UNIVERSITY OF TARTU

DEPARTMENT OF ENGLISH STUDIES

EFFECTIVENESS OF STILL IMAGES VS SHORT VIDEO CLIPS AS VISUAL AIDS IN VOCABULARY LEARNING AND RETENTION - A COMPARATIVE STUDY AMONG 8-9TH GRADE EFL STUDENTS

MA thesis

JÜRI ROOTS

SUPERVISOR: *Assoc. Prof.* Jane Klavan

TARTU

2024

ABSTRACT

The present master's thesis suggests that images and videos can be utilised as information organisers and mnemonic aids to enhance the aspects of attention, memorisation and recall in the learning process of previously unknown vocabulary. The aim of this thesis is to compare these two media and evaluate their relative effectiveness as aids in the context of learning new vocabulary. A quasi-experiment in three phases with 50 students from Estonian school was conducted where participants watched slideshows with images and short videoclips with text captions on screen. Their ability to recall previously unknown words after watching the slideshows was documented.

The thesis consists of an introduction, two core chapters and an overview of the quasi-experiment results. The aim and research question are presented in the introduction. Chapter one gives an overview of the literature and addresses the most relevant theoretical considerations related to the thesis. Chapter two focuses on the methodology of images and video clips used, the design of the quasi-experiment with immediate pre-, post- and delayed post-test data analysis, followed by the section where the results are summarised.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Following abbreviations are used in the current thesis:

EFL - English as a foreign language

CEFR - The Common European Framework of Reference for Languages

CALL - Computer assisted language learning

ICT - Information and communication technologies

GLT - Generative learning theory

CLT - Cognitive load theory

MMLT - Multimedia learning theory

DCT - Dual coding theory

MW - Merriam-Webster online dictionary

OED - Oxford English online Dictionary

SLA - Second Language Acquisition

INTRODUCTION

When studying a foreign language, mastering a large amount of vocabulary is one of the most crucial aspects of the language learning process (Milton 2008). As Groot (2000) points out, teachers should pay more attention to developments in the systematic theories concerning vocabulary teaching. This topic is being researched quite extensively with constant progress (Groot 2000:56).

From a teacher's perspective, the effectiveness of vocabulary teaching methods and tools that support learning is of paramount importance. One strategy to aid learners to memorise and recall new vocabulary is by using multimedia content. Mayer (2014) defines multimedia in this context as the combination of at least two different modalities when presenting new material to the learners - for example, visuals (still or animated) + words (audible or written). Although much research has been carried out in this area over the past 50 years, combinations of different modalities can make tracing their direct influence on specific cognitive aspects quite difficult to interpret and compare. In a metastudy of 41 research papers, Zhang and Zou (2021) note that most of the reviewed authors mention the need for further studies on the interrelated character of influencing variables when multimedia usage is concerned. There are studies that concentrate on the relationship between moving images (video) and still images as information organisers and mnemonic tools when learning new vocabulary, but their methodology and conclusions often vary from each other significantly (Ramezanali, Uchihara & Faez 2021). This constitutes a research gap that is still open to further studies.

The current thesis attempts to fill this gap by comparing the effectiveness of multimedia (captioned still images and short video clips without audio) as aids in new vocabulary meaning and form recall. A quasi-experiment (no separate control group) with pre-test, immediate post-test and delayed post-test was carried out. Participants were subjected to a test

environment with high cognitive load (large number of keywords, definitions and visuals in a short period of time) under which learning and remembering new vocabulary is considered difficult (Sweller 2005; Mayer et al 2024).

Short video clips vs more traditional still image content is an important subject of study because most of the participants belong to the generation Z (born between 1997-2012). As the first ‘digital natives' generation, they are constantly exposed to the short video content (15-60 seconds) on their phones. From the standpoint of research, this poses several intriguing questions about the effects of this type of multimedia content on a young person's retention and recall patterns.

This study tries to find an answer to the research question “is there a measurable difference in meaning and form recall of previously unknown vocabulary when comparing captioned still images and video clips without audio seen under the condition of high cognitive load?” As far as I am aware (at the time of writing this thesis) there has been no attempt to replicate similar studies or conduct this kind of comparative study in Estonian schools so far.

CHAPTER 1: LEARNING VOCABULARY WITH MULTIMEDIA

This chapter proceeds to present the theoretical basis for the current thesis in the following order: first the general overview of related approaches to second language acquisition according to Stephen Krashen's theory (Krashen 1982) is outlined in its wider theoretical context.

Second, a theoretical framework of vocabulary knowledge developed by Paul Nation is introduced (Nation 2000). Third, an exposition of most relatable research on cognitive processes follows, Cognitive Multimedia Learning Theory (CMLT) by Richard Mayer being the main focus (Mayer 2024). Fourth, an overview of the experimental research done on the topic most relevant to the current study is presented.

For practical reasons, studies related to specific questions of audio-visual perception, incidental vocabulary learning / acquisition or qualitative studies where student's opinions were surveyed, have been mostly (but not entirely) excluded. Some aspects of selected theories that were considered not fundamentally important to current thesis are omitted or intentionally described in lesser detail to conserve space.

1.1 Learning vs acquiring a language

Krashen's theory of second language acquisition (SLA) is worthy of closer examination because it conveys several key ideas that every teacher should be aware of when trying to teach either with or without the aid of multimedia. There are five interconnected hypotheses in Krashen's (1982) theory that form a comprehensive set of arguments about what constitutes language learning either as L1 or L2:

1) The acquisition-learning hypothesis

2) The natural order hypothesis

3) Monitor hypothesis

4) Input hypothesis

5) Affective filter hypothesis

The acquisition-learning hypothesis (Krashen 1982: 10) distinguishes between two independent processes / phenomena that are generally referred to as language acquisition and language learning. Acquisition refers to the immersive, intuitive and subconscious manner in which linguistic structures get internalised without the person being aware of the fact that they have gained new knowledge while being exposed to comprehensible input that slightly exceeds their current linguistic competence (expressed as i+1) by the help of contextual and other cues (Krashen 1981: 103). This is how children 'pick up' languages (either L1 or L2) - they are not concerned about the form of utterances, it happens naturally as a byproduct of comprehensible input and meaningful interaction (Krashen 1982: 46).

Given that one of the characteristics of the subconscious acquisition process is effortlessness, Mason & Krashen (2020: 6) call for practising teachers to strive to imitate conditions where students can expand their vocabulary knowledge in a way that is engaging, relevant, effortless and natural.

In contrast, learning is primarily associated with conscious effort and activities that revolve around processing either received or produced language content under the constant internal 'monitoring' - this requires processing capacity, which translates directly into cognitive effort and time spent to create an internal feedback loop before anything is internalised or expressed / produced either verbally or in written form (Krashen 1981: 3). Acquisition and learning are the only options available to the adult L2 learner in order to progress their L2 knowledge and their correlation depends largely on the amount of

comprehensible input - error correction, explicit teaching of grammatical rules etc. are largely (but not entirely) ineffective and function only as a Monitor (Krashen 1982: 17).

Criticism of Krashen's acquisition-learning hypothesis often includes arguments about the imprecise nature of the terminology used, since terms like 'consciousness' and 'learning' are already burdened with too many existing connotations across different disciplines - alternative options are suggested (as 'implicit' and 'explicit', for example) (McLaughlin 1990; Ellis 2005).

Others find drawing a sharp divide between learning and acquisition is too disruptive and unwarranted, advocating for a more balanced, continuum-like model (Gregg 1984; Brown 2000; Ellis 2005). There is also a considerable amount of experimental data that lends itself to interpretations which support the usefulness of explicit learning (VanPatten 1990; Green & Hecht 1992; Ellis 2005).

In contrast, Truscott (1998: 122) reports findings which confirm Krashen's hypothesis, claiming very low effectiveness to learning gains when formal grammar instruction is applied. Similarly, while Norris & Ortega (2000: 500) advocate the benefits of explicit instruction, they admit that the effects diminish returns as the time progresses. Agreeing with this sentiment, Jegerski (2021: 318) concludes that the fundamental distinction of acquisition and learning processes in Krashen's hypothesis is one of the most productive and widely accepted ideas in SLA research, albeit under other names. Lichtman and VanPatten propose the updated definition of Krashen's acquisition-learning hypothesis as follows:

> The complex and abstract mental representation of language is mainly built up through implicit learning processes as learners attempt to comprehend messages directed to them in the language. Explicit learning plays a more minor role in the language acquisition process, contributing to metalinguistic knowledge rather than mental representation of language. (Lichtman and VanPatten, 2021:288)

The natural order hypothesis posits that grammatical structures are acquired following a discernible sequence, with specific structures typically mastered before others. Krashen (1982) notes that the emergence of natural order of acquired language is most evident in

situations where test subjects are focusing on the communication itself, not the form (correctness) of it - as soon as they are required to start doing the written grammar-type tests, the order of errors in use of morphemes changes.

