1 had a complete score (i.e., 100%) on its performance fitting meaning recognition data, for which model c_2 (17.69 %) and model c_3 (0.00 %) provided markedly inferior fit. As for spoken form recognition responses, again, model c_1 had a high performance score of 79.15, with model c_2

being approximately 44 percentage points lower (34.92 %), while the latter was about 10 points higher than c_3 (25 %). Finally, Model c_1 had, again, a full score (i.e., 100%) on model performance fitting written form recognition data. As previously found, among the subset, only model c_2 additionally provided support to the data, and its overall performance was well under a fifth (17.25 %). In summary, the findings of the various metrics show a similar pattern in that contigduration was the strongest predictor of verbal-visual contiguity, while contigfrequency and contigratio were relatively insignificant.

4.6 Discussion

This study contributes new knowledge about the impact of imagery in L2 captioned video on incidental vocabulary learning. Its specific objective was to assess the effect of contiguity between words' verbal forms and their visual referents, what I termed "verbal-visual contiguity" on incidental acquisition of knowledge of meaning recognition and recall and spoken and written form recognition from extensive viewing of two full-length seasons of L2 captioned documentary series. Part of the research aim was to explore whether contiguity learning effects could still be observed when the construct is measured in timeframes extending beyond 5 seconds from the verbal occurrence. The research undertaken here adds to the rapidly expanding field by investigating, for the first time, the effect of verbal-visual contiguity as measured through contigduration (the amount of time a visual referent is displayed on the screen), contigfrequency (the subset of verbal occurrences that have visual occurrences), and contigratio (contigfrequency relative to the subset of verbal occurrences). Three additional innovations are the inclusion of two timeframes ($\mp$7 seconds, $\mp$25 seconds) that are longer than what has been previously observed, the extension of the length of exposure to eight viewing hours in two full-length seasons of documentary series, while also considering different parts of speech. No predictions were made as to which contiguity measure or contiguity timeframe would contribute to successful learning. The findings suggest that contigduration is reliably predictive of response accuracy in all measures of word knowledge, with the bigger impact found in meaning recognition, and that a longer contiguity timeframe of up to 25 seconds has the greatest potential to capture such an effect. The study also expands on past findings by generating insight into the effect of contiguity in incidental instead of explicit context.

The following section builds a discussion of the reasons why the contiguity timeframe that extends to 25 seconds was found to provide a better fit for the model. This will be followed by an illustration of the ways in which weak visual referents and related forms are believed to moderate verbal-visual contiguity. Next, I will address the second question, concerned with the effect of the three contiguity constituents in light of recent works, followed by an attempt to theorise the finding that contigduration is the only measurement that adequately represents verbal-visual contiguity.

Is the effect of verbal-visual contiguity on incidental word learning from extensive viewing of L2 captioned documentary series moderated by the length of the timeframe and inclusion/exclusion of weak visual referents and related word forms?

This first research question focused on the distinctive ways in which contiguity can be measured in terms of the length of contiguity timeframe, the quality of the visual referents, and the type of verbal forms observed. As a simplifying decision, the effects of the potential moderators of timeframe, visual quality, and word form were explored within contigduration as the dependent variable. The relative strengths of each level of the moderating variables were compared in models via Akaike weights and evidence ratios. The resulting model structures were then used in analyses of the other dependent variables.

The Length of the Timeframe

In the final part of his paper, Rodgers (2018) called into question "… whether there is a limit to the amount of time an image can be separated from the presentation of the aural form before vocabulary learning is no longer supported" (p. 205). The best fitting and most parsimonious model of verbal-visual contiguity in the present study modelled contiguity within a ∓25 seconds timeframe. This was the most robust finding to emerge from the first research question analysis for all four knowledge measures, based on the goodness of fit and coefficient of determination. The result supports the hypothesis that it is useful to consider a timeframe that exceeds 5 seconds to investigate contiguity learning effects in authentic video. In what follows, I will interpret this finding in light of the impact of repetition, particularly (1) potential visual Hebb effect across and within extended contiguity timeframe, (2) cross-situational word learning, (3) and conditional effect.

By way of illustration, the superior fit of the ∓25 seconds model was most evident in the meaning recognition data. The best model was 33 times better supported by the data than a similar model including contiguity measured within seven seconds. In fact, expanding the timeframe by 18 seconds (i.e., from 7 seconds to 25 seconds) increased the marginal explained variance from 3.5% to 6.9%. The second-best model fitted the data approximately 46 times better than a similar model modelling contiguity of seven seconds timeframe, with the marginal explained variance being augmented by 3.8 percentage points.

Notwithstanding, it is important to bear in mind that the analysis conducted here is exploratory and observational; thus, it does not answer Rodgers' question. It does not establish causality or allow us to draw strong inferences about which contiguity timeframe, among the two observed, are most closely associated with the contiguity learning effect. Nonetheless, the present results support an argument favouring the use of a span of time that is longer than 5 seconds to capture a greater extent of the potential contiguity learning effects.

The present results qualify the currently held arguments that the briefest contiguity timeframe (when word and referent occurrences are nearly simultaneous) is the most optimal for incidental vocabulary learning in the context of L2 captioned documentary series (or authentic video in general). Previous researchers evaluating contiguity were conservative regarding the length of the contiguity timeframe being observed (∓ 5 seconds ≥). In his descriptive study, Rodgers (2018) quantified verbal-visual contiguity in documentary series by capturing referents within two timeframes as short as ∓2 seconds and ∓5 seconds. Likewise, in her empirical investigation of the effect of this phenomenon, Peters (2019) utilised the ∓5 seconds timeframe to code referents in a procedure that was drawn from Rodgers's. Currently, only these two methods exist for the measurement of verbal-visual contiguity in authentic video. Rodgers based his criteria for selection on subtitling rules. The ∓5 seconds category was derived from the results of d'Ydewalle et al. (1987), which concluded that no subtitle or caption should appear for longer than six seconds. The ∓2 seconds category was based on the observation, from the same study, that two- and six-seconds presentation duration showed similar language processing levels.

The reported data are congruent with the previously reviewed studies that motivated the use of long timeframes. That is, ∓ 7 seconds timeframe (e.g., Bagget, 1984) and ∓ 25 seconds timeframe (e.g., Baddeley & Levy, 1971). Caution must be applied, however, for the obvious difference between the experimentally controlled presentation of stimuli in the STM literature and the variable distribution of the stimuli in authentic multimodal input. Essentially, extra durations marked from a prolonged contiguity timeframe (i.e., 25 seconds) better accounted for participants' accuracy scores due to repetition effects. For example, Peterson and Peterson (1959) found that STM of verbal items reached 18 seconds when rehearsal was prevented before recall, while Baddeley & Levy (1971) prevented rehearsal by blocking STM using a 20 seconds long distractor task. The two studies indicate that repetition is likely to extend the STM of stimuli beyond limits of 18 or 20 seconds, suggesting that contiguity learning effect can be achieved at long timeframes when repetition is involved. Here are some of the mechanisms behind this.

First, shorter timeframes might fail to account for a potential Hebb repetition effect. Immediate recall of a target sequence of stimuli improves if the same sequence is repeated, unannounced, with intervening filler sequences. This phenomenon in implicit learning is known as the Hebb effect (Hebb, 1961) and maintains for recall of verbal and visual stimuli (Hitch, Flude, & Burgess, 2009; Johnson & Miles, 2019). In contiguity terms, consecutive co-occurrence of forms with visual referent candidates can be conceived of as a sequence of target and filler trials that could be expected to elicit a visual Hebb effect. The shorter the timeframe, the less the captured trials at the level of the single timeframe, as well as across multiple timeframes.

In a similar vein, cross-situational learning (Yu & Smith, 2007) might have brought the positive effect observed for contigduration within 25 seconds timeframe. Along one contiguity timeframe lies a temporal relationship between verbal occurrences and between visual occurrences. To illustrate, a word occurring at 7 seconds of its referent could also have another referent at 15 seconds. Likewise, a referent occurring at 7 seconds of its word could also be a referent to the same word occurring at 15 seconds. A long timeframe allows to mark instances when a word consecutively appears with its proper and candidate referents within and across the timeframe. Recurrence of correct word-referent pairs over adjacent trials (situations)

is part of natural world setting, including documentary series: "… due to the nature of the genre, where viewers are educated in-depth on a topic which potentially leads to multiple occurrences of the related vocabulary" (Rodgers, 2018, p. 204). Such repetition increases the segregation of pairs, that is, the ability to distinguish repeated and unrepeated pairs, thus, promoting learning (Kachergis, Yu, & Shiffrin, 2009). "In the real world of real physics, there is likely to be considerable overlap between the objects present in a scene from one moment to the next" (p. 1709). The authors found that word-referent pairs achieved 30% and 66% accuracy if overlapped once and three times, respectively. Thereby, shorter timeframes might have masked the considerable degree of overlap between trials that aided correct inferences.

Third, the findings raise intriguing questions regarding the nature of the contiguity learning effect, specifically, whether contiguity learning effect for one verbal-visual occurrence is dependent on the presence and quality on the preceding fellow occurrences. A referent would exert a more powerful effect if it were in close proximity of fellow former referents or fellow verbal-visual co-occurrences. Therefore, the contiguity timeframe is possibly not absolute but conditional on preceding cumulative co-occurrences' frequency and functionality. To illustrate, a referent (R1) supposedly occurs at the first instance along with its form (i.e., perfect contiguity). A contiguity learning effect for a subsequent occurrence (R2) of the same referent may be more likely to emerge if it is fairly close to the form (e.g., at 10 seconds of its form) due to the limited prior occurrences available. On the other hand, a late referent (R7) may well produce a contiguity learning effect even if it is distant from the form (e.g., 20 seconds) due to the numerous preceding fellow occurrences which extend memory, especially if these possess a high level of strength. In sum, due to the absence of preceding occurrences, a contiguity learning effect might initially be elicited within a narrow timeframe for first encounters. However, reoccurrences may progressively increase the length of the contiguity timeframe of subsequent referents.

In sum, a contiguity learning effect was best captured at 25 seconds. The study's evidence implies an inevitable interdependence between the subtle effects of consecutive occurrences. The increased frequency of verbal occurrence appears to facilitate learning from viewing (Peters & Webb, 2018; Rodgers & Webb, 2019). In this study, increased verbal and visual occurrence (e.g., Hebb, overlapping, and

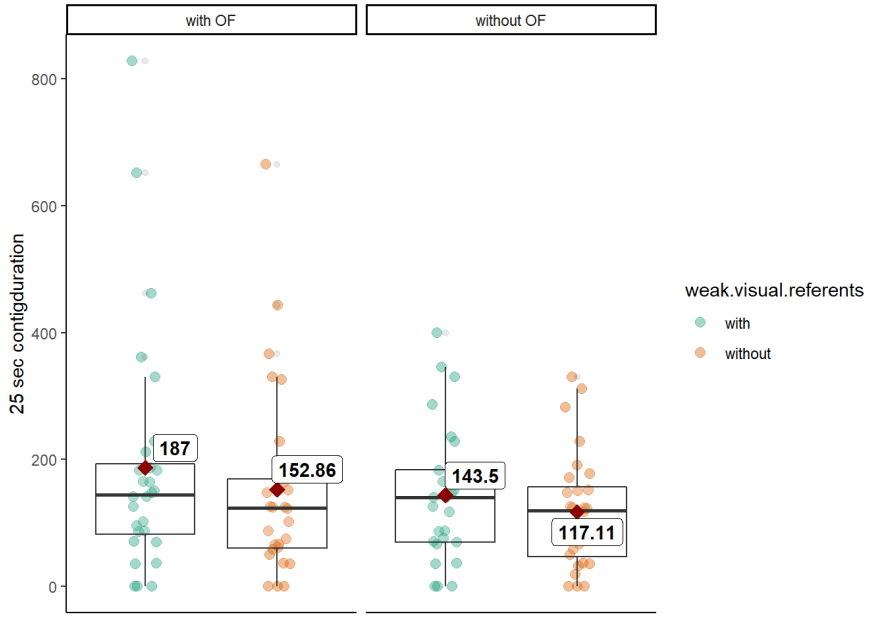
conditional effect) might have reinforced recognition. A drawback of a short timeframe is that it does not account for the successive encounters that are likely to enhance memory and support learning. Yet, the current analysis does not take us far in specifying the absolute limits of contiguity learning effects timeframe. A referent occurring at 25 seconds of its form may not entail that it is at the end of a contiguity learning effect spectrum. Such inference must be drawn from the presence of other forms and referents within or across contiguity spectra; it is all conditional.

The Quality of the Visual Referent

Weak referents accounted for about 18 % of the total visual duration (see Figure 4.10). Among the set of the compared contiguity models, the best fitting model was the one including weak visual referents. This result was found for meaning recall and recognition and spoken form recognition data. However, a similar model excluding weak referents displayed only a slightly worst fit to the data (meaning recall, $\Delta_{aicc} = 0.61$; meaning recognition; $\Delta_{aicc} = 0.75$; spoken form recognition, $\Delta_{aicc} = 0.37$) while also the effect of contigduration remained constant throughout. This held true for all models regardless of the length of timeframe or the type of the target word forms. The finding thus indicates that weak referents (i.e., 700 seconds ≤) had a marginal effect. The result shows that, irrespective of the referent quality, long contiguity durations predict learning. The lack of an adverse effect of weak referents, as was indicated by the absence of poor fit (i.e., high AIC) leads to various interpretations, including (1) the powerful depiction of imagery, (2) narrative saliency, (3) and familiarity.

Figure 4. 10

Mean Contigduration Measures (in seconds) Within ∓25 Seconds Timeframes, With and Without Other Word Forms and Weak Visual Referents



Note. Boxplots of contigduration for 28 target items in 8 hours of two documentary series. OF = other forms. For illustration, data are presented for 25s timeframe only, for both exact target words and other related forms. Medians are represented in the figure by the red points. Maximum contigduration = 828 s for the word *cosmic*. Large differences between points depict outliers and non-normal distribution.

First, the result could be attributed to the powerful capacities that pictures possess in depicting meaning. The visual referents were classed as weak for being thought to lack the required properties needed for accurate depiction, such as large size, high visual saliency, non-verbal signs, gestures etc. As Wollheim put it (1980, 2001, 2003), a weak image is one that does not necessarily lead its viewer to believe it is actually there; this could be due to its size or distance, for instance. Therefore, the results of this study perhaps reveal the need to appreciate the power of imagery, since even the smallest or ambiguous referent could carry with it powerful depiction. Second, recognition of weak referents must have been enhanced by narrative saliency. Such a claim can be substantiated by the strong performance found in the narration group in Study 1. Third, familiarity with weak referents due to high frequency is another contributing factor. Continuously encountering referents that are verbally contextualised might eventually transform a visual from a weak referent to a strong referent (one that is easily perceived and recognised). That is, these

visuals might have been immediately, but only temporarily, weak referents to the participant. As a visual referent becomes more familiar with every additional encounter, what was perceived weakly previously becomes gradually stronger.

Also, a referent was classed as weak for the ambiguity of the meaning it refers to (e.g., photon embodied in aurora, as seen in Section 4.4.2). The result could point out that participants might have perceived and recognised these images better than initially thought. Unfortunately, the follow-up debriefing did not look into this, and such an explanation remains speculative. Finally, the robustness of the coding criteria might have yielded the results and could hopefully lend itself well for use by future researchers.

In contrast to the findings of meaning recall and recognition and spoken form recognition, the model for written form recognition was not improved by including weak referents. Although the R^2 for a model without weak referents was (4.1%) lower than the R^2 for the inclusion model (8.6%), the relatively low values of AIC indicated that the higher explanation of variance was not necessarily robust. The result indicates that weak visual referents were not critical for predicting written form recognition data. This is not surprising given the ease with which written forms can be recognised compared to other aspects of word knowledge. Captions afford instant recognition of words' spelling. On the other hand, a visual referent depicts what a word refers to rather than how it is spelt, providing only an association effect for form learning to take place. This combination of findings further supports a conceptual premise that, among many aspects of word knowledge, pictures have the strongest effect on learning meanings.

Related Forms

Due to uncertainty as to whether receptive knowledge of the basewords extends to that of their derivatives (Laufer & Cobb, 2019; Kremmel 2016), the current study adopted the flemma as the main word counting unit while attempted to evaluate the potential impact of including and excluding compounds and derivatives in this count.

The best-supported model was the one in which the contiguity predictor includes words that are compounds and derivatives of target words, irrespective of the length of timeframe or the quality of images. This finding was true for meaning

recall and recognition and spoken form recognition, but it was most evident in meaning recognition results; the best model was approximately 10 times better supported by the data than a similar model excluding these related word forms. The results strongly support the presence of related word forms in the count of verbal frequency of words. For written form recognition results, however, maintaining related word forms was poorly supported compared to the model with a contiguity predictor excluding them. The results show that the model does not need compounds and derivatives as a counting unit to explain written form recognition data.

The models posit that the predictability of participants scoring accurately on a target word as a result of long contiguity duration is conditioned by the account or not of the duration of related forms' visual referents. According to the data, this account of extra durations played a positive role in predicting meaning recall and spoken form recognition, whilst it had the strongest impact on meaning recognition and went well against the prediction of written form recognition response.

This finding supports my earlier view of a possible moderating effect (requiring careful attention) of compounds and derivatives' occurrences on the association between verbal frequency and learning, hence, verbal-visual contiguity and learning. The result is in accord with a wealth of evidence corroborating the notion that prior linguistic knowledge, including morphological awareness, contributes to learning new words. For instance, although James' word-neighbour manipulation on children and adults was pertinent to the influence of local neighbourhood and not derivatives, she showed that "having one related word-form in vocabulary may be sufficient to facilitate recall of a new word" (2019, p. 110) based on an experiment with adults. Researchers are not aware of whether their participants fully master the morphological patterns of the English language. Including or excluding related word forms can, therefore, be interpreted as either an overestimation or underestimation of the learners' knowledge. As a result, an essential methodological implication of my result for future experimental vocabulary research on verbal encounters is that more information on target words' reoccurrences via other forms would help establish a greater degree of accuracy and thus provide more definitive evidence.

Interestingly, the study raises the possibility that contigduration of related forms exerts a moderating influence that varies according to aspects of word knowledge. These results relating to verbal-visual encounter accord with previous observations on verbal encounters suggesting that the required number of exposure in order for word learning to take place is knowledge dependent (e.g., reading, Pellicer-Sanchez & Schmitt, 2010). For instance, drawn from extensive data, meaning recall is now generally accepted as the most difficult knowledge to acquire (e.g., Pellicer-Sánchez, 2016). In addition, the current data suggest that repeated verbal-visual exposure to related forms reduces the learning burden at the semantic level more than at the form level, in contrast to recent findings on verbal encounters in authentic texts (cf. Godfroid et al., 2018). This is, however, unsurprising given the differences in the modality of occurrence. Visual referents contribute to the acquisition of meaning more than form because, as the name implies, they represent the primary input source of meaning the word refers to and provide no additional information to how it is pronounced or spelt; hence, the result. Relevant to this result is the set of evidence available that cross-situational word learning technique is inextricably linked with learning the meaning of words (e.g., Berens, Horst, & Bird, 2018; Hendrickson & Perfors, 2019; Vong & Lake, 2020). Moreover, written forms are plainly visible from the captions. Consequently, they can be immediately recognised, which in turn clarifies the detrimental impact the related forms had on the predictability of the written form models.

In addition, the solid positive result found in meaning recognition can be explained by the massive exposure to visual referents that are exclusive to related forms. For example, at the 25 seconds contiguity timeframe, derivatives and compounds accounted for a duration that is up to 1361 and 1109 seconds, if weak referents were included and excluded, respectively, which might have improved model performance.

What is the effect of three verbal-visual contiguity measures: contigduration, contigfrequency, and contigratio on incidental vocabulary learning from extensive viewing of L2 captioned documentary series?

To the best of our knowledge, the prevalence of verbal-visual contiguity in documentary series has only been explored in Rodgers (2018) and its effect on incidental L2 vocabulary learning has been partially investigated in three other studies (Ahrabi et al., 2021; Peters, 2019; Pujadas Jorba, 2019). None of the experimental studies have taken into consideration contigfrequency, contigduration, and contigratio as measures of verbal-visual contiguity. For all four dependent measures of vocabulary knowledge, the second research question revealed that, with good certainty, a contiguity learning effect is principally seen when operationalised as contigduration but diminishes and loses significance when it is operationalised as contigfrequency or contigratio. This result signals that the novel contiguity construct of contigduration significantly contributes to current research and is likely an improvement on previous operationalisations.

Only contigduration was positively associated with response accuracy on form and meaning tests. While contigfrequency and contigratio accounted for no variation in response accuracy, the duration of visual referents that were in proximity to verbal occurrences remained a significant predictor of all vocabulary outcomes. With every one unit increase in the on-screen duration of the visual referent, words meanings were 63% more likely to be recalled (SE = 0.18) and 87% more likely to be recognised (SE = 0.16).

The present study used a parsimonious model to obtain efficient estimates as indicated by the low standard errors. Controlling for more word-related covariates caused these to increase to as large as 0.28 (which is a commonly reported figure in empirical studies) and increased the odds of accuracy for meaning but not for form results. This further analysis showed that, for each additional unit of presentation duration, participants were two times (2.47) more likely to recall and three times (3.46) more likely to recognise the meaning of words. The larger standard errors indicate decreased certainty about the estimates.

The superior scores for meaning recognition support earlier observations that retrieving information is more demanding. For instance, Pellicer-Sánchez (2016) maintained that meaning recall is much more challenging to acquire than receptive knowledge of meaning as measured through recognition. In the form recognition tests, the odds favouring a correct spoken form response were 30% (SE=0.11) higher while the odds of a correct written form response were 53% (SE =0.11) higher when duration was one unit longer.

The results fulfil Rodgers' (2018) recommendation to replicate the analysis of verbal-visual contiguity in an authentic video and empirically investigate a potential effect of the resulting measures on incidental L2 vocabulary learning. As he noted specifically, "in a documentary, the same referent is likely to be talked about for a prolonged period, thus providing repeated encounters with the same [word] all the while showing the referent on-screen" (p. 205). The current study experimentally demonstrated the potential effect of the referent's duration, which is an important characteristic of imagery. Despite the inconsistency between my study and that of Peters (2019) and followers, the present finding, in general, seems to concur fairly well with hers in that both highlight just how vital imagery is. Nonetheless, the findings here offer novel empirical evidence for the effect of verbal-visual contiguity as measured by contigfrequency, contigduration, and contigratio.

The study by Peters (2019) showed that words with visual referents were almost three times more likely to be learnt incidentally compared to words without visual referents, for both form recognition (SE = 0.25) and meaning recall (SE = 0.14). However, given the different experimental methodologies used, direct comparisons of the two studies can be difficult to make. The present study differs from Peters' in that its results are drawn from watching 8 hours of two full-length documentary series, in contrast to Peters' which used a short excerpt of 12 minutes. Another fundamental methodological distinction lies in the length of the observed contiguity timeframe. This study's timeframe was 20 seconds longer than Peters'; hence, differing results should not be surprising. Most basically, when attempting to define and explain the construct of verbal-visual contiguity, Peters examined as a variable of interest whether a word is accompanied by a visual referent, regardless of

how often it appeared. Clearly then, none of the continuous contiguity variables observed in my study relate to the categorical variable analysed in Peters'.

Given the restricted use of a categorical variable based on whether the target word did or did not have at least one visual referent, it is not inconceivable that chance might have also played a role in Peters' results. To explain, ignoring the frequency of visual occurrences using binary coding leads to a loss of information, and it is unknown what results would have been obtained with a different measure of contiguity. Vocabulary researchers study the verbal frequency of occurrence when predicting learning as a function of verbal encounters. In the same vein, predicting learning as an outcome of contiguous visual referents requires examining the verbal-visual frequency of occurrence.

In addition, greater exposure to language may play a role in increasing comprehension, hence, learning. Although the present thesis' participants were university students, Peters had Dutch-speaking secondary school students. These are known for speaking better English and for having familiarity with the English language captioned TV. This implies that they may have had an advantage over participants in Algeria in which English language is not spoken outside the confines of school.

The current results perhaps indirectly contribute to a growing body of evidence demonstrating the major effect of L2 captions on learning written form (Aini et al., 2018). While no interaction was found between captions and imagery in Peters' study, the current data come from exposure to L2 captioned video, implying that the lower estimates for written form recognition are more likely to be due to the presence of captions. Further, the study broadly supports findings from previous eye-tracking research into L2 captioned video which pointed out the effectiveness of subtitle processing (e.g., Perego et al., 2010). To illustrate, the current meaning and form results might indicate that participants' acquisition of words occurs as a function of many sources of input, depending on the type of word knowledge (i.e., images for meaning, captions for form). This accords with previous observations that watching L2 captioned video does not necessarily impose a trade-off between the processing of captions and images.

Another possible explanation for the much lower outcomes obtained for form recognition in the present study relative to Peters may be the differences in the test format. While Peters adopted a yes/no combined test in which the target form is heard and read simultaneously, in my study, knowledge of written form and spoken form were measured via two separate multiple-choice tests. Thus, participants were more likely to exhibit some difficulty in answering. Also, the multiple-choice test was inclusive of "I don't know" option, which has been found to minimise guessing and partial knowledge (Zhang, 2013). Moreover, initial processing of target items before the experiment was strictly reduced through multiple means: by not testing meaning knowledge, by projecting the written form to a screen; by exposing items one at a time, and by interleaving 28 filler items to 28 target words, in contrast to Peters whose 36 items test were filled with only eight filler items. Though the one-week interval she implemented had surely lessened any testing effect, an interval of almost two weeks was used in the present study.

Another interesting contrast between Peters and the present study is her finding of an equal contiguity learning effect for form and meaning, contrary to the current data, which shows a clear advantage of contigduration on acquiring knowledge of meaning when compared to knowledge of forms. As Godfroid et al. (2018) maintained, form is "… perhaps a more shallow type of word knowledge that can be picked up more easily through simple repetition and implicit learning mechanisms" (p. 36). Thus, such difference in results of different aspects of word knowledge is, as indicated previously, quite reasonable.

The present study implemented deception in that the recruited university students were made to believe that they were participating in an experiment on comprehension and that the vocabulary pre-test was a placement test. Therefore, it is reasonably safe to claim that the learning that took place was incidental to exposure to the multimodal input, particularly to achieving comprehension, and not to the desire to score well in an anticipated test. While Peters only hid information about the upcoming delayed test, she perhaps considered such risk as negligible for secondary school students who are not alerted to experimental procedures. Thus, should not necessarily be treated with the utmost caution as with university students.

In summary, the second research question showed that, the effect of verbal-visual contiguity on incidental vocabulary learning from L2 documentary series is mostly manifested in contigduration. Contigfrequency and contigratio, on the other hand, did not show significant effect on learning.

What are the relative strengths of the three predictors (contigduration, contigfrequency, and contigratio) in incidental vocabulary learning from extensive viewing of L2 captioned documentary series?

The findings of Research Question 3 have concretely demonstrated that the different operationalisation of verbal-visual contiguity as contigduration, contigfrequency, and contigratio account for different amounts of variance across dependent variables. Contigduration was found to be the contiguity dimension that accounts for the greatest amount of variance. That is, incidental acquisition of knowledge of words from L2 captioned documentary series changes not as function of how often a verbal occurrence had visual occurrences (i.e., contigfrequency) but rather of the amount of time they occurred during the adopted contiguity timeframe. Finally, longer durations of visual referents aided learning of both word form and word meaning but were especially beneficial for learning meanings.

Based on Akaike estimates, the contigduration model fit the data better than the other evaluated models. The study indicated that, concerning meaning results, the contigduration hypothesis was, respectively, 10 and 20 times more likely than contigfrequency and contigratio for recall outcomes but 200 and 500 times more likely for recognition outcomes. With respect to form results, however, contigduration model was about 4 and 17 times as probable for spoken form recognition and 61 and 245 times for written form recognition. Analysis of the coefficient of determination interestingly corroborated these results. Contigduration accounted for 4.4 % of the variation in meaning recall responses, 6.8 % in meaning recognition, 0.4 % in spoken form recognition, and 1.3 % in written form recognition. On the other hand, the highest variation that contigfrequency and contigratio could explain was, respectively, 1.6 % for meaning recognition and 0.5 % for written form recognition data. Overall, there is a clear relationship between the duration of words' visual referents on the screen and the ability to score well in the vocabulary tests. Although only contigfrequency was a competitor, it remained

consistently insignificant relative to contigduration, while in none of the analyses was any importance for contigratio found.

The finding that contigduration is the most potent predictor of score accuracy is substantiated by two facts. First, it was interesting to find that the contigduration hypothesis was 500 times more likely than the contigratio hypothesis for meaning recognition. This result further underscores the plausibility of the contigduration model. Second, contiguity involves a linear relationship between contigfrequency and contigduration, as evidenced by the high Pearson correlation coefficient of 0.80. Thus, the result that only contigduration was associated with word learning has strengthened the confidence in contigduration as a strong predictor of learning. The results are explained in terms of eye fixations and variance in informativeness.

The current finding is consistent with and has important implications for studies of eye-tracking in L2 captioned video. The finding that contigduration is the only contiguity predictor that is significantly associated with incidental acquisition of words is explained by the link it has with eye fixations on the visual referent. In other words, the longer the duration of a visual referent, the more time available for the learner to fixate on the referent. Such fixation "… is needed to perceive, identify and encode objects and entities into memory" (Conklin, Pellicer-Sánchez, & Carrol, 2018, p. 113).

Another explanation for the significance of contigduration could be that contiguity measure is more informative than contigfrequency and contigratio. On the one hand, the contigfrequency measure was of limited information; it accounted only for the number of instances a verbal occurrence was accompanied by one or several visual occurrences, irrespective of how long the visual referent(s) stayed on the screen. On the other hand, there was greater variability in contigduration measure in terms of duration and the disparity in visual referents of a single word (as a result of variability in episodes). Therefore, in accordance with cross-situational learning theory, the contigduration measure captures more the learner's opportunities to disambiguate the correct word-referent mapping and isolate the item meanings in the presence of other visual or verbal candidates. This account also implies that the contigfrequency predictor might have gained in explanatory strength if the words' durations varied less.

In summary, several contributing factors must have enabled incidental acquisition of knowledge of the target items. Among the contiguity factors set out in this study, only contigduration was shown to be significant. From a theoretical perspective, contigduration comes out ahead of contigfrequency for being more informative. It has greater explanatory power in describing the opportunities available for the learner to disambiguate and reinforce correct word-referent pairings.

4.7 Conclusion

While Study 1 raised doubts on the importance of imagery in developing word knowledge, Study 2 helped dispel these. Incidental vocabulary learning from extensive viewing of L2 captioned documentary series is constrained by several intrinsic and extrinsic properties of the words in the video material. Among the critical extrinsic properties is the frequency of verbal encounters, which has been shown to be positively associated with vocabulary learning. Simultaneous processing of verbal forms and their corresponding visual referents (i.e., contiguity) exerts a greater influence on vocabulary learning; however, support for this claim is based mainly on studies oriented to materials developments and is difficult to find in studies on incidental vocabulary learning from authentic videos. Recently, verbal-visual encounter of words in a video at a 5-second timeframe was found to have a facilitative effect on vocabulary learning. Nevertheless, the finding's generalisability was limited by the use of a short documentary excerpt and shortcomings in the operationalisation of contiguity, indicated by the use of a dichotomous variable which may well have masked much of the variability in the predictor under analysis.

The current study is pertinent to extensive viewing of documentary series and carefully controls contigduration, contigfrequency, and contigratio as three different measures of verbal-visual contiguity, on two potentially useful contiguity timeframes: 7 seconds and 25 seconds. The results provide further insight into contiguity in this study area, indicating that the duration of verbal-visual encounters with the target word, at a maximum contiguity timeframe of 25 seconds is predictive of incidental vocabulary learning from two full-length seasons of documentary series. This result is clearly marked for knowledge of meaning recognition. The

study provides the novel evidence that the association of verbal-visual contiguity learning effects with incidental vocabulary learning from documentary series should not necessarily be limited to the number of co-occurrences of word forms with visual referents (i.e., contigfrequency). Instead, it is more associated with the length of the duration of these referents (i.e., contigduration). In sum, incidental acquisition of words is characterised by the possible involvement of several contributing factors. When many of these come into play, the verbal-visual co-occurrence in itself may have little or no effect on learning as compared with its duration. In contrast, operationalising contiguity as contigratio is not associated with incidental vocabulary learning.

It is worth noting that the current study is based on eight-hour viewing of two BBC documentary series. It is not certain whether contigduration would show a similar effect in other materials. Support for this assumption could be found in the latest developments regarding image memorability (e.g., Bylinskii, Isola, Bainbridge, Torralba, & Oliva, 2015; Isola, Parikh, Torralba, & Oliva, 2011; Isola, Xiao, Parikh, Torralba, & Oliva, 2013). Accordingly, variation in the memorability of images was found to be consistent across subjects, suggesting that independent of the observer, certain images are intrinsically more memorable, hence, more likely to improve prediction.

Another challenge for future research is to investigate the maximum amount of time a word can be separated from its referent before learning is no longer successful. Doing so will be difficult due to the variable distribution patterns of verbal and visual occurrences in authentic video. As previously discussed, the verbal-visual contiguity timeframe for vocabulary learning in videos is not absolute; it depends on the number of prior visual or verbal-visual occurrences. The more a referent is preceded by fellow occurrences, the longer its contiguity timeframe. Hence, future studies on this particular issue are perhaps required to address it through an experimental manipulation by which temporal proximity between verbal forms and between visual referents is held constant both within and across all contiguity spectra under investigation, which appears to be difficult to design. It should also be noted that little variance was explained by the simplified models in this study. Controlling for additional word-related covariates such as concreteness or

parts of speech might well explain variability in response accuracy beyond what is captured by the simplified contiguity model.

Chapter 5

Study 3. Spacing Effects in Learning

The two previous studies in this thesis have looked, in two different ways, at the potential effects of imagery in L2 captioned documentary series on incidental L2 vocabulary learning from extensive viewing. Study 1 (Modality effects in learning) employed a between-subjects design in which participants watched two full seasons of documentary series, either with multimodal input (image, text, sound) or with bimodal verbal input (text, sound). No significant difference was detected in overall learning gains between the two conditions. Study 2 (Contiguity effects in learning) employed a within-subjects design to test whether three measurable dimensions of verbal-visual contiguity (contigduration, contigfrequency, or contigratio) influenced learning gains within the viewing group. Participants were found to be two times more likely to recall and three times more likely to recognise word meaning when contigduration is one unit longer.

In addition to modality and verbal-visual contiguity, another factor that is likely to influence incidental L2 vocabulary learning from L2 captioned video is the spacing of the target items. In the two studies summarised above, documentary episodes were spaced over four sessions at about two-week intervals, and verbal occurrences of target learning items were either spaced over the sessions or massed in one single session. This raises the question of whether the observed vocabulary gains in the two experimental groups had also changed due to the distribution pattern of target words occurrences (across vs. within sessions). Studies of spacing in vocabulary research have primarily focused on explicit teaching as opposed to incidental learning. More importantly, research on the subject has been limited to unimodal input (reading) and bimodal input (listening-while-reading). The question of whether spacing contributes to learning from considerable multimodal input, mainly input provided from extensive viewing of documentaries, has yet to be examined. The present study has two aims. The primary aim is to investigate whether spaced repetitions of words in two full-length seasons of documentary series

facilitate incidental vocabulary acquisition. Specifically, to test whether words repeated across spaced diverse episodes will be better learnt than words repeated within one or two massed episodes, at four levels of word knowledge: meaning recall and recognition, and spoken and written form recognition. The second aim is to assess whether the effect of spacing conditions varies as a function of the presence of imagery (View vs. Non-View conditions).

The present study thus asks whether a word is more effectively learnt if verbal encounters are successively situated in a two-episode session (massed) or when encounters are spread over multiple two-episode sessions (spaced). The potential answers to this question derive from the spacing literature which provides strong evidence of an advantage for spaced presentations. In experimental psychology, the spacing effect, or the distributed practice effect ". . . refers to the fact that for a given amount of study time, spaced presentations yield substantially better learning than do massed presentations . . ." (Dempster, 1988, p. 627). But does the advantage of spacing extend to incidental learning from extensive viewing, particularly documentary series?

Throughout this chapter, the terms spacing effect and massing effect refer to the spacing advantage and massing advantage, respectively. Spacing conditions is used as an umbrella term for both conditions. The first section aims to give historical weight to the spacing phenomenon in learning and an account of the most prominent explanatory mechanisms behind its effect. The second section will provide a review of previous research on spacing effects in L2 vocabulary learning. This will be followed by a discussion of the few classroom-based studies that have been carried out under incidental conditions, and which have tended to focus on reading. The final main section will point to the lack of empirical evidence for the effectiveness of spacing in learning from extensive viewing, and the need to compare differences in incidental learning of words repeated across and within viewing sessions.

5.1 Distributed and Massed Occurrences Across and Within Extensive Documentary Viewing Sessions

5.1.1 The phenomenon

The effect of spacing has long been a question of great interest in the field of learning. The earliest account of this phenomenon dates to 1885 (reprinted in 1964) when Ebbinghaus put forward the view that "with any considerable number of repetitions, a suitable distribution of them over a space of time is decidedly more advantageous than the massing of them at a single time" (p. 89). His view was based on his finding that the distribution of 38 repetitions of 12-syllable series over three days produced an effect that was equal to that of introducing 68 repetitions successively. The finding was later substantiated by Jost's work, in 1897, who proposed that "if two associations are of equal strength but of different age, a new repetition has a greater value for the older one" (McGeoch, 1943, p140) – what came to be known as the Jost's law. Around the 1900s, research and case studies about spacing and distribution practice in learning through verbal memory tasks began to emerge (Dempster, 1988; James, 1901; Perkins, 1914; Pyle, 1913; Dearborn, 1910; Donovan & Radosevich, 1999; Edwards, 1917; Glenberg, 1979; Glenberg & Smith, 1981; Greene, 1989; Hintzman, 1976; Melton, 1967, 1970; Peterson, Wampler, Kirkpatrick, & Saltzman, 1963; Starch, 1912; Woodworth, 1938)

Various theories have been proposed to understand the underlying mechanism of spacing effects that accounts for a relation between learning and spacing. Two theories have been the most promising in the literature. The first rests on a deficient processing assumption (e.g., Bregman, 1967; Callan & Schweighofer, 2010; Challis, 1993; Cuddy & Jacoby, 1982; Dellarosa & Bourne, 1985; Gerbier & Toppino, 2015; Greene, 1989; Greeno, 1970; Hintzman, 1976; Jacoby, 1978; Johnston & Uhl, 1976; Krug, Davis, & Glover, 1990; Marquardt & Snee, 1975; Pavlik Jr & Anderson, 2005; Zechmeister & Shaughnessy, 1980). Proponents of this theory hypothesised that the negative effect of massed repetitions is associated with deficits in attention that are likely to result from successive encounters. Extremely small spacing between repetitions will likely breed a sense of instant familiarity with the item. This comfort of familiarity that the subject enjoys will adversely create an illusion that the item is already learnt, which in turn leads the subject to devote less attention to the item, whose "… presentation may still be activated in short-term

memory or readily accessible" (Koval, 2019, p. 1106). The second theory is known as a contextual (encoding) variability hypothesis (Bower, 1972; Gartman & Johnson, 1972; Glenberg, 1976, 1979; Greene, 1989; Landauer, 1975; Maddox, 2016; Madigan, 1969; Melton, 1970; Raaijmakers, 2003; Sobel, Cepeda, & Kapler, 2011). Based on this hypothesis, the context surrounding spaced learning is subject to change over time, from one session to another (e.g., activities, instructor's intonation or clothing, weather, unique situations, special circumstances etc.). This change in context augments the number and variety of contextual cues available to the learner to encode the item for later retrieval, in what could be seen as an associative form of learning.

Over the past twenty years, studies addressing the spacing phenomenon have continued to corroborate the robust evidence found in initial research (e.g., Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Delaney, Verkoeijen, & Spirgel, 2010; Gerbier & Toppino, 2015; Kornell & Bjork, 2008; Lotfolahi & Salehi, 2017; Pavlik Jr & Anderson, 2005; Rohrer & Pashler, 2007; Serrano, 2011; Suzuki & DeKeyser, 2017). In particular, one meta-analysis of more than 100 years of spacing in verbal learning research demonstrated that about 95 percent of 271 comparisons of retention performance showed that distributed practice generated more accurate final-test results relative to massed presentation (Cepeda et al., 2006). The literature on spacing phenomenon is substantial on what it is, what it does, as well as why it does it. Though the findings have been obtained from various perspectives, the current chapter expands the review on only one aspect: second language vocabulary acquisition. Based on learning context, the scope of spacing research in the field of second language vocabulary acquisition can be divided into two main categories: intentional and incidental.

5.1.2 Intentional L2 Vocabulary Learning

There has been considerable research on spacing effects within L2 vocabulary research. What we know about the topic is largely based on findings from deliberate decontextualised learning (e.g., Alfotais, 2019; Bahrick & Phelphs, 1987; Bloom & Shuell, 1981; Bolger & Zapata, 2011; Callan & Schweighofer, 2010; Goossens, Camp, Verkoeijen, Tabbers, & Zwaan, 2012; Kang, Lindsey, Mozer, & Pashler, 2014; Karpicke & Bauernschmidt, 2011; Kornell, 2009; Küpper-Tetzel, Erdfelder, &

Dickhäuser, 2014; Lotfolahi & Salehi, 2017; Nakata, 2015; Nakata & Suzuki, 2019; Nakata & Webb, 2016; Pashler, Zarow, & Triplett, 2003; Pavlik Jr & Anderson, 2005; Schuetze, 2015). Publications that concentrate on intentional learning have most frequently adopted the paired-associate paradigm where participants are instructed to memorise the form and meaning of target words and the majority of these studies have clearly supported a positive effect of spacing on deliberate learning.