According to Krashen (1982:13), the average acquisition order of English morphemes in ESL for L1 and L2 for children and adults is: 1. Progressive '-ing', plural and copula ('to be'), 2. Auxiliary (progressive, as in 'he is going'), article (a, the), 3. Irregular past, 4. Regular past, 3rd singular -s, possessive -s.

Gregg (1984) assesses claims in Krashen's hypothesis critically, noting that projecting conclusions from small amounts of English morphemes into the wider context of the whole SLA without taking into account substantial part of morpheme studies is an inaccurate approach and should be backed by much wider research before making that claim. Similarly, McLaughlin (1987), Gass and Selinker (2008) refer to their studies that challenge the presumed existence of such order. Expanding the previously presented observations, Zafar (2009) highlights experimental results where L2 learners acquire grammatical structures in a totally random succession depending on their L1, the influence of which Krashen does not even register. Despite all these substantial criticisms, by now the existence of the morpheme orders is well established by extensive amounts of research and accepted as such (Long 1990; Hawkins 2019; Lichtman & VanPatten 2021).

In his Monitor hypothesis, Krashen presumes the necessary existence of a cognitive process / system that can be described as a Monitor - an acquired knowledge repository that functions as a corrective filter before any language content is produced in verbal or written form (Krashen 1982: 15-17). As acquisition forms the basis for the cognitive system that initiates the process of language content to be produced, Monitor steps in only after the fact - to edit, to self-correct. In other words, conscious learning as an activity is unthinkable in any form outside of Monitor. At the same time, Krashen also stresses that Monitor enables

language users to access linguistic data which they have not acquired at the moment. Three conditions have to be fulfilled before the Monitor can function, (Krashen 1982: 16) asserts:

a) availability of time in order to process the usage of the right grammatical rules

b) focusing on form (the correctness of the produced language content)

c) familiarity with grammatical rules.

It is worth noting that according to Krashen (1982), the availability of time alone is insufficient - a person might be so preoccupied with monitoring the content of verbal output that they cannot simultaneously self-correct for all necessary grammatical rules. Monitoring is more effective on easier rules - it is nearly impossible to stimulate noticeable use of Monitor outside of proper grammar assessment situations (Krashen 1982: 18).

Questioning the whole Monitor model as an untestable and undeveloped concept regarding experimentally observable content, several researchers have contested its core assumptions (McLaughlin 1987; Long 1990; Ellis 1990; Mitchell & Myles 2004). In defence of the Monitor as a general concept, Ellis (2005: 96) points out that despite its theoretical shortcomings, monitoring as a phenomenon can be used when reinterpreted as a learning strategy. Lichtmann and VanPatten (2021) state it was the Krashen's Monitor model and the arguments it evoked that furthered the discussion about what a proper SLA theory should be in essence, inspiring several new approaches and theories (Lichtmann & VanPatten 2021: 283).

Krashen's (1982) input hypothesis tries to answer the very fundamental question about how language is acquired in principle. He elaborates that if his hypothesis is valid and the acquisition process is functionally superior to learning, teachers should strive to foster acquisition specifically, the issue of knowing how acquisition process works becomes central (Krashen 1982: 20). Input hypothesis is organically tied to the acquisition hypothesis, as the increase in linguistic competence can be achieved only by being exposed to input that

contains comprehensible content with elements that are beyond our current knowledge (the i+1 model). Although it seems contradictory at first, the key to acquiring next level language content through the input directed at us lies in our innate ability to deduct full meaning from messages by taking advantage of extra-linguistic cues (context, general knowledge etc) with the focus on meaning, not form. This is diametrically opposite to the view according to which learning structures first, then constantly repeating them, leads to the fluency in communication. Input hypothesis stipulates that exactly the inverse holds true. We acquire only by focusing on the meaning first, the structure follows naturally through the use (Krashen 1982: 20).

The initial criticism of Krashen's multifaceted input hypothesis revolved around the imprecise articulation of the concept of i+1 (McLaughlin 1987; Ellis 2003; Brown 2007), making it prone to an exceedingly wide array of interpretations. Parallel to theoretical debate, Swain (1985) points to her experimental studies where, in spite of being provided ample input, the students' progress was only marginal. The overall verdict by her is that the process of acquiring foreign language is far more complex than Krashen's model assumes.

Diverging from critical views about Krashen's input theory, Schwartz (1993) and Zobl (1995) report studies that confirm the superiority of an input-based approach. Brown (2007) rejects Krashen's input hypothesis based on purely theoretical considerations, since it relegates learners into a passive role without being actively engaged in the learning process. This contradiction inspired many researchers to develop new theories with a more balanced relationship between input and output (Swain 1985; DeBot 1996; Schmidt 1990; Ellis 2003).

Commenting on this trend, Lichtman and VanPatten describe the newest developments in recent complexity theory as reframing attempts of the word 'input' - for example they offer the term 'affordances' instead as describing potential 'opportunities for learning' which

learners may seize according to their subjective perceived need or interest regarding the input (Lichtman & VanPatten 2021: 297).

Affective filter hypothesis. The idea of affective filter in Krashen's hypothesis is borrowed from Dulay and Burt (1977). Krashen explains that elements like motivation, self-confidence and anxiety all relate directly to acquisition and far less to the learning model in SLA, since the common denominator is communication fluency. The filter hypothesis can be used to explain situations where learners are not progressing to the next level of linguistic competence in spite of being provided with ample comprehensive input, since they lack motivation, are afraid or suffer from anxiety (Krashen 1982: 33).

The affective filter is a variable that teachers need to take into account when planning and designing their lessons - as Krashen (1982) indicates, all of his hypotheses regarding SLA can be concentrated into one sentence: the effective L2 teacher makes input comprehensible while lowering the affective filter proactively (Krashen 1982: 32). Proposed concept of the affective filter is visualised in Figure 1:

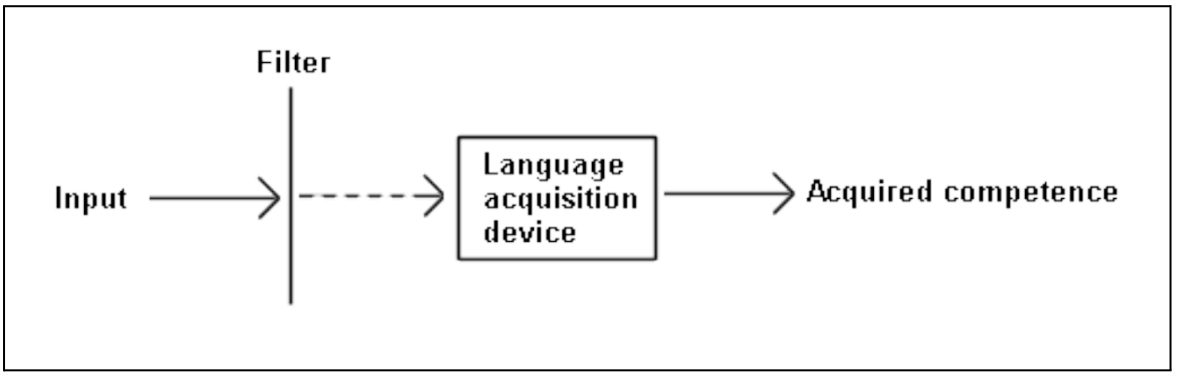


Figure 1. Operation of the affective filter (Krashen 1982:32)

Gass and Selinker (2008) are not quite convinced by these arguments. They claim that the hypothesis fails to account for a proper and verifiable description of the process. Similarly, Brown (2007) argues that according to the Krashen's hypothesis adult L2 learners are never

able to achieve language proficiency similar to native L1 speakers, since the filter hinders comprehensible input, but there are more than enough examples to prove exactly the opposite - adult L2 learners can achieve native-like proficiency. Contrary to these opinions, Chen (Chen 2022: 2154) concludes that in retrospect the ramifications of Krashen's ideas still inspire new research and his affective filter hypothesis can be considered one of the most promising areas of current investigations in SLA. Lichtman & VanPatten (2021: 283) agree that many of Krashen's theories on SLA live on even 40 years since their inception, although under different names and with modified content.

1.2 Knowing a word

Miller (1999) raises the question regarding word knowledge in the simplest possible manner: "What does a person who knows a word know?" (Miller 1999:3). In order to answer that question, a framework for describing 'knowledge' is needed. One of the most widely accepted frameworks uses terms 'vocabulary breadth' and 'vocabulary depth', where the first encompasses the number of words known on at least superficial level and second addresses the contextual understanding and competence in using the word (Anderson & Freebody 1981; Nation 2000; Schmitt 2014). Vermeer (2001) argues that due to the complex nature of the mental lexicon, precise distinction between these two correlated categories is not easy to apply, hence the usefulness of the construct is debatable in some cases. Furthermore, Schmitt (2014) extends the dimension of vocabulary depth into four interconnected subdivisions: active recall (form recall), passive recall (meaning recall), active recognition (form recognition), and passive recognition (meaning recognition).

Based on the research of Webb (2008), Laufer and Goldstein (2004), Schmitt (2014) presents data for a common pattern according to which form recall is significantly more

challenging - this means form recall reflects a bigger vocabulary depth in test results. This, however, depends on how strict are the form recall requirements - according to Webb (2008), if the form recall test accepts some deviation from the 100% correctness, the difference between the two is smaller.

Nation (2000) agrees that vocabulary knowledge is not only knowing the form or meaning of a word, but also having an adequate link between these two in one's memory available for retrieval (a form-meaning connection). Approaching this from a functional perspective, Nation indicates that the vocabulary knowledge can also be described on a receptive-productive scale of vocabulary skills. Furthermore, he describes receptive skills as an aspect of comprehensible input, whether through listening, reading or attempting to understand received information, while productive skills are connected to the speaking, writing and other outbound directed communication.

While this is a useful distinction, Nation accepts the potentially misleading character of these terms, since in some instances the conceptual difference is near impossible to draw - for example reading and listening contain both, when inspected more closely (Nation 2000:37). Other alternatives (as 'passive' and 'active' skills), proposed and used by some scholars (Laufer 1998; Meara 1990; Corson 1995) are no less confusing, because aforementioned listening and reading as activities can also be considered anything but passive (Nation 2000:38). Nevertheless, Meara (1990) develops the idea of passive vs active vocabulary into a rather coherent argument from the standpoint of how words are activated in the lexical network in one's memory. Active vocabulary presumes a large number of different associations between known words, leading to easier 'internal' activation, while passive vocabulary can only be activated by external stimuli, since it lacks any other connections between words except their written or audible form (Meara 1990).

Nation's receptive-productive vocabulary skills scale (Nation 2000) is largely accepted as a good and concise model to describe the aspects of word knowledge that fall into three main categories: form, meaning and use. Each aspect is then divided into three subdivisions in order to address further nuances of receptive or productive skills. This model is visualised here in Figure 2:

Form	Spoken	R P	What does the word sound like? How is the word pronounced?
	Written	R P	What does the word look like? How is the word written and spelled?
	Word parts	R P	What parts are recognizable in this word? What word parts are needed to express the meaning?
Meaning	Form and meaning	R P	What meaning does this word form signal? What word form can be used to express this meaning?
	Concept and referents	R P	What is included in the concept? What items can the concept refer to?
	Associations	R P	What other words does this make us think of? What other words could we use instead of this one?
Use	Grammatical functions	R P	In what patterns does the word occur? In what patterns must we use this word?
	Collocations	R P	What words or types of words occur with this one? What words or types of words must we use with this one?
	Constraints on use (register, frequency…)	R P	Where, when, and how often would we expect to meet this word? Where, when, and how often can we use this word?

Figure 2. Division of vocabulary knowledge types and categories. In column 3, R = receptive knowledge, P = productive knowledge. Table and contents adapted from Nation (2000:41).

1.3 Theories Regarding Use of Multimedia in Language Learning

The research that led to current theories on visuals and multimedia use in educational environments was concerned with the several experimentally confirmed observations during the last 50 years of research in cognitive psychology and applied linguistics. The most recognisable of these was the finding that learners formed stronger memory links between word-object pairs than word-word pairs (Wimer and Lambert 1959). It became known generally as the 'picture superiority effect' and was confirmed by numerous following studies in many emerging disciplines that researched human vision, memory and cognition (McBride & Dosher 2002; Defetyer et al 2009; Curran & Doyle 2011). Subsequent experimental data indicated that visual-pictorial and verbal-auditory stimuli might be processed and coded not in one, but in two separate, dedicated channels which both strengthened memory trace with representational links in the brain - this formed the basis of dual coding theory (DCT) (Paivio 1971). Three most relevant and interrelated theoretical frameworks regarding multimedia and instructional design emerged: dual coding theory (DCT), developed by Allan Paivio (2010), cognitive load theory (CLT) by John Sweller (2019) and cognitive theory of multimedia learning (CTML) by Richard E. Mayer (2024).

1.4 Dual coding theory

DCT was developed as a general cognition theory by Paivio (1971) and constantly revised over the following years. Although originally meant to address the questions of cognitive psychology in regard to verbal and nonverbal imagery in memory, it became remarkably influential in the fields of applied linguistics and language learning. Asserting that all human cognition is dependent on two functionally different, but interdependent systems,

Paivio states that there is a system for nonverbal objects ('objects' in the most generalised sense, including events) and a system related to language-related representations (Paivio 2010:207). In other words, there are two channels for processing two types of input - one channel deals with language related content and the other channel with non-linguistic information. When processing information, the human brain can create representational links to data simultaneously in both channels, strengthening associations and creating memory traces that translate into better information recall. In each channel this processing is done by the help of basic information units called 'logogens' (in the verbal system) and 'imagens' (in the nonverbal system) respectively (Paivio 2010) as seen in Figure 3 below:

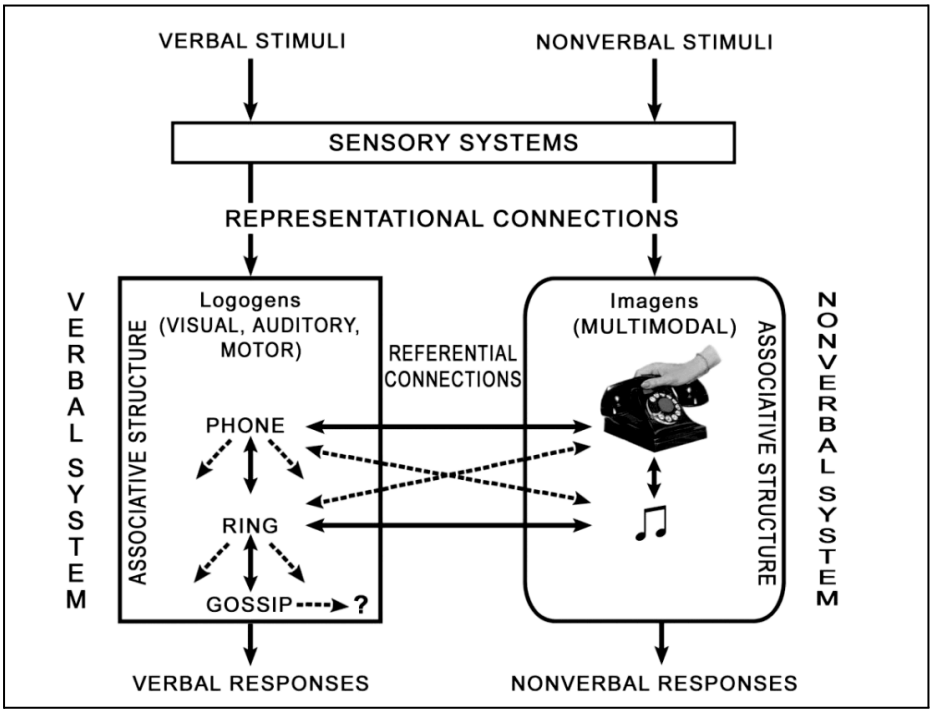


Figure 3. Visual representation of multimodal dual coding model (Paivio 2010:209)

The implication of these findings serve as a useful starting point when drawing connections to visual content usage in learning context. That is what Mayer does, when he builds his theory of cognitive multimedia learning (CTML) (Mayer, 2024).