Before I proceed to give an account of the current literature, I note that although seemingly similar, spacing studies in vocabulary research can be distinguished theoretically based on the study's design. Two approaches to research the effectiveness of spacing on vocabulary learning can be identified in the literature: individual items versus repeated occurrences of individual items. The individual items' approach focuses on the distinction between the distribution of multiple vocabulary items over many spaced sessions and the massing of the vocabulary items within one single session. For example, the teaching of "abstract" and "experiment" in a first session, then the teaching of "endeavour" and "success" in a second session as compared with teaching all the four items in a single session. The second approach, which this study is concerned with, focuses on the distinction between the distribution of repeated occurrences of an item over many spaced sessions and the massing of repeated occurrences of the item within one single session (e.g., Nakata & Elgort, 2020). For example, the occurrence of an item six times in three spaced sessions (i.e., 2 + 2 + 2) compared with the occurrence of another item of equal difficulty six times within a single session. Studies of this type are fewer in number. Therefore, studies of the first approach will serve as a good first pass for this review.

Among the recent investigations into spacing effects on intentional L2 vocabulary learning showed a benefit for spaced learning over massed learning irrespective of word class or participants' preference for a condition over the other (Alfotais, 2019). In a within-subjects experiment, first-year Saudi EFL university students practised the meaning of 30 new words spaced over four sessions, with one practice per each word in every session, and another 30 new words massed within one session, with four practices per each word in the session. Interpretation of the result is currently limited since the full results have not yet been published. Except

practice frequency, Al Fotais's abstract does not state whether spaced and massed items were carefully matched in terms of learnability of items themselves.

In a novel design departing from the study of semantic clustering and spacing effects in deliberate vocabulary learning contexts, Nakata and Suzuki (2019) hypothesised that spacing might benefit learning of semantically related words by alleviating interference between them. The study adopted a between-subjects design in which 48 English words were divided into massing and spacing groups. Words were paired with semantically related or unrelated Japanese words (i.e., translation equivalents). The results from 133 Japanese University students showed that, contrary to expectations, spacing benefits were remarkable for semantically unrelated than related words, indicating that interference (i.e., extra attention, effort etc.) was not necessarily detrimental to learning. Nakata and Suzuki concluded that the effects of semantic clustering and spacing in incidental learning are an important avenue for future research.

Interestingly, eye tracking has also recently been used to test the mechanisms of deficient processing account of spacing effects in deliberate but contextualised L2 vocabulary learning (Koval, 2019). In a within-subjects study, 40 adult English language speakers read 24 Finish words within English sentence contexts. Words were divided into spaced and massed conditions while being matched in terms of their length (i.e., number of letters) to control learnability. A short spacing interval of 6 minutes (with a distractor math task) was implemented. Koval found that, in line with the deficient-processing theory, massed words were remembered less than spaced words and also received less attentional processing.

Thus, the previous studies have clearly demonstrated that spacing benefits learning in intentional learning contexts. Koval (2019) commented that incidentally oriented learning contexts could possibly generate similar positive outcomes. She concluded that:

> "… even when a learner is not trying to commit a word to memory but only processes it for recognition and comprehension, repeated exposures that are close together may receive less attentional processing and may, therefore, be

not as useful for learning as they would be if they were more widely spaced (p. 1131)".

In the following sections, I will highlight two important areas in spacing research about which information is still lacking. I will present a series of experiments on the spacing effects in incidental vocabulary learning, in general, and in extensive viewing of documentary series, in particular.

5.1.3 Incidental L2 Vocabulary Learning

The present study was conducted in response to recent second language research on spacing effects in deliberate vocabulary learning that highlighted the importance (e.g., Koval, 2019) and indicated the need (e.g., Nakata & Suzuki, 2019) to study this phenomenon in incidental contexts. Given the prevalence of the spacing effect in the teaching and learning literature, it is worth knowing whether the phenomenon observed in conscious vocabulary learning could extend to contexts where vocabulary acquisition occurs incidentally due to prolonged exposure to input. The problem to date has received scant attention and a search of the literature revealed only a few recent studies that gave inconclusive results. As will be seen in the present section, current investigations of the spacing effect in incidental vocabulary learning have maintained a focus on reading (Çekiç & Bakla, 2019; Elgort, Brysbaert, Stevens, & Van Assche, 2018; Elgort & Warren, 2014; Nakata & Elgort, 2020) with two other studies considering bimodal input, that is, listening-while-reading (Serrano & Huang, 2018; Webb & Chang, 2015). This body of research suggests that our scientific understanding of the spacing effect on incidental learning remains limited to reading and listening, while the impact of spacing in learning from viewing has been understudied. In particular, it is still not known whether the reoccurrence of learning items across multiple episodes of documentary series, viewed in the form of multimodal input (L2 captioned video), yields better learning than words reoccurring within a single episode or viewing session.

The current study aims to examine the association between spacing and unintentional acquisition of L2 vocabulary in a realistic simulation of incidental L2 learning. Examples of the cognitive mechanisms underlying spacing effect in incidental word learning contexts have historically been central to laboratory-based memory research (e.g., Challis, 1993; Greene, 1989; Greene & Stillwell, 1995;

Russo & Mammarella, 2002; Russo, Parkin, Taylor, & Wilks, 1998; Toppino & Bloom, 2002; Toyota, 2013). Early empirical evidence for the positive effect of spaced presentation of words in incidental conditions was provided by Greene (1989) through a series of six experiments. Participants were divided into two groups: participants who studied a list of words at different spacing intervals under intentional learning condition (announced test) and participants who studied the same word list under incidental learning condition (unannounced test). Words were presented one at a time on a computer at a 10-second rate. For the incidental condition, participants were asked to determine the order in which words were presented. The results on a free recall test demonstrated spacing effects regardless of the intentionality of learning, whereas, in cued-memory tasks, spacing effects were found for the intentional condition only. However, Toppino & Bloom (2002) observed contrasting free recall results, in which a spacing effect was found in intentional learning but not in incidental learning. In a parallel study by Russo and Mammarella (2002), incidental learning was maintained by asking participants to evaluate the structural features of items. The results showed spacing effects in incidental conditions for non-words displayed in sequence for 3 seconds at 0.5 seconds interval. Furthermore, Verkoeijen, Rikers, & Schmidt (2005) showed that intentional learning generates more significant spacing effects and longer interstudy intervals than incidental learning. Despite these early results of distributed incidental learning, they have arguably no correspondence to real incidental learning contexts.

While insights derived from laboratory-based studies could be helpful in our understanding of the impact of spacing in L2 vocabulary acquisition, they are not drawn from learning outcomes that are incidental to input comprehension (e.g., memorising word order). Results are thus not directly transferable to L2 acquisition research, which describes incidental vocabulary learning as the "picking up" of new words as a result of engagement in meaningful listening, reading, speaking, or writing activities (Rott, 2012). Furthermore, in natural incidental learning settings, words being observed are not depicted one by one on a computer screen. Longer interstudy intervals also characterise real learning environments and thus, are in contrast with the usage of short gaps in laboratory-based studies.

Past reports noted that "the relative lack of applied research in educational settings is, from an educational perspective, the most serious shortcoming of

research on the spacing effect" (Dempster, 1988, p. 631). Regardless, classroom-based research on spacing is still in its infancy, as indicated by the recency of deliberate learning examples previously cited in section 5.1.2. Until fairly recently, the only real learning account of spacing had been Bloom & Shuell's study (1981) in intentional learning condition. Afterwards, Sobel et al. (2011) were the first to draw attention to the need to add real-classroom studies to the spacing literature due to the difficulty of generalising existing findings to actual educational settings. Over the past 10 years, the study of spacing in authentic educational contexts has been attracting much interest; nonetheless, the spacing advantage has been specifically documented for intentional vocabulary learning, while little data have been published on incidental vocabulary learning.

As can be seen from the above set of studies and also remarked by other researchers (e.g., Carpenter et al., 2012), much of the existing literature on spacing effects in incidental vocabulary learning has been drawn from the field of memory and cognition, and thus developed in controlled laboratory settings. There have been relatively few real-world classroom investigations into spaced word learning in incidental contexts compared with the number of studies on explicit teaching and learning. Consequently, further research is needed to build the evidence base of whether incidental vocabulary learning from engaging in single-modality texts (e.g., listening, reading) or from multimodal exposure (e.g., viewing television series) is more successful when encounters are distributed across multiple presentations (i.e., texts, episodes) compared to when they are massed within a single presentation.

Some studies have begun to examine the spacing effect on vocabulary acquisition from reading in authentic incidental learning experiences. The most recent work in this line of research, conducted by Nakata and Elgort (2020), employed a within-subjects design. The study included 66 Japanese speaking participants who were higher-intermediate to advanced English language learners. Participants encountered 48 pseudowords embedded in three informative English sentences either in spaced or massed fashion. They were instructed to infer the meaning of the pseudoword from the linguistic context. The study results revealed an advantage of the spaced condition for both knowledge of meaning recognition, as operationalised by a meaning-form matching test, and meaning recall, but not for tacit knowledge (semantic priming).

Other than the input mode, two points of difference between the former and the present study concerns feedback and exposure length. In Nakata and Elgort's study, each response was followed by immediate feedback on the correct meaning. Though this method adds to the spacing literature by informing the way the learners could get the full benefit of the context, the current study intends to examine the spacing effect on incidental vocabulary learning from viewing when context is the sole source of input for encounters. In fact, one potential concern in Nakata and Elgort's method is that it leaves open the question of whether learning was at some points primarily driven by the provision of feedback. For instance, it was found that successful third inference attempts the during treatment phase predicted accuracy in the spaced condition but not in massed condition. However, this result needs to be interpreted with caution. A likely explanation is that vocabulary gains were largely attributed to receiving the correct meaning twice prior to the third attempt instead of exposure to informative sentences. Hence, due to the presence of feedback as a confounder in the study, its findings do not reflect the actual relationship between contextual support and incidental vocabulary learning. The study could be repeated by adding another group that helps to isolate the effects of context and feedback. Another point of difference is that this study gave evidence for short-term exposure (96 minutes treatment) to contextualised vocabulary. The present research, in contrast, explores the effects of spacing longitudinally (an exposure of about eight hours over a six-week period of two-week intervals).

More than the methodological concerns, Nakata and Elgort' study provokes important theoretical discussions. First, the present study operationalises incidental vocabulary learning as the inevitable unintended acquisition of word knowledge as a result of exposure to input. Precisely, it is the accidental by-product of planned input comprehension. However, Nakata and Elgort misused the term incidental learning to refer to intentional learning, specifically, explicit inductive learning which results from an inductive procedure of guessing from context (Nation, 2001, p. 395). To explain, participants in their study were instructed to infer the meaning of pseudowords from the linguistic context. Although the researchers did not explicitly instruct the learners to commit these pseudowords to memory, the guessing-from-context strategy involved in the study was an utterly teacher-led activity geared to vocabulary learning, making the context of learning intentional rather than

incidental. Learning was not consequential but a primary product of an external reinforcer to infer meaning. Based on DeKeyser' set dimensions (2003), learning language from input with help from a teacher falls into the category of explicit inductive instruction. According to Qi and Lai (2017) this learning can be described as an "inductive guided discovery supplemented with the provision of explicit rules of the target features" (p. 27). In fact, while incidental learning may well involve inductive processes (e.g., guessing from context), the reverse is not true; inductive learning does not consist of incidental learning, unless learning is a secondary outcome that the learner does not intend to achieve.

My reference to this improper use of the term incidental learning to refer to inductive learning is motivated by a difference in the mechanisms that underlie the two processes. When word meaning is learner-driven and occurs only naturally as a secondary learning outcome, the learner is less aware of the learning opportunities. However, when learning is prescribed as part of a classroom activity, the learners are more able to think about how the language works and more conscious of the learning opportunities available to them from the linguistic context. That is usually lacking in real incidental contexts with which the present study is concerned. Hence, more robust learning and spacing effects might be expected in empirical studies with an instructed induction practice.

Contrary to the previous findings, an advantage was found for massed repetitions in a within-subject design study characterised by non-instructed incidental learning (Elgort & Warren, 2014). A total of 48 advanced and high-intermediate learners of English language learners encountered 48 repeated pseudowords within selected texts of a nonfiction book. Participants were asked to keep a reading log and not use a dictionary throughout the reading period of 10 days. Meaning recall results showed a benefit for repetitions occurring within one chapter over repetitions across multiple chapters, while in students with lower proficiency, incidental learning was only observed for within-chapter repetitions.

In a follow-up study, two texts from the above nonfiction book served as the stimulus for a two-day eye-tracking study by Elgort et al. (2018). The study revealed contrasting results for form and meaning. It included 40 Dutch-speaking university students who were higher-intermediate to advanced English language learners.

Participants encountered 14 target words with a frequency of occurrence that was either spaced over two days (N= 5 words) or massed within one of the two days (N = 9 words), though the latter condition was termed “short spacing” to reflect the authors’ interest in the lag effect. Offline tests and eye movement measures suggested that knowledge of meaning recall for words encountered across two days may be acquired more successfully than for words encountered within the same day, while the reverse might be true for knowledge of form.

In contrast to the previous design, Çekiç and Bakla (2019) considered the effects of spacing patterns between subjects, and through reading short texts supported with incidental exposure to electronic glosses. The study comprised 189 Turkish speaking participants who were intermediate English language learners. They encountered 20 target words, at a constant frequency of nine, in 36 short texts, where the distribution of encounters varied among learners according to spacing patterns. Pattern 9 was a fixed spaced group of a nine-week treatment period (i.e., nine sessions). Participants read, every week, four passages in one session, during which study items were presented only once. Pattern 7 was a spaced massing group of seven weeks of two-week and three-week intervals (three sessions). Participants read 12 passages within each session (massed), during which items were encountered three times per session. Pattern 3 group followed the same procedure as Pattern 7 except that the interstudy intervals were one week long. Therefore, this three-week treatment group was named spaced massing with fixed intervals in contrast to Pattern 7 that was of expanding intervals. Vocabulary Knowledge Scale (VKS) results showed that the scores of Pattern 9 and Pattern 3 increased almost equally and were significantly better than Pattern 7, with the long-term fixed spacing group being the most conducive to learning.

In addition to reading, two other studies on bimodal input (i.e., listening-while-reading) have been identified in the literature. Webb and Chang (2015) showed no influence of spacing on incidental learning. A total of 61 Taiwanese-speaking secondary school students with similar English proficiency encountered 100 target words in an extensive English language reading program. Participants read and listened to 10 graded readers. Meaning recognition, operationalised as a form-meaning matching test, showed no association between the distribution of occurrence and vocabulary gains. These results differed from those obtained

following exposure to the same text multiple times (Serrano & Huang, 2018). Their study adopted a between-subjects design, including 71 Taiwanese high school students studying English as a foreign language. Participants encountered 36 target words while reading and listening to an English text. The massed group was exposed to the exact text intensively (i.e., once every day for five consecutive days). The spaced group read and listened to the same text once every week for five consecutive days. The results of meaning recognition (form-meaning matching test) showed that massed practice contributed to increased scores in the short-term (immediate posttest). No significant difference was found in the long run (delayed posttest) between the groups' vocabulary gains. As for long-term retention during the period between the two tests, spaced practice was found to contribute to better retention.

Most of the studies above were concerned with the lag effect, that is, dealing with the question of short spacing versus long spacing. The present study implements two-week intervals between the treatment sessions. Although the length of the optimal interval is not experimentally addressed here, it is fundamental for the design of any study on spacing effects to ask: What are the trends and paths set out to us by the literature? There has been a consensus among researchers that the longer the spacing, the better the learning. Recent evidence suggests that the posttest schedule also conditions this benefit. Longer spacing is more advantageous at long-delayed schedules and shorter spacing is more beneficial at short-delayed schedules (e.g., Nakata & Webb, 2016). With this in mind, it remains that what is considered spacing in one study could be viewed as short in another, or as Nakata and Elgort (2020) put it, "… we are not really comparing apples with apples" (p. 5). As Cepeda et al. (2006) also maintained: "After more than a century of research on spacing… it is unfortunate that we cannot say with certainty how long the [interstudy interval] should be to optimize long term retention" (p. 370). One longitudinal study compared three spacing intervals: 14, 28, 56 days (Bahrick et al., 1993). Results indicated that the longer the spacing interval, the better the retention of learning items. Hence, the interval in the present study was specified as two weeks.

Lastly, based on a recent meta-analysis, studies implementing single session treatments have shown a trend towards greater learning than interventions with spaced sessions (Uchihara et al., 2019). Incidental vocabulary learning results from

subsequent noticing of forms of which meanings are unknown but gradually inferred with more encounters. Consistent with the authors' prediction, they explained that repeated encounters within one short span produce a positive cumulative effect that greatly increases chances of learning compared to distributed encounters (Webb, 2014).

The authors combined the results of 26 vocabulary intervention studies in which incidental word learning was studied as a function of frequency of occurrence. They attempted to understand the association between repeated exposure and incidental L2 vocabulary learning and obtain important information on how the distribution of occurrence interacts with frequency of occurrence. The studies varied in exposure type from entire novels to multiple TV episodes. Spacing between occurrences of items was one potential moderator of the frequency-learning relationship among 10 moderators that were examined (learner, treatment, and methodology related). Irrespective of treatment or content quantity, studies completed in one day were coded as massed (e.g., Hatami, 2017), while studies in which intervention sessions exceeded two days period were considered spaced (e.g., Daskalovska, 2016).

The meta-analysis showed a small frequency effect in studies presenting items in spaced fashion (r = .23) compared to studies presenting items in massed fashion (r = .38). In addition, two studies in massed condition but differing in exposure type revealed medium effect sizes: a single exposure study in which a single text was presented within one day (r = .33), and a repeated exposure study in which the same text was presented multiple times for more than a day, r = .46). On the other hand, two spaced studies in which participants were exposed to different texts, either in a controlled setting or at their own pace, showed a small and identical effect size (r = .19).

As the authors put it, the inconstancy of intervals between the posttest and experimental sessions in the spaced conditions might mask frequency effects on learning (Webb & Chang, 2015). Thus, studies conducted in spaced conditions were expected to be more likely to show a marginal frequency effect on incidental vocabulary acquisition. It is worth adding that, as noted earlier, the authors pointed out that the majority of studies on spacing have been completed in deliberate paired-

associate learning. Their discussion was void of any reference to incidental vocabulary learning interventions specifically comparing spaced to massed presentations. This supports the present study's claim that there is much less information about the spacing effect in incidental contexts.

In sum, little attention has been paid to the likely effect of spacing in incidental vocabulary learning, relative to deliberate learning. Classroom-based studies on spacing effects in incidental vocabulary acquisition have been limited to reading or listening-while-reading and have revealed inconsistent results. Importantly, further carefully controlled studies are necessary to compare differences between spaced and massed incidental vocabulary learning based on direct manipulation. Studies on reading provide a valuable account of how spacing contributes to incidental L2 vocabulary learning, notwithstanding that they represent potential spacing effects from only one source of input (i.e., written text), among many which the learner may encounter. In particular, learning from extensive viewing such as films and documentaries have become a major area of interest for vocabulary researchers for representing a valuable source of multimodal input. Nonetheless, it is still not yet clear how spacing intervenes in the process of learning from such input.

5.1.4 Extensive Viewing

Despite a history of interest in spacing on one hand and television viewing on the other, studies do not discuss spacing effects from this type of multimodal input. The current investigation is prompted by the lack of classroom-based research that compares distributed and massed presentations in the context of incidental L2 vocabulary learning from viewing. The remaining part of the review will stress the paucity of evidence in this area in second language research. It will first describe the two different ways that this research can be conducted. The section then will go on to present what is currently known in the literature before finally laying out possible reasons the research is understudied.

First, research on the effect of spaced and massed occurrence distribution on incidental vocabulary learning from viewing can take two main forms. The first type of study approach that could be used to identify spacing effects from viewing is the use of a single video to be watched repeatedly, either intensively (e.g., everyday for

three days) or under spaced conditions (e.g., every week for three weeks). This design is well-established in the reading literature; nonetheless, it does not reflect realistic scenarios of incidental learning from viewing. It suffers from a lack of ecological validity as it is unlikely that a person would choose to watch the same video content repeatedly, findings would thus have less significant implications. In addition, the approach necessitates a between-subjects design, which tends to impose limitations on spacing studies. In his classroom-based experiment, Snoder (2017) attributed the absence of evidence of an expanding learning schedule relative to intensive learning to the between-subjects design used in the study. He commented that this design is less methodologically robust and that "more research is needed to tease apart the effects of the variables in question using the more robust within-subjects designs" (p. 156).

The second approach to examine the impact of spaced and massed word occurrences in television viewing is adopted by the present study and is based on the distribution of occurrences within one video (massed) and across multiple videos of similar genre (spaced). This approach has several attractive features. As noted earlier, a between-subjects design may result in reduced sensitivity to significant differences; the multiple content videos approach permits the implementation of the more robust within-subjects design; thus, participants could serve as their own controls. More importantly, this approach enables comparison between learning from limited and multiple contexts, making the study more theoretically motivated for being firmly grounded in contextual variability theory (see Section 5.1.1).

At the time of this writing, only two studies were identified as being relevant but not similar to the present investigation (Pujadas Jorba 2019; Rodgers & Webb, 2019), with both studies showing a massing advantage. Rodgers & Webb (2019) examined incidental L2 vocabulary learning from +7 hours viewing of a television programme over 10 weeks (10 episodes). They did not compare distributed and massed occurrences but instead looked at the impact of the range of occurrences across multiple episodes. The study included 260 undergraduates who were pre-intermediate to intermediate learners of the English language. A total of 187 participants viewed 10 episodes of the TV series *Chuck* over 10 sessions, while 73 participants served as the Control group. The study used a one-week interval strategy with only a few sessions separated by two weeks. Each session consisted of

one episode with an average viewing time of 42 minutes and 49 seconds. The study included 60 target words with a range of occurrence frequency from 5 to 54 throughout the material and an average range of episodic occurrence of 3.7 episodes. Differences in item difficulty between words of different range were not considered. Episode range was a significant predictor of vocabulary gains on a tough test (β = -2.61, $t(57)$ = -2.25, p = .029). The test increased difficulty by creating multiple-choice item distractors that shared aspects of form or meaning with the accurate response. However, episode range did not predict vocabulary gains on a sensitive test (β = -2.41, $t(57)$ = -1.76, p = . 0.84) in which item distractors did not share the parts of speech nor the aspects of form and meaning with the target word. Relative frequency was defined as the total number of encounters of a target word in the overall input divided by the number of episodes. Further analysis of vocabulary gains and relative frequency revealed that learning was greater for target words that reoccurred within a single episode than target words that reoccurred across a range of episodes.

A massing advantage was also obtained by Pujadas Jorba (2019). However, her study was limited by its design. A total of 83 words out of 120 were massed items, that is more than twice the number of spaced items. Also, the number of occurrences of spaced items was reported as significantly greater than the number of massed occurrences. Moreover, the number of sessions over which items were spaced was not identical across all spaced words. Some words were spaced over a high number of episodes while others were spaced over few episodes. This also indicates that the interstudy intervals could be inconsistent across the spaced words. Lastly, the study did not control for the recency effect in learning massed words.

The present study aims to assess whether repeated occurrences distributed across multiple extensive TV viewing sessions facilitate incidental L2 vocabulary learning compared with repeated occurrences massed within a single session. This study differs from the two previous ones. It is the first to report on a controlled manipulation characterised by a between-items design in which spaced and massed word pairs are matched in terms of learnability, mainly, verbal frequency of occurrence. The length of the interstudy interval and the viewing sessions across which spaced occurrences were distributed were consistent for all spaced words. The interference of the recency effect was also considered.

In contrast to reading studies, authentic viewing studies pose particular methodological challenges when selecting target items. Reading research allows the simple substitution of real target words with pseudowords, which has a number of benefits: spaced versus massed pseudoword pairs only need to be matched orthographically (i.e., number of syllables, letters) and according to the frequency of occurrence in texts. Research in reading also permits the design (and the writing up) of texts manipulating the spacing of the target items (e.g., Çekiç & Bakla, 2019; Chen & Truscott, 2010). In contrast, investigating authentic viewing means it is not possible to use pseudowords. Consequently, stringent criteria are required to match spaced versus massed word pairs on all potential word properties (e.g., concreteness, cognate status) to control for extraneous variables. Importantly, when using authentic materials in viewing research, the spacing of items is also more 'opportunistic' and plays an additional role in the selection criteria for items. To explain, within each matched pair of words, one item needs to occur multiple times within one video. In contrast, the second matched item must occur the same number of times but across multiple videos of a similar genre. As a result, the difficulty in finding matched pairs under the two conditions means that only a smaller sample of items can usually be tested.

5.2 The Present Study

This third study examined whether repeated exposure to the same word across spaced episodes (i.e., viewing sessions) of two full-length seasons of documentary series, in the form of L2 captioned video, contributes to greater incidental acquisition of knowledge of meaning recall and recognition, and spoken and written form recognition, than repeated exposure that is massed in single viewing sessions. The question of whether the presence of imagery in the documentary series influences any potential effect was also addressed. The study adopted a between-group experimental design. The control group received no treatment while the experimental groups were exposed to two full-length seasons of documentary series which extended to eight viewing hours over six weeks at two-week intervals. The View group watched the episodes in the form of L2 captioned video. The Non-View group had the imagery removed from the video and therefore were exposed to L2 audio and captions only. The study complements previous research on spacing effects on vocabulary acquisition by considering the phenomenon in a different learning context. As noted above, the impact of the spacing effect on incidental acquisition of words in proper L2 learning contexts has been understudied. To the best of my knowledge, this is the first systematic study of the role of the distribution of word occurrences on incidental learning from viewing. The study differs from Rodgers and Webb (2019) in underlying methodology and analysis. Instead of examining random target words, eight spaced nouns were matched to eight massed nouns in learnability (all word-related covariates). Spacing was treated as a within-subjects, between-items categorical variable. Learnability was operationalised as the number of characters and syllables, concreteness, verbal frequency of occurrence, and cognateness of each item. Lastly, in comparison to Rodgers and Webb, viewing in the present study was more spaced; instead of one week, the study set a target two-week interval between sessions, and was more extensive as each session comprised two episodes of one hour (i.e., 2 hr, totalling 8 hr).

5.3 Method

5.3.1 Questions and hypotheses

Study 3 asked the following research questions:

***Research Question 1*:** Do repeated occurrences distributed across multiple extensive viewing sessions facilitate incidental L2 vocabulary learning from documentary series compared with repeated occurrences massed within a single session?
***Research Question 2*:** Does any spacing effect vary as a function of the presence of imagery?

5.3.2 Participants

One hundred seventy-three Algerian EFL learners in their third year of the Linguistics Bachelor programme at the University of Jijel, in the autumn semester in the 2017-2018 academic year, took part in Study 3. Of these, 29 participants were excluded: if they were absent in any session of the pretests and posttests and if they missed any session of the treatment phase. Data from 144 participants (131 females and 13 males) aged 21-23 years (M = 21.11) were kept for analysis. Participants were divided into three groups: Control (N = 34), View (N = 53), and Non-View (N = 57) using stratified random sampling. They were all native Arabic speakers with French as a second language, and with an intermediate to upper-intermediate English language level. The study was approved, and consent was obtained (See Chapter 3, Section 3.3.2 for full description of the participants and ethical considerations).

5.3.3 Materials

The materials were the two full-length seasons of the documentary series previously used in Study 1 and Study 2 which extended to eight viewing hours. Information on the series is detailed in Chapter 2, section 2.2.1.

5.3.4 Target Items

Sixteen nouns appearing in the documentary materials made up the target eight spaced versus massed word pairs for Study 3. The pairs are listed along with their values in Table 5.1. Spaced words occurred multiple times across all four sessions and massed words occurred multiple times within one session. The corresponding session for every massed word is presented in Table 5.2.

Table 5. 1

Target Spaced Vs. Massed Word Pairs (N = 8) Matched According to Learnability

Pairs		Verbal freq	Related forms[a]	Corpus Log Freq (Zipf)[b]	Length		Concreteness[c]	Contigduration	Cognate status
Spaced	Massed				Characters	Syllables			
supernova	manatee	15-14	0	3.08 - 2.08	9-7	4-3	3.78 - 4.66	125 - 228	Yes-no
constellation	hexagon	11 - 10	0-6	3.20 - 2.63	13-7	4-3	4.31 - 4.52	182 - 101	Yes
sphere	sulphur	16	31-2	3.68 - 3.35	6-7	1-2	4.44 - 4.23	462 - 150	Yes
spectrum	symmetry	12 - 11	0-8	3.80 - 3.25	8	2-3	2.97 - 2.79	165 - 164	Yes
particle	fusion	12 - 10	0-6	3.48 - 3.60	8-6	3-2	3.78 - 3.30	35 - 95	Yes
temple	pile	9 - 8	0	4.03 - 4.23	6-4	2-1	4.53 - 4.56	330 - 69	Yes-no
cosmos	photon	35 - 32	12-0	3.27 - 2.45	6	2	3.19 - 3.38	652 - 187	Yes
tide	moth	12- 11	6-0	4.25 - 3.64	4	1	4.10 - 4.69	212 - 147	No

Note. Freq = frequency; [a] Other forms were derivatives and compounds. [b] Measures were based on the SUBTLEX-UK word frequencies, presented in Zipf-values, a logarithmic scale: 1-3 = low frequency, 4-7 = high frequency (Van Heuven et al., 2014). [c] Measures were based on 40 thousand English lemma words on a 5-point rating scale going from abstract to concrete (Brysbaert et al., 2014).

Table 5. 2

Massed Words and Their Corresponding Session (from 1 to 4).

Item	Session
manatee	2
hexagon	2
sulphur	3
symmetry	2
fusion	4
pile	2
photon	1
moth	3

5.3.5 Procedure

The research schedule, procedure, and vocabulary tests for Study 3 are the same as that of Study 1 (See Section 3.3 of Chapter 3).

5.3.6 Scoring

Responses of participants meaning recognition and recall and spoken and written form recognition tests were scored in the way described in Study 1 and Study 2: "0" for incorrect, missing, and "I don't know" responses, and "1" for correct responses (see Chapter 3, Section 3.3.6 for details).

5.4 Analyses

5.4.1 Analysis Procedure

Data were analysed in R (R Core Team, 2018) using Rstudio (version 1.2; RStudio Team, 2018). Results were summarised using dplyr package (version 0.8.3; Wickham, François, Henry, & Müller, 2019). They were visualised using ggplot2 package (version 3.2.1; Wickham, 2016) for meaning recognition and recall results and ggpaired function of ggpubr package (version 0.2.4; Kassambara, 2019) for form recognition results (for having a pretest-posttest structure). Data were analysed with generalised linear mixed-effects (GLM) logistic regression models using glmer function of lme4 package (version 1.1-21; Bates, Maechler, Bolker, & Walker, 2015). The following part explains the sequence of procedures adopted to answer the two research questions for Study 3.

Research Question 1

The first research question examined whether repeated occurrences distributed across a range of extensive viewing sessions facilitate incidental L2 vocabulary acquisition compared with multiple occurrences crammed in one session. To answer this question, data for the View and Control groups were analysed. I initially conducted a GLM logistic regression analysis including all word-related explanatory variables that were theoretically meaningful and relevant to both the control and experimental groups. That is, the frequency of occurrence of target words and their related forms, which is treatment-related, was not included in the analysis. For meaning recall and recognition measures, which lacked a pretest, the baseline models specified posttest accuracy as the dependent variable, group as fixed effect,

written form pretest as a control variable, written form pretest × group as an interaction term, characters, concreteness, cognate (cognate = 1, noncognate = 0), and corpus frequency as control covariates, and participants and words as random effects, with random intercepts allowed to vary across participants and words (e.g., random = ~1 | word). For spoken and written form recognition measures, data were in a repeated-measures design (since participants had sat a pretest for these measures). The models specified response accuracy as the dependent variable, time × group as an interaction term, characters, concreteness, cognate, and corpus frequency as control covariates, and participants and words as random effects, with random slopes of time for each to denote that the effect of time varies across participants and words (e.g., random = ~Time | Word).

The significance of the main effect of spacing (spaced = 1, massed = 0) was then assessed for all dependent measures using likelihood ratio tests which compared the previously described models to identical models with spacing as an additive predictor. This additive model was then compared to an identical model with group × spacing interaction. To further investigate the effect of spacing, post-hoc pairwise comparisons were performed for significant interactions between the two levels of group and spacing, using Emmeans Package (1.4.4). The full model was prone to inflated standard errors and was therefore simplified by removing all word-related covariates. If the substantive results differed from the full models, only variables that did not significantly predict response accuracy were removed following a stepwise procedure for model comparisons using the likelihood ratio test. The three variables of group, spacing and cognate were automatically dummy coded by R software as a categorical variable, then releveled so that Control group (N = 34), spaced words (N = 8), and cognates (N = 12) were the reference level.

The large study sample size permitted the use of multilevel modelling and inclusion of theoretically meaningful covariates and maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013). This helped meet the independence assumption by controlling for individual variations among participants and across words (in both time points for form recognition). Group was not allowed to vary across words in the current study because it addresses whether the effect on spaced and massed words varies across groups.

Recency Effect.

While spaced words occurred throughout the four viewing sessions, massed words were encountered in single sessions. Because of the way these target words were distributed, it was impossible to examine the impact of spacing without further considering the interference of the recency effect, whereby words encountered in sessions closer to testing might be more likely to be recalled. For this purpose, an additional analysis of the data was performed.

The model in this analysis was identical to the simplified model used in the principal analysis except that *spacing* was replaced by *session*, a categorical variable of five levels (all, 1, 2, 3, 4). Each level indicated the position in session for massed words. Spaced words that were experienced in all sessions were set as the reference level (*all*); the model thus calculated the probability of response accuracy for words in each session in contrast to spaced words. The analysis generated inflated standard errors, thus, wide confidence intervals, due to the small number of items in each session; as an attempted solution, items of session 1 and session 2 were merged and made a single level (massed), while session 3 and 4 items made up a second level (massed recent). The substantive results were mostly preserved following the aggregating approach.

The significance of the main effect and interaction effect of the new spacing variable was then assessed using the same procedure previously described. However, post-hoc pairwise comparisons between the three levels of the spacing variable were run with Bonferroni adjustment for multiple comparisons, $\alpha = .017$.

Research Question 2

The second research question assessed whether the presence of imagery influenced any potential effect of spaced and massed occurrences on incidental L2 vocabulary learning. To answer this question, the subset of the data for the View and Non-View groups was analysed. The model from Research Question 1 analysis was implemented. Two predictors that were treatment-related (verbal frequency and related forms) and an interaction between each and the experimental condition (View/Non-View) were added to the model. To assess the extent spacing results differ as a function of imagery, the two-way interaction spacing × group was added to the model. The full model was prone to inflated standard errors and was therefore

simplified by removing all word-related covariates and comparing results in the same manner as described above.

5.5 Results

To answer the two research questions in this study, the results for the four dependent measures are reported separately and arranged in the following order: meaning recall and recognition and spoken and written form recognition for the reasons stated in Chapter 3, Section 3.5.3.

The mean scores for meaning recall and recognition posttests and written and spoken form recognition pretests and posttests for the View (N = 53), Non-View (N = 57), and Control (N = 34) groups on 16 items, Spaced (N = 8) and Massed (N = 8), are presented in the summary Table 5.3. Scores were plotted for every research question and are presented in figures.

Table 5. 3

Descriptive Statistics per Group for all Vocabulary Tests Scores (16 Items)

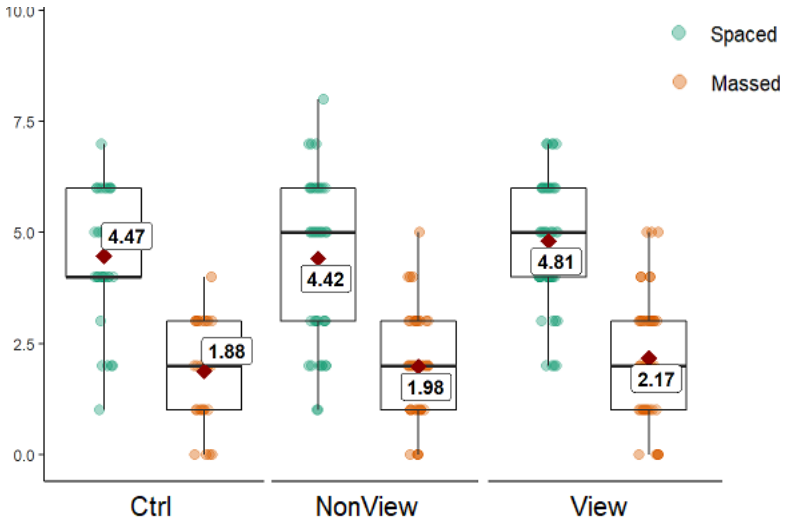
		Mean Scores											
		Pretest						Posttest					
		Spaced			Massed			Spaced			Massed		
		M (SD)	*Min*	*Max*	*M (SD)*	*Min*	*Max*	*M (SD)*	*Min*	*Max*	*M (SD)*	*Min*	*Max*
Meaning Recall	Control							1.41 (1.08)	0	5	0.47 (0.86)	0	3
	Non-View							3.18 (2.03)	0	8	1.79 (1.72)	0	6
	View							3 (1.53)	0	6	2.57 (1.87)	0	7
Meaning Recognition	Control							1.44 (1.08)	0	4	0.68 (0.84)	0	3
	Non-View							5.02 (1.96)	0	8	4.23 (2.20)	0	8
	View							4.62 (1.86)	1	8	4.70 (2.33)	0	8
Spoken Form Recognition	Control	2.95 (1.34)	1	6	2.38 (1.18)	0	5	3.15 (1.37)	1	6	2.05 (1.10)	0	5
	Non-View	3.27 (1.61)	0	6	2.55 (1.36)	0	5	4.79 (1.29)	2	8	3.50 (1.25)	1	7
	View	4 (1.56)	1	7	2.96 (1.54)	0	6	5.25 (1.37)	2	8	3.72 (1.43)	0	7
Written Form Recognition	Control	4.43 (1.48)	1	7	1.86 (1.06)	0	4	3.63 (1.55)	1	7	1.46 (1.15)	0	4
	Non-View	4.42 (1.63)	1	8	1.98 (1.04)	0	5	5.82 (1.27)	2	8	3.21 (1.24)	1	6
	View	4.81 (1.37)	2	7	2.17 (1.31)	0	5	6.19 (1.16)	8	2	3.06 (1.83)	0	8

Note. data %>% group_by(group, Time, Spacing) %>% summarise (mean = mean(Response), sd = sd(Response), max = max(Response), min = min(Response)).
M = mean. SD = standard Deviation. Maximum score = 16.

The coefficients of the final model fixed effects and random effects on response accuracy by participants are reported in tables. The first column provides the change in the log odds of response accuracy associated with change in the group and spacing conditions. A positive coefficient indicates an increase in accuracy and a negative coefficient indicates a reduction in accuracy, compared to the baseline category. Odds ratio is a measure of effect size. The predicted probabilities from models of the follow-up analysis (recency effect) were substantively similar to the original one and were hence plotted and depicted in figures to avoid duplication. Despite the careful matching procedure, pretest differences between spaced and massed items can be seen in written form recognition accuracy data comparing pretest performance between groups (Figure 5.1). There was a clear advantage for spaced words compared to massed words at the baseline reference. Fortunately, this advantage in the written form data was almost identical across the two groups. This made it rational to still include written form pretests as a baseline reference for meaning tests that lacked a pretest.

Figure 5. 1

Mean Accuracy in Written Form Recognition Pretest



Note. The boxplots show mean accuracy scores in written form recognition pretest by subject, across Control (N=34), Non-View (N = 57) and View (N = 53) groups for 16 nouns. Half items (N = 8) were spaced over four viewing sessions of documentary series, the second half of the items were massed in one of the session. Means are represented in the figure by the red points. Large differences between points depict outliers.

5.5.1 Research Question 1

Do repeated occurrences distributed across multiple extensive viewing sessions facilitate incidental L2 vocabulary learning from documentary series compared with repeated occurrences massed within a single session?

For all the four dependent measures in this first analysis, the word-related covariates were pruned from the model without effect on the significance of the variable of interest and overall substantive results. This was done to enable a more parsimonious model with better standard errors. To preview the results (Table 5.4), the four measures did not show an advantage of repeated occurrences that were distributed across multiple extensive viewing sessions over repeated occurrences that were massed within a single session, in the View group relative to the Control group. The results also indicated that there was no recency effect on the obtained spacing results. A detailed review of the results follows.

Table 5. 4

Summary of Research Question 1 Findings

Models	Main model	Recency model
Meaning recall	$\chi^2(1) = 2.29$	$\chi^2(2) = 2.78$
Meaning recognition	$\chi^2(1) = 2.4$	$\chi^2(2) = 2.62$
Spoken form recognition	$\chi^2(2) = 2.16$	$\chi^2(4) = 7.41$
Written form recognition	$\chi^2(2) = 3.48$	$\chi^2(4) = 3.60$

Note. Models included View and Control groups data; meaning results based on Spacing × Group interaction, form results based on Spacing × Group × Time interaction.

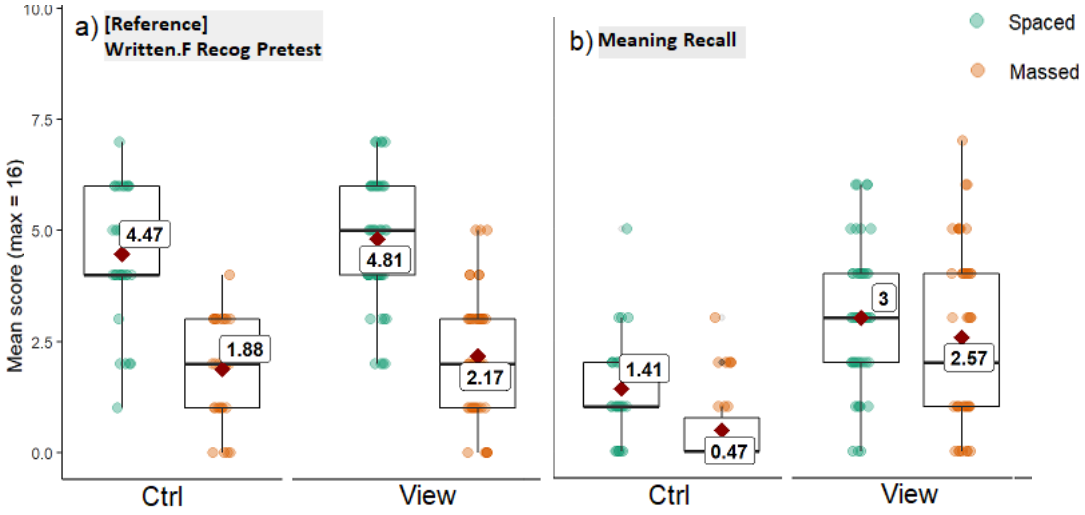
Meaning recall

The primary results showed a significant main effect of group in meaning recall accuracy of the sixteen nouns altogether, $\chi^2(1) = 41.58$, $p < .001$. The odds of a correct response in the View group were more than six times as high compared to the Control group (*OR* = 1/Exp(B) = 1/0.16 = 6.25, 95% CI [0.10, 0.27]).

The mean accuracy scores in meaning recall comparing performance on spaced and massed items for View and Control participants are shown in Figure 5.2. Results of GLM logistic regression on meaning recall data showed that spacing did not affect performance overall, $\chi^2(1) = 0.11$, $p = .742$. The interaction between group and spacing was also non-significant, $\chi^2(1) = 2.29$, $p = .130$. The coefficients estimates for response accuracy in meaning recall and recognition models are reported in Table 5.5.

Figure 5. 2

Mean Accuracy in Meaning Recall

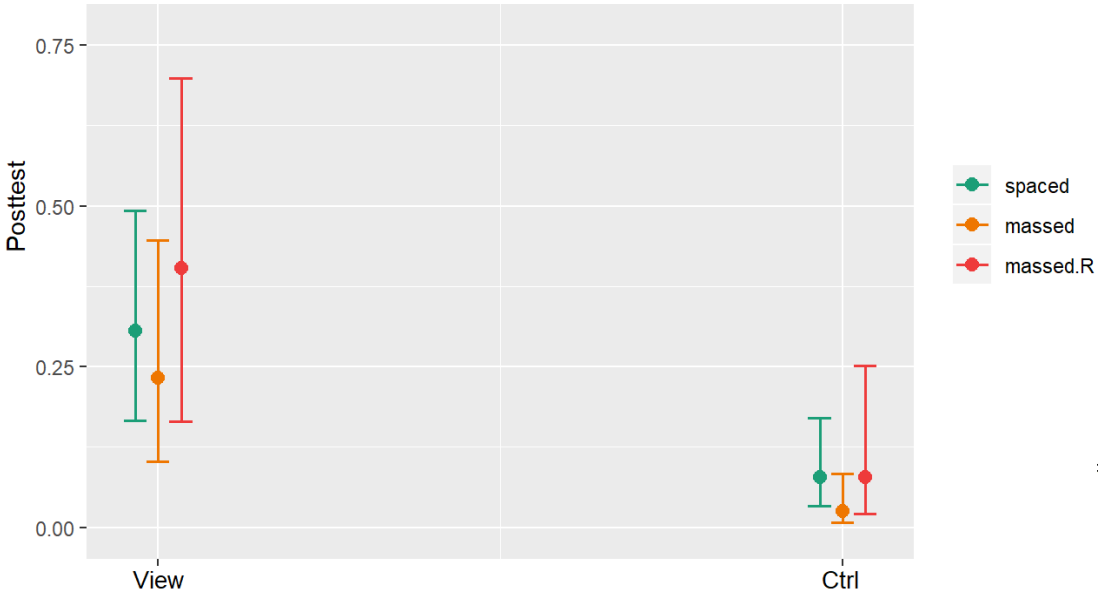


Note. The boxplots show mean accuracy scores in (a) written form recognition pretest and (b) meaning recall posttest by subject, across Control(N = 34) and View (N = 53) groups for 16 nouns. Half items (N = 8) were spaced over four viewing sessions of documentary series, the second half of the items were massed in one of the session. Means are represented in the figure by the red points. Meaning was not pretested to prevent prior exposure bias; the written form recognition pretest (a) was used as a baseline reference of prior knowledge.

The follow-up analysis of whether these results were affected by the recency of the massed items (Figure 5.3) still showed no effect of spacing in meaning recall, neither alone, $\chi^2(2) = 1.21$, $p = .545$, nor in a two-way interaction between group and spacing, $\chi^2(2) = 2.78$, $p = .249$. Taken together, these results demonstrate that meaning recall accuracy did not depend on whether items were spaced or massed.

Figure 5. 3

Spacing and Recency Effects in Meaning Recall



Note. The predicted probabilities plot shows the probability values for posttest accuracy on meaning recall of 16 nouns by spacing conditions, in View (N = 53) and Control(N = 34) groups, calculated from GLM logistic regression analysis. massed.R = massed recent. Half items (N = 8) were spaced over four viewing sessions of documentary series. Massed items (N = 5) occurred in the first two sessions, massed recent items (N = 3) occurred in the last two sessions.

Table 5. 5

GLM Logistic Regression Predicting Meaning Accuracy from Spacing

	Meaning recall					Meaning recognition				
Parameters	*B*	*SE*	*z*	*p*	*OR*	*B*	*SE*	*z*	*p*	*OR*
Fixed effects										
Intercept	−0.96	0.43	−2.21	*	0.38	0.30	0.40	0.75	.456	1.34
Group = Ctrl	−1.98	0.40	−4.90	***	0.14	−2.86	0.40	−7.08	***	0.06
Group = View										
Spacing = massed	−0.06	0.56	−0.12	.909	0.94	0.08	0.50	0.16	.876	1.08
Spacing = spaced										
Group (Ctrl) × spacing (massed)	−0.59	0.38	−1.54	.124	0.55	−0.57	0.36	−1.59	.112	0.56
Written. F	0.31	0.20	1.56	.118	1.38	0.36	0.20	1.79	.074	1.44
Group (Ctrl) × Written. F	0.77	0.39	1.99	*	2.15	0.36	0.36	0.99	.323	1.43
Random effects				Variance	*SD*				Variance	*SD*
Participant (intercept)				0.75	0.87				1.05	1.02
Item (intercept)				1.13	1.07				0.86	0.93

Note. Posttest ~ group × spacing + written form pretest × group + (1|participant) + (1|item). Model fitted to 1392 observations across 16 nouns. N = 87.
***p* < .01. ****p* <.001.

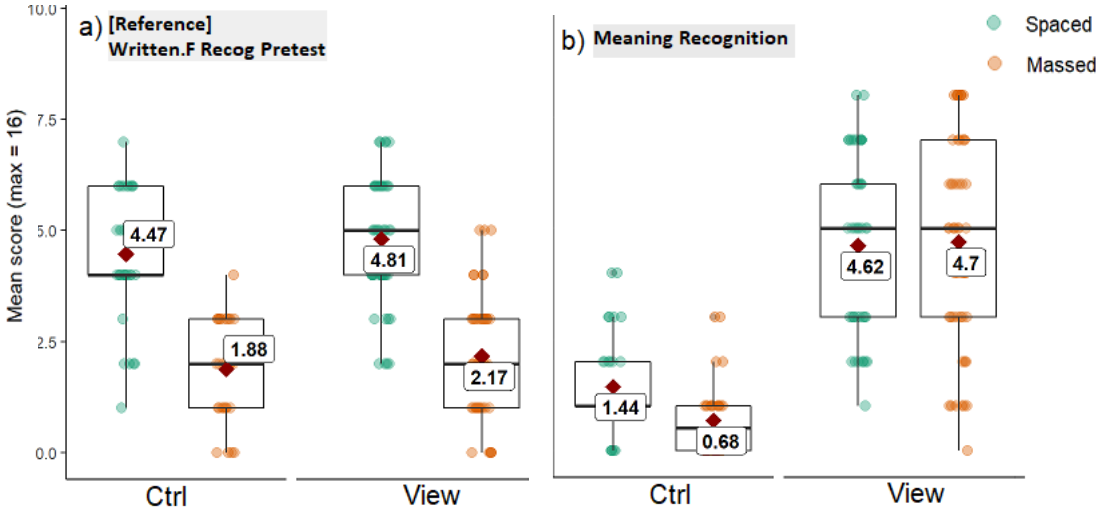
Meaning recognition

Overall, the data showed a significant main effect of group on meaning recall of the sixteen items, $\chi^2(1) = 76.74$, $p < .001$. View group participants were 20 times more likely to recognise a meaning of a word than Control participants (*OR* = 0.05, 1/Exp(B) = 1/0.05 = 20, 95% CI [0.03, 0.09]).

The mean accuracy scores in meaning recognition test comparing performance on spaced and massed items for View and Control participants are shown in Figure 5.4. The analysis revealed that neither the main effect of spacing on meaning recognition responses, $\chi^2(1) = 0.02$, $p = .898$, nor its interaction between group and spacing were significant, $\chi^2(1) = 2.4$, $p = .121$.

Figure 5. 4

Mean Accuracy in Meaning Recognition

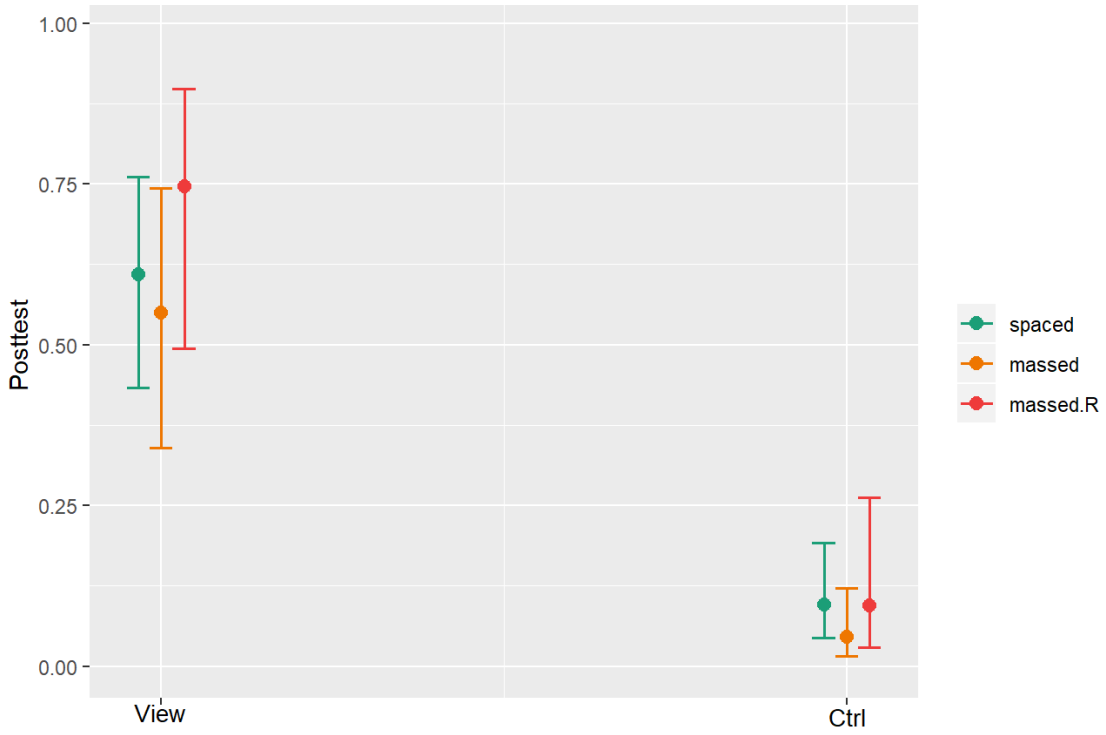


Note. The boxplots show mean accuracy scores in written form recognition pretest (a) and meaning recognition posttest (b), each by subject, across Control (N = 34) and View (N = 53) groups for 16 nouns. Half items (N = 8) were spaced over four viewing sessions of documentary series, the second half of the items were massed in one of the session. Means are represented in figure by the red points.

Similar findings were found in the follow-up analysis that examined whether recency of massed items influenced spacing results (Figure 5.5). Neither the main effect of spacing , $\chi^2(2) = 1.36$, $p = .506$, nor the interaction between group and spacing were significant, $\chi^2(2) = 2.62$, $p = .270$. These results suggest that spacing did not predict meaning recognition performance.

Figure 5. 5

Spacing and Recency Effects in Meaning Recognition



Note. Predicted probabilities plot shows the probability values for posttest accuracy on meaning recognition of 16 nouns by spacing conditions, in View (N = 53) and Control (N = 34) groups, calculated from GLM logistic regression analysis. massed.R = massed recent. Half items (N = 8) were spaced over four viewing sessions of documentary series. Massed items (N = 5) occurred in the first two sessions, massed recent items (N = 3) occurred in the last two sessions.

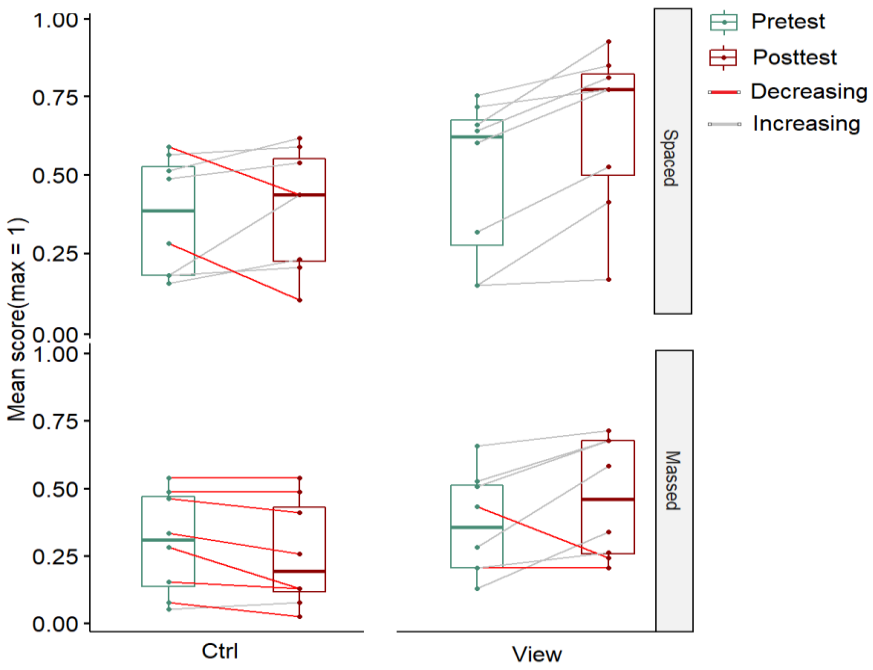
Spoken form recognition

Neither random slopes of time for items, $\chi^2(2) = 5.39$, $p = .067$, nor random slopes of time for participants, $\chi^2(2) = 5.14$, $p = .076$, contributed significantly to the model. Both, however, were retained in the model in favour of the recommended maximal random effects structure. Overall, there was a significant interaction between group and time, $\chi^2(1) = 17.06$, $p < .001$. The odds of the View group recognising the spoken form were two times as high compared to the Control group (*OR* = 1/Exp(B) = 1/0.48 = 2.08, 95% CI [0.34, 0.68]).

The spoken form recognition data comparing accuracy between spaced and massed items for Control and View participants performance (Figure 5.6) showed a marked drop in performance on massed words for the Control group. Recognition in the View group for spaced words was higher but showed a similar pattern than for massed words.

Figure 5. 6

Mean Accuracy in Spoken Form Recognition



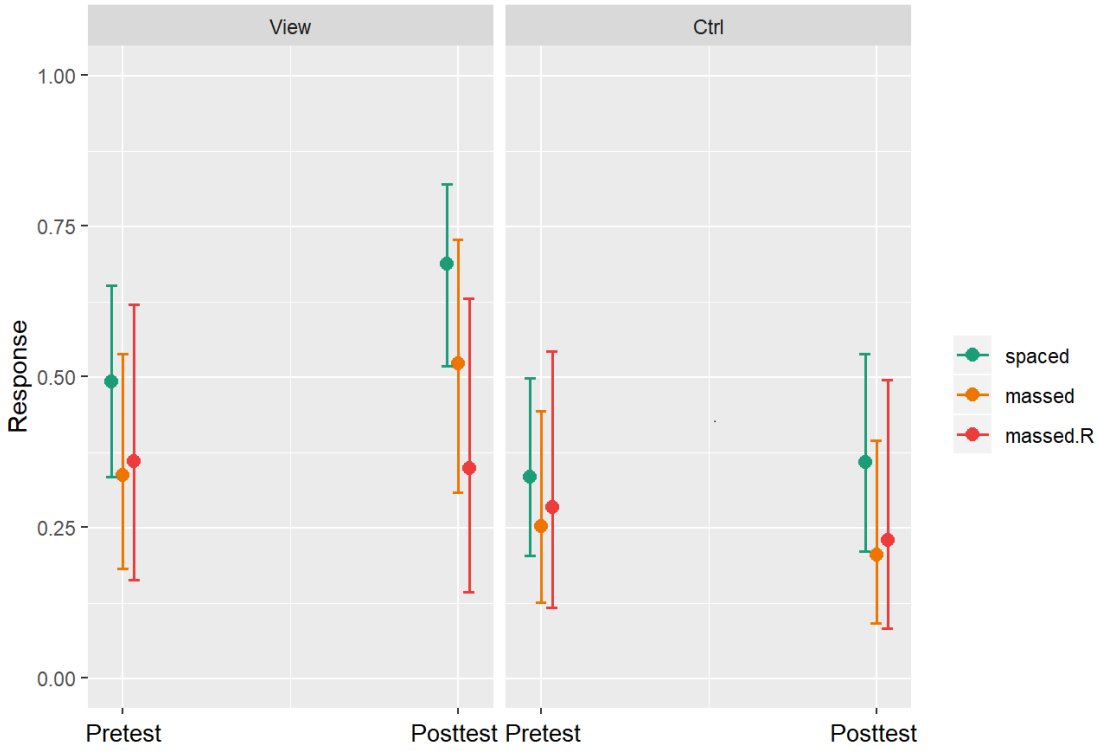
Note. The paired boxplots (by word) show mean accuracy scores of spoken form recognition across Control (N = 34) and View (N = 53) groups for 16 nouns. Half items (N = 8) were spaced over four viewing sessions of documentary series, the second half of the items were massed in one of the session. Grey and red lines match mean scores from pretest to posttest.

The analysis did not show a spacing advantage, as was indicated by the non-significant interaction between spacing and time of test, $\chi^2(1) = 2.56$, $p = .110$. There was also no evidence of a three-way interaction between spacing, time, and group, $\chi^2(2) = 2.16$, $p = .339$. Coefficients estimates on response accuracy from spoken form recognition and written form recognition models are reported in Table 5.6.

Again, the follow-up analysis indicated that recency did not change prediction results for spacing, neither in its interaction with time, $\chi^2(2) = 5.53$, $p = .063$, nor in its interaction with time and group, $\chi^2(4) = 7.41$, $p = .116$ (Figure 5.7). The results indicate that recognition of spoken form was not affected by spacing conditions.

Figure 5. 7

Spacing and Recency Effects in Spoken Form Recognition



Note. The predicted probabilities plot shows the probability values for accuracy of response on spoken form recognition of 16 nouns by Time and spacing conditions, in View (N = 53) and Control (N = 34) groups, calculated from GLM logistic regression analysis. massed.R = massed recent. Half items (N = 8) were spaced over four viewing sessions of documentary series. Massed items (N = 5) occurred in the first two sessions, massed recent items (N = 3) occurred in the last two sessions.

Table 5. 6

GLM Logistic Regression Predicting Form Accuracy from Spacing

	Spoken form recognition					Written form recognition				
Parameters	*B*	*SE*	*z*	*p*	*OR*	*B*	*SE*	*z*	*p*	*OR*
Fixed effects										
Intercept	−0.03	0.34	−0.09	.933	0.97	0.30	0.40	0.75	.456	1.34
Group = Ctrl	−0.66	0.19	−3.42	***	0.52	−2.86	0.40	−7.08	***	0.06
Group = View										
Time = Posttest	−0.82	0.19	4.41	***	2.27	0.08	0.50	0.16	.876	1.08
Time = Pretest										
Spacing = massed	−0.61	0.47	−1.31	0.19	0.54					
Spacing = spaced										
Group (Ctrl) × Time (Posttest)	−0.72	0.25	−2.94	**	0.49	−0.57	0.36	−1.59	.112	0.56
Group (Ctrl) × spacing (massed)	0.27	0.25	1.13	.259	1.31	0.36	0.20	1.79	.074	1.44
Time (Posttest) × spacing (massed)	−0.35	0.26	1.36	.174	0.70	0.36	0.36	0.99	.323	1.43
Group (Ctrl) ×Time (Posttest) × spacing (massed)	−0.03	0.35	−0.09	.933	0.97					
Random effects				Variance	*SD*				Variance	*SD*
Participant = intercept				0.19	0.44				1.05	1.02
Participant = Posttest				0.00	0.05				0.86	0.93
Item = Intercept				0.78	0.88					
Item = Posttest				0.07	0.27					

Note. Posttest ~ group × time × spacing + (Time|participant) + (Time|item). Model fitted to 2944 observations across 16 nouns. N = 92.
**p < .01. ***p <.001.

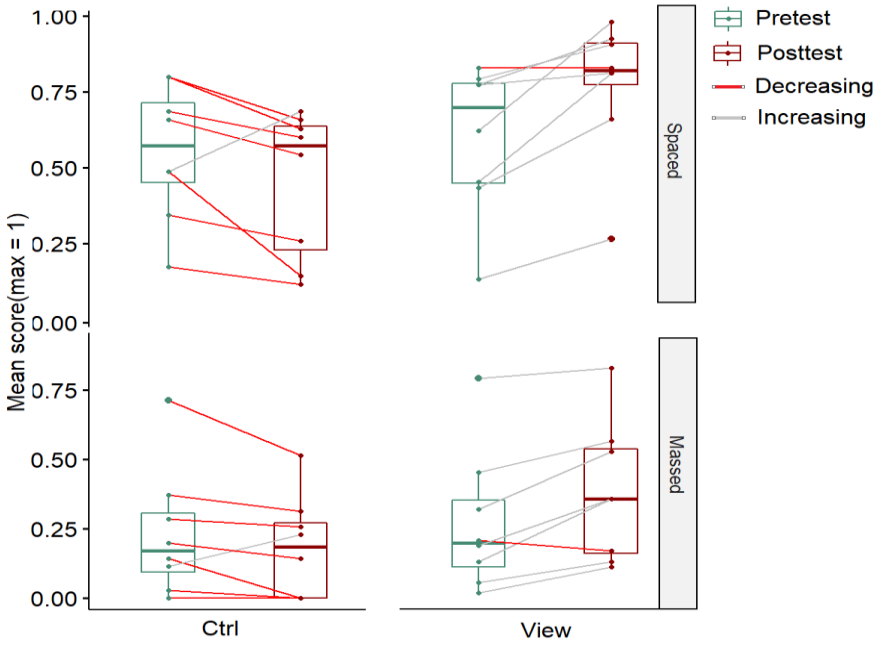
Written form recognition

Random slopes of time for items did not contribute significantly to the model, $\chi^2(2) = 4.06$, $p = .132$, whereas random slopes of time for participants did, (2) = 25.16, $p < .001$, both were retained in the model opting for a maximal random effects structure. Overall, written form accuracy depended on documentary viewing, as was indicated by a significant interaction between group and time, $\chi^2(1) = 45.55$, $p < .001$. The odds of recognising a written form were more than four times higher in the View group compared to the Control group (*OR* = 1/Exp(B) = 1/0.24 = 4.17, 95% CI [0.15, 0.34]).

The mean accuracy scores comparing written form recognition performance between spaced and massed items (see Figure 5.8) showed, for the Control group, a marked drop in scores for spaced and massed words. The View group showed a high level of accuracy in both conditions.

Figure 5. 8

Mean Accuracy in Written Form Recognition



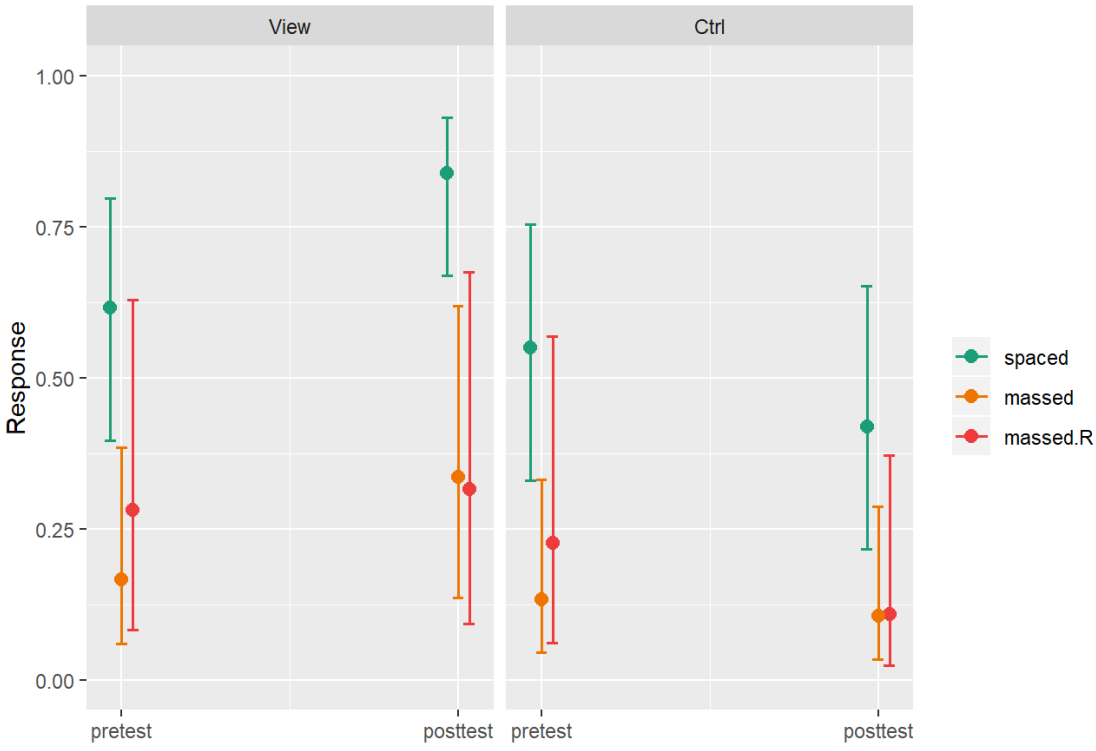
Note. The paired boxplots (by word) show mean accuracy scores of written form recognition across Control (N =34) and View (N = 53) groups for 16 nouns. Half items (N = 8) were spaced over four viewing sessions of documentary series, the second half of the items were massed in one of the session. Grey and red lines match mean scores from pretest to posttest.

The analysis results (Table 5.6) showed that the interaction between spacing and time, $\chi^2(1) = 1.11$, $p = .291$. and between spacing, time, and group, $\chi^2(2) = 3.48$, $p = .175$, were non-significant.

Similarly, the follow-up of whether recency of massed items influenced results (Figure 5.9) showed that revealed no significant interactions, neither between spacing and time, $\chi^2(2) = 5.82$, $p = .054$, nor between spacing, time, and group, $\chi^2(4) = 3.60$, $p = .463$. That is, the spacing conditions did not predict accuracy in the written form test.

Figure 5. 9

Spacing and Recency Effects in Written Form Recognition



Note. The predicted probabilities plot shows the probability values for accuracy of response on written form recognition of 16 nouns by Time and spacing conditions, in View (N = 53) and Control (N = 34) groups, calculated from GLM logistic regression analysis. massed.R = massed recent. Half items (N = 8) were spaced over four viewing sessions. Massed items (N = 5) occurred in the first two sessions, massed recent items (N = 3) occurred in the last two sessions.

5.5.2 Research Question 2

Does any spacing effect vary as a function of the presence of imagery?

Following the analysis procedure detailed above, models with and without word related covariates were compared in an attempt to maintain model parsimony and lower standard errors. Unlike Research Question 1, however, removing all covariates caused a change in the significance of the variable of interest (spacing). Due to the high number of covariates previously considered, only treatment-related variables (i.e., verbal frequency of occurrence of words and related forms) and their interaction with group were retained in the model; this structure maintained the substantive results from the full models. To provide an overview of the findings (Table 5.7), meaning recall and meaning recognition measures revealed a spacing advantage that was stronger in the Non-View group compared to the View group, and a massing advantage that was stronger in the View group compared to the Non-View group. The results held even when the potential effect of recent items was isolated.

Table 5. 7

Summary of Research Question 2 Findings

Models	Main model	Recency model
Meaning recall	**$\chi^2(1) = 9.71$****	**$\chi^2(1) = 18.03$*****
Meaning recognition	**$\chi^2(1) = 8.12$****	**$\chi^2(1) = 18.09$*****
Spoken form recognition	$\chi^2(2) = 0.89$	$\chi^2(4) = 1.08$
Written form recognition	$\chi^2(2) = 4.53$	$\chi^2(4) = 4.53$

Note. Models included View and Non-View groups data; meaning results based on Spacing × Group interaction, form results based on Spacing × Group × Time interaction.

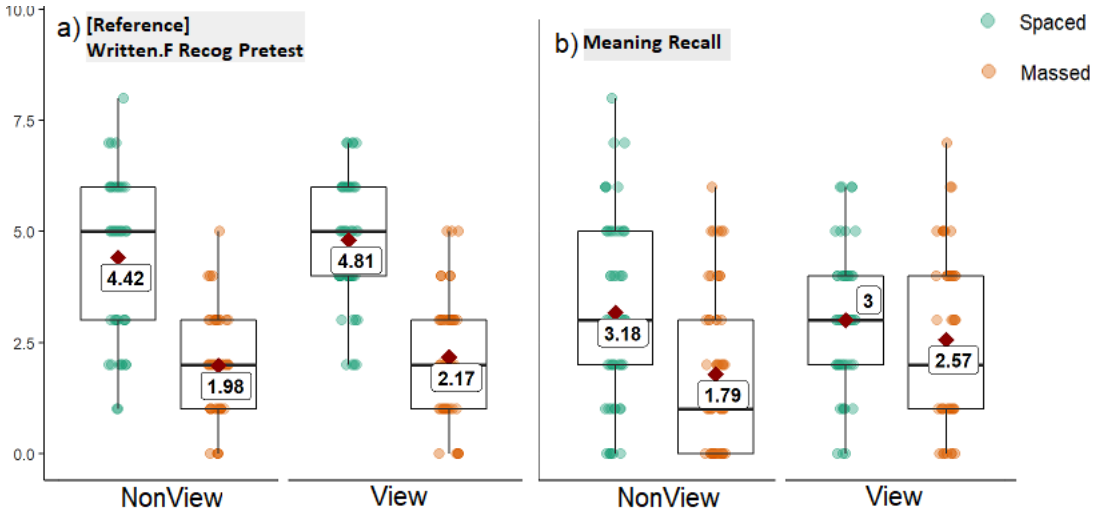
Meaning recall

The analysis showed no statistical difference between the two groups (View and Non-View) in the overall meaning recall, $\chi^2(1) = 1.06$, $p = .304$.

The mean accuracy scores in meaning recall comparing performance on spaced and massed items for the View and Non-View groups are shown in Figure 5.10. The analysis showed that the presence of spacing did not affect meaning recall overall, $\chi^2(1) = 1.92$, $p = .166$. However, there was a significant improvement in model prediction following the addition of the two-way interaction between group and spacing, $\chi^2(1) = 9.71$, $p = .002$, indicating that the spacing effect was different between groups. A negative estimate for group (Non-View) × spacing (massed) indicated that a spacing advantage was significantly stronger in the Non-View group, compared to the View group, while a massing advantage was significantly robust in the View group. This finding was confirmed by an additional analysis in which massed items were taken as the reference category. Coefficients estimates for response accuracy from meaning recall and recognition models are presented in Table 5.8.

Figure 5. 10

Mean Accuracy in Meaning Recall



Note. The boxplots show mean accuracy scores in written form recognition pretest (a) and meaning recall posttest (b), each by subject, across Non-View (N = 57) and View (N = 53) groups for 16 nouns. Half items (N = 8) were spaced over four viewing sessions of documentary series, the second half of the items were massed in one of the session. Means are represented in the figure by the red points. Meaning was not pretested to prevent prior exposure bias, written form recognition pretest (a) was used as a baseline reference.

Table 5. 8

GLM Logistic Regression Predicting Meaning Accuracy from Spacing

	Meaning recall					Meaning recognition				
Parameters	*B*	*SE*	*z*	*p*	*OR*	*B*	*SE*	*z*	*p*	*OR*
Fixed effects										
Intercept	−1.52	0.57	−2.66	**	0.22	−0.26	0.53	−0.50	.619	0.77
Group = Non-View	1.01	0.40	2.53	*	2.74	1.09	0.38	2.85	**	2.99
Group = View										
Spacing = massed	−0.21	0.44	−0.48	.632	0.81	0.04	0.40	0.09	.925	1.04
Spacing = spaced										
Written. F	0.44	0.20	2.21	*	1.55	0.41	0.20	2.10	*	1.51
Verbal freq	0.06	0.03	1.99	*	1.06	0.05	0.03	1.76	.078	1.05
Related forms	−0.06	0.03	−1.93	.054	0.94	−0.03	0.03	−1.28	.199	0.97
Group (Non-View) × spacing (massed)	−0.80	0.26	−3.14	**	0.45	−0.68	0.24	−2.87	**	0.50
Group (Non-View) × Written. F	0.48	0.27	1.78	.076	1.61	−0.18	0.26	−0.70	.483	0.83
Group (Non-View) × verbal freq	−0.07	0.02	−4.18	***	0.93	−0.05	0.02	−3.05	**	0.95
Group(Non-View) ×related forms	−0.01	0.02	−0.29	.775	0.99	0.00	0.02	0.17	.865	1.00
Random effects				Variance	*SD*				Variance	*SD*
Participant (intercept)				1.07	1.04				1.19	1.09
Item (intercept)				0.62	0.79				0.50	0.71

Note. Posttest ~ group × spacing + written form pretest × group + (1|participant) + (1|item). Model fitted to 1760 observations across 16 nouns. N = 110.
**p < .01. ***p <.001.

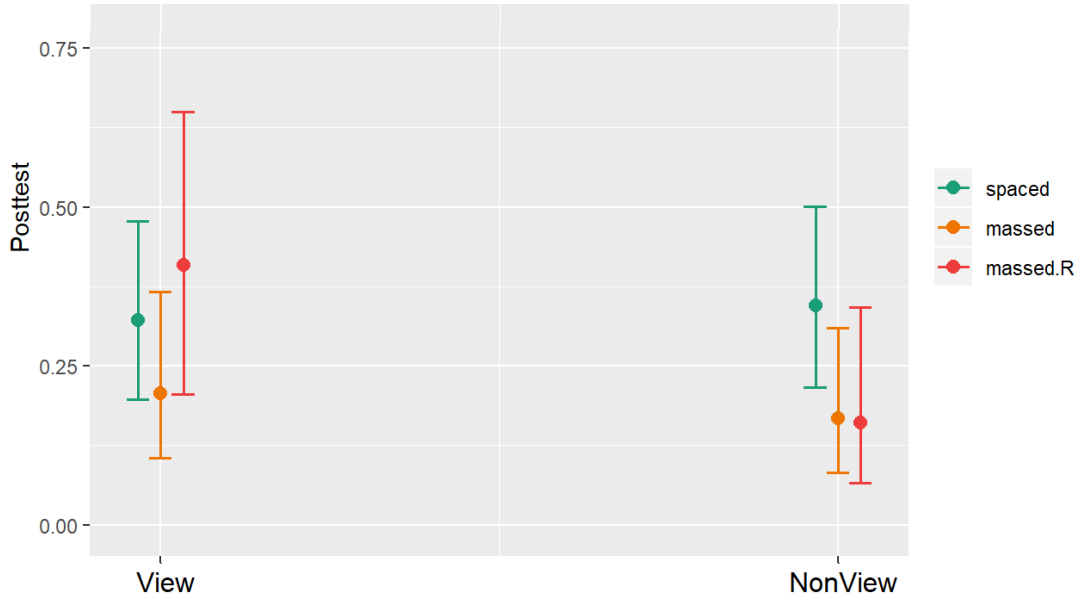
The follow-up pairwise analysis of the interaction between group and spacing revealed that meaning of spaced and massed words was recalled equally well in the View group (B = 0.21, SE = 0.44, z = 0.48, p = .632). There was, however, a significant advantage for spaced words in the Non-View group (B = 1.02, SE = 0.45, z = 2.28, p = .023).

Similar to the previous finding, the follow-up analysis of whether the results were affected by items occurring in sessions close to the posttest revealed a non-significant main effect of spacing, $\chi^2(1) = 2.6$, p = .272, and a significant interaction of spacing with group, $\chi^2(1) = 18.03$, $p < .001$. However, there was no significant group difference in learning spaced items and massed non-recent items (B = -0.37, SE = 0.30, z = -1.24, p = .214). Pairwise comparisons on the interaction as well did not show a significant difference between learning items in these two conditions, within View (B = 0.60, SE = 0.49, z = 1.22, p = .221) and Non-View groups (B = 0.60, SE = 0.57, z = -0.66, p = .790) (The p-values for these comparisons were compared against a Bonferroni-corrected α of .017).

On the other hand, learning spaced and massed recent items was significantly different in the two groups (B = -1.39, SE = 0.33, z = -4.23, $p < .001$). A negative estimate for the interaction term group (Non-View) × spacing (massed) indicated that the probability of a correct response for spaced items compared to massed recent items was higher in the Non-View group compared to the View group while the probability of a correct massed recent item, relative to spaced items, was higher in the View group than in the Non-View group (see Figure 5.11). Although pairwise comparisons on the interaction did not reveal a significant difference in learning spaced and massed recent items in either of the two groups; View (B = -0.37, SE = 0.57, z = -0.66, p = .511) and Non-View (B = 1.01, SE = 0.58, z = 1.75, p = .081) (The p-values for these comparisons were compared against a Bonferroni-corrected α of .017). According to these results, findings could be affected by the recency of items. To explain, the significant interaction formerly obtained had its origin in a group difference regarding the difference between learning spaced items and massed items encountered in the last two sessions (i.e., sessions 3+4).

Figure 5. 11

Spacing and Recency Effects in Meaning Recall

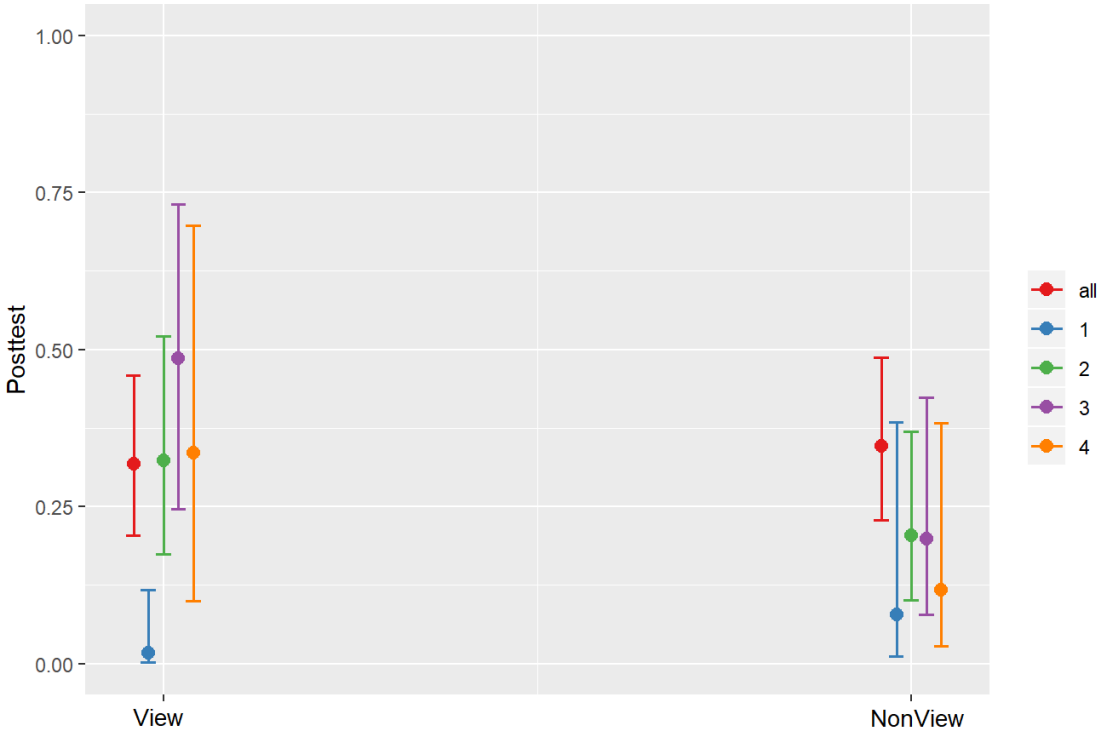


Note. The predicted probabilities plot shows the probability values for posttest accuracy on meaning recall of 16 nouns by spacing conditions, in View (N = 53) and Non-View (N = 57) groups, calculated from GLM logistic regression analysis. massed.R = massed recent. Half items (N = 8) were spaced over four viewing sessions of documentary series. Massed items (N = 5) occurred in the first two sessions, massed recent items (N = 3) occurred in the last two sessions.