1.5 Cognitive Theory of Multimedia Learning

During its development over four decades, Mayer's Cognitive Theory of Multimedia Learning (CTML) has gone through ten name changes and the author has shifted the main focus of the theory several times as new experimental evidence has emerged from relevant fields of research (Mayer 2024). In CTML, Mayer (2014) tries to conceptualise processes and conditions under which meaningful learning can occur while aided by multimedia instructional messages. He develops the theory on three basic assumptions under which the brain processes all incoming information. These are:

1) Dual channel assumption (describes twofold coding of information in memory)
2) Limited capacity assumption (describes limits of cognitive resources)
3) Active processing assumption (describes how information is processed and stored)

Mayer (2014) builds his dual channel assumption on Paivio's dual coding theory (Paivio 2010) and Baddeley's model of working memory (Baddeley 2007), it is theoretically assumed that information is processed in two separate, interconnected channels (auditory-verbal and visual-pictorial). Mayer's limited capacity assumption (2014) borrows its main concepts from Sweller's cognitive load theory (Sweller, 1994; Sweller et al 2011), this assumption states that the working memory and two channels for information processing are limited in their capacity. Furthermore, they are easily overloaded. Only a part of any given visual or verbal input will be registered for processing due to the limitations of cognitive resources available (Mayer 2014). As Mayer elaborates (2024:18), three types of processes occur when attempts at understanding poorly communicated verbal or visual instructions are encountered by the learner (extraneous processing), mental representations of complex material are created in

working memory (essential processing), or when the learner tries to organise incoming information into coherent and retrievable collections of data (generative processing) - the latter being greatly dependent on the motivation to acquire new information by the learner. The cognitive processes 'compete' and override access to available resources - when one process has access to more resources, it automatically means that the other processes have less (Mayer 2024:18).

From the perspective of the learning environment, these three demands on limited cognitive resources have to be addressed by appropriate instructional strategies if the goal of successful learning outcome is considered. It can be done by a) minimising extraneous processing, b) controlling essential processing and c) encouraging generative processing (Mayer 2024:18). And lastly, active processing assumption states that information in long term memory has to be accessed, retrieved and processed in the working memory, all in order to restructure and combine new and existing data into a coherent knowledge that is stored in the long-term memory again. The whole process is illustrated in Figure 4:

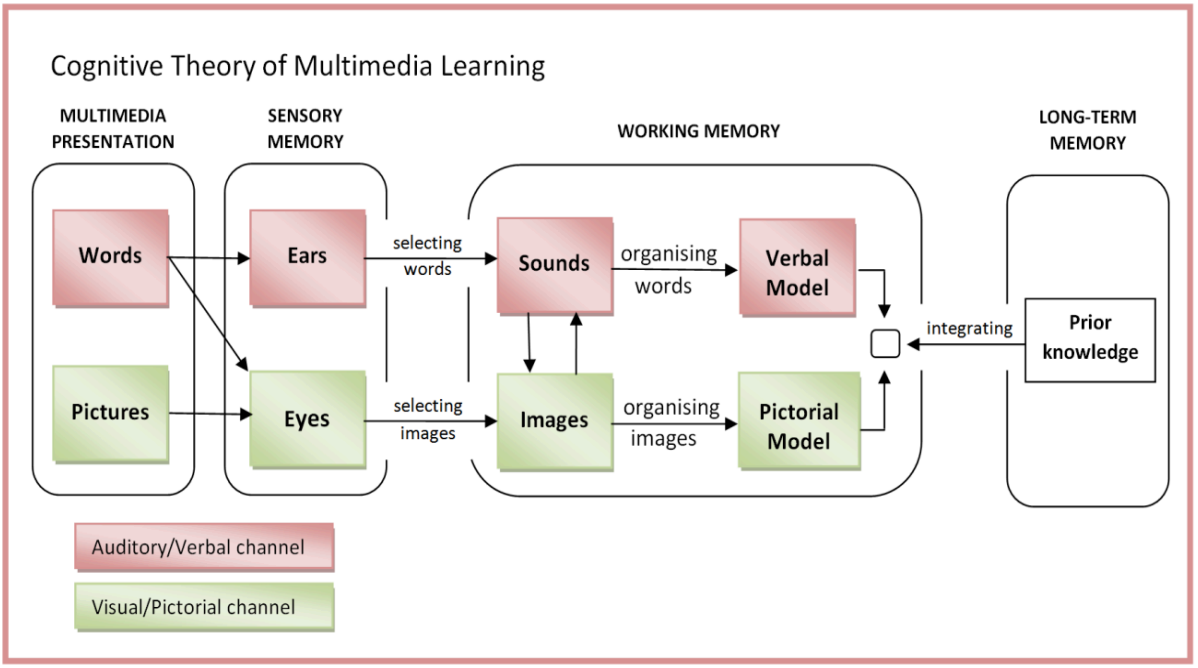


Figure 4. Graphical illustration of CTML elements and processes (Mayer 2014:52).

Plass and Jones (2005) propose a variation on the same theme, integrating known elements from previous SLA studies with Mayer's (2014: 52) work as shown in Figure 5:

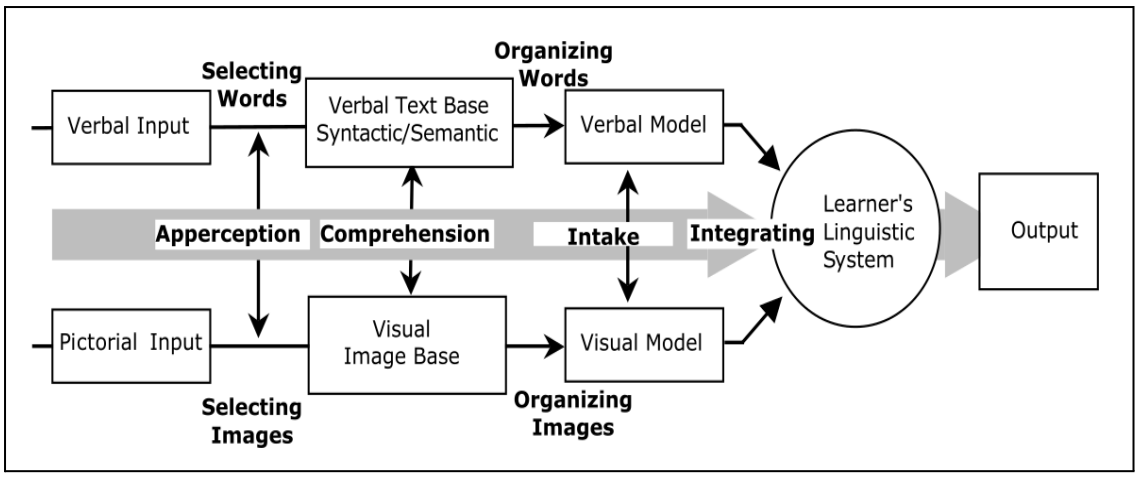


Figure 5. Integrated model of second-language acquisition with multimedia (Plass & Jones 2005: 471).

It is worth noting the crucial element of 'apperception' in Plass and Jones' (2005) integrated approach as seen in Figure 5. The idea of 'apperception' or 'noticing' is borrowed here from the work of Chapelle (1998) who points out that not everything can be considered input, since only those parts that get noticed ('appreceived'), have the potential to become input in the proper sense (Chapelle 1998:22-23). This nuance is important when all previously mentioned SLA-related cognitive theories are reviewed in the light of multimedia theories that follow.

1.6 Cognitive load theory

Sweller developed his cognitive load theory (CLT) with the premise that human working memory has "/…/ no logical central executive available when required to organise novel information" in contrast with long term memory (Sweller 2005:19). Sweller contends that since all new incoming information is disorganised and without meaningful structure, extra

processing is required to organise it or test it out for effectiveness, this requires a considerable amount of cognitive resources (Sweller 2005). It follows that due to limited cognitive resources available, excessive amounts of simultaneous stimuli hamper understanding of the learned material and undermine our ability to create meaningful connections between pieces of information, therefore negating the possibility of later retrieval (Sweller 1994). Agreeing with this, Mayer et al argue that when functional limitations of the human mind are ignored, multimedia content quickly becomes a distractor rather than an asset (Mayer, Heiser & Lonn 2001).

Sydorenko (2010) provides diverging ideas to these claims based on her experiments - she found that when presented with a choice, most students prefer to have access to as many modalities as possible, since they want to feel in control of the learning process and maintain referential access to different cues simultaneously (Sydorenko 2010:62-63). Her test subjects were exposed to three modalities at once in L2 and performed better than the group with input in only two modalities, making purported 'redundancy' into an aid rather than a distraction (Sydorenko 2010:62-63). This does not agree with the CLT redundancy principle as a cognitive load that Sweller (2005) proposes. According to Sweller, it is not beneficial to present the same material through two different modalities synchronously in native language, since it does not support successful information processing for L1 learners (Sweller 2005). Sydorenko (2010) also concludes that L1 and L2 process information they are exposed to very differently (depending whether it is L1 or L2 input).