However, this supposition can be disconfirmed by the first pre-aggregating analysis conducted on the probability of a correct score on items occurring in one session (massed: 1, 2, 3, 4) compared to items occurring in all sessions (spaced: all) Although session 3 was held towards the end of the experiment, it occurred two weeks before testing; thus, it was not entirely recent. The pre-aggregating analysis which showed a significant interaction between group and session, $\chi^2(4) = 25.04$, $p < .001$, revealed a significant group difference between knowledge of spaced items and items massed in session 3 ($B = -1.46$, $SE = 0.39$, $z = -3.73$), indicating that the results previously obtained still hold true and were not affected by the recency of items. Although the analysis reported here generated wide confidence intervals due to the small number of items in each session, it isolated the effect of recent items and provided a more detailed report of the significant group difference in spacing and massing advantage (see Figure 5.12).

Figure 5. 12

Session Effects in Meaning Recall



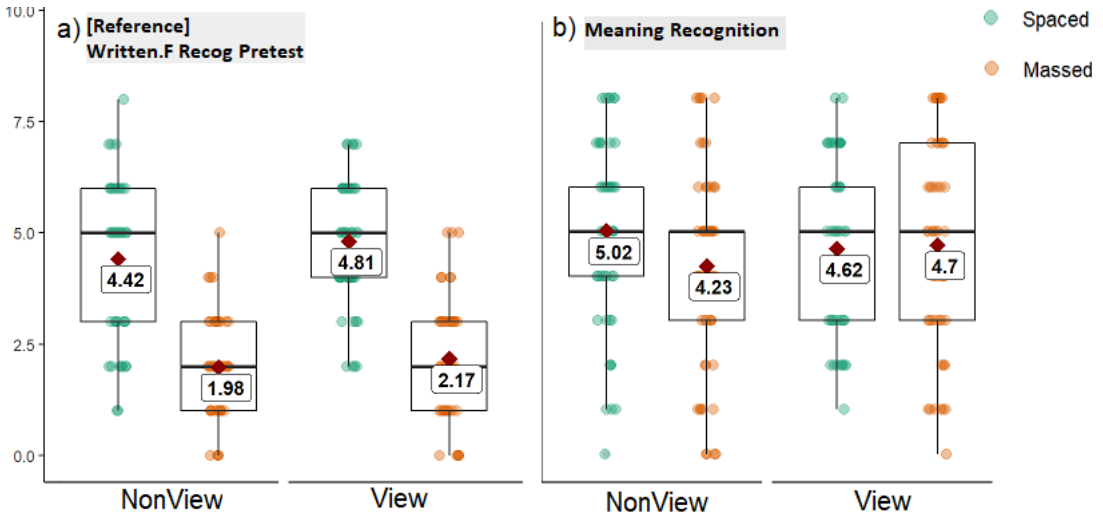
Note. The predicted probabilities plot shows the probability values for posttest accuracy on meaning recall of 16 nouns by session, in View (N = 53) and Non-View (N = 57) groups, calculated from GLM logistic regression analysis. all (N = 8) were spaced over four viewing sessions of documentary series. Session 1 (N = 1), session 2 (N = 4), session 3 (N = 2), session 4 (N = 1).

Meaning recognition

There was no significant difference in overall response accuracy between the View and Non-View groups, $\chi^2(1) = 0.02$, $p = .886$.

Figure 5. 13

Mean Accuracy in Meaning Recognition



Note. The boxplots show mean accuracy scores in written form recognition pretest (a) and meaning recognition posttest (b), each by subject, across Non-View (N = 57) and View (N = 53) groups for 16 nouns. Half items (N = 8) were spaced over four viewing sessions of documentary series, the second half of the items were massed in one of the session. Means are represented in the figure by the red points.

The data for meaning recognition accuracy comparing performance on spaced and massed items for the two experimental groups are shown in Figure 5.13. The analysis showed that spacing was not significant as a main effect, $\chi^2(1) = 0.72$, $p = .397$, but was involved in a significant two-way interaction with group, $\chi^2(1) = 8.12$, $p = .004$. The interaction model (Table 5.8) showed that the spacing effect was different between the groups (the Non-View group was more likely to recognise the meaning of spaced words relative to massed words than the View group). Although pairwise comparisons showed that meaning of words in the two spacing conditions was recognised equally well in the View group (B = -0.04, *SE* = 0.40, *z* = -0.09, *p* = .925) and the Non-View group (B = 0.65, *SE* = 0.40, *z* = 1.62, *p* = .110).

Similarly, the follow-up analysis of the recency effect showed no significant main effect of spacing, $\chi^2(2) = 1.42$, $p = .492$, but a significant interaction between spacing and group, $\chi^2 (2) = 18.09$, $p < .001$. The analysis revealed that there was no significant difference between groups in gained spacing advantage against gained massing advantage of non-recent items (i.e., items occurring in the first two sessions) ($B = -0.29$, $SE = 0.27$, $z = -1.07$, $p = .284$). Pairwise comparisons between the two conditions were not significant either within both the View ($B = 0.34$, $SE = 0.45$, $z = 0.75$, $p = .454$) and Non-View groups ($B = 0.63$, $SE = 0.45$, $z = 1.41$, $p = .160$) (The p-values for these comparisons were compared against a Bonferroni-corrected α of .017).

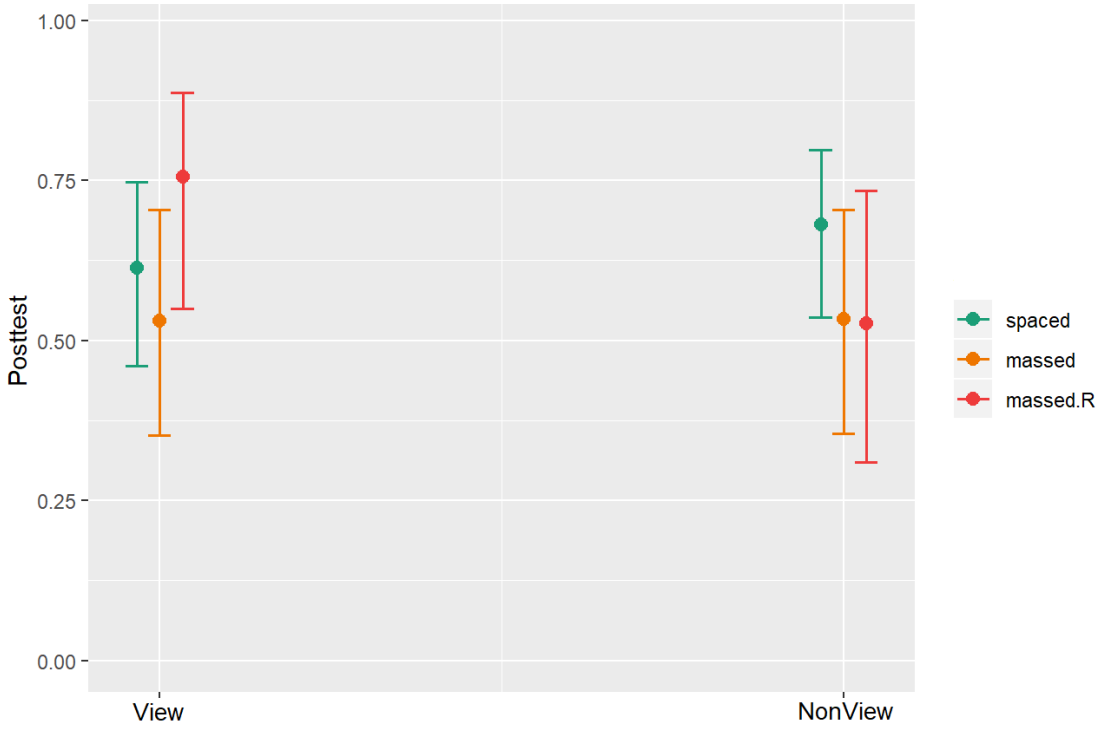
However, the significant interaction was explained by a significant group difference in learning spaced and massed recent items (i.e., items occurring in the last two sessions) ($B = -1.32$, $SE = 0.31$, $z = -4.21$, $p < .001$). The probability of recognising spaced words meanings (compared to massed recent words) was higher in the Non-View group compared to the View group, while the probability of a correct response on massed recent items (compared to spaced items) was higher when imagery was retained in the video (View group). Though pairwise comparisons of the interaction showed that meaning of spaced and massed recent words was recognised equally well in the View group ($B = -0.67$, $SE = 0.53$, $z = -1.25$, $p = .210$) and the Non-View group ($B = 0.65$, $SE = 0.53$, $z = 1.24$, $p = .214$) at the Bonferroni-corrected level ($\alpha = .017$).

Similar to the results of meaning recall, the recency analysis results for meaning recognition data suggest that the significant interaction in the original model occurred due to a significant difference between knowledge of spaced words and knowledge of words massed within recent sessions (sessions 3+4). In other words, the spacing results were affected by the recency of items. However, session 3 was not completely recent; it occurred two weeks before testing. In the initial pre-aggregation analysis, the model specified spacing predictor as a categorical variable of five levels (all, 1, 2, 3, 4) indicating massed words' positions in sessions, and showed a significant interaction between group and session, $\chi^2(4) = 26.61$, $p < .001$. The analysis showed that the advantage of massed words in session 3 over spaced words was significantly stronger in the View group compared to the Non-View

group (*B* = -0.93, *SE* = 0.38, *z* = -2.48. *p* < .050). Although this analysis using session models generated wide confidence intervals due to the small number of items in each session, it suggested that there was a difference between the two groups in learning spaced and massed items (see Figure 5.14).

Figure 5. 14

Spacing and Recency Effects in Meaning Recognition



Note. The predicted probabilities plot shows the probability values for posttest accuracy on meaning recognition of 16 nouns by spacing conditions, in View (N = 53) and Non-View (N = 57) groups, calculated from GLM logistic regression analysis. massed.R = massed recent. Half items (N = 8) were spaced over four viewing sessions of documentary series. Massed items (N = 5) occurred in the first two sessions, massed recent items (N = 3) occurred in the last two sessions.

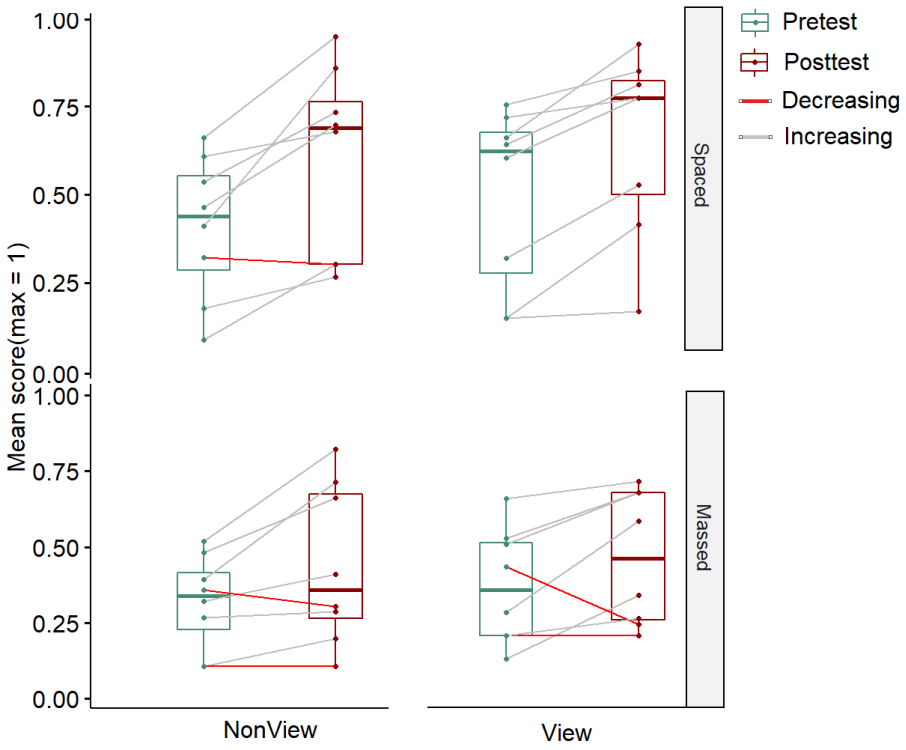
Spoken form recognition

Random slopes of time for items contributed significantly to the model, $\chi^2(2) = 19.05$, $p < .001$, while random slopes of time for participants did not, $\chi^2(2) = 3.69$, $p = .158$. Both random slopes were retained in the model in favour of the recommended maximal random effects structure. Overall, the two groups did not differ in spoken form recognition accuracy as the interaction between group and time was non-significant, $\chi^2(1) = 0.73$, $p = .394$.

The mean accuracy scores for spoken form recognition (Figure 5.15) showed that accuracy was generally similar for spaced and massed words. The Non-View and View groups demonstrated comparable performance. The analysis results showed that spacing conditions did not affect the change in spoken form recognition

Figure 5. 15

Mean Accuracy in Spoken Form Recognition



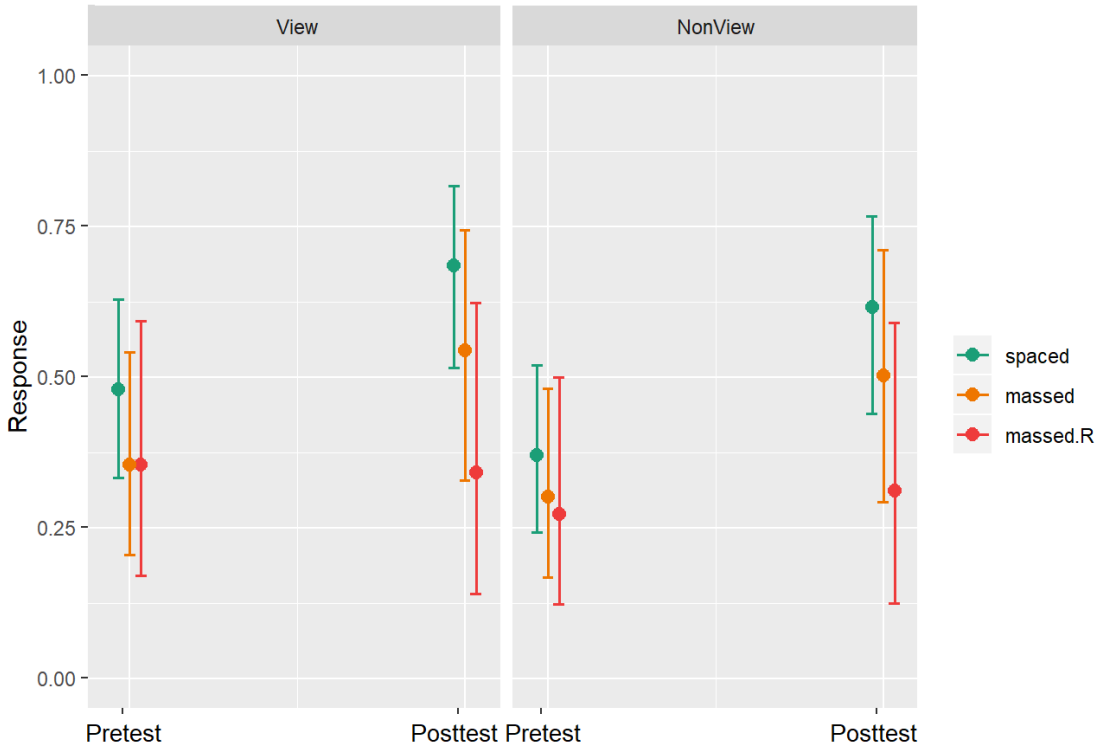
Note. The paired boxplots (by word) show mean accuracy scores of spoken form recognition across Non-View (N = 57) and View (N = 53) groups for 16 nouns. Half items (N = 8) were spaced over four viewing sessions of documentary series, the second half of the items were massed in one of the session. Grey and red lines match mean scores from pretest to posttest.

accuracy overall, $\chi^2(2) = 2.75$, $p = .253$. Three-way interaction of spacing with group and time was also non-significant, $\chi^2(2) = 0.89$, $p = .640$. As such, no further analysis was warranted. The model results for spoken form recognition and written form recognition models are reported in Table 5.9.

The follow-up analysis of the potential influence of imagery on the spacing effect revealed similar results (Figure 5.16). Spacing did not interact with time, $\chi^2(4) = 6.61$, $p = .158$, nor with time and group, $\chi^2(4) = 1.08$, $p = .898$. Hence, a benefit of spacing or massing in any of the two groups could not be identified in spoken form recognition analysis.

Figure 5. 16

Spacing and Recency Effects in Spoken Form Recognition



Note. The predicted probabilities plot shows the probability values for accuracy of response on spoken form recognition of 16 nouns by Time and spacing conditions, in View (N = 53) and Non-View (N = 57) groups, calculated from GLM logistic regression analysis. massed.R = massed recent. Half items (N = 8) were spaced over four viewing sessions of documentary series. Massed items (N = 5) occurred in the first two sessions, massed recent items (N = 3) occurred in the last two sessions.

Table 5. 9

GLM Logistic Regression Predicting Form Accuracy from Spacing

	Spoken form recognition					Written form recognition				
Parameters	*B*	*SE*	*z*	*p*	*OR*	*B*	*SE*	*z*	*p*	*OR*
Fixed effects										
Intercept	−0.39	0.54	0.73	.468	1.48	0.32	0.88	0.36	.719	1.37
Group = Non-View	−0.27	0.25	−1.12	.268	0.76	−0.18	0.28	−0.66	.508	0.83
Group = View										
Time = Posttest	−0.87	0.23	3.72	***	2.39	1.22	0.26	4.69	***	3.38
Time = Pretest										
Spacing = massed	−0.51	0.44	−1.18	.237	0.60	−1.79	0.71	−2.53	*	0.17
Spacing = spaced										
Verbal freq	−0.05	0.03	−1.56	.119	0.96	−0.02	0.05	−0.36	.721	0.98
Related forms	0.04	0.03	1.45	.146	1.04	0.08	0.05	1.56	.120	1.07
Group (Non-View) × Time (Posttest)	0.14	0.23	0.59	.554	1.15	−0.08	0.25	−0.31	.756	0.93
Group (Non-View) × spacing (massed)	0.15	0.22	0.70	.485	1.17	0.13	0.25	0.51	.611	1.14
Time (Posttest) × spacing (massed)	−0.35	0.26	1.36	.174	0.70	0.45	0.37	−1.21	.228	0.64
Group (Non-View) × verbal freq	−0.01	0.01	−0.94	.350	0.99	−0.02	0.01	−1.26	.210	0.98
Group (Non-View) × related forms	−0.01	0.01	−0.55	.586	0.99	0.02	0.01	1.75	.081	1.02
Group (Non-View) ×Time (Posttest) × spacing (massed)	−0.00	0.31	−0.01	.996	1.00	0.41	036	1.15	.250	1.51
Random effects				Variance	*SD*				Variance	*SD*
Participant = intercept				0.27	0.82				0.45	0.67
Participant = Posttest				0.12	0.34				0.02	0.13
Item = Intercept				0.62	0.79				1.75	1.32
Item = Posttest				0.20	0.45				0.25	0.50

Note. Posttest ~ spacing × time × group + verbal freq × group + related forms × group (Time|participant) + (Time|item). Model fitted to 3488 observations for spoken form (N = 109) and 3520 observations for written form (N = 110), across 16 nouns.

**p < .01. ***p <.00

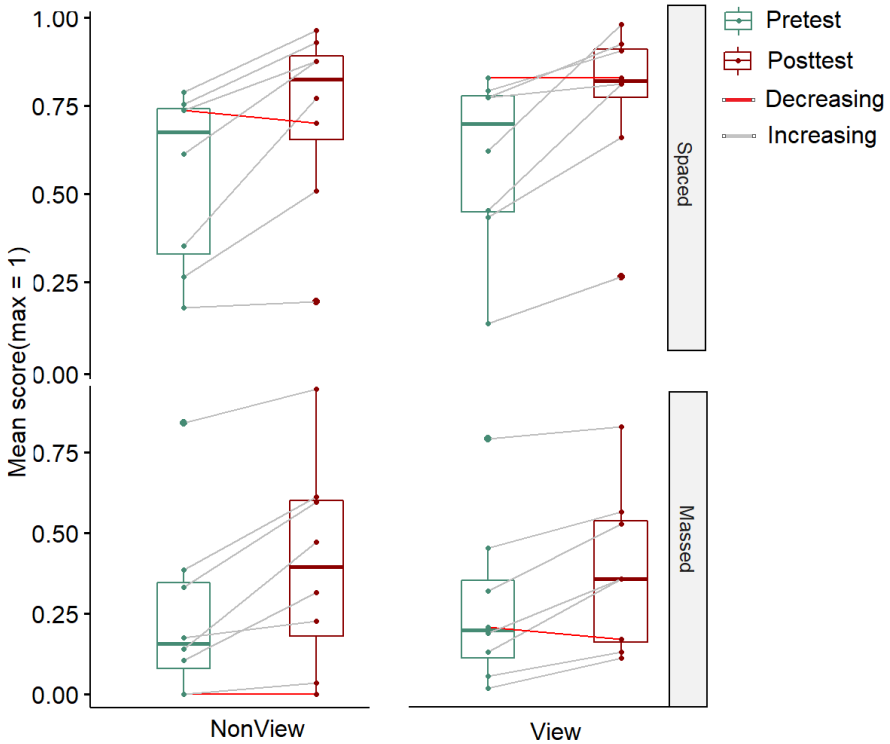
Written form recognition

Random slopes of time were significant for items, $\chi^2(2) = 13.98$, $p = .001$, but not for participants, $\chi^2(2) = 0.37$, $p = .832$. Both were retained in favour of the maximal random effects structure. The difference in accuracy between the two times of test did not differ between the View and Non-View groups, $\chi^2(1) = 0.80$, $p = .371$.

Similar to spoken form, the mean score accuracy data for written form recognition (Figure 5.17) demonstrated improved performance on the posttest for both spaced and massed words. The performance appeared to be equally well in the View and Non-View groups. The analysis did not show an effect of spacing between the pretest and the posttest, $\chi^2(1) = 0.48$, $p = .490$. The three-way interaction between spacing, time, and group also did not reach significance, $\chi^2(2) = 4.53$, $p = .104$.

Figure 5. 17

Mean Accuracy in Written Form Recognition

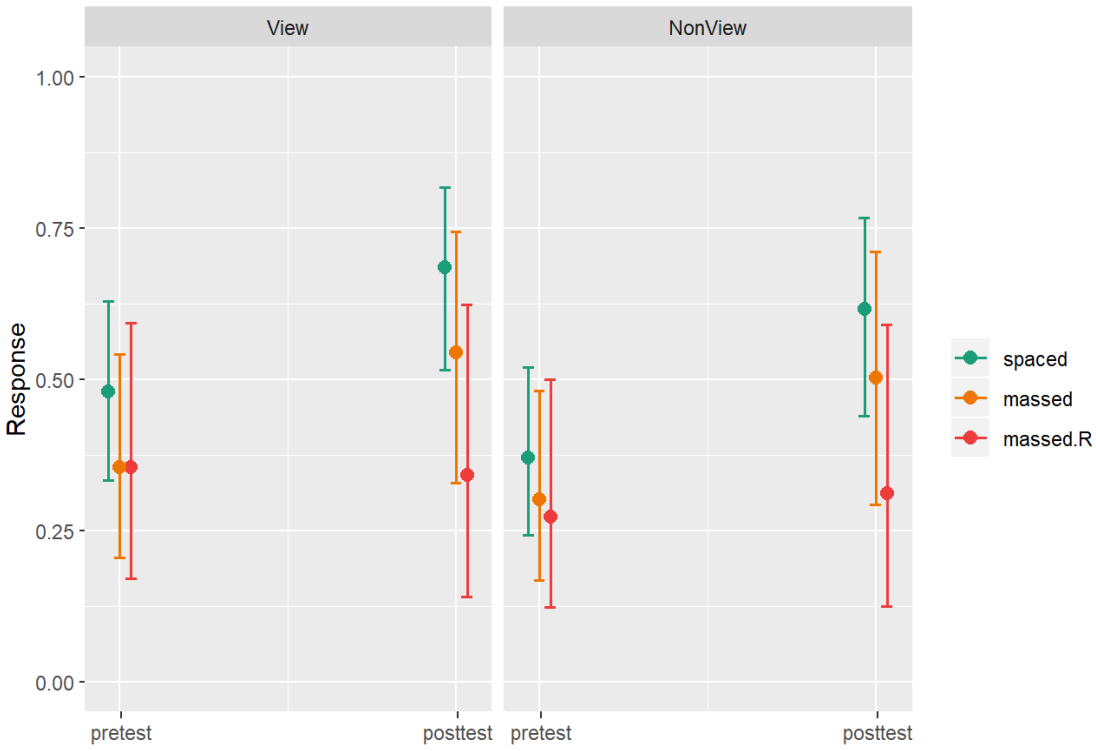


Note. The paired boxplots (by word) show mean accuracy scores of written form recognition across Non-View (N = 57) and View (N = 53) groups for 16 nouns. Half items (N = 8) were spaced over four viewing sessions of documentary series, the second half of the items were massed in one of the session. Grey and red lines match mean scores from pretest to posttest.

Similar results were obtained from the follow-up analysis (Figure 5.18) which examined the spacing effect whilst controlling for the recency of massed items. The interaction between spacing conditions and time was non-significant, $\chi^2(2) = 3.48$, $p = .176$, as was the interaction between spacing, time, and group, $\chi^2(4) = 4.53$, $p = .338$.

Figure 5. 18

Spacing and Recency Effects in Written Form Recognition



Note. The predicted probabilities plot shows the probability values for accuracy on written form recognition of 16 nouns by Time and spacing , in View (N = 53) and Non-View (N = 57) groups, calculated from GLM logistic regression analysis. massed.R = massed recent. Half items (N = 8) were spaced over four viewing sessions of documentary series. Massed items (N = 5) occurred in the first two sessions, massed recent items (N = 3) occurred in the last two session.

5.5.3 Summary of Findings

In summary, the first research question results suggest that the spaced and massed distribution of repeated occurrences across and within extensive viewing sessions, respectively, do not impact incidental acquisition of words. This result is based on comparisons between the View and Control groups data of four measures of word knowledge. The second research question result indicated that the effect of spacing conditions on response accuracy depends on the presence of imagery. This holds true for knowledge of meaning recognition and meaning recall. Participants who had imagery removed (Non-View group) were more likely to benefit from a spacing advantage in comparison with View participants who were, on the other hand, more likely to have an advantage of the massing condition.

5.5.4 Exploratory Analysis: Contigduration

Additional analyses of the subset of the data for meaning recall for the View and Control group (Research Question 1) were carried out. This was done because a massing advantage was obtained from Research Question 2 analysis but not from Research Question 1 analysis. The analysis aimed to explore whether the lack of a massing effect in Research Question 1 analysis resulted from the powerful effect of imagery masking the potential effect of massed distribution. To achieve this, the simplified model from Research Question 1 analysis was implemented, and the effect of contigduration[14] was isolated by entering to the model contigduration and its interaction with group as additional predictor variables. Hence, the model specified posttest accuracy as the dependent variable, group as fixed effect, written form pretest and contigduration as control variables, written form pretest × group and contigduration × group as interaction terms, and participants and words as random effects, with random intercepts allowed to vary across participants and words. The significance of the main effect of spacing and its interaction with group was then assessed using likelihood ratio tests in the same way described in Section 5.4.1.

The results showed that spacing did not affect performance overall, $\chi^2(1) = 0$, $p = .94$. Nevertheless, the interaction between group and spacing was significant,

[14] the amount of time a visual referent is displayed on the screen. See Study 2 in Chapter 4 for details about why and how contigduration was measured.

$\chi^2(1) = 8.47$, $p = .004$, indicating that the effect of spacing conditions was different between groups. A negative estimate for group (Control) × spacing (massed) indicated that a massing advantage was significantly larger in the View group than the Control group, despite the fact that pairwise comparisons showed that meaning of words in the two spacing conditions was recalled equally well in the View group (B = -0.22, *SE* = 0.64, *z* = -0.34, *p* = .732) and the Control group (B = 0.98, *SE* = 0.7, *z* = 1.39, *p* = .170).

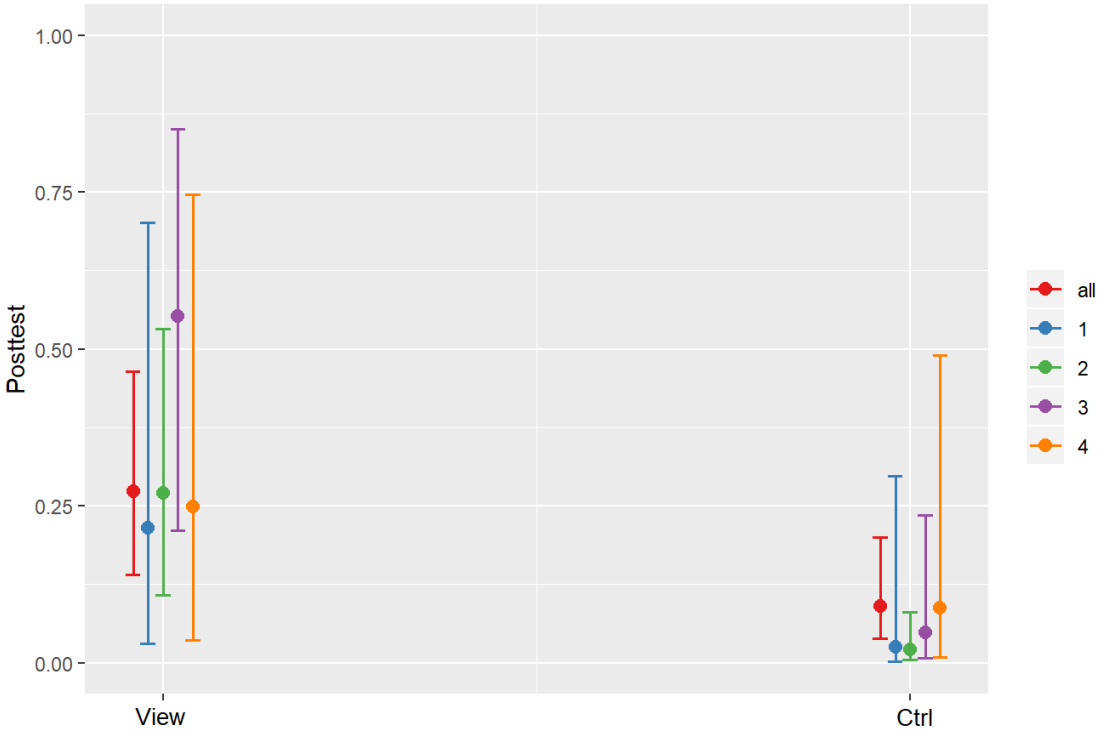
The follow-up analysis of whether the results were affected by the recency of items occurring in the last two sessions revealed a non-significant main effect of spacing, $\chi^2(2) = 1.13$, $p = .569$, but a significant interaction of spacing with group, $\chi^2(2) = 9.03$, $p = .011$. The negative estimate indicated that the View group was more likely to recognise the meaning of massed non-recent words relative to spaced words than the Control group. Though pairwise comparisons showed equal recall of meaning of spaced words and massed non-recent words for both the View group (*B* = 0.09, *SE* = 0.68, *z* = 0.14, *p* = .891) and the Control group (*B* = 1.52, *SE* = 0.81, *z* = 1.88, *p* = .060) (The p-values for these comparisons were compared against a Bonferroni-corrected α of .017).

Interestingly, more exploratory pairwise comparisons using the session model further revealed a lack of a recency effect on spacing results. To explain, comparisons revealed that no significant difference exists between View and Control group in their acquisition of session 4 (i.e., recent) items (*B* = 1.19, *SE* = 0.63, *z* = 1.87. *p* = .061). A similar result was attained when contigduration was eliminated from the model (*B* = 0.67, *SE* = 0.61, *z* = 1.09. *p* = .274). This is a rather surprising outcome since View participants were expected to be more likely to recall the meaning of more recently encountered massed words than Control participants. It must be pointed out, however, that the small number of items per session reduced the statistical power of this type of analysis. These findings are presented in Figure 5.19.

Figure 5. 19

Session Effects in Meaning Recall: contigduration model

Note. The predicted probabilities plot shows the probability values for posttest accuracy on meaning recall of 16 nouns by session, in View (N = 53) and Control (N = 34) groups, calculated from GLM logistic regression analysis. all (N = 8) were spaced over four viewing sessions of documentary series. Session 1 (N = 1), session 2 (N = 4), session 3 (N = 2), session 4 (N = 1).



5.6 Discussion

This study contributes to our understanding of the spacing effect in the context of incidental vocabulary learning from extensive viewing. The study's main aim was to compare the effects of spaced (by 2-week intervals) and massed distributions of repeated word occurrences on incidental L2 vocabulary learning from viewing two full-length seasons of L2 captioned documentary series. The study further examined whether any potential spacing advantage could be influenced by the presence of imagery in the documentary series. This work adds to the growing body of spacing research in several ways: by expanding recent efforts to study spacing effect under incidental learning contexts as opposed to explicit teaching, by shifting focus from unimodal/bimodal input to multimodal input (i.e., viewing), by determining whether spacing effect changes as a function of the presence of imagery, by extending exposure to eight hours of viewing, and by measuring vocabulary learning at the level of four aspects of word knowledge (meaning recall, meaning recognition, spoken form recognition, and written form recognition).

GLM logistic regression analysis showed that, overall, participants' learning following extensive viewing was equivalent for spaced and massed items. Interestingly, however, the follow-up analysis of whether any spacing effect varies as a function of the presence of imagery revealed that Non-View participants, who experienced the documentary series without imagery, demonstrated a spacing advantage relative to the View group. The converse was also true in that the View group demonstrated an advantage on massed words relative to the Non-View group. These results were evident at meaning levels but not at form levels. Exploratory analyses further revealed that, in comparison with the Control group, there was an advantage in the View group for massed words relative to spaced words, but conditioned upon controlling verbal-visual referents of words employing contigduration variable as measured in Study 2.

Do repeated occurrences distributed across multiple extensive viewing sessions of documentary series facilitate incidental L2 vocabulary compared with repeated occurrences massed within a single session?

The first question in this study sought to determine whether better results are found for incidental L2 vocabulary learning when word occurrences are repeated across multiple sessions of extensive viewing (i.e., spacing by 2 week intervals) than when occurrences are repeated within a single session (i.e., massing). GLM logistic regressions were conducted to examine whether spacing conditions predicted the learning. The results confirmed that learning occurred overall, relative to the Control group: participants who viewed eight hours of documentary series were six times and twenty times more likely to recall and recognise meaning, respectively, as well as two times and four times more likely to recognise spoken form and written form, respectively, than participants who only sat the tests.

Surprisingly, however, the results of all measures did not show any statistically significant difference between the two groups regarding their difference in learning spaced words and massed words. Similar results were found when the analysis controlled for the potential effect of items occurring in the last two sessions. This finding that participants from the two groups did not learn spaced words and massed words differently does not support an advantage of spacing over massing. The result supports the work of recent studies in this area, suggesting a lack of spacing effect in the context of incidental vocabulary learning (Rodgers & Webb, 2019; Uchihara et al., 2019). Nonetheless, in contrast to these previous investigations, the first analysis in this study did not demonstrate a massing effect either, within the context of incidental learning from extensive viewing.

The findings from Research Question 1 extend the limited existing studies in the area of research in which it was argued that the spacing effect may not generalise to the context of incidental learning. Webb (2014) postulated that, "although repetition is a factor in incidental learning, its effects may be greatest when repeated encounters occur within a short span" (p. 2). As he argued, the 2-week intervals employed in this study between encounters could have been too long and potentially caused "a decay in knowledge". However, the null result obtained in this first analysis does not support the conclusions of previous studies linking a massed

distribution to better incidental vocabulary learning outcomes. These studies have argued that, continuously encountering of unknown words over a short period would ultimately produce a positive cumulative effect on incidental acquisition (Uchihara et al., 2019).

The presence of imagery might explain the null spacing result. The possibility of immediate recognition of plenty word meanings through visual referents meant that learners were less likely to undergo context-dependent retrieval processes at every new encounter. This in turn suggests that acquisition was less related to spacing and variability in context. Hence, findings for the first analysis suggest that watching the documentary series might have been sufficient to increase students' level of vocabulary irrespective of whether words were spaced over multiple sessions or massed within single sessions. Specifically, the findings lend credence to the hypothesis that a spacing advantage is less likely to be observed in the presence of powerful effects of imagery.

An explanation for the lack of a spacing or massing advantage may lie within a limitation in the between-items design. The observed variation in difficulty among spaced words and massed words, based on the written form pretest, could present a minor shortcoming in the design that might explain the lack of an association between spacing conditions and accuracy in posttests. Although attempts were made to ensure that words in each condition were equally difficult by matching eight spaced nouns to eight massed nouns in terms of learnability, preliminary results showed that participants recognised the written form of spaced words more than the massed words prior to treatment. This finding generates two concerns related to learnability. The first is the opportunity to learn: based on pretest results, a spacing advantage was lacking because participants might have had more room to learn massed words than spaced words. Nonetheless, this difference in opportunities was approximately equal among groups. It is, therefore, difficult to assume that the findings were limited by the negative effect of variation in the opportunity to learn. The second concern is the difficulty to learn. A massing advantage might have not been identified because massed words were more challenging to learn than spaced words as indicated by the pretest results of written form recognition.

Does any spacing effect vary as a function of the presence of imagery?

The second question in this study aimed to examine whether imagery impacted the effect of spaced (by 2-week intervals) and massed word occurrences on incidental L2 vocabulary learning from extensive viewing. GLM logistic regressions were performed to assess the significance of adding an interaction between the treatment group and spacing condition. Participants who viewed eight hours of documentaries with imagery preserved (View group) did not significantly differ in response accuracy from participants who had imagery hidden from view (Non-View). This result was obtained for all four measures of word knowledge (all *ps* > .300). Nevertheless, the advantage of spaced or massed condition was found to be dependent on whether participants viewed or did not view imagery because, in interactions, treatment groups and spacing conditions predicted whether or not participants scored accurately in meaning recall and meaning recognition tests. Negative estimates for the interaction terms group (Non-View) × spacing (massed) indicated that the difference in response accuracy between spaced words and massed words was bigger in the Non-View group than the View group for both meaning recall and meaning recognition, while a massing advantage was stronger in the View group.

The Non-View Group

The result that spacing benefits for incidental vocabulary learning can be achieved from listening-while reading is consistent with the one obtained for long-term retention in Serrano and Huang's (2018) study of incidental learning from listening-while-reading. Meaning recognition results showed that spaced practice contributed to better vocabulary gains. It differs, nonetheless, from the result for short-term retention in the same study which showed that performance of massing group participants was superior to the spacing group. Webb and Chang (2015) also showed no influence of spacing on incidental learning of knowledge of meaning recognition from listening-while-reading. However, variation in participants (secondary school), sample size (N = 61), target words (N = 100), and materials (10 graded readers) make direct comparisons difficult. The result is also in contrast to previous propositions that occurrences that are repeated within a short span are more

likely to be associated with incidental vocabulary learning (Uchihara et al., 2019; Webb, 2014).

This finding for the Non-View group is in line with the theory of contextual variability. Participants in this group listened to and read L2 captions of the episodes and encoded the unknown target word in memory along with its pertaining context. This context could be exemplified by the class setting, the script itself, the linguistic and situational context in which the word occurred. These factors might have helped to preserve the spaced words' memory traces. By the end of the fourth session, the memory traces of words had been more distinct, thus, greatly contributed to the construction of knowledge of meaning recall and recognition. The result corresponds with Koval's (2019) supposition that, when learners process unknown words for comprehension purposes, they are more likely to benefit from repeated occurrences that are widely spaced even if they do not attempt to commit the word to memory. As will be explained next, while both View and Non-View groups experienced changes in contexts, only the Non-View seem to have benefited from the latter. Longer contigdurations in the View group which was previously shown to affect vocabulary learning positively (see Study 2) might have overridden the effect of variation in contexts.

The View Group

The finding that massed distribution enhances incidental vocabulary learning from extensive viewing of documentary series reflects Rodgers and Webb's (2019). The authors found that improved performance is likely for words when occurrences are repeated within a single episode rather than across a range of episodes. The result was also reported by Pujadas Jorba (2019).

The present findings are significant in at least one major respect: the effect of spaced and massed practice in incidental vocabulary learning seems to be input-dependent. The superiority of the spacing advantage in the Non-View group and massing advantage in the View group suggests that previous propositions that the massed condition is superior to the spaced condition in incidental vocabulary learning (Uchihara et al., 2019; Webb, 2014) do not hold for all types of input exposure. The spacing effect may relate to input aspects such as the type of materials in which input occurs or the length of input exposure. Most importantly,

differences in the modes of input, specifically, the presence and absence of imagery, might influence spacing results as was evidenced in the current study.