While both Sweller (2005) and Sydorenko (2010) report on valid, experimentally observed phenomena, they seem to miss a reference to two different theoretical approaches of redundancy categorisation. Albers et al (2023) report on experimental study where they tested participants in an environment where information presented was designed for two kinds of redundancies:

a) contentual redundancy (overlap of content from two sources repeated, both are separately comprehensible)

b) modal redundancy (poorly structured simultaneous multimedia content in different modalities)

Contentual redundancy was conducive to better learning results and seemed to mitigate negative effects of cognitive load, while modal redundancy affected results conversely (Albers 2023: 339,341). As Mayer et al have theorised and experimentally shown, a potent combination to evoke modal redundancy is video + narration + text (of the same narration). As the text and narration are competing for attention and processing, they produce a 'split attention effect' - the learner has to spend limited cognitive resources to mentally compare them (Mayer et al 2001: 191).

1.7 Subtitles and Captions

Mayer provides evidence from empirical studies that images or videos with text help to understand content much more effectively than visuals or text alone; this is also well documented as an instructional design principle in CMLT (Mayer 2021 as quoted in 2024: 20). Visuals with text were noted to help memorisation accuracy better than words alone (Curran & Doyle 2011: 1259).

According to Reynolds et al (2022: 15), textual addition to video content can be described as subtitling or captioning. Distinct features of subtitling are related to presence of sound and synchronous translation of audible dialogue. Captioning is deployed as a textual representation of other sounds in addition to dialogue in cases the receiver of the video content is, for example, hearing impaired and benefits from additional cues in textual form (Reynolds et al 2022: 15). The relationship between L1 and L2 both subtitles and captions,

including their main types in SLA research can be divided into several varieties:

1) intralingual subtitles (L1 audio to L1 on screen text), can be keywords only for research purposes. Captions of this variety follow the same logic with addition of indication of different speakers, additional sound effects etc being marked down in textual form;

2) interlingual subtitles (L2 to L1 on screen text), can be keywords only for research purposes. Captions of this variety follow the same logic with addition of indication of different speakers, additional sound effects etc being marked down in textual form;

3) bilingual subtitles (audio is displayed as simultaneous translation of two different languages on screen texts simultaneously, while one of these translations is often in the same language as the audio track - for example, L2 audio to L1 and L2 texts)

4) glossed keyword subtitles / captions (intra- and interlingual) add a gloss (a short written definition) to the displayed keywords.

5) full captions with highlighted keywords (intra- and interlingual) add a highlighting of particular words (for research purposes) (Reynolds et al 2022:2-3).

An experimental study by Sydorenko (2010) regarding videos with captions states that captions were the most prominent element participants paid attention to (video and audio gaining attention after that), suggesting that in some instances captions could be considered superior to audio layer in multimedia (Sydorenko, 2010: 63). Reynolds et al agree with this conclusion, recommending teachers to use intralingual captions if the goal is to learn vocabulary (Reynolds et al 2022:16).

1.8 Research results where video + text was superior

Relevant to the question of moving vs still images, cognitive load and working vs. long term memory, a well-documented small scale study by Matthews et. al (2007) showed that

very short video clips with fluid movement seemed to be superior to still images when recognition memory was concerned. The superiority persisted through extended recall intervals (Matthews, Benjamin & Osborne, 2007: 991).

A number of studies have been done in Tartu University regarding videos as vocabulary teaching aids. Sallert (2016) adopted well-known tv-series as the base material and cut individual seasons into 10-minute coherent video clips and showed these during a normal 45-minute lesson to students. According to her, the results showed that when used as a tool to facilitate learning of unfamiliar vocabulary, this vocabulary learning method was relatively efficient - 70% of participants learned over half of the unfamiliar words they encountered (Sallert 2016:34). Video content also encompasses computer games, which provide strong motivational clues even for those who keep observing the gaming activity from aside and are not manually / directly involved. Mehine (2021: 10) draws a conclusion that there seems to be a positive effect on intrinsic motivation to learn new vocabulary when it is encountered while playing a videogame - it is necessary to understand the storyline and / or reach the winning condition. Lin and Tseng (2012: 352) deducted from their experiments that video + text was very effective for learning difficult vocabulary

In general terms these are important clues in favour of the video content, but the generalisation cannot be extended too far - as Sweller et al (2011) have pointed out, when video is combined with added modalities in the form of text and simultaneous narration, different framework needs to be applied in order to understand the mechanics of attention and memorisation - in this case the hypotheses of CLT and CMLT are helpful.

1.9 Studies where image + text was more effective

Multimedia content in its various forms is notorious for providing researchers with the most diverse results, therefore making comparative data analysis extremely challenging. There are numerous studies where diverse data in favour of still images instead of videos was considered more effective from the vocabulary learning standpoint or information recall in general. A number of other research results implied that visuals hindered the recognition of word forms or made no remarkable difference (Chun & Plass 1996; Boers et al 2017; Peters 2019). Confirming these findings, Kim (2019) concluded in a study regarding pictorial content and the question of image + text combination, that images combined with text outperformed videos with texts especially when the texts consisted of single words. Kim reports controversial findings where for single word memorisation image + text seemed to outperform videos with captions, but were not as effective when phrasal verb recall was expected (Kim 2019: 19).

1.10 Theoretical implications

To sum it up: the progress and research in the cross-disciplinary studies (applied linguistics, cognitive psychology, neurosciences, neurocomputing etc) during the last 50 years has given us a variety of useful theories and models to conceptualise the relationship between learning languages and multimedia use in classroom context (as DCT and CTML). Yet with the growth of experimental data available, the question of exact variables influencing the learning outcomes is still a matter of much debate. Interpreting the conflicting results is not easy, since there are several competing theoretical approaches one has to be able to navigate before drawing any substantiated conclusions. Even if we have seemingly moved on from the

earlier, foundational models that initiated a paradigm shift in linguistics (nativist theories, Monitor model etc), they are still influential and keep reminding us of the fundamental questions embedded in human cognition, pedagogy and linguistics in general. These questions are often simply rephrased, but not easily answered in full. Experimental data we accumulate is often fragmented and details on methodologies used often leave several key questions hanging in the air. Learning and teaching a foreign language is utmostly complex and multifaceted endeavour, where many factors influence the outcome simultaneously. It means all research purporting to reveal some validated cause-effect truths in this domain needs to be carried out with utmost sensibility to several competing theories at once, while at the same time being capable of extrapolating generalisable conclusions in practical, applicable terms.

CHAPTER 2: QUASI-EXPERIMENT: IMAGES AND VIDEO CLIPS

2.1 Research Aim and the Research Question

The aim of this research was to compare the effectiveness of still images and short video clips with captions as mnemonic aids in learning new vocabulary under conditions of high cognitive load. The efficacy of both multimedia types was measured as percentage difference of the form and meaning recall tests results between immediate post-test and delayed post-test. The quasi-experiment without a separate control group focused on the visual component only - still images and video clips without auditory component (sound). The research question was formulated as "is there a measurable difference in meaning and form recall of previously unknown vocabulary when comparing captioned still images and video clips without audio seen under the condition of high cognitive load?"

2.2 Sampling and Participants

The participants (n=50) were 13-14 year old secondary school students (8-9th grades) in Estonia. Their approximate CEFR proficiency level was B1-B1+. It was assessed by their constant previous monitored performance during a two and a half year period in school by their teacher (the author of the current thesis) while studying from the corresponding study material. Most participants also took a trial state examination of English for ninth grade the same year the study was conducted. The results showed the majority of them had achieved at

least level B1 in their usage of English. Sampling was non-random (school classes), consisting of the students that the author was teaching on a regular basis for several years.

2.3 Research Design

Sato and Loewen (2019) remark that researchers should strive towards greater ecological validity in their experiments. In line with this, current research was conducted in a school during real lessons with participants (students) being in their natural environment (English classes). Several authors have mentioned the difficulties that arise when attempting to maintain control of the test environment and simultaneously reduce the level of artificiality that is inherently introduced by experiments in highly controlled settings (Spada 2005; Leung & Valdés 2019).

To counter this, current research attempted to keep a balance between the amount of control over experiment settings vs. ecological validity by reducing the possible influencing factors to a minimum. For example, audio content was removed from the slideshows. Such a decision in research design was based on the well documented phenomena called split attention, which is listed as a separate principle in Mayer's CMLT (Mayer, Heiser & Lonn, 2001). This has also been tested as a modal redundancy problem by Albers et al (2023).

Mayer's CMLT predicts an emergence of a learning environment conducive to cognitive overload when learners are presented with rapidly changing information (the pace of which they cannot control) in more than one modality, forcing them to split their attention between written text, visuals and audio (Mayer, 2014). Since a high number of dynamically changing slides with written text, still images and videos present in the current study already constituted two separate modalities according to dual coding theory (Paivio 2010), the

conditions for high cognitive load (Sweller 2019) were already met and no further extension of this factor was deemed necessary.