The result for the View group may partly be explained by the retrieval-effort hypothesis. Research has shown that active retrieval of previously learned information leads to substantial long-term retention, what is often referred to as the testing effect or retrieval-based practice (Bae, Therriault, & Redifer, 2019; Roediger & Butler, 2011; Roediger & Karpicke, 2006). A key finding of the testing effect is the retrieval-effort hypothesis, that learning conditions that introduce difficulties can improve long-term retention (Bjork, 1975; Bjork & Kroll, 2015; Carpenter, 2009; Carpenter & DeLosh, 2006), including retention of L2 vocabulary (Schneider, Healy, & Bourne Jr, 2002). Importantly, cued recall as a means of retrieval practice was found to facilitate long-term retention of new meanings of familiar L2 vocabulary that is unintentionally acquired from reading storybooks or textbooks (Hulme, 2018). It can thus be suggested that, since a spacing advantage was found in the Non-View group but not in the View group, then the absence of imagery might have imposed cued recall for Non-View participants and thus produced some degree of those "desirable difficulties" (Bjork, 1994). To explain, when a word form is recognised in a second context that is lacking visual referents, but meaning recall fails, then the exerted efforts to infer the meaning of the unknown word, after a period of forgetting (2-week interval) and without the assistance of visual cues, constitute a valuable retrieval practice that would likely make the word more salient in every upcoming session. This proposed explanation of the result also accords with the finding that there is an association between fewer retrieval cues and improved long-term retention (Carpenter & DeLosh's, 2006). They suggested that elaborative retrieval processing increases as cue support decreases. As Bjork & Kroll (2015) put it, "The difficulties introduced by variation, spacing, interleaving, and so forth are desirable because responding to those difficulties (successfully) engages the very processes that support learning, comprehension, and remembering." (p. 242).

On the other hand, the lack of a spacing advantage in the presence of imagery may then be the outcome of the relatively good correlation between visual referents and learning which, in turn, entails minimum efforts to retrieve meaning at every new encounter (i.e., session). To illustrate, the target words in this study were often

marked by longer on-screen durations of their visual referents, ranging from 35 seconds to up to 652 seconds (i.e., +10 minutes). It is therefore likely that prolonged contigduration, which is characterised by immediate exposure to the meaning of unknown words via visual referents, might have strengthened and maintained memory traces for form and meaning and the latter had become readily accessible. As such, the effect of spacing was of marginal significance since participants did not necessarily have to encode contextual features at every new session and learning was not, therefore, context-related.

Exploratory analysis: contigduration

An exploratory analysis was perhaps the most useful in furthering our understanding of the nature of the interaction between imagery and spacing. The finding that the View group benefited from a massing advantage that was significantly greater than the Control group when contigduration is added to the model contradicts results from Research Question 1 (Control vs. View, without contigduration). Nonetheless, it is consistent with the second analysis results (Non-View vs. View). This interaction between spacing and the presence of imagery could be seen only when the duration of verbal-visual referents was controlled (i.e., held constant). Therefore, it could be hypothesised that, for View participants, the association between verbal forms and visual referents was strong enough to offset the spacing effect. In other words, verbal-visual contigduration might have masked the massing effect in learning in Research Question 1 analysis. Whilst a massing advantage was observed when contigduration was held constant for all words, it is impossible, in real-world viewing, for visual referents of unknown words to be all of equal duration. Hence, from a theoretical perspective, there seems to be a massing effect in learning from extensive viewing. Nevertheless, these results are not necessarily reflective of everyday practice and perhaps of limited importance in learning from authentic video materials. Interestingly, the finding ruled out the previous supposition that a spacing advantage did not emerge in the View group due to possible shortcomings in the design.

5.7 Conclusion

The spacing effect is the finding that spacing learning farther apart in time tends to reinforce knowledge more strongly than massing learning in a single phase. It is one of the most remarkable phenomena in the field of education; however, most of the body of literature has been documented in laboratory research. The past 10 years have brought a renewed focus on the impact of spacing in real-world educational settings on learning, including the domain of L2 vocabulary instruction. Nonetheless, the phenomenon has been mostly explored in relation to intentional learning, while there have been little published data on its effect in incidental contexts. In particular, only two studies have examined spacing effects in learning from extensive television viewing. The researchers found that word occurrences crammed in single episodes are more salient and likely to be learned than occurrences distributed across multiple TV episodes. However, the finding was based upon data on form-meaning connection and on the relative frequency of occurrence; the number of encounters of the item in the overall episodes in which the item occurs divided by the number of episodes (i.e., range).

To the best of my knowledge, the present investigation is the first controlled manipulation to investigate the spacing effect in learning from extensive TV viewing, particularly viewing two full-length seasons of L2 captioned documentary series. It compares differences in learning when word occurrences are spaced across all experimental sessions and when word occurrences are massed in a single session only. It takes account of learnability by matching items' characteristics in the two conditions (e.g., verbal frequency), the recency effect by applying stricter accountability for the variability in episodes, and specificity in word knowledge by testing meaning recall, meaning recognition, spoken form recognition, and written form recognition. The work also represents the first attempt to determine whether or not spacing effect on incidental L2 vocabulary learning is sensitive to changes in input modalities, specifically, to the presence and absence of imagery in extensive listening-while-reading. The study results are discussed in light of two critical theoretical accounts of spacing effects: contextual variability and retrieval effort.

This investigation extends our knowledge and understanding of the powerful effect of imagery in L2 captioned video. It shows that spacing is a significant

determinant of the ability to recognise and recall newly learnt meanings following extensive exposure to English language (captioned) documentaries. It provides evidence that the presence and absence of imagery in this listening-while-reading context exerts an influence on the spacing effect on the acquisition of knowledge of meaning. Specifically, it identifies the spacing advantage as being more likely without imagery than with imagery, while a massing advantage is likely to be obtained when imagery is retained. The study gives more profound insights into the mechanisms behind the spacing effects in incidental contexts, by suggesting that spacing is especially effective when fewer cues are available. Its effect tends to be achieved in contexts when the learner does not enjoy sufficient support for learning.

The study further accounts for the impact of the presence of imagery using contigduration variable. One of the significant findings to finally emerge from the study is that the massing effect in the View group is influenced by imagery. The presence of visual referents plays a role in the weak link that was initially observed between the massing condition and incidental acquisition of words (in Research Question 1). This is because the positive impact of a massed distribution emerged only when verbal-visual contigduration was included as input to separate its effect. Thus, the overall findings do not compel the conclusion that massed occurrences across extensive viewing sessions augment learning because, in reality, the duration of visual referents in documentary series or an equivalent source of input is never the same for individual vocabulary items. Furthermore, while the study reports almost similar results for knowledge of meaning recall and meaning recognition, spoken form and written form data do not show any significant findings. This suggests that, under extensive bimodal and multimodal input conditions, the phenomenon might be confined to knowledge of meaning only.

Chapter 6

Conclusions and Final Remarks

This final chapter will present a general conclusion to the thesis. I will summarise the findings of each study. I will then introduce the ways in which this thesis contributes to theoretical and methodological knowledge and understanding. I will end the chapter by offering implications for L2 learning and educational practice, acknowledging limitations, and making recommendations for further research.

6.1 Summary of Findings

6.1.1 Chapter 2

The two-part norming study of this thesis was conducted on two samples of students with similar characteristics to tertiary EFL university learners in the Linguistics Bachelor programme, who were targeted in the present thesis. The study showed that a total of 28 items occurring in two full-length seasons of documentary series, Wonders of the Universe and Forces of Nature, were unknown to the sample. These were therefore selected to serve as the target words for the studies of the thesis. Moreover, the study established that the documentary episodes selected for the thesis were within the lexical competence of the target population.

6.1.2 Chapter 3

The first study presented in this thesis assessed the effect of sustained exposure to L2 captioned documentary episodes on incidental acquisition of L2 vocabulary. Third-year EFL university students in the Linguistics Bachelor programme watched two full-length seasons of documentary series, extending to eight viewing hours, via 2 hour long sessions over six weeks at two-week intervals. Measures of word knowledge revealed robust benefits at the level of meaning recall, meaning recognition, spoken form recognition, and written form recognition. Secondly, the learners who watched the episodes in L2 captioned video format were hypothesised to outperform in tests of meaning and spoken form, while learners who were exposed to L2 captions and audio only, were predicted to outperform in written form test. Data painted a different picture from what was speculated as the two groups acquired

words equally well on all dependent measures. The View and Non-View groups demonstrated good comprehension of the episodes as well. Evidence suggests that familiarity with L2 input could have played a role in the comprehension of the episodes and acquisition of target words. In addition, the study showed that the presence of L2 captions does not hamper the viewing process and prolonged exposure to L2 captioned documentary series, with or without imagery, enhanced learners' motivation to learn.

I concluded that the results of Study 1 did not fully explain the role of imagery on incidental vocabulary learning. There were two possible causes for the unexpected null results of Study 1: (1) that imagery did not influence incidental L2 vocabulary learning, (2) that imagery did influence incidental L2 vocabulary learning, but obscuring imagery in the Non-View group stimulated different learning strategies that were equal in strength to the effect of imagery in the View group. In Study 2, I attempted to disambiguate these possibilities by testing hypothesis 1, that imagery did not influence learning, through an examination of contiguity effects in learning.

6.1.3 Chapter 4

In Study 2, I tested my view that the null result in Study 1 was not likely to be attributed to the ineffectiveness of imagery. The study took an alternative approach to examine the role of imagery in vocabulary learning from L2 captioned by assessing the effect of verbal-visual contiguity within episodes. I identified three constructs of interest when attempting to define and explain this effect. The findings clearly supported my proposal in Study 1 that imagery influences incidental L2 vocabulary learning from extensive TV viewing.

The results revealed ∓ 25 seconds as the optimum contiguity timeframe relative to ∓ 7 seconds. This result may be explained by the Hebb repetition effect and cross-situational learning. A long timeframe is more likely to capture learning that is induced by the correct segregation of form-referent pairs resulting from repetitive encounters. This explanation supports the conceptual premise that the maximum timespan before verbal-visual contiguity effect is no longer successful for

a particular verbal-visual co-occurrence is conditioned upon the frequency and quality of previous cumulative verbal-visual encounters with the same pair. For every verbal-visual co-occurrence, the more the preceding identical encounters, the longer it can be separated from its referent and still induce the contiguity effect. The study also showed that the cumulative length of the verbal-visual encounters (contigduration) is a better predictor of incidental L2 vocabulary learning compared with the frequency of the verbal-visual encounters (contigfrequency), while the proportion of verbal occurrences that are accompanied by an image (contigratio) failed to show any effect at all. The importance of contigduration is clearly supported by spoken and written form recognition and meaning recall findings, but is most prominently observed in meaning recognition results.

6.1.4 Chapter 5

The last study presented in this thesis fills gap in spacing research and also highlights the powerful effect of imagery in a third way. Study 3 explored the impact of spaced and massed occurrences in episodes of documentaries series on incidental L2 vocabulary learning through an experimental manipulation of the encountered items.

Non-View participants enjoyed a marked advantage over View participants in their ability to acquire spaced words compared to massed words. The lack of visual support in one session may have created a difficult cueing condition for the learner in the following session. In line with the retrieval-effort hypothesis, this may have stimulated desirable difficulties that bring about growth in vocabulary knowledge. Contextual variability theory, on the other hand, suggests that the lack of visual cues may have encouraged the encoding of contextual features in every new encounter. Overall, the result indicates a role for the absence of cues in producing and enhancing the spacing effect on incidental L2 vocabulary learning. View participants did not demonstrate any spacing advantage, neither when they were compared to those in the Control group nor when compared to those in the Non-View group. The finding suggests that these participants may have benefited from prompt recognition of meaning with the assistance of verbal-visual contigdurations. This factor which was shown to enhance learning in Study 2 might have strengthened memory traces for form and meaning. Learning was, therefore,

less dependent on contextual encoding. However, a massing advantage was significant in the View group compared to the Non-View group but not when compared to the Control group. These results produced a set of conflicting evidence regarding whether or not the View group benefited from massed occurrences compared to spaced occurrences.

I considered the possibility any benefits of massed occurrences were masked by the effectiveness of imagery; this perhaps led me to revisit the first analysis through controlling the contigduration of items. This re-analysis provided a basis for disambiguating the massing advantage in the View group as massing came out as significant compared to the Control group. There are some limitations to the practical implications of this result, however. In authentic audio-visual input, contigduration is never constant (in contrast to the re-analysis when it was statistically controlled), suggesting that a massing effect is less likely in practice as a spacing phenomenon is one of many factors that affect learning. Finally, the combination of findings showed that any influence of distributed and massed occurrences across and within extensive sessions of multimodal or bimodal input exposure is likely to be exerted on knowledge of meaning rather than form.

The arguments above may be reduced to one quintessential point: the impact of spacing conditions in incidental vocabulary learning is input-dependent. The massing advantage is less likely to be observed in the presence of visual cues while a spacing advantage is very likely in their absence. These results further substantiate the conclusions made throughout the chapters of the thesis, that imagery in L2 captioned video is significant to word learning and its absence positively stimulates vocabulary learning behaviours.

6.2 Theoretical Contributions

The importance and originality of this thesis to SLA research lies in the presentation of three studies (Chapters 3, 4, 5) that each highlights the positive influence of imagery on L2 word learning from viewing, in ways that have not been shown before. Study 1 provides new evidence of intermediate learners' potential to acquire L2 words and comprehend input in documentary episodes of 2 hr, presented in L2 captioned video format, which is triple the commonly-adopted length in viewing

research (30 to 40 minutes). The study is, therefore, the first to document incidental acquisition of L2 vocabulary from typical out-of-class TV exposure and weakens the consensus among researchers that prolonged sessions come at the expense of attention or comfort. The evidence is based on 8 hr exposure at two-week intervals and prevails even in the absence of imagery and for meaning recall, meaning recognition, spoken form recognition, and written form recognition measures.

Study 2 offers the most insight to the literature into the effect of imagery in L2 captioned video. It actualises the construct of verbal-visual contiguity in authentic audio-visual input by introducing three sets of quantitative measures which rationalise its effect. This is the first study of substantial duration that pinpoints the association between these measures and incidental L2 word learning of different parts of speech. The principal theoretical implication of this study is that contigduration is the strongest predictor of successful L2 word learning, especially for knowledge of meaning recognition. In the presence of long contigdurations, contigfrequency and contigratio are not of any decisive importance to learning. This result also substantiates, in a new way, that L2 learners do, in fact, process imagery in the presence of captions. The study challenges the assumption underpinning previous designs; that the most optimal verbal-visual contiguity timeframe in audio-visual input is the shortest. The study shows, for the first time, that a $\mp$25 timeframe is the most promising in capturing verbal-visual contiguity effect on incidental L2 vocabulary learning from L2 captioned documentary series. That is 20 seconds longer than the currently employed timeframe in the literature. It also offers viable theoretical mechanisms behind this result, most interestingly, the conditional effect. I posit that there is no absolute timeframe for verbal-visual contiguity effect on vocabulary learning in audio-visual input. For every verbal-visual co-occurrence, the timeframe extends or narrows depending on the frequency and quality of preceding fellow verbal-visual encounters.

Lastly, there are several important areas where Study 3 makes an original contribution to the literature. The study appears to be the first to report a spacing advantage in incidental L2 vocabulary learning from extensive exposure to documentary episodes in the form of bimodal input (captions + audio), based on a carefully planned between-items design. The finding contradicts the recently-held

notion that a massing advantage is likely to dominate in incidental learning contexts. That is, previous results do not generalise to all types of incidental learning. The most significant contribution in Study 3 is the finding that the impact of spaced and massed occurrences in incidental learning is input-dependent. For bimodal input, my study provides evidence that the lack of imagery and visual cues optimises encoding of contextual features and imposes longer processing time of input (four spaced sessions). The retrieval-effort effect may be more pronounced when there are fewer cues to retrieve the meaning of newly known words, generating the spacing advantage. For multimodal input, however, the lack of a spacing advantage indicates that learners do not necessarily process target words regularly in every session. Instead, the presence of visual cues might speed up the learning process during the initial sessions. Also, an advantage for massed occurrences in viewing is significant in theory but not in practice by virtue of superior imagery strength which tends to mask it (contigduration).

An important conclusion to emerge from the overall thesis is that presenting learners with different input sources may stimulate them to acquire words in different ways but not necessarily in different amounts. In sum, the thesis provided the first comprehensive investigation of the role of imagery on incidental L2 vocabulary learning from extensive exposure to television series in the format of L2 captioned video.

6.3 Methodological Contributions

This thesis makes six methodological contributions. First, it highlights that a note of caution is due in the design of experiments. Attempts to study the effect of a specific variable by eliminating it may yield inconclusive results. Second, I established a multidimensional quantitative measurement of verbal-visual contiguity that includes duration, frequency, and ratio, which increasingly advances current studies. Third, the method section in Study 2 fully describes the procedures performed to measure contiguity in the series, unlike former studies. This was done to ensure reproducibility and generalisability of results. Also, setting criteria for what constitutes a visual referent (see Chapter 4, Section 4.4.2) is a clear coding improvement in other research. Specifically, I divided visual referents into two categories: strong and weak referents, and evaluated the impact of including the

latter on the results before carrying out the principal analysis. The results were in favour of the adoption of weak referents. This methodology indicates the robustness of the results and it could hopefully lend itself well for use in upcoming studies. Fourth, the between-items design that was employed for studying the spacing effect is innovative in viewing research. Fifth, I adopted the flemma as the main counting unit of verbal-visual co-occurrences. Nevertheless, I also considered the potential moderating effect of verbal-visual co-occurrences of compounds and derivatives of target words on the contiguity effect results. The inclusion of these related forms in contiguity measurement was found to be promising. Finally, in contrast to previous studies, this thesis handled the interference of the recency effect by accounting for the variability of episodes.

6.4 Pedagogical Implications

The findings of this thesis have several important implications for future practice in the EFL class. A key practice priority may be placed for documentary series. Incidental L2 vocabulary acquisition from viewing L2 captioned documentary series is facilitated when the duration of visual referents that occur close to their corresponding word forms (i.e., contigduration) is higher. Because verbal-visual contiguity is more frequent in documentaries than narrative television such as films and drama series (Rodgers, 2018), documentary viewing should be more often implemented in EFL classes. There is also the motivational advantage. Extensive exposure to documentary series is beneficial for EFL learners because it plays a critical role in increasing their motivation to learn the language, as was evidenced from the debriefing questionnaire.

The thesis findings have direct implications for the role of the EFL teacher. One fascinating result in the context of the present investigation was that the majority of participants were found to be already experiencing a sufficient amount of multimodal input outside the confines of the classroom, especially films and drama series. Based on Webb's out-of-class viewing programme (2015) and the positive learning outcomes in this research, teachers could encourage students to watch more L2 captioned documentary series for recreational purposes (i.e., home viewing). An even more fruitful practice may be to give students viewing-based assignments that promote incidental vocabulary learning. Examples could be answering

comprehension questions on the basis of what is implied in the series, writing weekly reports of single episodes or monthly reports of multiple episodes and present them in front of the class, and lastly, the replication of self-selected scenes in front of the class (i.e., role plays). The combination of these comprehension-based tasks can be perceived as instrumental to the attainment of incidental word learning from viewing. Moreover, although the findings indicate that the presence of imagery fosters motivation and assists learners in rapid engagement with the content, this research does not support recommendations to substitute multimodal mode for bimodal mode. Both approaches to input presentation are functional in the development of word knowledge.

The thesis results go against the consensus among practitioners that extensive exposure to input in a single session would overload EFL learners of intermediate proficiency. Teachers can incorporate sessions of extensive documentary viewing within the intermediate EFL programme without concerns about concentration difficulties. This can be done by devoting classes as long as 2 hr to this type of activity, every week or fortnight, for instance, but keeping in mind three important precautions to alleviate concerns over students' inability to focus. (1) The programs must meet students' preferences. This can be achieved by proposing some titles and selecting one based on the learners' votes. (2) The input must be within learners' lexical capacities; thus, teachers have to analyse the lexical coverage of episodes. (3) viewing must be interspersed with comprehension tasks. In addition, using episodes of a similar genre is highly recommended to adhere to the narrow viewing principle (Rodgers & Webb, 2011), which the current results support.

The findings are also of interest to content developers. The study highly recommends the spacing practice for activities that present unknown words with fewer cues for meaning recognition. For instance, if learners are to be presented with texts throughout an EFL program, it is better to make low-frequency words reappear moderately in multiple texts, using an optimal interval of one to two weeks than appear extensively in one single text. Developers could design four different texts to be studied over one month. Texts (1) and (2) share the same low-frequency items and also texts (3) and (4). In the textbook, however, texts (1) and (3) should be encountered in the first two weeks, followed by texts (2) and (4) in weeks 3 and 4.

This periodic repetition is likely to enhance learning even when the teacher does not draw students' attention to items (i.e., incidental learning). Based on the study results, neither spacing nor massing is required to support incidental word learning in the presence of explicit contextual support such as visual cues.

Moreover, this thesis highlights the significance of L2 captions to EFL learners as well as their enthusiasm to learn from them. Notwithstanding, the lowest gains in the spoken form recognition test in both experimental groups, compared to the three other measures, indicate that the presence of L2 captions possibly has a detrimental effect on spoken form acquisition. These findings, therefore, strongly recommend judicious use of L2 captions on the part of the teachers and learners. For example, teachers may periodically request students to turn off L2 captions during out-of-class viewing.

6.5 Limitations and Suggestions for Future Research

This thesis indeed laid itself open to certain limitations; however, most of these are not specific to this investigation but instead to vocabulary and viewing research, in general. The first source of weakness is the limited number of items imposed by the nature of the authentic audio-visual materials and the fact that target words needed to occur at least twice in every session and a minimum of eight times in the four sessions. The materials selection phase of the Norming Study clearly shows my objective to target a large group of words. Transcripts of 18 BBC documentary episodes were analysed, resulting in 54 potential target words. However, 26 of these were found to be known by a sample of participants similar to the target population and were, thus, excluded. Notwithstanding the relatively low number of target words, this limitation was somewhat mitigated, given the high quality of items: 20% occurred from 24 to 40 times, 60% from 10 to 17 times, and 20% from 8 to 9 times.

It was unfortunate that many Control participants dropped out of the experiment (posttest and delayed posttest). I surmised that these students lacked motivation for not being selected in the experimental conditions. My supposition was confirmed by informal conversations outside the confines of the study. Control participants approached me and questioned why they received the traditional teaching instead of being screened the documentaries similar to fellow students. For

this reason, I recommend that researchers present Control participants with activities that generate an excitement level similar to that of experimental activities (e.g., episodes of another television program) but that are void of any target words addressed in the study. This method will likely help ensure fairness to students and preserve the sample size.

Another uncontrolled variable in this research is participants' interaction outside the experimental sessions. It is not known how many students from different groups have talked about the TV series episodes over the experimental period. For instance, there is the possibility that target items had come up in students' conversations and this could influence the study results.

Moreover, one-week delayed posttests were conducted to establish a greater degree of accuracy regarding acquisition. However, the sessions coincided with students' preparation for exams; hence, there were notable absences in the three groups and data were not analysed. It was not possible to reschedule the tests at the time because it was the end of the winter term, neither after holidays because I would have exceeded the permitted study period at the institution.

In addition, I posttested participants immediately after the fourth treatment. It could be argued that significant learning gains arose from the recency effect. However, attempts were made to partially rectify this inevitable limitation, by selecting the two episodes with the lowest verbal frequency of items to be presented in the fourth session. Only one massed item among eight occurred in the last session and only three spaced words among twenty had their highest verbal frequency in the last session. Also, a similar study could administer the debriefing survey following each experimental session to record changes in perceptions on information-processing and motivation throughout the treatment period.

Furthermore, I measured knowledge of meaning through only posttests to reduce the risk of acquiring correct written forms of target words in advance of the treatment. The use of posttest-only design was also encouraged by the fact that the target words were not known to a sample with similar characteristics to the target students. The absence of pretests raised concerns about the extent to which significant scores in meaning posttests in Study 1 were attributed to the treatment

itself. Nevertheless, the fact that contigduration and spaced occurrences explained changes in meaning responses in Study 2 and Study 3, and for View and Non-View groups, respectively, help us reject the former hypothesis. The study is also limited by the lack of eye-tracking information on students' processing of visual referents and L2 captions. This helpful method could have informed how the contigduration effect was affected by the split attention imposed by captions. This method was not used due to the inaccessibility of devices in the context of research as well as the large sample size. Another limitation is that measuring verbal-visual contiguity for 536 occurrences at $\mp$25 timeframe was not done without difficulty, in terms of time and energy.

With regards to Study 2, an essential next step in validating the effect of verbal-visual contiguity in videos on word learning is to compare the effect of co-occurrences (as studied in this thesis) to the effect of asynchronous occurrences using an identical length of timeframe. Though this is a possible study, it is not an easy one due to methodological difficulties since referents usually occur close to forms in authentic videos. In addition, due to the considerable overlap of verbal-visual occurrences in videos at random distributing patterns, it could be hard to identify the exact length of the contiguity effect timeframe for word learning since it is conditional on the presence of other occurrences. Addressing this question here may, in fact, be of less significance due to the nature of audio-visual input, in general. Nonetheless, studying the effect of the sequence of occurrence of word forms and referents might be unhelpful for TV viewing research but an extremely beneficial research topic for designers of educational videos where they have control over the position of forms and referents. In addition, the present study shows that contigduration in documentary series augments learning based on 8 hr exposure. The study should be repeated using drama series or films to understand whether imagery plays a different role in different genres of TV viewing. Finally, contigduration effect could be usefully explored in further research to determine whether the current results apply equally or differentially to participants of other English language proficiency levels as well as other second languages.

A similar study on learners with lower L2 proficiency is strongly recommended. In the present study, learners in bimodal and multimodal input

conditions showed equal vocabulary gains independently of the presence of imagery. This may put the thesis participants into the category of maximal users (Vanderplank, 2019) who intend to consciously pick up aspects of the language while enjoying authentic content. Therefore, it is unknown whether similar experimentation into beginners will yield the same result. Also, while intermediate EFL learners can continuously keep a somewhat steady focus on L2 content for up to 2 hr, exploring beginners' relative effectiveness in achieving this would be a fruitful area for further work. A further study could also assess whether the lack of an advantage for spaced and massed occurrences in TV viewing is specific to the documentary genre only. Finally, if the debate is to be moved forward, researchers could examine the effect of the distribution of the target items instead of the occurrences of target items in viewing research.

6.6 Conclusion

This thesis has foregrounded the significance of imagery in L2 captioned video in incidental L2 vocabulary development. It has done so by assessing modality, contiguity, and spacing effects in incidental L2 vocabulary learning from extensive viewing of documentary series. The results indicated that watching two full-length seasons of BBC documentaries totalling to 8 hr over 2 hr length sessions in a six-week period of two-week intervals, in the format of L2 captioned video (i.e., multimodal input), significantly promotes intermediate L2 learners' acquisition of four aspects of word knowledge: meaning recall and recognition and spoken form and written form recognition. Obscuring imagery (i.e., bimodal input) does not necessarily hinder learning for this proficiency group but rather opens up other paths to learning. Furthermore, verbal-visual contiguity in viewing is a multifaceted construct that includes contigduration, contigfrequency, and contigratio. However, only the former stands out as a decisive factor in incidental vocabulary learning from viewing documentaries. Precisely, while eye-tracking studies have shown that the time spent looking at a word predicts its learning, the current study revealed that also does the time spent looking at its visual referent occurring within a ∓25 second timeframe relative to the form, especially for knowledge of meaning recognition. Finally, acquisition of words occurs from extensive viewing of L2 captioned documentary episodes independently of the distribution of words' occurrences. In the absence of visual cues and referents, however, knowledge of meaning would be

learnt better when verbal occurrences of words are spread across viewing sessions (spaced condition), compared to appearing within a single session (massed condition). Taken together, the thesis has assessed the strength of imagery in extensive viewing of L2 captioned documentaries. It has demonstrated the significance of verbal-visual contiguity in incidental L2 word learning, has shown the consequences of its presence and absence, as well as the ways in which it interacts with spacing.

References

Ahrabi Fakhr, M., Borzabadi Farahani, D., & Khomeijani Farahani, A. A. (2021). Incidental Vocabulary Learning and Retention from Audiovisual Input and Factors Affecting Them. *English Teaching and Learning*. 45, 167–188. https://doi.org/10.1007/s42321-020-00066-y

Aini, N., Jelani, M., & Boers, F. (2018). Examining incidental vocabulary acquisition from captioned video Does test modality matter ? *ITL - International Journal of Applied Linguistics*, *169*(1), 169–190. https://doi.org/10.1075/itl.00011.jel

Akhtar, N. (2002). Relevance and early word learning. *Journal of Child Language*, *29*(3), 677–686.

Akhtar, N., & Montague, L. (1999). Early lexical acquisition: The role of cross-situational learning. *First Language*, *19*(57), 347–358.

Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2006). Analysing Tests of Reading and Listening in Relation to the Common European Framework of Reference: The Experience of The Dutch CEFR Construct Project. *Language Assessment Quarterly*. https://doi.org/10.1207/s15434311laq0301_2

Alfotais, A. (2019). *Investigating the effect of spaced versus massed practice on vocabulary retention in the EFL classroom*. University of Essex. Retrieved from http://repository.essex.ac.uk/id/eprint/25062

Allan, D. (2004). *Oxford Placement Test 2: Marking Kit with User's Guide and Diagnostic Key*. Oxford University Press.

Alshumrani, H. (2019). *L2 incidental vocabulary learning and retention through different modalities of audio-visual input.* [Doctoral dissertation, University of Southampton].

Ashcroft, R. J., Garner, J., & Hadingham, O. (2018). Incidental Vocabulary Learning through Watching Movies. *Australian Journal of Applied Linguistics*, *1*(3), 135–147.

Assaf, A. G., Tsionas, M., & Tasiopoulos, A. (2019). Diagnosing and correcting the effects of multicollinearity: Bayesian implications of ridge regression. *Tourism Management, 71*, 1–8. *Average TV viewing time by country and age*. (2018, October). Statista. https://www.statista.com/statistics/276748/average-daily-tv-viewing-time-per-person-in-selected-countries/

Baart, M., & Vroomen, J. (2010). Do you see what you are hearing? Cross-modal effects of speech sounds on lipreading. *Neuroscience Letters*, *471*(2), 100–103. https://doi.org/10.1016/j.neulet.2010.01.019

Baddeley, A. D., & Levy, B. A. (1971). Semantic coding and short-term memory. *Journal of Experimental Psychology*, *89*(1), 132.

Bae, C. L., Therriault, D. J., & Redifer, J. L. (2019). Investigating the testing effect: Retrieval as a characteristic of effective study strategies. *Learning and Instruction*, *60*, 206–214. https://doi.org/https://doi.org/10.1016/j.learninstruc.2017.12.008

Baggett, P. (1984). Role of temporal overlap of visual and auditory material in forming dual media associations. *Journal of Educational Psychology*, *76*(3), 408–417. https://doi.org/10.1037/0022-0663.76.3.408

Baggett, P., & Ehrenfeucht, A. (1983). Encoding and retaining information in the visuals and verbals of an educational movie. *Educational Communication and Technology Journal*. https://doi.org/10.1007/BF02765208

Baghaei, P., & Amrahi, N. (2011). Validation of a multiple choice English vocabulary test with the Rasch model. *Journal of Language Teaching & Research*, *2*(5).

Bahrick, H. P., & Phelphs, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(2), 344.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, *4*(5), 316–321. https://doi.org/10.1111/j.1467-9280.1993.tb00571.x

Baltova, I. (1994). The impact of video on the comprehension skills of core French students. *Canadian Modern Language Review*, *50*(3), 507–531.

Barak, M., & Dori, Y. J. (2011). Science education in primary schools: Is an animation worth a thousand pictures? *Journal of Science Education and Technology*, *20*(5), 608–620.

Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition*, *27*(3), 387–414. https://doi.org/10.1017/S0272263105050175

Barnard, F. R. (1927). One picture is worth a thousand words. Printers' Ink 10 March.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/https://doi.org/10.1016/j.jml.2012.11.001

Barton, K. (2019). MuMIn: Multi-Model Inference. Retrieved from https://cran.r-project.org/package=MuMIn

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bates, D., Maechler, M., Bolker, B., &Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package (Version 1.1-21, Version 1.1-26) [Computer software]. http://CRAN.R-project.org/package=lme4

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279.

Beglar, D. (2010). A Rasch-based validation of the vocabulary size test. *Language Testing*, *27*(1), 101–118. https://doi.org/10.1177/0265532209340194

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Detecting and assessing collinearity. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, 85–191.

Berens, S. C., Horst, J. S., & Bird, C. M. (2018). Cross-situational learning is supported by propose-but-verify hypothesis testing. *Current Biology*, *28*(7), 1132–1136.

Bernsen, N. O. (2008). Multimodality theory. In *Multimodal User Interfaces* (pp. 5–29). Springer.

Bhutto, F. (2019, September 13). How Turkish TV is taking over the world. *The Guardian*. https://www.theguardian.com/tv-and-radio/2019/sep/13/turkish-tv-magnificent-century-dizi-taking-over-world

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. https://doi.org/10.1017/S0142716402004022

Birulés, J., & Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. *PLoS ONE*, *11*(6). https://doi.org/10.1371/journal.pone.0158409

Bisson, M. J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014a). The Role of Repeated Exposure to Multimodal Input in Incidental Acquisition of Foreign Language Vocabulary. *Language Learning*, *64*(4), 855–877. https://doi.org/10.1111/lang.12085

Bisson, M. J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014b). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, *35*(2), 399–418. https://doi.org/10.1017/S0142716412000434

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Hillsdale, NJ: Erlbaum.

Bjork, R. A., & Kroll, J. F. (2015). Desirable difficulties in vocabulary learning. *The American Journal of Psychology*, *128*(2), 241. https://doi.org/https://doi.org/10.5406/amerjpsyc.128.2.0241

Blanc, S. S. (1953). Vitalizing the classroom: Pictorial materials. *School Science and Mathematics*, *53*(2), 150–153.

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research*, *74*(4), 245–248. https://doi.org/10.1080/00220671.1981.10885317

Blything, L. P., & Cain, K. (2016). Children's processing and comprehension of complex sentences containing temporal connectives: The influence of

memory on the time course of accurate responses. *Developmental Psychology*, *52*(10), 1517.

Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2004.06.006

Bolger, P. A., & Zapata, G. (2011). Semantic Categories and Context in L2 Vocabulary Learning. *Language Learning*, *61*(2), 614–646. https://doi.org/10.1111/j.1467-9922.2010.00624.x

Bolibaugh, C., Vanek, N., & Marsden, E. J. (2021). Towards a credibility revolution in bilingualism research : Open data and materials as stepping stones to more reproducible and replicable research. *Bilingualism: Language and Cognition*.

Bordwell, D. (2006). *The way Hollywood tells it: Story and style in modern movies*. Univ of California Press.

Borrás, I., & Lafayette, R. (1994). Effects of multimedia courseware subtitling on the speaking performance of college students of French. *The Modern Language Journal*, *78*(1), 61–75. https://doi.org/10.1111/j.1540-4781.1994.tb02015.x

Bower, G. H. (1972). Stimulus sampling theory of encoding variability. AW Melton & E. Martin. *Coding Processes in Human Memory. New York: Winston*, *3*, 85–123.

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering Pictures: Pleasure and Arousal in Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/0278-7393.18.2.379

Bregman, A. S. (1967). Distribution of practice and between-trials interference. *Canadian Journal of Psychology*, *21*, 1–14. https://doi.org/10.1037/h0082962

Brindley, G., & Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment. *Language Testing*. https://doi.org/10.1191/0265532202lt236oa

Brown, J. H. (2010). Seeing things in pictures. In Catharine Abell Katerina Bantinaki (ed.), Philosophical Perspectives on Depiction. Oxford University Press. pp. 208--36.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, *20*(2), 136–163.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Burnham, K. P., & Anderson, D. R. (2002). A practical information-theoretic approach. *Model Selection and Multimodel Inference, 2nd Ed. Springer, New York*.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304.

Burridge, K., & Stebbins, T. N. (2016). *For the love of language: An introduction to linguistics*. Cambridge University Press.

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*, *116*, 165–178. https://doi.org/10.1016/j.visres.2015.03.005

Callan, D., & Schweighofer, N. (2010). Neural correlates of the spacing effect in explicit verbal semantic encoding support the deficient-processing theory. *Human Brain Mapping*, *31*(4), 645–659. https://doi.org/10.1002/hbm.20894

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., … David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*(5312), 593–596.

Carbo, M. (1981). Making books talk to children. *The Reading Teacher*, *35*(2), 186–189.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1563. https://doi.org/https://doi.org/10.1037/a0017021

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*(2), 268–276. https://doi.org/https://doi.org/10.3758/BF03193405

Carpenter, S. K., & Geller, J. (2020). Is a picture really worth a thousand words? Evaluating contributions of fluency and analytic processing in metacognitive judgements for pictures in foreign language vocabulary learning. *Quarterly*

Journal of Experimental Psychology, *73*(2), 211–224. https://doi.org/10.1177/1747021819879416

Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 92.

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*(3), 369–378.

Çekiç, A., & Bakla, A. (2019). The effects of spacing patterns on incidental L2 vocabulary learning through reading with electronic glosses. *Instructional Science*, *47*(3), 353–371. https://doi.org/10.1007/s11251-019-09483-4

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354.

Challis, B. H. (1993). Spacing Effects on Cued-Memory Tests Depend on Level of Processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 389. https://doi.org/10.1037/0278-7393.19.2.389

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*. https://doi.org/10.1207/s1532690xci0804_2

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction British Journal of Educational Psychology, 62 (2), pp. 233-246.

Chang, A. C. S. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System*, *37*(4), 652–663. https://doi.org/10.1016/j.system.2009.09.009

Chang, A. C. S., & Millett, S. (2014). The effect of extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal*, *68*(1), 31–40. https://doi.org/10.1093/elt/cct052

Chang, A. C. S., & Millett, S. (2015). Improving reading rates and comprehension through audio-assisted extensive reading for beginner learners. *System*, *52*, 91–102. https://doi.org/10.1016/j.system.2015.05.003

Chang, A. C. S., & Renandya, W. A. (2019). The effect of narrow reading on L2 learners' vocabulary acquisition. *RELC Journal*, 0033688219871387.

Charles, T. J., & Trenkic, D. (2015). Speech segmentation in a second language : The role of bi-modal input. *Subtitles and Language Learning: Principles, Strategies, and Practical Experiences*, *44*, 173–198. Retrieved from https://www.researchgate.net/publication/303445600

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, *31*(5), 693–713. https://doi.org/10.1093/applin/amq031

Chomsky, C. (1976). After decoding: what? *Language Arts*, *53*(3), 288–314.

Chomsky, C. (1978). "When You Still Can't Read in Third Grade: After Decoding, What?" *What Research Has to Say about Reading Instruction*, S. Jay Samuels, Ed., pp. 13-30. Newark, Del.: International Reading Association.

Cobb, T. (n.d.). Compleat web vp (Version 2.5) [Computer software]. https://www.lextutor.ca/vp/comp/

Cobb, T. (n.d.). The compleat word lister (Version 3.2) [Computer software]. https://www.lextutor.ca/freq/comp/

Cobb, T. (n.d.). Vocab size test [Computer software]. https://www.lextutor.ca/tests/vst/

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*. https://doi.org/10.3758/BF03204258

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge University Press.

Cooter, S., & Dyas, M. (Directors), & Cox, B. (Presenter) (2016). Forces of nature [TV documentary series; DVD]. BBC Studios, PBS, & France Télévisions.

Cooter, S., Holt, C., & Lachmann, M. (Directors), & Cox, B. (Presenter) (2011). Wonders of the universe [TV documentary series; DVD]. New York, NY: BBC America.

Coxhead, A., & Walls, R. (2012). TED Talks, vocabulary, and listening for EAP. *TESOLANZ Journal*, *20*(1), 55–67.

Crookes, G., & Schmidt, R. W. (1991). Motivation: Reopening the research agenda. *Language Learning*, *41*(4), 469–512.

Cruttenden, A. (1997). *Intonation*. Cambridge University Press.

Crystal, D. (1987). The Cambridge encyclopedia oflanguage. *Cam Bridge: CambridgeUniversityPress*.

Cuddy, L. J., & Jacoby, L. L. (1982). When forgetting helps memory: an analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior, 21,* 451–467. https://doi.org/10.1016/S0022-5371(82)90727-7

Cunningham, A. E. (2005). Vocabulary growth through independent reading and reading aloud to children. *Teaching and Learning Vocabulary: Bringing Research to Practice*, 45–68.

Cutajar, A. (2017). *Motivation, engagement and understanding in history: A study of using moving-image sources in a Maltese secondary history classroom.* [Doctoral dissertation, University of York]. White Rose Libraries. https://etheses.whiterose.ac.uk/20138/

d'Ydewalle, G., Van Rensbergen, J., & Pollet, J. (1987). Reading a message when the same message is available auditorily in another language: The case of subtitling. In J.K. O'Regan & A. Lévy-Schoen (Eds.), *Eye movements: From physiology to cognition* (pp. 313–321). Amsterdam: Elsevier Science. https://doi.org/10.1016/B978-0-444-70113-8.50047-3

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, *33*(1), 66–76. https://doi.org/10.1016/j.esp.2013.08.001

Daskalovska, N. (2016). Acquisition of three word knowledge aspects through reading. *The Journal of Educational Research*, *109*(1), 68–80. https://doi.org/10.1080/00220671.2014.918530

De Groot, A. M. B. (2006). Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. *Language Learning*. https://doi.org/10.1111/j.1467-9922.2006.00374.x

De Groot, A. M., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language learning, 50*(1), 1-56.

Deaner, R. O., & Smith, B. A. (2013). Sex differences in sports across 50 societies. *Cross-Cultural Research, 47*(3), 268–309.

Dearborn, W. E. (1910). Experiments in learning. *Journal of Educational Psychology, 1,* 373–388.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Malden: Blackwell.

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In *Psychology of learning and motivation* (Vol. 53, pp. 63–147). Elsevier.