Due to practical and ethical reasons the research was conducted as a quasi-experiment (no control group) with five classes of students during a two-day period. Different classes were considered one group (n=50) in the data analysis due to the fact that all conditions for the participants were nearly identical throughout the experiment (room, presentation technique, pre- and post-tests, dates).

The null hypothesis was 'no significant difference occurs between still images vs. video as a tool to help memorise, recall and produce previously unknown vocabulary under condition of high cognitive load in written form'. Since the same group of participants was exposed to more than one type of treatment conditions (video with captions and still images with captions) the test type fell under the repeated measures t-test design category (also known as within subjects design, as all participants took part in every test condition).

The test consisted of three phases: 1) written pretest (meaning recall), 2) immediate post-test after slideshow with captions (word meaning recall), 3) 24-hour delayed post-test after slideshow without captions (word form recall).

Before the experiment the school principal was informed and asked for a signature on a dedicated letter of consent (see Appendix 1). All students were informed that they can choose to participate or decline from participation in the experiment in due time. Each participant was given a written consent form to be signed by the student and their adult representative. In case the students forgot to return the written consent form with their adult representatives' signature, the latter were asked to reply to the electronic version of consent question with a simple "I agree / I disagree" message (see Appendix 2).

Students who did not qualify as participants with expressly given consent were excluded from the experiment (2 students). The first phase of the experiment consisted of an

immediate pre-test. Twenty four words without definitions were printed on a sheet of paper in a column (see Appendix 3). Detailed description of the criteria for selecting the words is provided in the following sub-chapter. Participants were asked to write down the definitions for the words they already knew, their answers were collected and analysed. The words known by participants (they were able to provide logical and coherent definitions in their own words) were marked for removal from the final data analysis on a granular level (per participant).

During the second phase of the experiment a slideshow of images and videos with the captioned definitions of the keywords were projected on a screen (Appendix 4). The choice of a slideshow as a medium seemed reasonable based on the aspect of ecological validity - projecting multimedia content on the screen during the language lessons is seemingly universal and widely practised. Nevertheless, when a metastudy of 30 experiments on multimedia usage was consulted (Zhang & Zou 2021), it became clear that only one study out of 30 used slideshow as a primary tool to mediate content.

Captions were shown under the visuals as black boxes with white text. The keywords (vocabulary items tested for recall) were incorporated in the sentences that were explaining the meaning of the said words. Vocabulary items under study were written in capital letters and short explanations / example sentences in lower case.

In order to increase the variety and avoid possible bias, the words presented in the slideshow were representative of three categories - nouns, verbs and adjectives (Appendix 5). Each category included 8 words, of which half (4 words) were accompanied by a still image and the second half (4 words) by a video clip. In total, the slideshow consisted of 24 visuals: 12 images and 12 video clips. The order of images and videos shown was randomised (still image illustrating a noun was followed mostly by a video depicting a noun, then a still image illustrating an adjective was shown etc). All words were presented on screen as one

continuous Google Slides slideshow. Each item was projected for 7-13 seconds (videos as source material from pexels.com were not of equal length).

An immediate post-test was administered after the slideshow. It was identical to the pre-test, consisting of one sheet of paper with 24 words in a table column. All answer sheets were collected and digitised.

The final phase took place 24 hours after the previous test. The same slideshow was projected on the screen but this time no captions were provided with the visuals. Immediately after the slideshow a final test was administered in a digital format - students used their phones in order to access flexiquiz.com, where they were presented with a 'fill in the blanks' task (form recall) (see Appendix 6). The task consisted of 24 sentences (definitions from the slideshow) with keywords removed and alternative synonyms offered in brackets after the blank / gap. All results were exported into a spreadsheet and analysed. The whole research design is described in Figure 7:

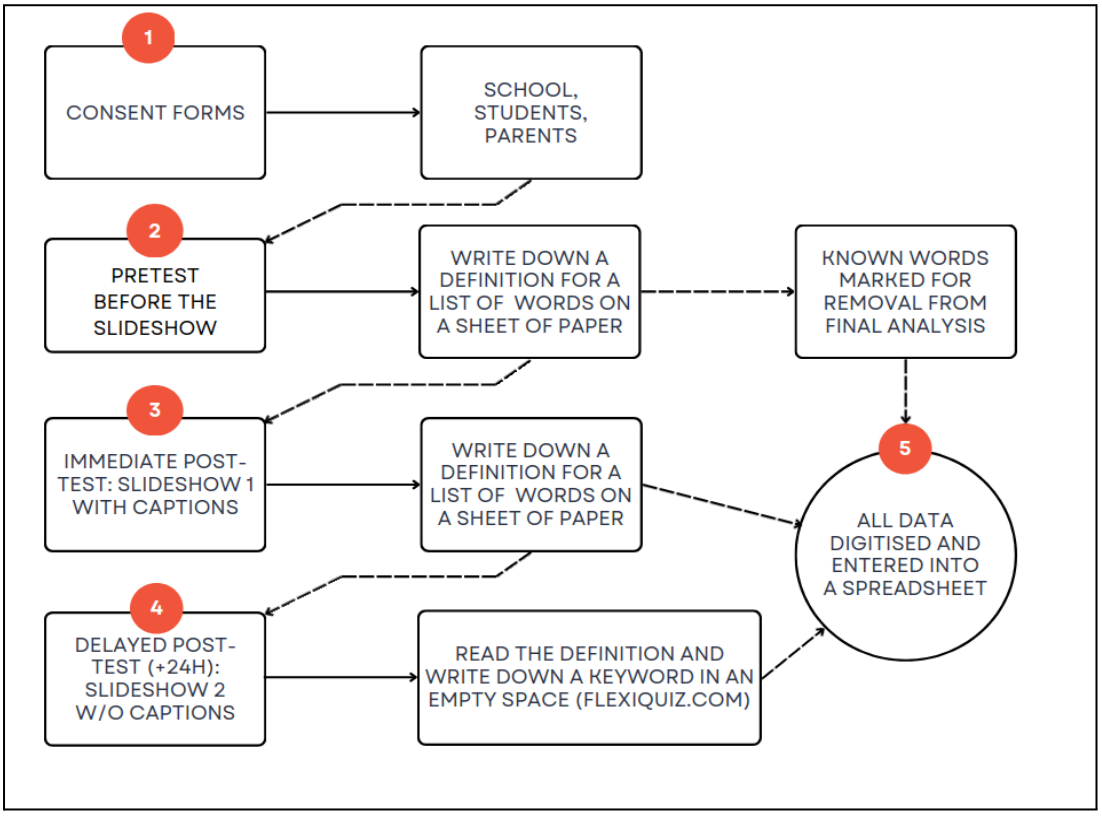


Figure 7. Research design and data collection tools used. See Appendixes 3 and 6 for recall test forms used. For the visual design of slideshow slides see Appendix 4.

2.4 Data Collection and Analysis

Data was collected with immediate pre-test and post-test answer sheets, then digitised by entering the collected values into a spreadsheet (Google Sheets) where final results were calculated. All identifiable data was anonymised by replacing the names of participants with numerical codes. Based on the pre-test data, the words that were known by the participants were marked separately in a spreadsheet and removed from the final post-test word count for these specific participants and submitted to comparative analysis (if the participant knew 4 out of 24 words in the pretest, then this particular student scored 20 words as unknown and 4 words were removed from calculations regarding them specifically).

2.5 Materials

The vocabulary items were semi-randomly chosen by using https://relatedwords.org or thesaurus.com (choosing from synonyms). One aspect influencing the word choice was availability of visual material on pexels.com. If there was no correlating visual material available or it was not deemed very descriptive, another word was chosen instead.

All words were checked for occurrence statistics on Oxford English Dictionary online (oed.com). Attempt was made to include words with no more than 3 occurrences per million words in modern English, desirably significantly less (0.05-1) according to OED. There were few vocabulary items that exceeded this indicator (2 words) or for which the occurrence data was not available (2 words). Word definitions were all created by the author and checked for close resemblance in meaning with MW and dictionary.com. Visual material was acquired from pexels.com. One image (referring to “singlet”) was generated with online AI tools at dezgo.com (prompt: “one wrestler wearing singlet hands raised”). Captions were created in Google Slides and consisted of a black box with a white text. The keyword was written with

all letters capitalised. Pre-test and immediate post-test questionnaires were printed on A4 papers and given to participants immediately before and after the intervention (slideshow). The delayed post-test was executed digitally in the flexiquiz.com test environment. All results were anonymised and exported to Google Sheets.

2.6 Results

2.6.1 Pretest results

Pretest results showed that 13 words out 24 were known by participants in total. The criteria for "known" status consisted of meaning recall test results where participants were able to give coherent short definitions or proper associations with examples of usage of the words. All words known were removed from the later test results on a granular level (per each participant separately). The words detected during pretest are shown in Table 1:

Word	Instances (n= 50)	% of total (24)
Apron	18	75%
Rodent	13	54%
Crevice	12	50%
Livid	12	50%
To vault	8	33%
Singlet	2	8%
To avert	2	8%
Despondent	1	4%
Spry	1	4%
To dart	1	4%
Aglet	1	4%
Dilapidated	1	4%
To hurl	1	4%

Table 1. Words known by number of individual instances and percentage of the whole set.

2.6.2 Immediate post-test results

Immediate post-test raw cumulative data showed slight superiority of the still images vs videos with captions. Raw cumulative average calculation for still images (meaning recall) was 5.34 words out of 24 and for videos 4.5 words out of 24. Comparison of cumulative sums and averages, difference in percentages is shown in Table 2:

	Images with captions	**Videos with captions**
Cumulative sum	267	225
Average (/50)	5.34	4.5
Recall rate (average)	22.25%	18.75%
	Difference: 17.07%	

Table 2. Comparison of immediate post test results for written meaning recall

2.6.3 Delayed post-test results

Delayed post-test was carried out 24 hours later and measured form recall of the words with visual and verbal cues separated. The slideshow was shown without captions. After that a digital test was administered (flexiquiz.com). Participants were tested for form recognition with definitions being given without the keywords (blanks in the text). The cues in the form of possible synonyms to keywords were provided. Results can be seen in Table 3:

	Images with captions	Videos with captions
Cumulative sum	123	113
Average (/50)	2.46	2.26
Recall rate (average)	10.25%	9.41%
	Difference: 8.47%	

Table 3. Comparison of delayed post-test results for written form recall

2.6.4 Immediate post-test and delayed post test results comparison

When comparing immediate post-test vs delayed post-test raw cumulative data, the difference was converted into a ratio expressed as a decimal number. If two out of four initially encountered words (meaning recall during immediate post-test) were recalled in their written form during the delayed post-test, it translated accordingly into 50% or 0.5 value. If one word out of four was remembered, that translated into 25% or 0.25 value on the hypothetical 24-hour forgetting curve. When summed, the results indicated the effectiveness of one type of media versus the other as a mnemonic aid as seen in Table 4.

	Recall: images	**Recall: videos**
Recall ratio between two tests	22.14	24.36 Difference: 8.47%

Table 4. Comparison of immediate post-test and delayed post-test results by media type.

DISCUSSION

When comparing test results, a notable difference (17.07%) can be observed between image and video content recall in the initial, immediate post-test. Results seem to affirm the superiority of image content effect on recall when tested immediately after first exposure as has been also observed by Kim (2019).

That initial difference declined to 8.47% when measured 24 hours later in the delayed post-test, which hints at the possibility of image content memory-trace potential being weaker than video content's ability to retain its trace 'value' - more image-related vocabulary is forgotten than those words that were associated with video content. There might be another possible explanation for this: according to Albers (2023), the content redundancy can have positive effects under certain conditions when exposed to multimedia. Since participants were exposed to the same textual content repeatedly (including the delayed post-test itself) and the factor for modality redundancy was excluded from multimedia presentation (visuals + text - audio), a possible effect of positive content redundancy cannot be ruled out.

In addition, a question of emotional valence as an affective component in regard to videos and vocabulary (Arriagada-Mödinger 2022) must be taken into account. It could be assumed that a dynamic content with vivid depictions of strong emotions or unusual activity (facial expressions, activities depicting near-extreme sports or situations etc in some clips) has cultural / contextual connotations on the positive / negative continuum that evoke stronger emotions than still images with 'boring' objects in the minds of teenagers. Although this topic is underexplored in the current thesis for practical and ethical reasons (the author cannot claim any substantial knowledge in the fields of neuroscience or cognitive psychology), it needs to be pointed out nevertheless.

A small difference in immediate and delayed post-test results seems to partially confirm the findings of similar studies that report video and text combination being slightly more effective for long-term memory recall in delayed post-test results. One such example would be a study by Kim (2019), although in that case similar results were reported while testing for recall of phrasal verbs and not single words. Al-Seghayer's (2001) study comes closest to the current thesis' attempt at comparing videos vs images with the conclusion that videos+text is more effective. Unfortunately there are a number of key differences with both studies mentioned (cognitive load, temporal aspect, text type, monitor vs. projector screen etc), so comparisons have to be generalised to quite a large extent - this diminishes the value of said comparisons and makes extracting meaningful conclusions difficult.

Video content (short video clips) might be considered a slightly more effective mnemonic device for learning new vocabulary even under high cognitive load, provided that the visual content is accompanied by the on-screen captions that include prominently perceivable keywords. In order to measure the statistical significance of the collected data, a T-Test calculator for 2 dependent means at https://www.socscistatistics.com (Stangroom 2018) was used. Treatment 1 field was filled with data from image-related results and treatment 2 field with video-related results (immediate post-test and delayed post-test difference values). Significance level was set to 0.5 and hypothesis type as "one-tailed". First, the calculations for difference scores are described:

Mean: 0.04

$\mu = 0$

$S^2 = SS/df = 12.22/(50-1) = 0.25$

$S^2_M = S^2/N = 0.25/50 = 0$

$S_M = \sqrt{S^2_M} = \sqrt{0} = 0.07$

T-value calculation: $t = (M - \mu) / S_M = (0.04 - 0) / 0.07 = 0.63$ Result: The value of t was 0.628624. The value of p was .26626. The result was not significant at $p < .05$.

Some limitations of the current study need to be addressed. Although an attempt to find a balance between ecological validity and experimental control in this study was central, few aspects of it are open to debate. As with any small-scale, quasi-experimental and non-randomised research, proceeding with generalisation of these results outside of reported parameters is not advisable. To begin with, the delayed post-test did not compare the performance of participants in the same way as the immediate post-test did. Although it was done in order to vary the types of cues given and measure performance of participants from more than one angle (meaning recall vs form recall), these differences between two tests were more likely than not able to influence the final outcome when comparing the results. Furthermore, there is some research data that calls for caution due to possible test-modality congruency when participants who are provided with on-screen captions are asked to answer in written form only. When participants are presented with textual content (like captions) and tested later in a written form for recall, they tend to score higher than those who have been exposed to only video + audio content. The reverse also applies (Mohd Jelani & Boers 2018; Sydorenko 2010).

Future research on this topic is highly recommended. In order to increase our understanding of the effects of visual content on learning vocabulary, new studies should incorporate larger sample sizes, better filters for words and video content. In addition, expansion of the research horizon by measuring emotional (hedonic) valence (the level of ‘pleasantness’ or ‘unpleasantness’ associated with specific words, affecting memorisation of vocabulary) (Arriagada-Mödinger 2022) might illuminate some underexplored dependencies between the visual content and affective influence on memorised vocabulary.

CONCLUSION

The current thesis concurs with many of the already established theoretical assumptions about multimedia usage as a useful tool to aid learners in learning new vocabulary. This research confirms that visual aids (still images and videos) with L2 captions and clearly distinguishable keywords help to improve the written form and meaning recall among students of 7-9th grade students in Estonia.

A nominal difference between the effects of still images and video clips effect on recall was observed when simple, cumulative datasets were compared. At first, images with captions outperformed video clips in immediate post-test results of meaning recall by 17%. This could indicate that still images are slightly better visual aids from the perspective of a shorter timeframe - they are easier to process and require potentially less cognitive resources. However, when more permanent memorisation goals are considered, short video clips with captions might be considered more effective aids regarding retrieval of word forms from long-term memory due to their possible affective influence: in the delayed post-test, videos outperformed still images by 10%.

Although results of this study are not conclusive, short video clips with concise L2 definitions might serve a useful role as mnemonic aids to teach new vocabulary in an EFL class.

REFERENCES:

Anderson, Richard C., & Freebody, Peter. 1981. Vocabulary knowledge. In Guthrie, J.T. (Ed.), Comprehension and teaching: Research reviews (pp. 77–117). Newark, DE: International Reading Association.

Albers, Fabian; Trypke, Melanie; Stebner, Ferdinand; Wirth, Joachim; Plass, Jan L. 2023. Different types of redundancy and their effect on learning and cognitive load. *British Journal of Educational Psychology*, 93:2, 339–352. https://doi.org/10.1111/bjep.12592

Al-Seghayer, Khalid. 2001. The Effect of Multimedia Annotation Modes on L2 Vocabulary Acquisition: A Comparative Study. *Language Learning & Technology* 5 (1): 202–232.