Dellarosa, D., & Bourne, L. E. (1985). Surface form and the spacing effect. *Memory & Cognition*, *13*(6), 529–537. https://doi.org/10.3758/BF03198324

Dempster, F. N. (1988). The Spacing Effect: A Case Study in the Failure to Apply the Results of Psychological Research. *American Psychologist*, *43*(8), 627–634. https://doi.org/10.1037/0003-066X.43.8.627

Deno, S. L. (1968). Effects of words and pictures as stimuli in learning language equivalents. *Journal of Educational Psychology*, *59*(3), 202.

Dodd, B. (1980). Interaction of auditory and visual information in speech perception. *British Journal of Psychology*, *71*(4), 541–549. doi:10.18637/jss.v067.i01.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, *84*(5), 795–805. https://doi.org/10.1037/0021-9010.84.5.795

Douglas, T. (2011, March 25). Winners – 37th BPG television and radio awards. Broadcasting Press Guild. http://www.broadcastingpressguild.org/2011/03/winners-37th-bpg-television-and-radio-awards/

Dowhower, S. L. (1987). Effects of Repeated Reading on Second-Grade Transitional Readers ' Fluency and Comprehension Author (s): Sarah Lynn Dowhower Published by : Wiley on behalf of the International Literacy Association Stable URL : http://www.jstor.org/stable/747699 REFEREN. *Reading Research Quarterly*, *22*(4), 389–406.

Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System*, *88*. https://doi.org/https://doi.org/10.1016/j.system.2019.102166

Durso, F. T., & Shore, W. J. (1991). Partial knowledge of word meanings. *Journal of Experimental Psychology: General*, *120*(2), 190.

Ebbinghaus, H. (1985). *Memory: A contribution to experimental psychology*. New York: Dover.

Edwards, A. S. (1917). The Distribution of Time in Learning Small Amounts of Material. *Studies in Psychology: Titchener Commemorative Volume*, 209–213.

Eitel, A., & Scheiter, K. (2015). Picture or text first? Explaining sequence effects when learning with pictures and text. *Educational Psychology Review*, *27*(1), 153–180.

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. https://doi.org/10.1111/lang.12052

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. https://doi.org/10.1017/S0272263117000109

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. *Handbook of Second Language Acquisition*, 63–103.

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, *43*(4), 559–617.

Ellis, R. (1997). *SLA research and language teaching*. Oxford University Press.

Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *Quarterly Journal of Experimental Psychology*, *63*(4), 639–645. https://doi.org/10.1080/17470210903469864

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, *42*(3), 499–523. https://doi.org/10.1017/S0272263119000494

Feyereisen, P., & De Lannoy, J.-D. (1991). *Gestures and speech: Psychological investigations*. Cambridge University Press.

Flom, P. L. (1999). Multicollinearity diagnostics for multiple regression: A Monte Carlo study.

Florax, M., & Ploetzner, R. (2010). The influence of presentation format and subject complexity on learning from illustrated texts in biology. In *Learning in the Disciplines: ICLS 2010 Conference Proceedings - 9th International Conference of the Learning Sciences.*

Froeberg, S. (1918). Simultaneous versus successive association. *Psychological Review*, *25*(2), 156.

Frumuselu, A. D., De Maeyer, S., Donche, V., & Colon Plana, M. del M. G. (2015). Television series inside the EFL classroom: Bridging the gap between teaching and learning informal language through subtitles. *Linguistics and Education*, *32*, 107–117. https://doi.org/10.1016/j.linged.2015.10.001

Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, *28*(2), 241–265.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. Arnold.

Gardner, R. C. (2001). Integrative motivation and second language acquisition. *Motivation and Second Language Acquisition*, *23*(1), 1–19.

Gardner, R. C. (2010). *Motivation and Second Language Acquisition: The Socio-educational Model*. Peter Lang. Retrieved from https://books.google.co.uk/books?id=Ky15oSCIfLwC

Gartman, L. M., & Johnson, N. F. (1972). Massed versus distributed repetition of homographs: A test of the differential-encoding hypothesis. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 801–808.

Garza, T. J. (1991). Evaluating the Use of Captioned Video Materials in Advanced Foreign Language Learning. *Foreign Language Annals*, *24*(3), 239–258. https://doi.org/10.1111/j.1944-9720.1991.tb00469.x

Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*(1), 66–87. https://doi.org/10.1016/0006-8993(91)91275-6

Gerbier, E., & Toppino, T. C. (2015). The effect of distributed practice: Neuroscience, cognition, and education. *Trends in Neuroscience and Education*, *4*(3), 49–59.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 1–16.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*(2), 95–112. https://doi.org/10.3758/BF03197590

Glenberg, A. M., & Smith, S. M. (1981). Spacing repetitions and solving problems are not the same. *Journal of Verbal Learning and Verbal Behavior*, *20*(1), 110–119. https://doi.org/10.1016/S0022-5371(81)90345-5

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., … Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism*, *21*(3), 563–584. https://doi.org/10.1017/S1366728917000219

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*. https://doi.org/10.1016/S0346-251X(99)00060-3

Goldberg, F. (1974). Effects of imagery on learning incidental material in the classroom. *Journal of Educational Psychology*, *66*(2), 233.

Grabe, W., & Stoller, F. (1997). Reading and vocabulary development in a second language: A case study. *Second Language Vocabulary Acquisition*, 98–122.

Granger, S. (1993). Cognates: an aid or a barrier to successful L2 vocabulary development? *ITL-International Journal of Applied Linguistics*, *99*(1), 43–56.

Greene, R. L. (1989). Spacing Effects in Memory: Evidence for a Two-Process Account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 371–377. https://doi.org/10.1037/0278-7393.15.3.371

Greene, R. L., & Stillwell, A. M. (1995). Effects of encoding variability and spacing on frequency discrimination. *Journal of Memory and Language*, *34*(4), 468–476. https://doi.org/10.1006/jmla.1995.1021

Greeno, J. G. (1970). Conservation of information-processing capacity in paired-associate memorizing. *Journal of Verbal Learning and Verbal Behavior*, *9*, 581–586. https://doi.org/10.1016/S0022-5371(70)80105-0

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25.

Guthrie, E. R. (1933). Association as a function of time interval. *Psychological Review*, *40*(4), 355.

Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *ITL-International Journal of Applied Linguistics*, *166*(2), 278–306. https://doi.org/https://doi.org/10.1075/itl.166.2.04gyl

Hadar, U., & Butterworth, B. (1997). Iconic gestures, imagery, and word retrieval in speech. *Semiotica*, *115*(1–2), 147–172. https://doi.org/10.1515/semi.1997.115.1-2.147

Hadar, U., & Pinchas-Zamir, L. (2004). The semantic specificity of gesture: Implications for gesture classification and function. *Journal of Language and Social Psychology*, *23*(2), 204–214. https://doi.org/10.1177/0261927X04263825

Hair, J. F., Anderson, R. E., Babin, B. J., & Black, W. C. (2010). *Multivariate Data Analysis: A Global Perspective* ((Vol. 7).). Upper Saddle River, NJ: Pearson.

Harmer, J. (1991). The practice of English teaching. *London & New York: Longman*.

Hatami, S. (2017). The differential impact of reading and listening on L2 incidental acquisition of different dimensions of word knowledge. *Reading in a Foreign Language*, *29*(1), 61–85. https://doi.org/10125/66728

Hebb, D. O. (1949). Organization of behavior. New York: Wiley. *J. Clin. Psychol*, *6*(3), 307–335.

Hendrickson, A. T., & Perfors, A. (2019). Cross-situational learning in a Zipfian environment. *Cognition*, *189*(May 2017), 11–22. https://doi.org/10.1016/j.cognition.2019.03.005

Hernandez, S. S. (2004). *The effects of video and captioned text and the influence of verbal and spatial abilities on second language listening comprehension in a multimedia learning environment.* [Doctoral dissertation, New York University]. ProQuest.

https://www.proquest.com/openview/59c3e1508c41353eccf8be65e9f40ed4/1?pq-origsite=gscholar&cbl=18750&diss=y

Hiebert, E. H., & Kamil, M. L. (2005). *Teaching and learning vocabulary: Bringing research to practice*. Routledge.

Hintzman, D. L. (1976). Repetition and memory. In *Psychology of learning and motivation* (Vol. 10, pp. 47–91). New York: Academic Press.

Hitch, G. J., Flude, B., & Burgess, N. (2009). Slave to the rhythm: Experimental tests of a model for verbal short-term memory and long-term sequence learning. *Journal of Memory and Language, 61*(1), 97–111.

Holsanova, J., Holmberg, N., & Holmqvist, K. (2009). Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology*. https://doi.org/10.1002/acp.1525

Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: acquiring second language vocabulary through reading. *Reading in a Foreign Language*. Retrieved from http://eric.ed.gov/ERICWebPortal/recordDetail?accno=EJ577617%5Cnpapers2://publication/uuid/0954784E-875B-4DED-845C-CF67D6F78102

Hot Docs (2018). 2018 documentary audience research. http://assets.hotdocs.ca.s3.amazonaws.com/doc/HD18_Doc-Audience-Report_rev1.pdf

Hsieh, Y. (2020). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning*, *33*(5–6), 567–589.

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. https://doi.org/10.2190/RG06-LYWB-216Y-R27G

Huddleston, R. (1984). *Introduction to the Grammar of English*. Cambridge University Press.

Hulme, R. C. (2018). *Incidental learning of new meanings for familiar words*. [Doctoral dissertation, University College London]. UCL Discovery. https://discovery.ucl.ac.uk/id/eprint/10061270.

Hulme, R. C., Barsky, D., & Rodd, J. M. (2019). Incidental Learning and Long-Term Retention of New Word Meanings From Stories: The Effect of Number of Exposures. *Language Learning*, *69*(1), 18–43. https://doi.org/10.1111/lang.12313

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *51*(3), 539–558.

IBM Corp. (2017). IBM SPSS statistics for windows (Version 25.0) [Computer software]. Armonk, NY: IBM, Corp.

Imdad, M. U. & Aslam, M. (2018). mctest: Multicollinearity Diagnostic Measures. Retrieved from https://cran.r-project.org/package=mctest

Imdadullah, M., Aslam, M., & Altaf, S. (2016). Mctest: An R package for detection of collinearity among regressors. *The R Journal*, *8*(2), 495–505. Institutional Repository. https://eprints.soton.ac.uk/434623/

Isola, P., Parikh, D., Torralba, A., & Oliva, A. (2011). Understanding the intrinsic memorability of images. In *Advances in neural information processing systems* (pp. 2429–2437).

Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2013). What makes a photograph memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(7), 1469–1482.

Ivarsson, J., & Carrol, M. (1998). Subtitling. Simrishamn. *Sweden: TransEdit*.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*(6), 649–667.

James, E. L. (2019). *Understanding individual differences in learning and consolidating new vocabulary*. [Doctoral dissertation, University of York]. White Rose Libraries. https://etheses.whiterose.ac.uk/23489/.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). New York: springer.

James, W. (1901). *Talks to teachers on psychology." And to students on some of life's ideals.* New York: Holt.

Janebi Enayat, M., Amirian, S. M. R., Zareian, G., & Ghaniabadi, S. (2018). Reliable Measure of Written Receptive Vocabulary Size: Using the L2 Depth of

Vocabulary Knowledge as a Yardstick. *SAGE Open*, *8*(1), 1–15. https://doi.org/10.1177/2158244017752221

Jarząbek, P. (2018, December 26). Are new movies longer than they were 10, 20, 50 year ago? *Towards Data Science*. https://towardsdatascience.com/are-new-movies-longer-than-they-were-10hh20-50-year-ago-a35356b2ca5b?gi=34e4fce27d08#:~:text=The%20most%20popular%20runtime%20is,is%2080%E2%80%93120%20minutes%20long

Johnson, A. J., & Miles, C. (2019). Visual Hebb Repetition Effects: The Role of Psychological Distinctiveness Revisited. *Frontiers in Psychology*, *10*, 17.

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology. Applied*, *18*(2), 178–191. https://doi.org/10.1037/a0026923

Johnston, W. A., & Uhl, C. N. (1976). The contributions of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(2), 153.

Joklová, K. (2009). *Using pictures in teaching vocabulary*. [Bachelor dissertation, Masarykova univerzita]. https://is.muni.cz/th/uc2nv/

Kachergis, G., Yu, C., & Shiffrin, R. M. (2009). Proceedings of the Annual Meeting of the Cognitive Science Temporal Contiguity in Cross-Situational Statistical Learning.

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In *The Cambridge handbook of multimedia learning, 2nd ed.* (pp. 247–262). New York, NY, US: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.013

Kang, S. H. K., Lindsey, R. V., Mozer, M. C., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin and Review*, *21*(6), 1544–1550. https://doi.org/10.3758/s13423-014-0636-z

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1250–1257. https://doi.org/10.1037/a0023436

Kashyap, V. (2011). AutoTC: *Automatic Time-Code recognition for the purpose of synchronisation of subtitles in the broadcasting of motion pictures using the SMPTE standard.* [Master dissertation, Massey University]. Massey Research Online. https://mro.massey.ac.nz/bitstream/handle/10179/3779/02_whole.pdf

Kass, R. A., & Tinsley, H. E. A. (1979). Factor analysis. *Journal of Leisure Research*, *11*, 120–138.

Kassambara, A. (2019). ggpubr: "ggplot2" Based Publication Ready Plots. R package (Version 0.2.4) [Computer software]. https://cran.r-project.org/web/packages/ggpubr/index.html

Kayaoglu, M. N., & Akbas, D. (2011). A small scale experimental study: Using animations to learn vocabulary. *Turkish Online Journal of Educational Technology-TOJET*, *10*(2), 24–30.

Ke, F., Lin Kun Shan, H., Ching, Y. H., & Dwyer, F. (2006). Effects of Animation on Multi-Level Learning Outcomes for Learners with Different Characteristics: A Meta-Analytic Assessment and Interpretation. *Journal of Visual Literacy*, *26*(1), 15–40. https://doi.org/10.1080/23796529.2006.11674630

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Kester, L., Kirschner, P. A., & Van Merriënboer, J. J. G. (2005). The management of cognitive load during complex cognitive skill acquisition by means of computer-simulated problem solving. *British Journal of Educational Psychology*. https://doi.org/10.1348/000709904X19254

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*(2), 153–194.

Konnikova, M. (2014, December 10). Excuse me while I kiss this guy. The New Yorker. https://www.newyorker.com/science/maria-konnikova/science-misheard-lyrics-mondegreens

Kopstein, F. F., & Roshal, S. M. (1954). Learning foreign vocabulary from pictures versus words. *American Psychologist*, *9*(1), 407–408.

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(9), 1297–1317. https://doi.org/10.1002/acp.1537

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, *19*(6), 585–592.

Koval, N. G. (2019). Testing the deficient processing account of the spacing effect in second language vocabulary learning: Evidence from eye tracking. *Applied Psycholinguistics*, *40*(5), 1103–1139. https://doi.org/10.1017/S0142716419000158

Krashen, S. D. (1981). The case of narrow reading. *TESOL Newsletter*, *15*(6), 23.

Krejtz, I., Szarkowska, A., & Krejtz, K. (2013). The effects of shot changes on eye movements in subtitling. *Journal of Eye Movement Research*, *6*(5), 1–12. https://doi.org/10.16910/jemr.6.5.3

Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: Challenging conventions. *Tesol QUARTERLY*, *50*(4), 976–987.

Kroll, N. E., Parks, T., Parkinson, S. R., Bieber, S. L., & Johnson, A. L. (1970). Short-term memory while shadowing: Recall of visually and of aurally presented letters. *Journal of Experimental Psychology*, *85*(2), 220.

Krug, D., Davis, T. B., & Glover, J. A. (1990). Massed versus distributed repeated reading: A case of forgetting helping recall? *Journal of Educational Psychology*, *82*(2), 366–371.

Kruger, J.-L., Hefer, E., & Matthew, G. (2013). Measuring the impact of captions on cognitive load, *1*(August), 62–66. https://doi.org/10.1145/2509315.2509331

Küpper-Tetzel, C. E., Erdfelder, E., & Dickhäuser, O. (2014). The lag effect in secondary school classrooms: Enhancing students' memory for vocabulary. *Instructional Science*, *42*(3), 373–388. https://doi.org/10.1007/s11251-013-9285-2

Landauer, T. K. (1975). Memory without organization: Properties of a model with random storage and undirected retrieval. *Cognitive Psychology*, *7*(4), 495–531.

Lang, A., Dhillon, K., & Dong, Q. (1995). The Effects of Emotional Arousal and Valence on Television Viewers' Cognitive Capacity and Memory. *Journal of*

Broadcasting & Electronic Media. https://doi.org/10.1080/08838159509364309

Lang, A., Newhagen, J., & Reeves, B. (1996). Negative video as structure: Emotion, attention, capacity, and memory. *Journal of Broadcasting & Electronic Media*. https://doi.org/10.1080/08838159609364369

Laufer, B. (1997a). The Lexical Plight in Second Language Reading. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition* (pp. 20–34). Cambridge: Cambridge University Press.

Laufer, B. (1997b). What's in a word that makes it hard or easy? Intralexical factors affecting the difficulty of vocabulary acquisition. *Vocabulary Description, Acquisition and Pedagogy*, 140–155.

Laufer, B., & Cobb, T. (2019). How Much Knowledge of Derived Words Is Needed for Reading? *Applied Linguistics*, (Nation 2013), 1–29. https://doi.org/10.1093/applin/amz051

Lavery, M. R., Acharya, P., Sivo, S. A., & Xu, L. (2019). Number of predictors and multicollinearity: What are their effects on error and bias in regression? *Communications in Statistics-Simulation and Computation*, *48*(1), 27–38.

Lenzner, A., Schnotz, W., & Müller, A. (2013). The role of decorative pictures in learning. *Instructional Science*, *41*(5), 811–831.

Leona, N. L., van Koert, M. J. H., van der Molen, M. W., Rispens, J. E., Tijms, J., & Snellings, P. (2021). Explaining individual differences in young English language learners' vocabulary knowledge: The role of Extramural English Exposure and motivation. *System*, *96*, 102402.

Levitt, S. D., & List, J. A. (2011). Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. American Economic Journal: Applied Economics, 3(1), 224–238.

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, *10*(1), 105–129.

Llanes, À., Tragant, E., Pinyana, À., & Cerviño-Povedano, E. (2016). The impact of reading modality on reading fluency and comprehension in English as a foreign language: the case of children. *In Paper presented at the 'Multimodal input in second language learning' Symposium, Barcelona, Spain.*

Lotfolahi, A. R., & Salehi, H. (2017). Spacing effects in vocabulary learning: Young EFL learners in focus. *Cogent Education*, *4*(1). https://doi.org/10.1080/2331186X.2017.1287391

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.

Lüdecke, D., Waggoner, P., & Makowski, D. (2020). Performance: Assessment of Regression Models Performance. Retrieved from https://cran.r-project.org/package=performance

Maddox, G. B. (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: A case for encoding variability and study-phase retrieval. *Journal of Cognitive Psychology*, *28*(6), 684–706.

Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*(6), 828–835.

Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental Acquisition of Multiword Expressions Through Audiovisual Materials. *Studies in Second Language Acquisition*, 1–24. https://doi.org/10.1017/S0272263121000036

Markham, P. L. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals*, *32*(3), 321–328. https://doi.org/10.1111/j.1944-9720.1999.tb01344.x

Marquardt, D. W., & Snee, R. D. (1975). Ridge regression in practice. *The American Statistician*, *29*(1), 3–20.

Marquaridt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, *12*(3), 591–612.

Mayer, R. E. (2014). The Cambridge Handbook of Multimedia Learning (2nd ed., Cambridge Handbooks in Psychology). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139547369.

Mayer, R. E. (2001). *Multimedia Learning*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164603

Mayer, R. E. (2009). *Multimedia Learning. Second Edition*. Cambridge University Press.

Mayer, R. E., & Anderson, R. B. (1991). Animations Need Narrations: An Experimental Test of a Dual-Coding Hypothesis. *Journal of Educational Psychology*, *83*(4), 484–490. https://doi.org/10.1037/0022-0663.83.4.484

Mayer, R. E., & Anderson, R. B. (1992). The Instructive Animation: Helping Students Build Connections Between Words and Pictures in Multimedia Learning. *Journal of Educational Psychology*, *84*(4), 444–452. https://doi.org/10.1037/0022-0663.84.4.444

Mayer, R. E., & Gallini, J. K. (1990). When Is an Illustration Worth Ten Thousand Words? *Journal of Educational Psychology*. https://doi.org/10.1037/0022-0663.82.4.715

Mayer, R. E., & Sims, V. K. (1994). For Whom Is a Picture Worth a Thousand Words? Extensions of a Dual-Coding Theory of Multimedia Learning. *Journal of Educational Psychology*. https://doi.org/10.1037/0022-0663.86.3.389

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing Constructivist Learning from Multimedia Communications by Minimizing Cognitive Load. *Journal of Educational Psychology*. https://doi.org/10.1037/0022-0663.91.4.638

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development*. https://doi.org/10.1007/BF02300480

Mazahery, S., Hashemian, M., & Alipour, J. (2021). Vocabulary Learning by Iranian Adult L2 Learners via Extensive Viewing of Subtitled and Captioned TV Series. *Teaching English as a Second Language (Formerly Journal of Teaching Language Skills), 40*(1), 83–115. https://doi.org/10.22099/jtls.2021.39209.2921

Mazerolle, M. J. (2019). AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). (R package Version 2.2-2). Retrieved from https://cran.r-project.org/package=AICcmodavg.

McGeoch, J. A. (1943). *The psychology of human learning*. New York: Longmans Green.

McGrath, M. (1985). *An examination of cues for visual and audio-visual speech perception using natural and computer-generated faces*. [Doctoral dissertation, University of Nottingham]. https://ci.nii.ac.jp/naid/10014688774/

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica: Biochemia Medica*, *22*(3), 276–282.

McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, *39*(6), 823–845.

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, *19*(6), 741–760.

MED-EL. (2020). The Speed of Hearing. Retrieved February 19, 2020, from https://blog.medel.com/the-speed-of-hearing/

Melton, A. W. (1967). Repetition and Retrieval from Memory. *Science*, *158*(3800), 532. https://doi.org/10.1126/science.158.3800.532-b

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*, 596–606.

Menard, S. (2002). *Applied logistic regression analysis* (Vol. 106). Sage.

Meyer, V. (1982). Prime-O-Tec: A successful strategy for adult disabled readers. *Journal of Reading*, *25*(6), 512–515. Retrieved from https://www.jstor.org/stable/40029109

Michas, I. C., & Berry, D. C. (2000). Learning a Procedural Task: Effectiveness of Multimedia Presentations. *Applied Cognitive Psychology*, *14*(6), 555–575. https://doi.org/10.1002/1099-0720(200011/12)14:6<555::AID-ACP677>3.0.CO;2-4

Microsoft Corporation. (2012). Windows movie maker (Version 16.4.3528.0331) [Computer software].

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Mishan, F. (2005). *Designing authenticity into language learning materials*. Intellect Books.

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PloS one*, *4*(11).

Montero Perez, M. (2019). Pre-learning vocabulary before viewing captioned video: an eye-tracking study. *The Language Learning Journal*, *47*(4), 460–478. https://doi.org/10.1080/09571736.2019.1638623

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. https://doi.org/10.1111/modl.12215

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, *18*(1), 118–141.

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. https://doi.org/10.1016/j.system.2013.07.013

Montgomery, A. A., & Jackson, P. L. (1983). Physical characteristics of the lips underlying vowel lipreading performance. *The Journal of the Acoustical Society of America*, *73*(6), 2134–2144.

Moravec, J. (2016, March 29). A theory for invisible learning. *Education Futures.* https://www.educationfutures.com/blog/post/theory-invisible-learning

Moreno, R., & Mayer, R. E. (1999). Visual presentations in multimedia learning: Conditions that overload visual working memory. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *1614*, 793–800. https://doi.org/10.1007/3-540-48762-x_98

Morin, F. (2015, March 9). Pourquoi les programmes durent-ils 52 minutes à la television? [Why programs last for 52 minutes?]. *Le Figaro.* https://tvmag.lefigaro.fr/le-scan-tele/actu-tele/2015/03/09/28001-20150309ARTFIG00332-pourquoi-les-programmes-durent-ils-52-minutes-a-la-television.php

Müller, K. (2019). dplyr: A Grammar of Data Manipulation.Wickham, H., François, R., Henry, L., & Müller, K. (2019). dplyr: A Grammar of Data Manipulation. R package (Version 0.8.3) [Computer software].https://CRAN.R-project.org/package=dplyr

Muñoz, C. (2017). The role of age and proficiency in subtitle reading. An eye-tracking study. *System*. https://doi.org/10.1016/j.system.2017.04.015

Murdock, B. B. (1971). Four-channel effects in short-term memory *, *24*(4), 197–198.

Myers, R. (1990). *Classical and modern regression with applications* ((2nd ed)). Boston, MA: Duxbury.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142.

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition*, *37*(4), 677–711. https://doi.org/10.1017/S0272263114000825

Nakata, T., & Elgort, I. (2020). Effects of spacing on contextual vocabulary learning : Spacing facilitates the acquisition of explicit , but not tacit , vocabulary knowledge. *Second Language Research*. https://doi.org/10.1177/0267658320927764

Nakata, T., & Suzuki, Y. (2019). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition*, *41*(2), 287–311. https://doi.org/10.1017/S0272263118000219

Nakata, T., & Webb, S. (2016). Does studying vocabulary in smaller sets increase learning. *Studies in Second Language Acquisition*, *38*(3), 523–552. https://doi.org/10.1017/S0272263115000236

Nation, I. (n.d.).Vocabulary tests. Victoria University of Wellington. https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-tests

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, I. S. P. (2006). How Large a Vocabulary Is Needed for Reading and Listening? *The Canadian Modern Language Review / La Revue Canadienne Des Langues Vivantes*, *63*(1), 59–81. https://doi.org/10.1353/cml.2006.0049

Nation, I., & Beglar, D. (2007). A Vocabulary Size Test. *The Language Teacher*, *31*(7), 9–12. Retrieved from papers://342a052f-a8c2-46f8-9784-a05ceb151618/Paper/p10

Nation, P. (2007). The Four Strands. *Innovation in Language Learning and Teaching*, *1*(1), 2–13. https://doi.org/10.2167/illt039.0

Neuman, S. B., & Koskinen, P. (1992). Captioned Television as Comprehensible Input: Effects of Incidental Word Learning from Context for Language Minority Students. *Reading Research Quarterly*, *27*(1), 95–106. https://doi.org/10.2307/747835

Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *19*(2), 155.

Nickerson, R. S. (1968). A note on long-term recognition memory for pictorial material. *Psychonomic Science*, *11*(2), 58.

Nicole A. M. C. Goossens, Gino Camp, Peter P. J. L. Verkoeijen, Huib K. Tabbers & Rolf A. Zwaan. (2012). Spreading the words: A spacing effect in vocabulary learning. *Journal of Cognitive Psychology*, *24*(8), 965–971. https://doi.org/10.1080/20445911.2012.722617

Niikuni, K., & Muramoto, T. (2014). Effects of punctuation on the processing of temporarily ambiguous sentences in Japanese. *Japanese Psychological Research*, *56*(3), 275–287. https://doi.org/10.1111/jpr.12052

Nodine, C. F. (1969). Temporal variables in paired-associate learning: The law of contiguity revisited. *Psychological Review*, *76*(4), 351.

Noels, K. A., Pelletier, L. G., Clément, R., & Vallerand, R. J. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning*, *50*(1), 57–85.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Nurmukhamedov, U. (2017). Lexical Coverage of TED Talks: Implications for Vocabulary Instruction. *TESOL Journal*, *8*(4), 768–790. https://doi.org/10.1002/tesj.323

Owens, P., & Sweller, J. (2008). Cognitive load theory and music instruction. *Educational Psychology*. https://doi.org/10.1080/01443410701369146

Oxford, R. (1990). Language learning strategies. *New York*, *3*.

Oxford, R., & Nyikos, M. (1989). Variables affecting choice of language learning strategies by university students. *The Modern Language Journal*, *73*(3), 291–300.

Oxford, R., Park-Oh, Y., Ito, S., & Sumrall, M. (1993). Learning a language by satellite television: What influences student achievement? *System, 21*(1), 31–48.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*(1), 1–4.

Page, M., Cumming, N., Norris, D., Hitch, G. J., & McNeil, A. M. (2006). Repetition learning in the immediate serial recall of visual and auditory materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 716.

Pain, M. T. G., & Hibbs, A. (2007). Sprint starts and the minimum auditory reaction time. *Journal of Sports Sciences*. https://doi.org/10.1080/02640410600718004

Paivio, A. (1971). *Imagery and verbal processes*. New York, NY: Holt, Rinehart, and Winston

Paivio, A. (1986). *Mental representations : a dual-coding approach.* New York : Oxford

Paivio, A. (2014). *Mind and its evolution: A dual coding theoretical approach*. Psychology Press.

Paivio, A., & Desrochers, A. (1979). Effects of an imagery mnemonic on second language recall and comprehension. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *33*(1), 17.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*(1p2), 1.

Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1051–1057. https://doi.org/10.1037/0278-7393.29. 6.1051

Pavlik Jr, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, *29*(4), 559–586.

Pawlicz, A., & Napierala, T. (2017). The determinants of hotel room rates: an analysis of the hotel industry in Warsaw, Poland. *International Journal of Contemporary Hospitality Management*.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. Studies in Second Language Acquisition. *Studies in Second Language Acquisition*, *38*(1), 97–130. https://doi.org/10.1017/S0272263115000224

Pellicer-Sanchez, A., & Schmitt, N. (2010). Incidental Vocabulary Acquisition from an Authentic Novel: Do. *Reading in a Foreign Language*.

Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Llanes, À., & Serrano, R. (2018). *L2 reading and reading-while-listening in multimodal learning conditions: An eye tracking study*. *ELT Research Papers*. Retrieved from https://www.teachingenglish.org.uk/sites/teacheng/files/pub_H191 ELT_L2 reading and reading-while-listening in multimodal_FINAL.pdf

Perego, E., del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology*, *13*(3), 243–272. https://doi.org/10.1080/15213269.2010.502873

Perkins, N. L. (1914). The value of distributed repetitions in rote learning. *British Journal of Psychology*, *7*(2), 253.

Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL - International Journal of Applied Linguistics*, *169*(1), 142–168. https://doi.org/10.1075/itl.00010.pet

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, *53*(4), 1–25.

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. https://doi.org/10.1017/s0272263117000407

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, *63*, 134–148. https://doi.org/10.1016/j.system.2016.10.002

Peterson, L. R., Wampler, R., Kirkpatrick, M., & Saltzman, D. (1963). Effect of spacing presentations on retention of a paired associate over short intervals. *Journal of Experimental Psychology*, *66*(2), 206.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*(3), 193.

Pinheiro, J.C. & Bates, D. M. (2000). *Mixed-effects Models in S and S-Plus.* Springer,NewYork.

Pinker, S. (1984). Language learnability and language development. Cambridge, MA: Harvard University Press.

Pinker, Steven. (1994). The Language Instinct: How the Mind Creates. *Language. New York: Harper Collins*.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, *31*(6), 459–470.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33.

Pociask, F. D., & Morrison, G. R. (2008). Controlling split attention and redundancy in physical therapy instruction. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-007-9062-5

Price, K. (1983). Closed-captioned TV: An untapped resource. *Matsol Newsletter*, *12*(2), 1–8.

Prisko, L.-H. (1963). Short-term memory in focal cerebral damage. McGill University Libraries.

Puimège, E., & Peters, E. (2019). Learners' English Vocabulary Knowledge Prior to Formal Instruction: The Role of Learner-Related and Word-Related Variables. *Language Learning*, 1–35. https://doi.org/10.1111/lang.12364

Pujadas Jorba, G. (2019*). Language learning through extensive TV viewing. A study with adolescent EFL learners*. [Doctoral dissertation, Universitat De Barcelona]. Dipòsit Digital. http://hdl.handle.net/2445/146118

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: a study of L2 vocabulary learning by adolescents. *Language Learning Journal*, *47*(4), 479–496. https://doi.org/10.1080/09571736.2019.1616806

Pyle, W. H. (1913). Economical learning. *Journal of Educational Psychology*, *4*(3), 148.

Qi, X., & Lai, C. (2017). The effects of deductive instruction and inductive instruction on learners' development of pragmatic competence in the teaching of Chinese as a second language. *System*, *70*, 26–37. https://doi.org/https://doi.org/10.1016/j.system.2017.08.011

R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org

Raaijmakers, J. G. W. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science*, *27*(3), 431–452.

Ramezanali, N., & Faez, F. (2019). Vocabulary learning and retention through multimedia glossing. *Language Learning and Technology*, *23*(2), 105–124.

Rasinski, T. V. (1990). Effects of repeated reading and listening-while-reading on reading fluency. *The Journal of Educational Research*, *83*(3), 147–151.

Reeves, B., Lang, A., Kim, E. Y., & Tatar, D. (1999). The Effects of Screen Size and Message Content on Attention and Arousal. *Media Psychology*. https://doi.org/10.1207/s1532785xmep0101_4

Reitsma, P. (1988). Reading practice for beginners: Effects of guided reading, reading-while-listening, and independent reading with computer-based speech feedback. *Reading Research Quarterly*, *23*(2), 219–235. Retrieved from http://www.jstor.org/stable/747803

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, *16*(3), 550–554.

Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions.* [Doctoral dissertation, Victoria University of Wellington]. University Library Papers and Theses. http://hdl.handle.net/10063/2870

Rodgers, M. P. H. (2016). Extensive listening and viewing: the benefits of audiobooks and television. *The European Journal of Applied Linguistics and TEFL*, *5*, 43–57.

Rodgers, M. P. H. (2018). The images in television programs and the potential for learning unknown words: The relationship between on-screen imagery and

vocabulary. *ITL-International Journal of Applied Linguistics*, *169*(1), 191–211.

Rodgers, M. P. H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, *45*(4), 689–717. https://doi.org/10.5054/tq.2011.268062

Rodgers, M. P. H., & Webb, S. (2019). Incidental vocabulary learning through viewing television. *ITL-International Journal of Applied Linguistics*, *171*(2), 191–220. https://doi.org/10.1075/itl.18034.rod

Rodgers, T. S. (1969). On measuring vocabulary difficulty an analysis of item variables in learning russian-english vocabulary pairs. *IRAL - International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral.1969.7.4.327

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, *24*(2), 3–13.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., & Karpicke, J. D. (2006). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, *1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Rogers, J., Webb, S., & Nakata, T. (2015). Do the cognacy characteristics of loanwords make them more easily learned than noncognates? *Language Teaching Research*, *19*(1), 9–27.

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science*, *16*(4), 183–186.

Romaní, C. C., & Moravec, J. W. (2011). *Aprendizaje invisible: Hacia una nueva ecología de la educación* (Vol. 3). Edicions Universitat Barcelona.

Romanko, R. (2017). *The vocabulary demands of popular English songs*. (Publication No. 10681738) [Doctoral dissertation, Temple University]. ProQuest Dissertations & Theses.

Rott, S. (2012, November 5). Incidental Vocabulary Acquisition. *The Encyclopedia of Applied Linguistics*. https://doi.org/https://doi.org/10.1002/9781405198431.wbeal0531

Royal Television Society. (n.d.). RTS programme awards 2011. Retrieved from https://rts.org.uk/award/rts-programme-awards-2011

RStudio Team. (2018). RStudio: Integrated Development Environment for R. (Version 1.2) [Computer software]. Boston, MA: RStudio, Inc. http://www.rstudio.com

RStudio Team. (2018). RStudio: Integrated Development Environment for R. (Version 3.5.1) [Computer software]. Boston, MA: RStudio, Inc. http://www.rstudio.com

Russell Lenth (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package (Version 1.4.4) [Computer software].

Russo, R., & Mammarella, N. (2002). Spacing effects in recognition memory: When meaning matters. *European Journal of Cognitive Psychology*, *14*(1), 49–59. https://doi.org/10.1080/09541440042000133

Russo, R., Parkin, A. J., Taylor, S. R., & Wilks, J. (1998). Revising current two-process accounts of spacing effects in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(1), 161. https://doi.org/10.1037/0278-7393.24.1.161

Saragi, T. (1978). Vocabulary learning and reading. *System*, *6*(2), 72–78.

Schaffert, N., Janzen, T. B., Mattes, K., & Thaut, M. H. (2019). A review on the relationship between sound and movement in sports and rehabilitation. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2019.00244

Schmidt, R. W. (1990). The role of consciousness in second language learning1. *Applied Linguistics*, *11*(2), 129–158.

Schmidt, R., & Watanabe, Y. (2001). Motivation, strategy use, and pedagogical preferences in foreign language learning. *Motivation and Second Language Acquisition*, *23*(1), 313–359.

Schmidt, R., Boraie, D., & Kassabgy, O. (1996). Foreign language motivation: Internal structure and external connections. In R. Oxford (Ed.), *Language Learning Motivation: Pathways to the New Century* (pp. 9–70). Honolulu: University of Hawaii, Second Language Teaching & Curriculum Center.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Springer.

Schmitt, N., & Carter, R. (2000). The lexical advantages of narrow reading for second language learners. *Tesol Journal*, *9*(1), 4–9.

Schmitt, N., Nation, P., & Kremmel, B. (2020). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching*, *53*(1), 109–120. https://doi.org/10.1017/S0261444819000326

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.

Schmitz, H.-C. (2008). *Accentuation and interpretation*. Springer.

Schneider, V. I., Healy, A. F., & Bourne Jr, L. E. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language*, *46*(2), 419–440. https://doi.org/https://doi.org/10.1006/jmla.2001.2813

Schreiber, P. A. (1980). On the acquisition of reading fluency. *Journal of Reading Behavior*, *12*(3), 177–186.

Schuetze, U. (2015). Spacing techniques in second language vocabulary acquisition: Short-term gains vs. long-term memory. *Language Teaching Research*, *19*(1), 28–42.

Screencast-O-Matic. (2017). Web launch recorder (Version 2.0) [Computer software]. Seattle, WA: Big Nerd Software, LLC. https://screencast-o-matic.com/screen-recorder.

Seibert, L. . (1945). A study of the practice of guessing word meanings from a context. *Modern Language Journal*, *29*(4), 296–323. https://doi.org/10.1007/s12129-015-9500-5

Serrano R., Andriá M., & Pellicer-Sánchez, A. (2016). Reading-while-listening in primary school: Linguistic and non-linguistic outcomes. *In Paper presented at the 'Multimodal input in second language learning' Symposium, Barcelona, Spain*

Serrano, R. (2011). The time factor in EFL classroom practice. *Language Learning*, *61*(1), 117–145.

Serrano, R., & Huang, H. (2018). Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text? *TESOL Quarterly*, *52*(4), 971–994. https://doi.org/10.1002/tesq.445

Shelton, J., & Kumar, G. P. (2010). Comparison between Auditory and Visual Simple Reaction Times. *Neuroscience and Medicine*. https://doi.org/10.4236/nm.2010.11004

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 156–163. https://doi.org/10.1016/S0022-5371(67)80067-7

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review*, *76*(2), 179.

Sinyashina, E. (2019). The effect of repetition on incidental legal vocabulary learning through long-term exposure to authentic videos.

Sinyashina, E. (2020a). 'Incidental + Intentional' vs 'Intentional + Incidental' Vocabulary Learning: Which is More Effective? *Complutense Journal of English Studies*, *28*, 81–96. https://doi.org/10.5209/cjes.66685

Sinyashina, E. (2020b). Watching Captioned Authentic Videos for Incidental Vocabulary Learning: Is It Effective? *NJES Nordic Journal of English Studies*, *19*(1), 28–64. https://doi.org/10.35360/njes.539

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1–2), 39–91.

Skinner, C. H., Adamson, K. L., Woodward, J. R., Jackson, R. R., Atchison, L. A., & Mims, J. W. (1993). A comparison of fast-rate, slow-rate, and silent previewing interventions on reading performance. *Journal of Learning Disabilities*, *26*(10), 674–681. https://doi.org/10.1177/002221949302601005

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics, *106*, 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Snoder, P. (2017). Improving English learners' productive collocation knowledge: The effects of involvement load, spacing, and intentionality. *TESL Canada Journal*, *34*(3), 140–164. https://doi.org/10.18806/tesl.v34i3.1277

Soares, S., Atallah, B., & Paton, J. (2016). Midbrain dopamine neurons control judgment of time. *Science*, *354*(November), 1273–127.

Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology*, *25*(5), 763–767. https://doi.org/10.1002/acp.1747

software]. Vienna, Austria: R Foundation for Statistical Computing.

Specker, E. (2008). *L1/L2 eye movement reading of closed captioning: A multimodal analysis of multimodal use*. University of Arizona.

Spiess, A.-N., & Neumeyer, N. (2010). An evaluation of R 2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC Pharmacology*, *10*(1), 6.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*(2), 73–74.

Starch, D. (1912). Periods of work in learning. *Journal of Educational Psycholog3r*, *3*, 209–213.

Suárez, M. del M., & Gesa, F. (2019). Learning vocabulary with the support of sustained exposure to captioned video: do proficiency and aptitude make a difference? *Language Learning Journal*, *47*(4), 497–517. https://doi.org/10.1080/09571736.2019.1617768

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*(1273), 71–78.

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*(4), 747–790.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive Load as a Factor in the Structuring of Technical Material. *Journal of Experimental*

Psychology: General, *119*(2), 176–192. https://doi.org/10.1037/0096-3445.119.2.176

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50–73. https://doi.org/10.1021/ja0385134

Szarkowska, A., & Gerber-Morón, O. (2019). Two or three lines: a mixed-methods study on subtitle processing and preferences. *Perspectives: Studies in Translation Theory and Practice*, *27*(1), 144–164. https://doi.org/10.1080/0907676X.2018.1520267

Tangkakarn, B., & Gampper, C. (2020). The effects of reading-while-listening and listening-before-reading-while-listening on listening and vocabulary. *International Journal of Instruction*, *13*(3), 789–804. https://doi.org/10.29333/iji.2020.13353a

Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals*, *38*(3), 422–427.

Tegge, F. (2017). The lexical coverage of popular songs in English language teaching. *System*, *67*, 87–98. https://doi.org/10.1016/j.system.2017.04.016

Teng, F. (2016). Incidental vocabulary acquisition from reading- only and reading-while-listening: a multi- dimensional approach. *Innovation in Language Learning and Teaching*, *12*(3), 274–288. https://doi.org/10.1080/17501229.2016.1203328

Teng, F. (2021). *Language Learning Through Captioned Videos*. Routledge. https://doi.org/10.4324/9780429264740

Teng, M. F. (2020). *Language Learning Through Captioned Videos: Incidental Vocabulary Acquisition*. Routledge.

The Grierson Trust. (2017, November 6). The Winners of the 2017 Grierson Awards. https://griersontrust.org/about-us/news/2017/winners-announced.html

Thornbury, S. (2002). *How to teach vocabulary*. (J. Harmer, Ed.). Edinburgh Gate: Pearson Education Limited.

ThoughtCo Team. (2019, April 4). Terms of enrichment: How French has influenced English. https://www.thoughtco.com/how-french-has-influenced-english-1371255

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When Two Sensory Modes Are Better Than One. *Journal of Experimental Psychology: Applied*. https://doi.org/10.1037/1076-898X.3.4.257

Toppino, T. C., & Bloom, L. C. (2002). The spacing effect, free recall, and two-process theory: A closer look. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 437. https://doi.org/10.1037/0278-7393.28.3.437

Toyota, H. (2013). Significance of autobiographical episodes and spacing effects in incidental memory. *Perceptual and Motor Skills*, *117*(2), 402–410. https://doi.org/10.2466/22.10.PMS.117x19z1

Trafton, A. (2014). In the blink of an eye. MIT neuroscientists find the brain can identify images seen for as little as 13 milliseconds. MIT News Office. Massachusetts Institute of Technology.

Tragant Mestres, E., Llanes Baró, À., & Pinyana Garriga, À. (2018). Linguistic and non-linguistic outcomes of a reading-while-listening program for young learners of English. *Reading and Writing*, *32*(3), 819–838. https://doi.org/10.1007/s11145-018-9886-x

Tragant, E., & Pellicer-Sánchez, A. (2019). Young EFL learners' processing of multimodal input: Examining learners' eye movements. *System*, *80*, 212–223. https://doi.org/10.1016/j.system.2018.12.002

Tragant, E., Muñoz, C., & Spada, N. (2016). Maximizing young learners' input: An intervention program. *Canadian Modern Language Review*, *72*(2), 234–257. https://doi.org/10.3138/cmlr.2942

Tye-Murray, N., Sommers, M., & Spehar, B. (2007). Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends in Amplification*, *11*(4), 233–241.

Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., & Sommers, M. (2016). Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration. *Psychology and Aging*, *31*(4), 380.

Tzovaras, D. (2008). *Multimodal user interfaces: from signals to interaction*. Springer Science & Business Media.

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. https://doi.org/10.1111/lang.12343

Universite de Montreal. (2015, October 8). Machines have nothing on mum when it comes to listening. *Newswise.* https://www.newswise.com/articles/machines-have-nothing-on-mum-when-it-comes-to-listening.

Valentini, A., Ricketts, J., Pye, R. E., & Houston-Price, C. (2018). Listening while reading promotes word learning from stories. *Journal of Experimental Child Psychology*, *167*, 10–31. https://doi.org/10.1016/j.jecp.2017.09.022

Vallerand, R. J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. *Advances in Experimental Social Psychology*, *29*, 271–360.

Van Gog, T., Paas, F., & Sweller, J. (2010). Cognitive load theory: Advances in research on worked examples, animations, and cognitive load measurement. *Educational Psychology Review*, *22*(4), 375–378.

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190.

Van Merrienboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177.

Van Zeeland, H., & Schmitt, N. (2013a). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*(3), 609–624. https://doi.org/10.1016/j.system.2013.07.012

Van Zeeland, H., & Schmitt, N. (2013b). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457–479. https://doi.org/10.1093/applin/ams074

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics*. 24, 3-25. Cambridge University Press. https://doi.org/10.1017/s0267190504000017

Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT Journal*, *42*(4), 272–281.

Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching: Subtitles for the deaf and hard-of-hearing as tools for language learning.* Springer. https://doi.org/10.1057/978-1-137-50045-8

Vanderplank, R. (2019). Gist watching can only take you so far': attitudes, strategies and changes in behaviour in watching films with captions. *The Language Learning Journal*, *47*(4), 407–423. https://doi.org/10.1371/journal.pone.0007785

Velikovsky, J. T. (2012). Patterns in the Top 20 ROI Films: of Scenes and Film Duration/Screenplay Length and the emergence of Elliot Waves. Retrieved from https://storyality.wordpress.com/2012/12/23/storyality-53-2-patterns-in-the-top-20-roi-films-of-scenes-vs-film-durationscreenplay-length-and-elliot-waves/

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2005). Limitations to the spacing effect: Demonstration of an inverted u-shaped relationship between interrepetition spacing and free recall. *Experimental Psychology*, *52*(4), 257–263. https://doi.org/10.1027/1618-3169.52.4.257

Vogt, P., & Smith, A. D. M. (2005). Learning colour words is slow: A cross-situational learning account. *Behavioral and Brain Sciences*, *28*(4), 509–510.

Vonesh, E. F., Chinchilli, V. M., & Pu, K. (1996). Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics*, 572–587.

Vong, W. K., & Lake, B. M. (2020). Learning word-referent mappings and concepts from raw inputs. Retrieved from http://arxiv.org/abs/2003.05573

Vu, D. Van, & Peters, E. (2020). Learning Vocabulary from Reading-only, Reading-while-listening, and Reading with Textual Input Enhancement: Insights from Vietnamese EFL Learners. *RELC Journal*. https://doi.org/10.1177/0033688220911485

Wagner, E. (2010). Test-takers' interaction with an L2 video listening test. *System*. https://doi.org/10.1016/j.system.2010.01.003

Webb, S. (2014). Repetition in incidental vocabulary learning. In C. A. Chapelle (Ed.), The encyclopedia of applied linguistics (pp. 1–6). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781405198431.wbeal1425

Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159–168). New York, NY: Routledge.

Webb, S., & Chang, A. C. S. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review*, *68*(3), 267–290. https://doi.org/10.3138/cmlr.1204.1

Webb, S., & Chang, A. C. S. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*(6), 667–686. https://doi.org/10.1177/1362168814559800

Webb, S., & Rodgers, M. P. H. (2009a). The lexical coverage of movies. *Applied Linguistics*, *30*(3), 407–427. https://doi.org/10.1093/applin/amp010

Webb, S., & Rodgers, M. P. H. (2009b). Vocabulary demands of television programs. *Language Learning*, *59*(2), 335–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

Webb, S., Newton, J., & Chang, A. (2013). Incidental Learning of Collocation. *Language Learning*, *63*(1), 91–120. https://doi.org/10.1111/j.1467-9922.2012.00729.x

Webber, N. E. (1978). Pictures and words as stimuli in learning foreign language responses. *The Journal of Psychology*, *98*(1), 57–63.

Weissfeld, L. A., & Sereika, S. M. (1991). A multicollinearity diagnostic for generalized linear models. *Communications in Statistics-Theory and Methods*, *20*(4), 1183–1198.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, *53*(1), 13–40. https://doi.org/https://doi.org/10.3138/cmlr.53.1.13

Whaley, C. P. (1978). Word—nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*(2), 143–154.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. R package (Version 3.2.1). [Computer software]. Springer-Verlag NewYork. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., & Müller, K. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham, H., François, R., Henry, L., & Müller, K. (2019). dplyr: A Grammar of Data Manipulation. R package (Version 0.8.4) [Computer software].https://CRAN.R-project.org/package=dplyr

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. https://doi.org/10.1111/j.1540-4781.2013.01432.x

Winn, B. D., Skinner, C. H., Oliver, R., Hale, A. D., & Ziegler, M. (2006). The Effects of Listening While Reading and Repeated Reading on the Reading Fluency of Adult Learners. *Journal of Adolescent & Adult Literacy*, *50*(3), 196–205. https://doi.org/10.1598/jaal.50.3.4

Wollheim, R. (1980). Seeing-as, seeing-in, and pictorial representation. *Art and Its Objects*, *2*, 205–226.

Wollheim, R. (2001). *Richard Wollheim on the art of painting: art as representation and expression*. Cambridge University Press.

Wollheim, R. (2003). What makes representational painting truly visual? In *Aristotelian Society Supplementary Volume* (Vol. 77, pp. 131–147). Wiley Online Library.

Woodworth, R. S. (1938). *Experimental psychology*. London:Methuen.

Wright, S. (1954). The Death of Lady Mondegreen. *Harper's Magazine*, *209*(1254), 48–51.

Yu, C., & Smith, L. B. (2007). Rapid Word Learning Under Uncertainty via Cross-Situational Statistics, *18*(5), 414–420.

Zarei, A. A. (2009). The Effect of Bimodal, Standard, and Reversed Subtitling on L2 Vocabulary Recognition and Recall. *Pazhuhesh-e Zabanha-Ye Khareji*, *49*, 65–85.

Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, *15*, 41–44. https://doi.org/10.3758/BF03329756

Zhang, X. (2013). The I don't know option in the vocabulary size test. *TESOL Quarterly*, *47*(4), 790–811. https://doi.org/10.1002/tesq.98

Ziska, J. D. A. M. (2018). Art as Alchemy : The Bildobjekt Interpretation of Pictorial Illusion.

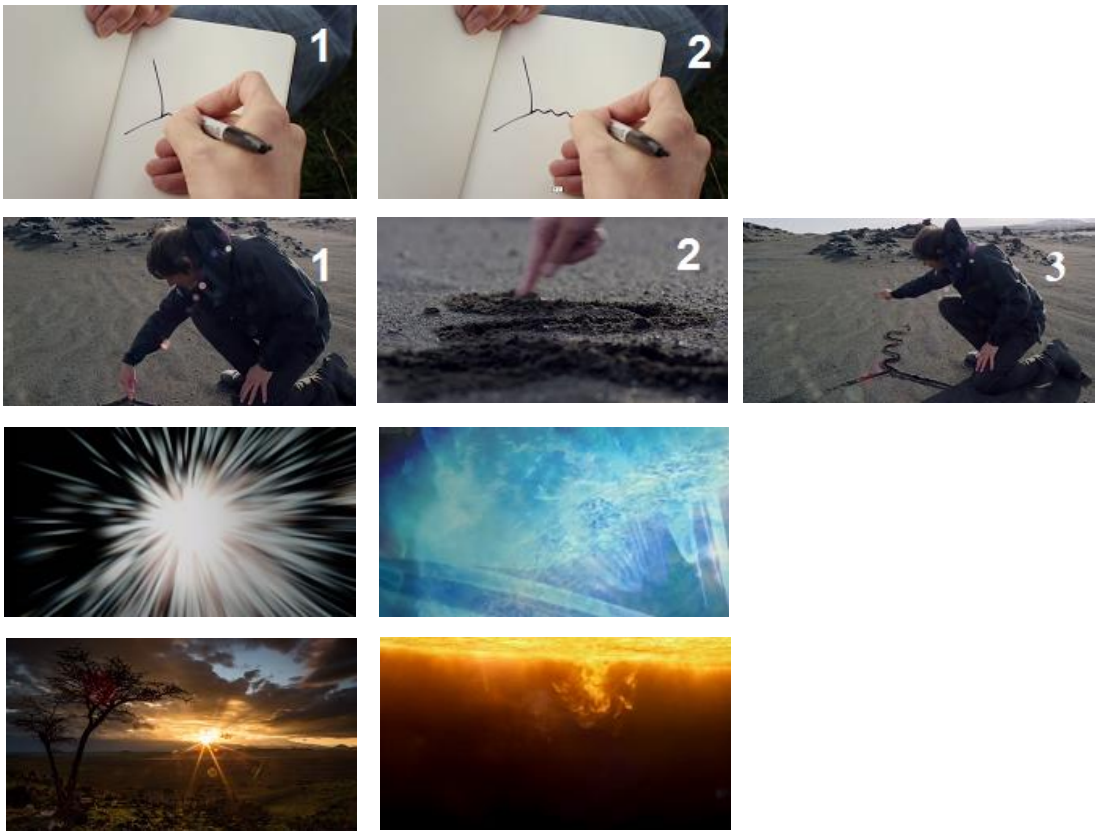
Appendices

Appendix A: Examples of Visual Referents

(Cooter et al., 2011; Cooter et al., 2016)

constellation

emit



Appendix A (continued)

manatee

Moth

Supernova

Appendix A (continued)

Cosmos

Strong

Weak

Fusion

Appendix A (continued)

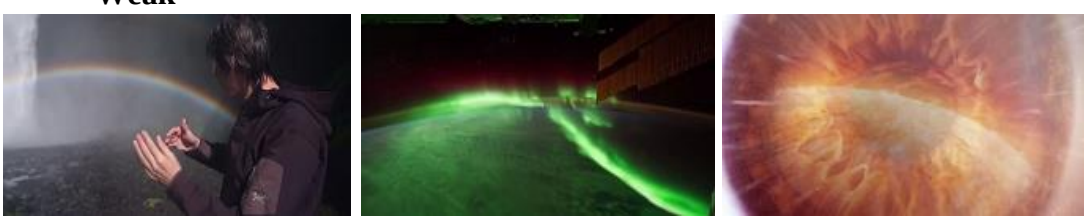
Faint

forge

Photon

Strong

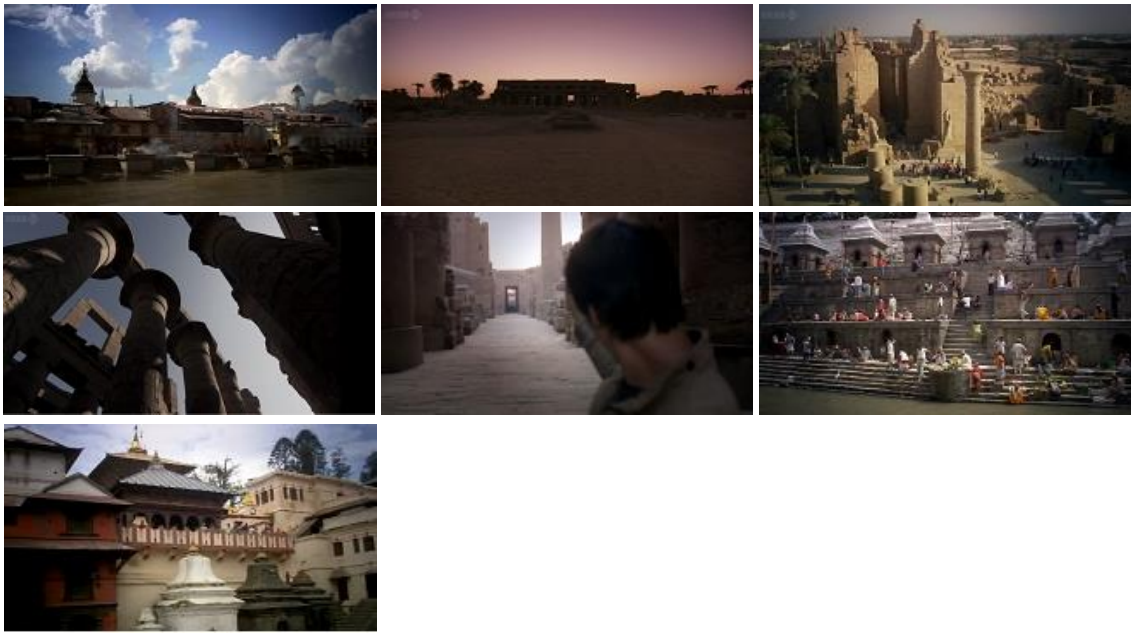
Weak



Appendix A (continued)

Hexagon

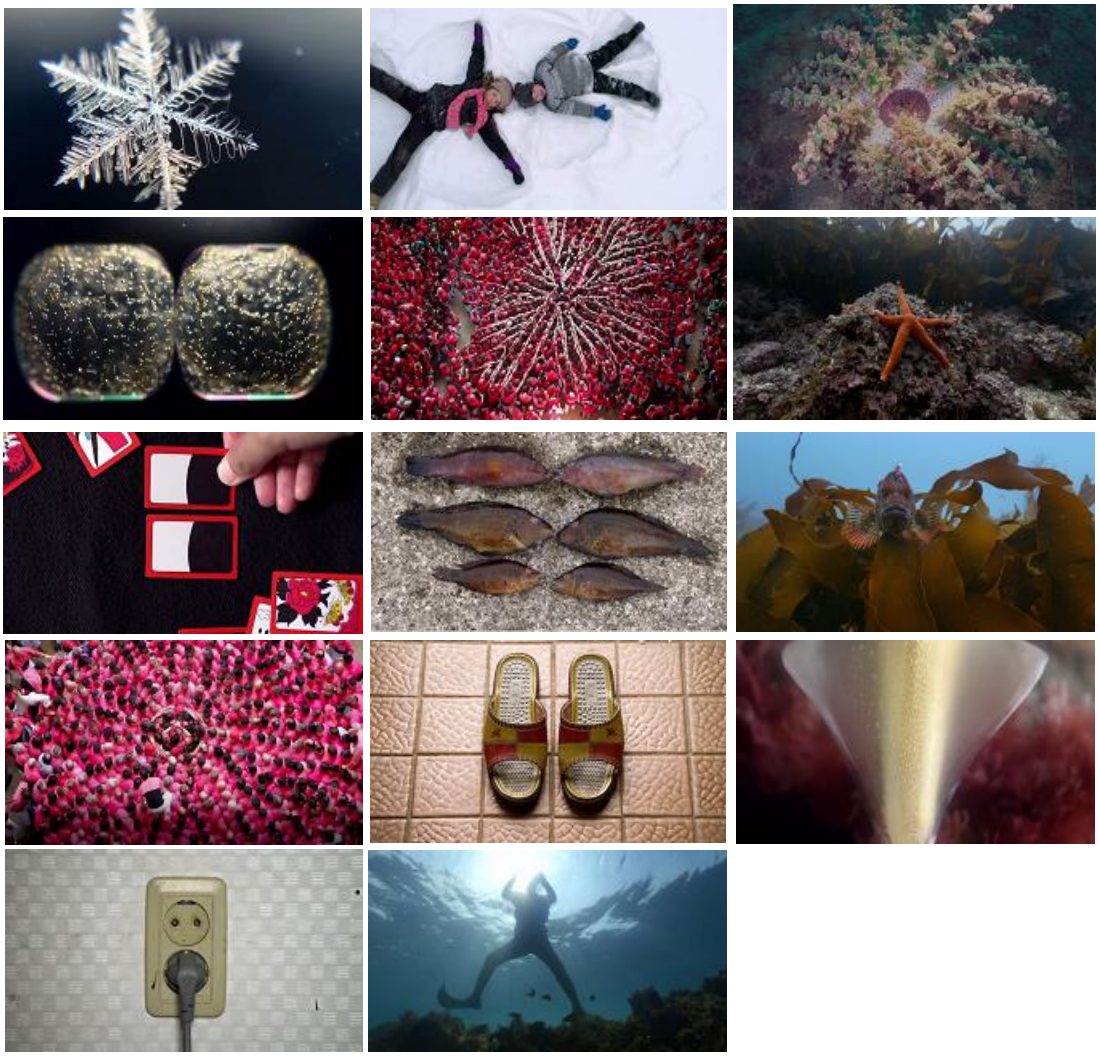
Temple



Appendix A (continued)

Symmetry

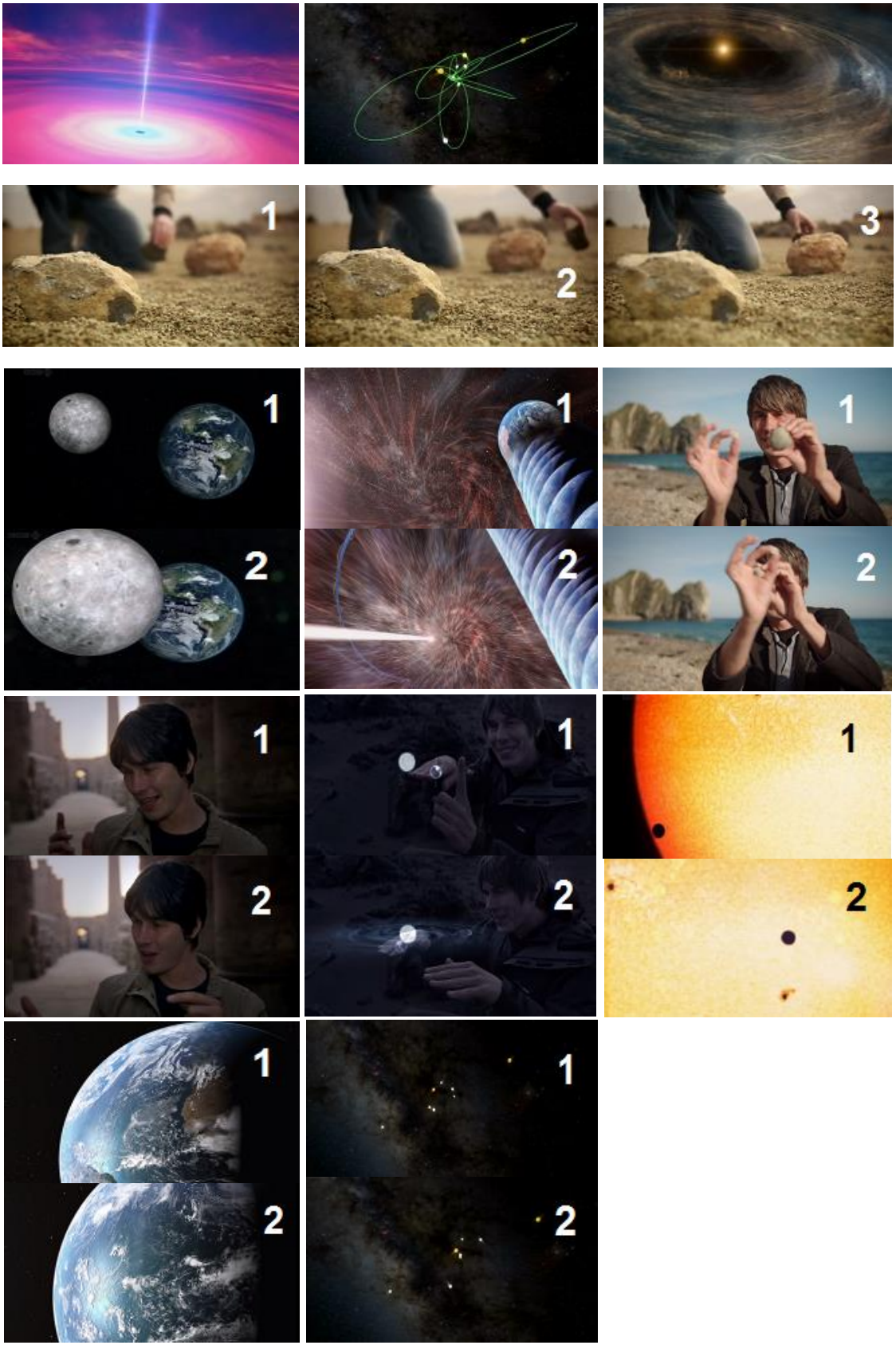
Strong



Weak

Appendix A (continued)

Orbit



Appendix A (continued)

Particles

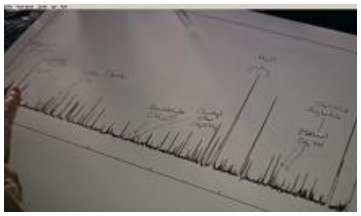


Pile

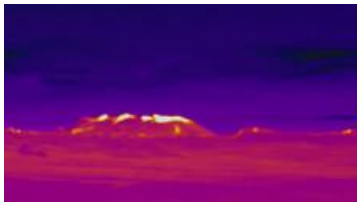
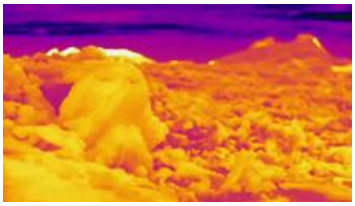
Strong **Weak**


Spectrum

Strong



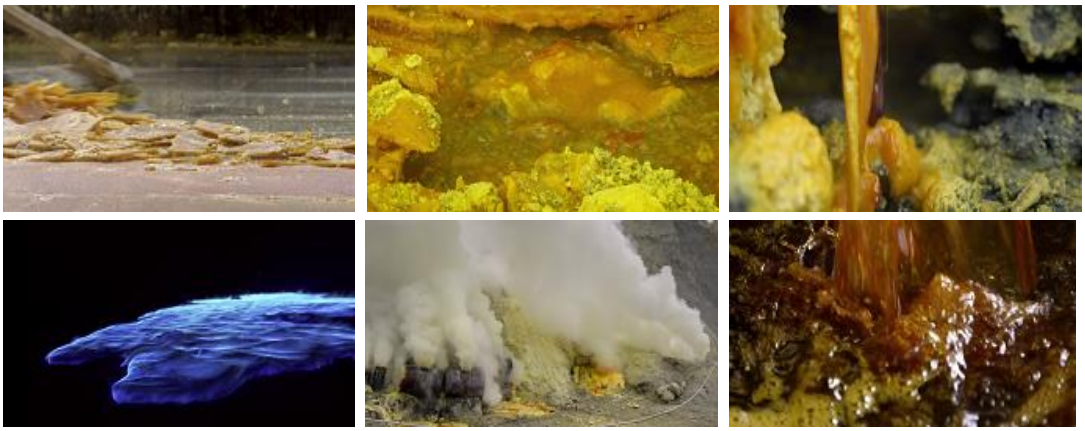
Weak

Appendix A (continued)

Stretch

Sulphur



Appendix A (continued)

Sphere

Strong

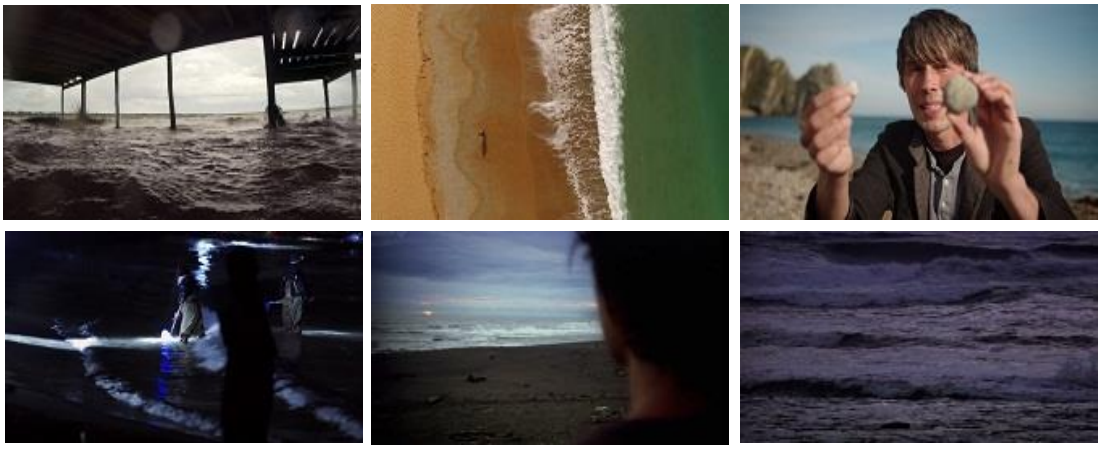
Weak

Appendix A (continued)

Tide

Strong

Weak



Appendix B

Distribution of Occurrence of 55 Potential Target Words in 18 Episodes of Documentary Series

Ep/words	U1	U2	U3	U4	F1	F2	F3	F4	SS1	SS2	SS3	L1	L2	L3	N	D	H1	H2
Entropy	13																	
Fusion		10																
snowflake					29													
iceberg					18													
manatee					14													
Symmetry					11													
Hexagon					10													
sulphur							16											
moth							11											
photon								32										
aurora								8										
iron							11											
pile	8																	
cosmos	17	1	8	6				3										
cosmic	2	3	1	4														
dense	2	2	6	2	1	1	1											
magnificent	4	1	3	4														
incredibly	4	8	4	2			1											
constellation	1	3	3	2				2										

(continued)

Appendix B (continued)

Ep/words	U1	U2	U3	U4	F1	F2	F3	F4	SS1	SS2	SS3	L1	L2	L3	N	D	H1	H2
primordial	1	1	1	1		1	1			1	1			1				
stellar	2	1	2	1					1							1		2
horizon	3		2	3		2												
rotate	1	1	3		3													
particle	1	1	1				1	8										
forge		3	2	1		1	1											
squash		1	2	2	2						1				2	1		
alien		1	3	1			1	3										
temple	1	2		7														
denser	1	2		5			1											
Orbit	1		4	3		2		1										
emit	1	2		1			1	6										
intricate	3		2	2						2		2	1			1		
supernova		9	2				1											
fuse		6		1					1				1	1	2			
nucleus		6	4		2		6	1										
spectrum		5		3				2										
sphere			2		11													
peer			2	5	1													
sculpt			2	1	7													
curve			12			1												
spontaneously	2				2					3							1	
ultimately	1	1	1	2	1			2		1			2				3	1
virtually	1			1								2	3				1	
literally	1	2	3	2			2											
relatively			1	2		1	1	2								1		

(continued)

Appendix B (continued)

Ep/words	U1	U2	U3	U4	F1	F2	F3	F4	SS1	SS2	SS3	L1	L2	L3	N	D	H1	H2
bounce	1	1						6										
dust	1	1	7	3		2	1	1										
faint	3	0	1	3	1											2		
float			2		9		2											
roughly			1	1	2						3				4	1		
tide	1		3			5	1											
stretch	1		5	8	1		1											
seemingly					3	2	1	1				1	2					
eventually	6	3	2		1	1	1											

Note. Ep = episode; words (9 adjectives, 10 adverbs, 12 verbs, and 24 nouns); 7 documentary series (U = Wonders of Universe ; F = Forces of Nature ; SS = Wonders of Solar System ; L = Wonders of Life ; N = Night with the Stars ; D = Science of Doctor Who; H = Human Universe).

Appendix C

Consent Form: Leadership Team

THE UNIVERSITY of York

DEPARTMENT OF EDUCATION
Heslington, York, YO10 5DD
Tel: + 44 (0) 1904 323 460
web: *http://www.york.ac.uk/education*

Leadership Team Information Page

(Head of the English Language Department)

Project: The Effects of Input Modality on Incidental Learning of Vocabulary

Dear Head of the Department,

I, Miss Souheyla Ghebghoub am currently carrying out a research project to investigate the effects of imagery on incidental vocabulary learning in the Algerian EFL class. I would like to request permission to conduct the study at your institution and have access to the language laboratories and/or the projector rooms.
Please ,carefully read the following, then complete the consent form in the next page.

The role of participants. Participants enrolled as tertiary students in the 2016/2017 academic year will be randomly recruited. They will be asked to self-report their knowledge of a set of words during no more than 15 minutes. This data will be used to adjust the research materials. 180 participants enrolled as tertiary students in the 2017/2018 academic year will be recruited using a proficiency test (about 40 minutes) and a stratified random sampling procedure. Participants will be assigned to one of three groups, a control group which will receive no treatment, a listening-while-reading group in which participants will read and listen to an authentic material, and a listening-while-reading *and viewing* group in which participants will watch a subtitled video. The study will be scheduled over 11 weeks: the experiment will consist of 4 sessions (each introducing an authentic material of about 120 minutes, with comprehension questions every 20 minutes). All participants will have to sit for a vocabulary pre-test and a post-test before and after the study, respectively. The tests should take no more than one hour to complete.

N.B. Students participation in this study is voluntary and they have the right to withdraw at any time up to one week after the end of the study when the data will be anonymised.

Anonymity and confidentiality. Students' names will be encoded into the format of letters and numbers so that they will not be identified as individuals. The anonymized data may be disseminated through seminars, conference presentations, journal articles, and other scholarly publications.

Storing and using your data: The collected data will be securely stored in locked files and only the researcher and her supervisor will be able to access the data. The data will be used as part of the PhD thesis and research publications but participants will not be identified as individuals in any of these. The data will be kept for approximately 6 years after which point it will be destroyed.

The research has been approved by the Department of Education, University of York.
If you have any question concerning students' involvement in this study, please contact:
Researcher : sg1339@york.ac.uk
Education Ethics Committee : education-research-administrator@york.ac.uk
Thank you for taking the time to read this information.

Yours sincerely,
Souheyla Ghebghoub

Appendix C (continued)

The following statements establish that you have read and understood what taking part in this research study will involve. Please tick the boxes that apply.

1- I confirm that I have read and understood the nature of the study in which students in my institution will take part. ☐

2- I confirm that I have read and understood the role of participants in this study. ☐

3- I understand that students' participation in this study is voluntary and that they are free to withdraw at any time up to one week after the end of the study. ☐

4- I understand that withdrawing from the study means that any data participants provided will be destroyed. ☐

5- I understand that the data students provide will be securely stored. ☐

6- I understand that any information students provide will be dealt with anonymously and that they will not be identified as an individual in the final thesis or any sort of publications. ☐

7- I confirm that I have had the opportunity to ask questions. ☐

8- I give the permission for the study to be conducted in this institution. ☐

Full name: ……

Institution:……

……

……..

Date: ……

Signature: ……

……..

……..

Appendix D

Consent Form: Norming Study Participants

Introduction

Dear Student,

Souheyla Ghebghoub is currently carrying out a research project about effects of input modality on Algerian tertiary EFL University students. You are kindly invited to participate in this online survey which aims to serve the initial stage of the research. You will be asked to indicate the extent to which you know a set of words. It should take approximately 10 minutes to complete. The data will be used to inform the appropriateness of the research materials; it will only be used for academic and research purposes.

Your participation in this survey is voluntary. If you agree to complete the questionnaire, you are free to exit the survey at any point without any penalty being imposed on you. Once the questionnaire is submitted, however, the data cannot be withdrawn as it is anonymous and there will be no way to identify your data. The anonymized data may be disseminated through seminars, conference presentations, journal articles, and other scholarly publications.

The research has been approved by the Dept of Education, University of York. For any questions or complaints, please contact (Souheyla, sg1339@york.ac.uk) or Chair of the Ethics Committee (education-research-administrator@york.ac.uk).

Please select your choice below.
Clicking on the “Agree” button indicates that:

- You have read the above information
- You voluntarily agree to participate
- You are 18 years of age or older

○ Agree
○ Disagree

Appendix E

Consent Form: Experiment Participants

THE UNIVERSITY of York

DEPARTMENT OF EDUCATION
Heslington, York, YO10 5DD
Tel: + 44 (0) 1904 323 460
Web: http://www.york.ac.uk/education

Participants Information Page

(2017/2018 Tertiary Students)

Project: The Effects of Input Modality on EFL Learners

Dear Participant

I, Miss Souheyla Ghebghoub, am currently carrying out a research project to investigate the effects that integration of different input modalities could have on EFL University students. You are kindly invited to participate in this study. Please carefully read the following, then complete the consent form in the next page.

The role of participants. I will give you a proficiency test to complete earlier before the start of the study, it will take no more than 40 minutes. Based on the results of the test and stratified random sampling, you will be allocated to either an experimental or a control group. If you will be assigned to the control group, then your role is to appear 2 times along the project period; during its first and last week, in order to complete assessment tasks. Each of the two sessions should take no more than one hour. If you will be assigned to an experimental group, then your task is to appear in 4 sessions every fortnight, other than the two ones mentioned earlier, that is, 6 times in total. During each experimental session, you will be exposed to authentic materials for content comprehension purposes. The overall material content for each session will be about 120 minutes and displayed in audio and written format or in audio, written, and pictorial format, depending on the group you will be allocated to. Breaks every 20 minutes will apply with a 10-minutes comprehension task.

N.B. Your participation in this study is voluntary. While your participation in all sessions is very important, you have the right to withdraw at any time up to one week after the end of the study. If you wish to do so, please notify the researcher, and any data that you provided will be destroyed.

Anonymity and confidentiality. The profile information that you will provide is your full name in consent form. You will be given numerical codes; each code is to be unique to each student so that you will not be identified as an individual. The anonymized data may be disseminated through seminars, conference presentations, journal articles, and other scholarly publications.

Storing and using your data: The collected data will be securely stored in locked files within internal and external hard drives and only the researcher and his supervisor will be able to access the data. The data will be used as part of the PhD thesis and research publications but you will not be identified as an individual in any of these. The data will be kept for approximately 6 years after which point it will be destroyed.

The research has been approved by the Department of Education, University of York.
If you have any questions, concerns, or complaints that you would like to raise before completing the form below, or after the data collection, please contact:

Researcher : sg1339@york.ac.uk
Education Ethics Committee : education-research-administrator@york.ac.uk

Thank you for taking the time to read this information.

Yours sincerely,
Souheyla Ghebghoub

Appendix E (continued)

The following statements establish that you have read and understood what taking part in this research study will involve. Please tick the boxes that apply.

1- I confirm that I have read and understood the nature of the study in which I will take part. ☐

2- I confirm that I have read and understood my role as a participant in this study. ☐

3- I understand that my participation in this study is voluntary and that I am free to withdraw at any time up to one week after the end of the study. ☐

4- I understand that withdrawing from the study means that any data I provided will be destroyed. ☐

5- I understand that the data I provide will be securely stored. ☐

6- I understand that data could be used for future analysis or other purposes, for up to six years, after which it will be destroyed. ☐

7- I understand that any information I provide will be dealt with anonymously and that I will not be identified as an individual in the final thesis or any sort of academic publications. ☐

8- I confirm that I have had the opportunity to ask questions. ☐

9- I agree to take part in this study. ☐

Full name: ……

Date: ……

Signature: ……

Contact detail: ……

Appendix F

Consent Form: Inter-coder

THE UNIVERSITY of York

DEPARTMENT OF EDUCATION
Heslington, York, YO10 5DD
Tel: + 44 (0) 1904 323 460
Web: http://www.york.ac.uk/education

Coder Information Page

Project: The Effects of Input Modality on EFL Learners

Dear coder,

I, Miss Souheyla Ghebghoub, am currently carrying out a research project to investigate the effects that integration of different input modalities could have on EFL University students. You are kindly invited to participate in this study. Please carefully read the following, then complete the consent form in the next page.

The role of the coder. To indicate whether a video segment constitutes a visual referent of a given word, and if it does, whether the visual referent is strong or weak based on certain coding criteria that will be provided in the inter-coding protocol. The process will be repeated for 97 short segments.

N.B. Your participation in this study is voluntary. You have the right to withdraw at any time. If you wish to do so, please notify the researcher, and any data that you provided will be destroyed.

Storing and using your data: The collected data will be securely stored in locked files within internal and external hard drives and only the researcher and her supervisor will be able to access the data. The data will be used as part of the PhD thesis and research publications, but you will not be identified as an individual in any of these. The anonymised data will be kept for approximately 6 years after which point it will be destroyed.

The research has been approved by the Department of Education, University of York.
If you have any questions, concerns, or complaints that you would like to raise before completing the form below, or after the data collection, please contact:

Researcher : sg1339@york.ac.uk
Education Ethics Committee : education-research-administrator@york.ac.uk

Thank you for taking the time to read this information.

Yours sincerely,
Souheyla Ghebghoub

Appendix F (continued)

The following statements establish that you have read and understood what taking part in this research study will involve. Please tick the boxes that apply.

1- I confirm that I have read and understood the nature of the study in which I will take part. ☐

2- I confirm that I have read and understood my role as a coder in this study. ☐

3- I understand that my participation in this study is voluntary and that I am free to withdraw at any time. ☐

4- I understand that withdrawing from the study means that any data I provided will be destroyed. ☐

5- I understand that the data I provide will be securely stored. ☐

6- I understand that data could be used for future analysis or other purposes, for up to six years, after which it will be destroyed. ☐

7- I understand that any information I provide will be dealt with anonymously and that I will not be identified as an individual in the final thesis or any sort of academic publications. ☐

8- I confirm that I have had the opportunity to ask questions. ☐

9- I agree to take part in this study. ☐

Full name: ……

Date: ……

Signature: ……

Contact detail: ……

Appendix G

Simplified Vocabulary Knowledge Scale (Online Version)

manatee

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

squash

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

roughly

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

Appendix H

Language Profile Questionnaire

Language Profile.
Please tick **ONLY ONE** (√) box

101

- Gender Male ☐ Female ☐
- N° of Years studying English language? ☐

PART 1:

- How often do you do the following in ENGLISH language?

	everyday	several times a week	a few times a week	couple times a month	rarely	never
watch videos with ENGLISH subtitles						
watch videos with Arabic/ French subtitles						
films						
documentaries						
popular TV series						
TV News						
sport games (e.g. football matches, tennis)						
listen to music						
listen to music while reading lyrics						
listen to audiobooks						
listen to audiobooks while reading script						
listen to talk radio						

PART2:

- How difficult to understand do you find the following in ENGLISH language?
  (You should place only one letter per box)

  a. Authentic videos without subtitles
  b. Authentic videos with ENGLISH language subtitles
  c. Radio programmes/audiobooks without script
  d. Radio programmes/audiobooks with script

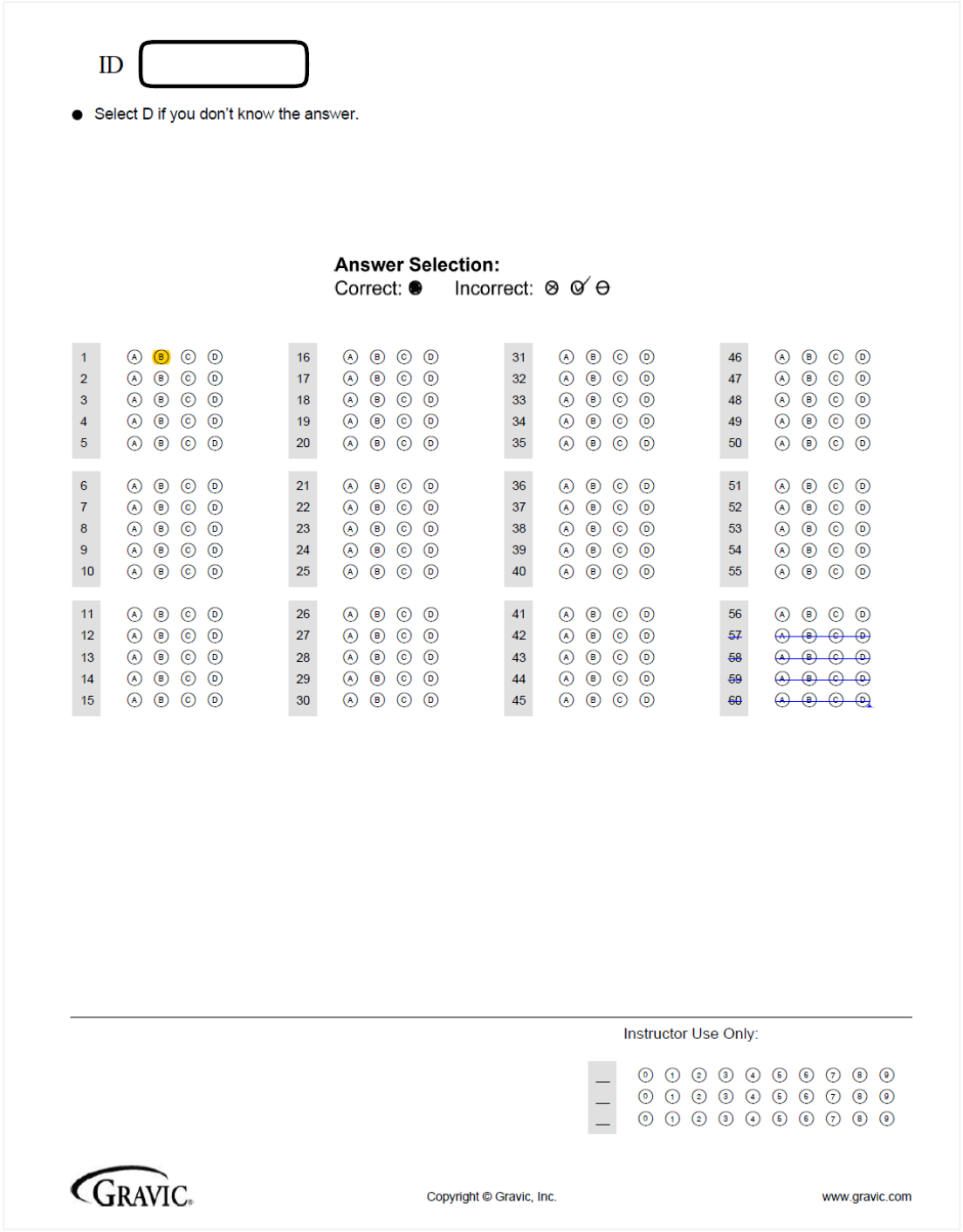
Easy (0) _____ : _____: _____: _____: _____ difficult (4)

Appendix I

Answer Sheets for Vocabulary Tests

Pretests

Section 1: Spoken Form Recognition



ID

● Select D if you don't know the answer.

Answer Selection:
Correct: ● Incorrect: ⊗ ✓ ⊖

1	A	**B**	C	D	16	A	B	C	D	31	A	B	C	D	46	A	B	C	D
2	A	B	C	D	17	A	B	C	D	32	A	B	C	D	47	A	B	C	D
3	A	B	C	D	18	A	B	C	D	33	A	B	C	D	48	A	B	C	D
4	A	B	C	D	19	A	B	C	D	34	A	B	C	D	49	A	B	C	D
5	A	B	C	D	20	A	B	C	D	35	A	B	C	D	50	A	B	C	D
6	A	B	C	D	21	A	B	C	D	36	A	B	C	D	51	A	B	C	D
7	A	B	C	D	22	A	B	C	D	37	A	B	C	D	52	A	B	C	D
8	A	B	C	D	23	A	B	C	D	38	A	B	C	D	53	A	B	C	D
9	A	B	C	D	24	A	B	C	D	39	A	B	C	D	54	A	B	C	D
10	A	B	C	D	25	A	B	C	D	40	A	B	C	D	55	A	B	C	D
11	A	B	C	D	26	A	B	C	D	41	A	B	C	D	56	A	B	C	D
12	A	B	C	D	27	A	B	C	D	42	A	B	C	D	~~57~~	~~A~~	~~B~~	~~C~~	~~D~~
13	A	B	C	D	28	A	B	C	D	43	A	B	C	D	~~58~~	~~A~~	~~B~~	~~C~~	~~D~~
14	A	B	C	D	29	A	B	C	D	44	A	B	C	D	~~59~~	~~A~~	~~B~~	~~C~~	~~D~~
15	A	B	C	D	30	A	B	C	D	45	A	B	C	D	~~60~~	~~A~~	~~B~~	~~C~~	~~D~~

Instructor Use Only:

__ 0 1 2 3 4 5 6 7 8 9
__ 0 1 2 3 4 5 6 7 8 9
__ 0 1 2 3 4 5 6 7 8 9

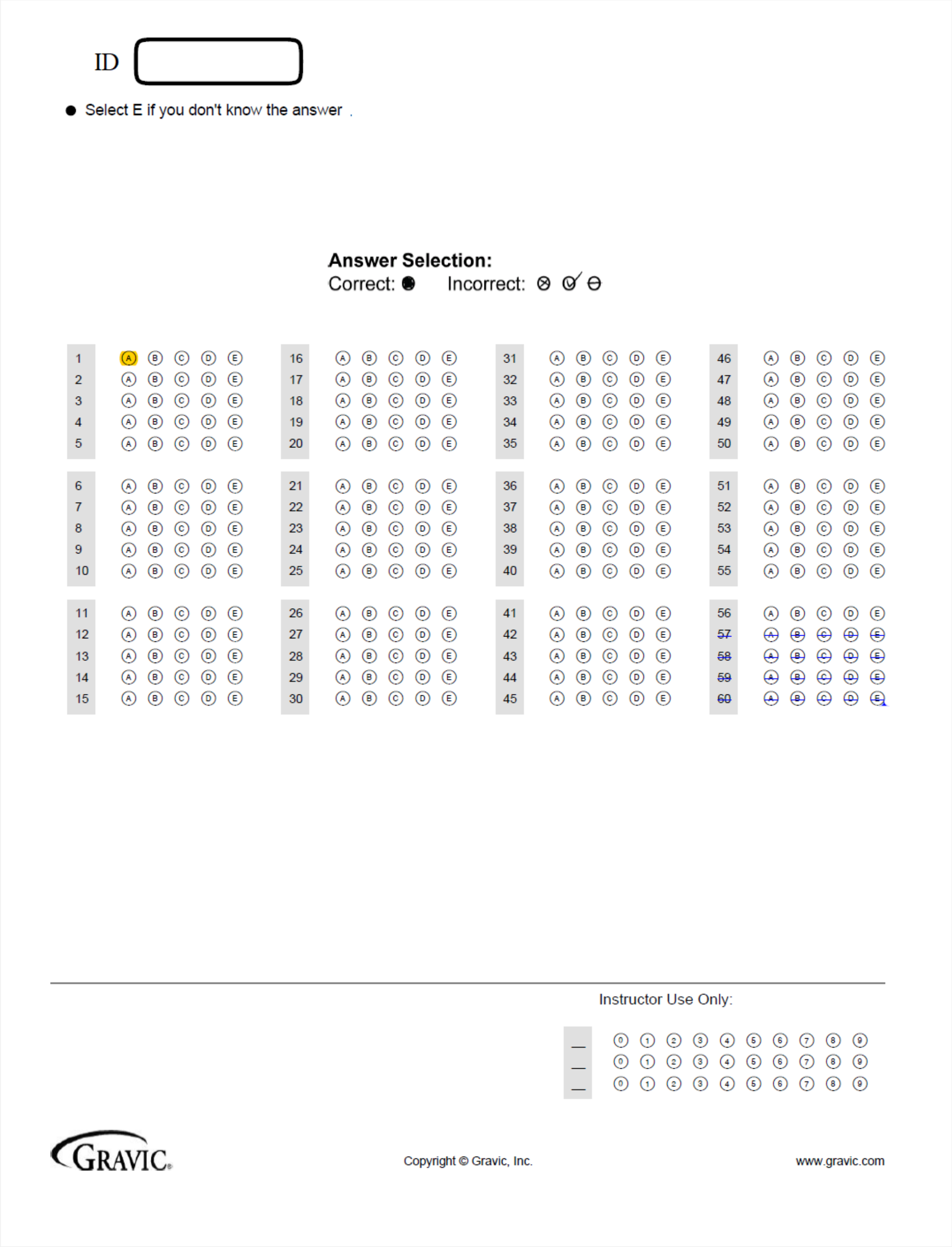
GRAVIC

www.gravic.com

Appendix I (continued)

Pretests

Section 2: Written Form Recognition



ID

- Select E if you don't know the answer.

Answer Selection:
Correct: ● Incorrect: ⊗ ✓ ⊖

1	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	16	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	31	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	46	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
2	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	17	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	32	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	47	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
3	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	18	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	33	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	48	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
4	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	19	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	34	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	49	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
5	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	20	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	35	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	50	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
6	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	21	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	36	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	51	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
7	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	22	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	37	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	52	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
8	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	23	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	38	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	53	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
9	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	24	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	39	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	54	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
10	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	25	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	40	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	55	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
11	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	26	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	41	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	56	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ
12	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	27	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	42	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	~~57~~	~~Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ~~
13	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	28	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	43	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	~~58~~	~~Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ~~
14	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	29	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	44	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	~~59~~	~~Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ~~
15	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	30	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	45	Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ	~~60~~	~~Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ~~

Instructor Use Only:

—	⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
—	⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
—	⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

GRAVIC

www.gravic.com

Appendix I (continued)

Posttests

Section 1: Spoken Form Recognition

ID

- Select D if you don't know the answer.

Answer Selection:
Correct: ● Incorrect: ⊗ ✓ ⊖

1	Ⓐ	Ⓑ	Ⓒ	Ⓓ	11	Ⓐ	Ⓑ	Ⓒ	Ⓓ	21	Ⓐ	Ⓑ	Ⓒ	Ⓓ	6	Ⓐ	Ⓑ	Ⓒ	Ⓓ
2	Ⓐ	Ⓑ	Ⓒ	Ⓓ	12	Ⓐ	Ⓑ	Ⓒ	Ⓓ	22	Ⓐ	Ⓑ	Ⓒ	Ⓓ	7	Ⓐ	Ⓑ	Ⓒ	Ⓓ
3	Ⓐ	Ⓑ	Ⓒ	Ⓓ	13	Ⓐ	Ⓑ	Ⓒ	Ⓓ	23	Ⓐ	Ⓑ	Ⓒ	Ⓓ	8	Ⓐ	Ⓑ	Ⓒ	Ⓓ
4	Ⓐ	Ⓑ	Ⓒ	Ⓓ	14	Ⓐ	Ⓑ	Ⓒ	Ⓓ	24	Ⓐ	Ⓑ	Ⓒ	Ⓓ	9	Ⓐ	Ⓑ	Ⓒ	Ⓓ
5	Ⓐ	Ⓑ	Ⓒ	Ⓓ	15	Ⓐ	Ⓑ	Ⓒ	Ⓓ	25	Ⓐ	Ⓑ	Ⓒ	Ⓓ	10	Ⓐ	Ⓑ	Ⓒ	Ⓓ
6	Ⓐ	Ⓑ	Ⓒ	Ⓓ	16	Ⓐ	Ⓑ	Ⓒ	Ⓓ	26	Ⓐ	Ⓑ	Ⓒ	Ⓓ					
7	Ⓐ	Ⓑ	Ⓒ	Ⓓ	17	Ⓐ	Ⓑ	Ⓒ	Ⓓ	27	Ⓐ	Ⓑ	Ⓒ	Ⓓ					
8	Ⓐ	Ⓑ	Ⓒ	Ⓓ	18	Ⓐ	Ⓑ	Ⓒ	Ⓓ	28	Ⓐ	Ⓑ	Ⓒ	Ⓓ					
9	Ⓐ	Ⓑ	Ⓒ	Ⓓ	19	Ⓐ	Ⓑ	Ⓒ	Ⓓ	29	Ⓐ	Ⓑ	Ⓒ	Ⓓ					
10	Ⓐ	Ⓑ	Ⓒ	Ⓓ	20	Ⓐ	Ⓑ	Ⓒ	Ⓓ	30	Ⓐ	Ⓑ	Ⓒ	Ⓓ					

Instructor Use Only:

_	0	1	2	3	4	5	6	7	8	9
_	0	1	2	3	4	5	6	7	8	9
_	0	1	2	3	4	5	6	7	8	9

GRAVIC

www.gravic.com

Appendix I (continued)

Posttests

Section 2: Written Form Recognition

ID

● Select E if you don't know the answer .

Answer Selection:
Correct: ● Incorrect: ⊗ ⊘ ⊖

1	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	11	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	21	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
2	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	12	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	22	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
3	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	13	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	23	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
4	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	14	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	24	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
5	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	15	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	25	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
6	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	16	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	26	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
7	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	17	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	27	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
8	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	18	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	28	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
9	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	19	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	29	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ
10	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	20	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	30	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ

Instructor Use Only:

__ ⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
__ ⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
__ ⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

GRAVIC

www.gravic.com

Appendix I (continued)

Posttests

Section 3: Meaning Recall

Questions:

1. **Write the equivalent L1 translation (Arabic) to the word you know. If you know 2 meanings, then please provide both.**
2. **Indicate its part(s) of speech (Ⓐ noun, Ⓑ verb, Ⓒ adjective).**

1-	_____	Ⓐ_Ⓑ_Ⓒ_	15-	_____	Ⓐ_Ⓑ_Ⓒ_
2-	_____	Ⓐ_Ⓑ_Ⓒ_	16-	_____	Ⓐ_Ⓑ_Ⓒ_
3-	_____	Ⓐ_Ⓑ_Ⓒ_	17-	_____	Ⓐ_Ⓑ_Ⓒ_
4-	_____	Ⓐ_Ⓑ_Ⓒ_	18-	_____	Ⓐ_Ⓑ_Ⓒ_
5-	_____	Ⓐ_Ⓑ_Ⓒ_	19-	_____	Ⓐ_Ⓑ_Ⓒ_
6-	_____	Ⓐ_Ⓑ_Ⓒ_	20-	_____	Ⓐ_Ⓑ_Ⓒ_
7-	_____	Ⓐ_Ⓑ_Ⓒ_	21-	_____	Ⓐ_Ⓑ_Ⓒ_
8-	_____	Ⓐ_Ⓑ_Ⓒ_	22-	_____	Ⓐ_Ⓑ_Ⓒ_
9-	_____	Ⓐ_Ⓑ_Ⓒ_	23-	_____	Ⓐ_Ⓑ_Ⓒ_
10-	_____	Ⓐ_Ⓑ_Ⓒ_	24-	_____	Ⓐ_Ⓑ_Ⓒ_
11-	_____	Ⓐ_Ⓑ_Ⓒ_	25-	_____	Ⓐ_Ⓑ_Ⓒ_
12-	_____	Ⓐ_Ⓑ_Ⓒ_	26-	_____	Ⓐ_Ⓑ_Ⓒ_
13-	_____	Ⓐ_Ⓑ_Ⓒ_	27-	_____	Ⓐ_Ⓑ_Ⓒ_
14-	_____	Ⓐ_Ⓑ_Ⓒ_	28-	_____	Ⓐ_Ⓑ_Ⓒ_

Appendix I (continued)

Posttests

Section 4: Meaning Recognition

STUDENT NUMBER			Immediate Posttest	TOTAL SCORE		

QUESTION: Match ↘ each word on the left side to its equivalent translation on the right side. Please note, one translation was added for distraction.

BLOCK 1

1.	1. جُسَيْمْ
2.	2. سْدَاسِيَ
3.	3. كَثِيف
4.	4. طَيْف
5.	5. مَدّ و جَزْر
6.	6. بَعَثَ
7.	7. شَكَّل
8.	8. انفجار نجمي
9.	9. مجموعة نجوم
10.	10. كَوْنِيَ
11.	11. فراشة
12.	12. غريب
13.	13. مَعْبَد
14.	14. تَمَدَّدَ
	15. أَكْثَفْ

BLOCK 2

1.	1. كومة
2.	2. تناظر
3.	3. مُعَقَّدْ
4.	4. اندماج
5.	5. سحق
6.	6. كَمُ مِن الضوء
7.	7. دَارَ
8.	8. بَعَثَ
9.	9. خروف البحر
10.	10. كَوْن
11.	11. تَحَتَ
12.	12. كبريت
13.	13. مجموعة نجوم
14.	14. خَافِتْ
	15. جسم كروي

Appendix J

Pretest Target and Filler Items

Spoken Form Recognition

(continued)

	Word	Correct	non-word	non-word
1	match	B. /matʃ/	A./ mɒtʃ/	C./mɛtʃ/
2	regent	A./ˈriːdʒ(ə)nt/	B./ˈreiʒ(ə)nt/	C.ˈrɒdʒ(ə)nt/
3	constellation	A./kɒnstəˈleɪʃ(ə)n/	B./kənstəˈleɪʃ(ə)n/	C./kɒntəˈleɪʃ(ə)n/
4	null	A./nʌl/	B./ nɒl/	C./ niːl/
5	fail	C. /feɪl/	A. /leɪf/	B. /lif/
6	temple	A. /tɛmp(ə)l/	B./rɛmp(ə)l/	C./vɛmp(ə)l/
7	standard	A./standəd/	B./stɒndəd/	C./stʊndəd/
8	thesaurus	C./θɪˈsɔːrəs/	A./θɪˈsɔːməs/	B./θɪˈrɔːsəs/
9	supernova	B./suːpəˈnəʊvə/	A. /suːpəˈni :və/	C./suːtəˈna :və/
10	alien	A./eɪlɪən/	B./ʌɪliən/	C./feɪlɪən/
11	reptile	C./rɛptʌɪl/	A./ˈrɛptɒl/	B./ˈ rɛbtʌɪl/
12	mumble	B./mʌmb(ə)l/	A./mʌb(ə)l/	C./mʌlb(ə)l/
13	knife	B./nʌɪf/	A./knʌɪf/	C./nɔɪf/
14	spectrum	A./spɛktrəm/	B./sɛktrəm/	C./ faktrəm/
15	dense	C./dɛns/	A./ rɛns/	B./vɛns/
16	stretch	A./strɛtʃ/	B./ strɒtʃ/	C./ strɪːtʃ/
17	pigtail	B./pɪgteɪl/	A./pɪgtoɪl/	C./bɪgteɪl/
18	particle	C./pɑːtɪk(ə)l/	A./pɑːtɪp(ə)l	B./pəːtɪk(ə)l
19	denser	B./dɛnsə/	A./rɛnsə/	C. /vɛnsə/
20	spleen	A./spliːn/	B./sliːn/	C./pliːn/
21	rouble	C./ruːb(ə)l/	A./reib(ə)l/	B./ri :b(ə)l/
22	forge	B./ fɔːdʒ/	A./fɑː dʒ /	C./v ɑː dʒ/
23	hexagon	A./hɛksəg(ə)n/	B./hɛksədʒən/	C./fɛksəʒən/
24	integrate	A./ɪntɪgreɪt/	B/ɪntrɪgreɪt/	C./ɪntɪgeɪt/
25	fusion	B./fjuːʒ(ə)n/	A./fjuːdʒ(ə)n/	C./fjuːz(ə)n/
26	emit	A./ɪˈmɪt/	B./ ɪˈmɪ θ/	C./ ɪˈmɪ p/
27	transcription	B./tranˈskrɪpʃ(ə)n/	A./tranˈsrɪpʃ(ə)n/	C./tranˈskrɪbʃ(ə)n/
28	sculpt	C./skʌlpt/	A./skrʌlpt/	B./skʌmpt/
29	pile	C./pʌɪl/	A./pʌɪm/	B./pɔɪl/
30	orbit	A./ˈɔːbɪt/	B./ˈɔːbɪf/	C./ˈɔːbɪd/

Appendix J (continued)

	Word	Correct	non-word	non-word
31	limpid	A./ˈlɪmpɪd/	B./lɪmbɪd/	C./lɪmɪd
32	sulphur	C./sʌlfə/	A./sɪlfə/	B./ sfɛː/
33	photon	A./fəʊtɒn/	B./f ɒ t ɪ n/	C./faʊntɒn/
34	devious	A./ˈdiːvɪəs/	B./diːvɪəʃ/	C./diːvəsɪəs/
35	maintain	A./meɪnˈteɪn/	B./meɪnˈseɪn/	C./meɪnˈdeɪn/
36	tide	C./tʌɪd/	A./mʌɪd/	B./ vʌɪd/
37	emeritus	A./ɪˈmɛrɪtəs/	B./ɪˈmɛrʃəs/	C./ɪˈmɛrtiəs/
38	cosmos	B./kɒzmɒs	A./kɒsmɒs/	C./kɒmsos/
39	trill	A./trɪl/	B./trɪt/	C./trid/
40	faint	A./feɪ nt/	B./f ɔːnt/	C./faː nt/
41	squash	A./skwɒʃ/	B./skwɒtʃ/	C./skwɒl/
42	File	B.fʌɪl/	A./sʌɪl/	C./ʌɪl/
43	moth	C./mɒθ/	A./məːθ/	B./mʌɪθ/
44	intricate	B./ɪ n t rɪkət/	A./ɪ n t rikeit/	C./ɪ n t rigət/
45	hutch	B./hʌtʃ/	A./fʌtʃ/	C./jʌtʃ/
46	cosmic	B./kɒzmɪk/	A./kɒsmɪk/	C./kɒsnɪk/
47	symmetry	C./s ɪ m ɪ tri/	A./s ɪ m əntri/	B./s ɪ m ɪstri/
48	aviary	C./eɪvɪəri/	A./ˈI :vɪəri/	B./ˈ ɔːvɪəri/
49	wonder	A./ wʌndə/	B./wʌnfə/	C.ˈ/wɒnfə/
50	manatee	A./manətiː/	B./manətəʊn/	C./manəfəʊn/
51	beagle	A./ˈbiːg(ə)l/	B./bɑg(ə)l/	C./bɪg(ə)l/
52	sphere	B./ sfɪə/	A./ sfɛː/	C./ sfɒr /
53	impale	A./ɪ mˈpeɪl/	B./ɪ mˈ feɪl//	C./ɪ mˈ pɔɪl//
54	tense	B./ tɛns/	A./rɛns/	C./mɛns /
55	lush	C./lʌʃ/	A./lɪʃ/	B./lɒʃ/
56	stealth	B./stɛlθ/	A./stɛlt/	C./tɛlθ/

Note. A, B, C is the sequence of presentation.
Options were presented in pseudo-random order in terms of the function of item (target/filler), its part of speech, and the position of the correct spoken form.

Appendix J (continued)

Written Form Recognition

	A.	B.	C.	D.
1	**thesaurus**	thesaumus	thauresus	thesomus
2	sperniva	superneve	sperneva	**supernova**
3	**alien**	feillen	feelian	alian
4	reptoil	rebtile	rebtial	**reptile**
5	muble	mulble	malble	**mumble**
6	**constellation**	consellation	contilation	contillation
7	noll	**null**	neell	nall
8	**sphere**	sphur	sphor	sphure
9	insicrate	**intricate**	entricate	intrigate
10	motch	mutch	meatch	**match**
11	**knife**	nife	noif	noife
12	sectrum	**spectrum**	fectrum	sactrum
13	rence	**dense**	vence	dence
14	sretch	sritch	sreatch	**stretch**
15	**particle**	partiple	purtical	purticle
16	rencer	vencer	dencer	**denser**
17	remple	vemple	**temple**	rample
18	rizent	**regent**	rogent	reegent
19	reible	**rouble**	reeble	rible
20	**farge**	vurge	varge	forge
21	hexapon	fexagon	fexagen	**hexagon**
22	pigtoil	pigteil	**pigtail**	bigtail
23	intregrate	**integrate**	integreat	integate
24	fusian	fewsion	**fusion**	fudjon
25	emith	vemit	imit	**emit**
26	**transcription**	transription	transruption	trunsription
27	scrulpt	**sculpt**	scumpt	scampt
28	**pile**	pime	poile	Pial
29	orbif	arbit	arbbit	**orbit**
30	limbid	**limpid**	limid	limbide

(continued)

Appendix J (continued)

	A.	B.	C.	D.
31	imershess	**emeritus**	emertious	Emershess
32	cozmos	**cosmos**	rosmos	rozmos
33	trit	trid	tril	**trill**
34	**faint**	faunt	feant	fent
35	squatch	**squash**	squitch	skwitch
36	sile	ile	**file**	fyle
37	ploton	photin	fotin	**photon**
38	devioush	devesious	divesious	**devious**
39	maintein	maindain	**maintain**	meinsain
40	tider	toid	tighd	**tide**
41	**moth**	meith	maith	mooth
42	sleen	pleen	spleegne	**spleen**
43	**hutch**	futch	jatch	fatch
44	**cosmic**	cozmic	rozmic	rozmic
45	**symmetry**	simmentry	semistery	semmentry
46	spher	sphur	**sulphur**	silpher
47	**impale**	imfail	impoil	imphale
48	rense	**tense**	mense	rince
49	manatone	manatea	manaphon	**manatee**
50	stelt	tealth	stealt	**stealth**
51	**wonder**	wonfer	onedare	onefare
52	lish	**lush**	laush	luash
53	**beagle**	beegle	bagle	bigle
54	**aviary**	eiviary	eviary	oviary
55	strondard	standed	staundard	**standard**
56	laif	lif	**fail**	feil

Note. Options were presented in pseudo-random order in terms of the function of item (target/filler), its part of speech, and the position of the correct written form. The item in bold is the correct form.

Appendix K

Comprehension Questions

Session 1

Questions 1-2.

1. The sun rising between the 2 pillars marks the summer solstice.

True /**False**

2. How many stars are in the Milky Way Galaxy?

A. 20 B

B. 200 B

C. 2000 B

Questions 3-4.

3. Andromeda is a spiral galaxy, i.e.,

A. It has red stars.

B. It is like a big egg filled with light.

C. It has a lighted centre encircled by ringed arms .

4. Every point of light in The Hubble Ultra Deep Field is not a star but a galaxy.

True / False

Questions 5-6.

5. What can you find at the Burgess Shale?

A. Fossils

B. Slate Mine

C. Meteorites

6. What erupted during the Cambrian Explosion?

A. Mount Etna

B. The sun

C. Life

Appendix K (continued)

Questions 7-8

7. The Skoga River tumbled after the ice sheets melted.

True / False

8. Our eyes cannot detect infrared light.

True / False

Questions 9-10

9. The rain came after _____

A. 4 months

B. 3 months

C. 6 months

10. Which of the following contributes to Para Kapooni's reunion with family?

A. light

B. photosynthesis

C. the sun

Questions 11-12

11. What does our planet look like from 6 billion kilometres away?

A. a pale green dot

B. a pale white dot

C. a pale blue dot

12. The pink colour of the aurora comes from oxygen atoms.

True / **False**

Session 2

Questions 1-2

1. Chankillo works as a calendar that tells the year.

True / **False**

2. What does Professor Brian Cox use as a metaphor for "the arrow of time"?

A. Perito Moreno Glacier

B. Namibia

C. The Milky Way

Appendix K (continued)

Questions 3-4

3. There are very few ways of rearranging the grains of the sandcastle without changing its structure, this means that:

A. the sandcastle has no entropy.

B. the sandcastle has low entropy.

C. the sandcastle has high entropy.

4. In the future, the universe will be less ordered.

True / False

Questions 5-6

5. Since Proxima Centauri burns very slowly, it will be the first dying star in the universe.

True / **False**

6. The sun will eventually become a dwarf.

A. a star of small size and red light

B. a star of big size and low luminosity

C. a star of low luminosity and white light

Questions 7-8

7. _____ gives the towers the strength against collapsing.

A. Gravity

B. the push inwards in all directions by people on the ground

C. David Merit

8. The bee produces 1 gram of wax by consuming more than six grams of honey.

True / False

Questions 9-10

9. A huge research paper of mathematics proved that bees build beehives from scratch using an instinctive behaviour.

True / **False**

10. The power station helps to provide warmth for people.

True / **False**

Appendix K (continued)

Questions 11-12

*11. Which of the following **is not** one of the 4 forces of nature needed to describe the snowflake's journey to the ground?*

A. molecules

B. electromagnetism

C. Gravity

12. Two snowflakes are:

A. alike

B. different

Session 3

Questions 1-2

1. What is the plane that simulates zero-gravity also known as?

A. Plane Insane

B. Vomit Comet

C. Float Boat

2. According to the documentary, gravity helps shaping the world.

True / False

Questions 3-4

3. _____ has a gravity that is hundred million times as strong as on Earth.

A. The Moon

B. Pluto

C. Neutron Star

4. Which planet could we mostly survive at its surface?

A. Wasp-8B

B. Neptune

C. Jupiter

Appendix K (continued)

Questions 5-6

5. Einstein's theory of photoelectric effect was used in this documentary.

True / **False**

6. There is a black hole at the centre of the Milky Way Galaxy.

True / False

Questions 7-8

7. What are the building blocks of the universe?

A. Protons

B. Atoms

C. Chemical elements

8. We are made of the same elements that make up the planet.

True / False

Questions 9.10

9. Why is water called "the universal solvent"?

A. It dissolves more substances than any other liquid.

B. It is the most complex substance.

C. It carries the ingredients of life.

10. Water is a nonpolar molecule.

True / **False**

Questions 11-12

11. Chemical reactions in squids release energy as a blue light

True / False

12. Humans are made of the basic ingredients of:

A. Chemistry

B. Life

C. Earth

Appendix K (continued)

Session 4

Questions 1-2

1. Our world is made up of how many elements?

A. 92

B. 72

C. 52

2. The rock in the Himalayas Mountains was originally formed in the ocean.

True / False

Questions 3-4

3. When a star runs out of hydrogen, it begins to shine.

True / **False**

4. The sun converts hydrogen into _____

A. lithium

B. helium

C. nitrogen

Questions 5-6

5. Complex chemistry is happening in Orion nebula.

True / False

6. What does "Betelgeuse" mean?

A. a full moon

B. a neutron star

C. a galaxy

Questions 7-8

7. What is the plane that flies twice the speed of sound known as?

A. Storm

B. Thunder

C. Typhoon

8. The sunrise appeared again as the plane accelerated.

True / False

Appendix K (continued)

Questions 9-10

9. The centrifugal force tries to _____ everything

A. throw

B. pull

C. rotate

10. As night falls, the beetles can no longer navigate.

True / **False**

Questions 11-12

11. Cox explained that because the earth moves, moments differ in location too.

True / False

12. Once we pass summers or winters,

A. they cease to exist in space-time.

B. they do not cease to exist in space-time.

Appendix L

Online Debriefing Questionnaire

Debriefing Survey

I have greatly valued your participation in my research.
Thank you very much for your efforts and precious time, thank you for being so wonderful. You are kindly invited to fill out this short survey.
Please select your choice below:

Question 1. To which group did you belong in this research?

- Reading-while-listening + viewing
- Reading-while-listening
- Tests only

Skip to end of survey if "To which group did you belong in this research?" = Tests only

Question 2. How good was your comprehension of the documentary episodes?

Question 3. How easy or difficult was it to process the input in the documentary series?

Question 4. How satisfied or dissatisfied were you with the length of the episodes?

Question 5. How likely are you to attempt to learn vocabulary by following more TV programs?

Question 6. How helpful or unhelpful was the written input (L2 captions) all over the episodes?

Question 7. Which of the following is the voice of the documentaries' presenter?

- Voice 1
- Voice 2
- Voice 3

Display Question 8. If "To which group did you belong in this research?" = Listening-while-reading + viewing

Question 8. How distracting was it to split your attention from the image area to the subtitle area?

Display Question 9. If "To which group did you belong in this research?" = Listening-while-reading + viewing images

Question 9. To what extent do you think that watching the episodes of the documentary series affected your intrinsic/integrative motivation as an EFL learner?

Display Question 10. If "To which group did you belong in this research?" = listening-while-reading

Question 10. How did you feel about following the episodes of the documentary series without imagery?

Question 11. To what extent do you think that following the episodes through Listening-while-Reading affected your intrinsic/integrative motivation as an EFL learner?

Note. The survey consisted of 3AFC items and 10-point Likert-scale based items.

Appendix M

Inter-Coding Protocol

Inter-Coding Protocol

Coder Profile

Gender: Male ☐ Female ☐

Age: .....

Mother tongue: .....

Second/foreign language(s): .....

IELTS test results: .....

Thank you for taking part in this inter-coding reliability assessment that is needed for my PhD research.

You will be presented with 97 short video segments (timespans), each corresponds to a specific word and lasts for a few seconds.

Please watch the segment carefully, and indicate on your answer sheet if you believe the segment consists of a visual referent for the given word. Use dichotomous coding (0, 1): "0" for "no visual referent" and "1" for "there is a visual referent".

If your code = "1", please indicate whether you perceive the referent as strong "1" or weak "0". The coding should be based on the following criteria:

Coding Criteria

Strong Referents

An image be it still (e.g., object) or dynamic (e.g., action, event) should be coded as a *strong* referent if it met the following:

- It is large enough to be noticeable.
- It conveys the meaning accurately, especially in the case of a dynamic image.
- It is immediately perceived as a visual referent for the target word due to its overall observable properties.
- It is small/distant but supported with non-verbal signs or high visual saliency.

Weak Referents

An image should be coded as *weak* referents if it has some degree of narrative saliency and meets the following:

- It is small and distant with distracting images.
- It might not be straightforward to discern if the image possessed the required properties to be a visual referent.

Appendix M (continued)

Inter-coder Sheet

Word	Timespan		Episode	Is there a visual referent?	Is it a strong referent?
	start	end			
particle	00:08:05	00:08:21	F4		
	00:08:30	00:08:49	F4		
supernova	00:30:09	00:30:10	U3		
	00:45:11	00:45:36	U2		
pile	00:31:22	00:31:35	U1		
	00:33:42	00:33:53	U1		
emit	00:01:20	00:01:25	F4		
	00:15:28	00:15:36	U2		
	00:35:34	00:35:37	U4		
forge	00:01:50	00:01:53	U3		
	00:09:09	00:09:16	U4		
	00:30:50	00:30:53	F2		
temple	00:02:35	00:03:08	U2		
	00:02:57	00:03:22	U1		
	00:03:00	00:03:25	U4		
squash	00:08:20	00:08:21	U2		
	00:12:02	00:12:07	F1		
	00:35:57	00:35:58	U4		
spectrum	00:07:40	00:07:49	F4		
	00:15:39	00:16:03	U2		
	00:35:37	00:35:45	U4		
	00:52:53	00:53:01	U2		
stretch	00:15:43	00:15:47	U2		
	00:24:04	00:24:07	F3		
	00:33:59	00:34:03	U4		
	00:35:34	00:35:37	U3		
faint	00:22:37	00:22:44	U1		
	00:29:39	00:29:51	U4		
	00:31:39	00:32:04	U4		
	00:44:00	00:44:01	U1		
sulphur	00:05:41	00:05:45	F3		
	00:08:40	00:08:52	F3		
	00:09:36	00:09:41	F3		
	00:11:10	00:11:12	F3		
dense/denser	00:19:37	00:19:55	F3		
	00:23:20	00:23:30	F2		
	00:27:34	00:27:38	U2		
	00:45:25	00:45:26	U4		

(continued)

Appendix M (continued)

Word	Timespan		Episode	Is there a visual referent?	Is it a strong referent?
	start	end			
fusion	00:25:24	00:25:35	U2		
	00:26:01	00:26:05	U2		
	00:27:00	00:27:01	U2		
	00:34:47	00:34:54	U2		
photon	00:08:29	00:08:34	F4		
	00:29:58	00:30:00	F4		
	00:49:27	00:49:38	F4		
	00:55:28	00:55:36	F4		
manatee	00:26:22	00:26:23	F1		
	00:27:41	00:27:51	F1		
	00:28:27	00:28:33	F1		
	00:31:08	00:32:02	F1		
tide	00:10:03	00:10:15	U1		
	00:21:45	00:22:10	F2		
	00:26:24	00:26:33	F2		
	00:44:08	00:44:11	F3		
moth	00:02:22	00:02:26	F3		
	00:02:28	00:02:34	F3		
	00:02:42	00:03:24	F3		
	00:46:58	00:47:03	F3		
hexagon	00:02:56	00:02:58	F1		
	00:23:29	00:23:31	F1		
	00:25:13	00:25:14	F1		
	00:25:35	00:25:47	F1		
	00:52:32	00:52:37	F1		
constellation	00:20:53	00:20:59	U1		
	00:24:49	00:24:55	U4		
	00:51:18	00:51:24	U2		
	00:55:44	00:55:50	F4		
	00:47:50	00:47:59	U2		
symmetry	00:07:59	00:08:01	F1		
	00:29:25	00:29:33	F1		
	00:33:55	00:34:00	F1		
	00:36:44	00:36:47	F1		
	00:36:59	00:37:10	F1		
	00:40:27	00:40:31	F1		
	00:40:56	00:41:00	F1		

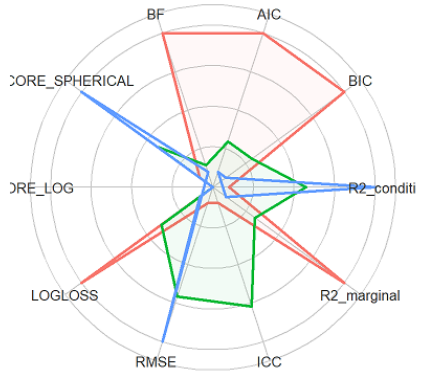
(continued)

Appendix M (continued)

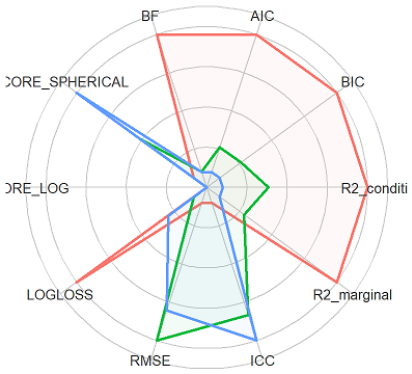
Word	Timespan		Episode	Is there a visual referent?	Is it a strong referent?
	start	end			
orbit	00:03:37	00:03:39	U4		
	00:19:08	00:19:11	U4		
	00:20:22	00:20:25	F2		
	00:37:39	00:37:55	U3		
	00:43:17	00:43:30	U3		
	00:50:14	00:50:23	U3		
	00:56:14	00:56:26	F4		
sphere	00:03:01	00:03:03	F4		
	00:03:45	00:04:09	F1		
	00:14:26	00:14:33	F1		
	00:30:36	00:30:57	F1		
	00:33:19	00:33:25	F4		
	00:36:24	00:36:31	F4		
	00:56:37	00:56:56	F4		
cosmos/cosmic	00:01:33	00:01:38	U4		
	00:07:16	00:07:27	U1		
	00:08:44	00:08:57	U1		
	00:30:37	00:30:49	U2		
	00:40:06	00:40:24	U1		
	00:31:41	00:31:48	U3		
	00:30:17	00:30:24	U2		
	00:50:43	00:51:03	U3		

Appendix N

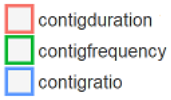
Spiderweb Plot of Performance Model Comparison

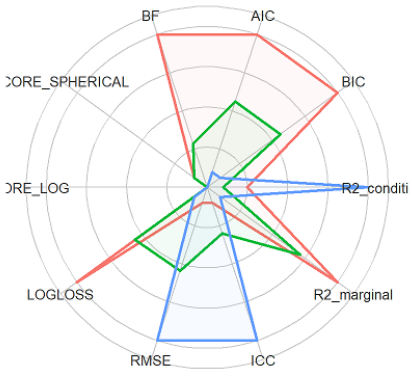


meaning recall

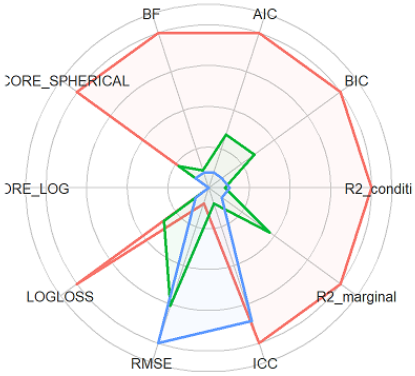


meaning recognition





spoken form recognition



written form recognition

Note. Based on means of 10 normalised statistical criteria. Other than AIC, $R^2_{conditional}$, $R^2_{marginal}$, and ICC, BF, metrics included BIC, RMSE, LOGLOSS, SCORE LOG, SCORE SPHERICAL, and PCP. Larger values indicate better performance, points closer to the centre indicate worse performance.
BIC: Bayesian Information Criterion.
RMSE: root mean squared error for (mixed-effects) models.
LOGLOSS: log loss for models with binary outcome.
SCORE LOG: score of logarithmic proper scoring rule.
SCORE SPHERICAL: score of spherical proper scoring rule.
PCP: percentage of correct prediction.