Arriagada-Mödinger, Francia; Ferreira, Roberto A. 2022. The Effect of Emotional Valence on Auditory Word Recognition Memory in English as a Foreign Language. *Journal of Psycholinguistic Research* 51: 309–322. https://doi.org/10.1007/s10936-022-09841-3

Baddeley, Alan David. 2007. Working memory, thought, and action. Oxford, England: Oxford University Press.

Boers, Frank; Warren, Paul; He, Lin; Deconinck, Julie. 2017. Does Adding Pictures to Glosses Enhance Vocabulary Uptake from Reading? *System 66*, 113–129.

Brown, Douglas H. 2000. Principles of language learning and teaching (4th ed.). New York: Addison Wesley Longman.

Brown, Douglas H. 2007. Principles of language learning and teaching (5th ed.). Addison Wesley Longman

Chen, Yuyang. 2022. A review of research on Krashen's SLA theory based on WOS database (1974-2021). *Creative Education*, 13:7, 2147-2156. https://doi.org/10.4236/ce.2022.137135

Chun, Dorothy M.; Plass, Jan L. 1996. Effects of Multimedia Annotations on Vocabulary Acquisition. *The Modern Language Journal* 80:2, 183–198. https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=1112&context=edupub Accessed: 17.02.2024

Corson, David. 1995. Using English Words. Dordrecht: Kluwer Academic Publishers.

Curran, Tim; Doyle, Jeanne. 2011. Picture superiority doubly dissociates the ERP correlates of recollection and familiarity. *Journal of Cognitive Neuroscience*. 23:5, 1247–1262. https://doi.org/10.1162/jocn.2010.21464

de Bot, Kees. 1996. The psycholinguistics of the Output Hypothesis. *Language Learning*, 46, 529–555.

Defetyer, Margaret Anne; Russo, Riccardo; McPartlin, Pamela Luise. 2009. The picture superiority effect in recognition memory: a developmental study using the response signal procedure. *Cognitive Development*, 24:3, 265–273. https://doi:10.1016/j.cogdev.2009.05.002

Dulay, Heidi; Burt, Marina K. 1977. Remarks on creativity in language acquisition. M. Burt, H.Dulay and M. Finnochiaro (Eds.). Viewpoints on English as a second language. New York: Regents Press, pp. 95–126.

Ellis, Nick C. 1995. Vocabulary acquisition: psychological perspectives and pedagogical implications. *The Language Teacher*, 19, 12-16.

Ellis, Rod. 1990. Instructed second language acquisition. Oxford: Blackwell.

Ellis, Rod. 2003. Task-based language learning and teaching. Oxford: Oxford University Press.

Ellis, Rod. 2005. Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition,* 27:2, 141-172.

Gass, Susan M. & Larry Selinker. 2008. Second language acquisition: An introductory course, 3rd edn. New York: Routledge.

Green, Peter; Hecht, Karlheinz. 1992. Implicit and explicit grammar: an empirical study. *Applied Linguistics*, 13, 168-184.

Gregg, Kevin R. 1984. Krashen's Monitor and Occam's Razor. *Applied Linguistics*, 5:2, 79–100.

Groot, Peter J.M. 2000. Computer assisted second language vocabulary acquisition. *Language Learning & Technology*, 4:1, 60-81.

Jegerski, Jill. 2021. Krashen and second language processing. *Foreign Language Annals*, 54:2, 318-323. https://doi.org/10.1111/flan.12557

Kim, SoHee. 2019. The effects of using images and video clips for vocabulary learning: Single words vs. phrasal verbs. *STEM Journal*, 20:2, 19-42.

Krashen, Stephen D. 1981. Second language acquisition and second language learning. Oxford: Pergamon. https://www.sdkrashen.com/content/books/sl_acquisition_and_learning.pdf Accessed 12.02.2024

Krashen, Stephen D. 1982. Principles and Practice in Second Language Acquisition. Pergamon Press.

Krashen, Stephen D. 2003. Explorations in Language Acquisition and Use. Portsmouth, Heinemann.

Laufer, Batia. 1998. The development of passive and active vocabulary: same or different? *Applied Linguistics*, 19, 255-271.

Leung, Constant; Valdés, Guadalupe. 2019. Translanguaging and the transdisciplinary framework for language teaching and learning in a multilingual world. *The Modern Language Journal*, 103:2, 348–370. https://doi.org/10.1111/modl.12568

Lichtman, Karen; Van Patten, Bill. 2021. Was Krashen Right? Forty years later. *Foreign Language Annals*. 54:2, 283-305. https://doi.org/10.1111/flan.12552

Long, Michael. H. 1990. The least a second language acquisition theory needs to explain. *TESOL Quarterly*, 24, 649–666. https://doi.org/10.2307/3587113

Mason, Beniko.; Krashen, Stephen. 2020. The promise of "Optimal Input." *Turkish Online Journal of English Language Teaching*, 5:3, 146-155. https://www.beniko-mason.net/content/articles/2020-10-20-revised-the-promise-of-optimal-input.pdf Accessed 22.04.2024.

Matthews, William. J.; Benjamin, Clare; Osborne, Claire. 2007. Memory for moving and static images. *Psychonomic Bulletin & Review* 14, 989–993. https://doi.org/10.3758/BF03194133

Mayer, Richard E. 1997. Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32:1, 1-19.

Mayer, Richard E. 2009. Multimedia learning (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Mayer, Richard E. (Ed.). 2014. The Cambridge handbook of multimedia learning. New York, NY: Cambridge University Press.

Mayer, Richard. E.; Heiser, Julie, & Lonn, Steve. 2001. Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, 93:1, 187–198. https://doi.org/10.1037/0022-0663.93.1.187

Mayer, Richard. E. 2024. The Past, Present, and Future of the Cognitive Theory of Multimedia Learning. *Educational Psychology Review* 36, 8. https://doi.org/10.1007/s10648-023-09842-1

McBride, Dawn M.; Dosher, Barbara Ann. 2002. A comparison of conscious and automatic memory processes for picture and word stimuli: a process dissociation analysis. *Consciousness and Cognition*. 11:3, 423–460. https://doi:10.1016/s1053-8100(02)00007-7

McLaughlin, Barry. 1987. Theories of second language learning. Edward Arnold.

McLaughlin, Barry. 1990. "Conscious" versus "unconscious" learning. *TESOL Quarterly*, 24:4, 617–638.

Meara, Paul. 1990. A note on passive vocabulary. *Second Language Research*, 6, 150-154.

Mehine, Indra. 2021. Motivation and study skills in learning English vocabulary through computer games in Estonia. Master's thesis. University of Tartu.

Miller, George A. 1999. On knowing a word. *Annual Review of Psychology, 50,* 1–19. https://doi.org/10.1146/annurev.psych.50.1.1

Mitchell, Rosamund., & Myles, Florence. 2004. Second language learning theories (2nd ed.). Arnold.

Milton, James. 2008. Vocabulary uptake from informal learning tasks. *The Language Learning Journal*, 36:2, 227–237. https://doi.org/10.1080/09571730802390742

Mohd Jelani, Nurul Aini; Boers, Frank. 2018. Examining incidental vocabulary acquisition from captioned video. Does test modality matter? *ITL-International Journal of Applied Linguistics*, 169, 169–190. https://doi.org/10.1075/itl.00011.jel

Nation, Paul I.S. 2000. Learning Vocabulary in Another Language. Cambridge: Cambridge University Press.

Paivio, Allan. 1971. Imagery and verbal processes. New York: Holt, Rinehart, and Winston.

Paivio, Allan. 2010. Dual coding theory and the mental lexicon. *The Mental Lexicon. 5*, 205-230. https://doi.org/10.1075/ml.5.2.04pai

Peters, Elke. 2019. The Effect of Imagery and On-Screen Text on Foreign Language Vocabulary Learning From Audiovisual Input. *TESOL Quarterly*, *53*:4, 1008–1032. http://www.jstor.org/stable/45238629 Accessed 17 Mar. 2024.

Ramezanali, Nasrin; Uchihara, Takumi; Faez, Farahnaz. 2021. Efficacy of Multimodal Glossing on Second Language Vocabulary Learning: A Meta-Analysis. *TESOL Quarterly 55, 105–133.* https://doi.org/10.1002/tesq.579

Reynolds, Barry Lee; Cui, Yanwen; Kao, Chian-Wen; Thomas, Nathan. 2022. Vocabulary Acquisition through Viewing Captioned and Subtitled Video: A Scoping Review and Meta-Analysis. *Systems*, 10:133. https://doi.org/10.3390/systems10050133

Sallert, Helena. 2016. Using a television series to facilitate advanced learners' vocabulary learning. Master's thesis. University of Tartu.

Sato, Masatoshi; Loewen, Shawn. 2019. Methodological strengths, challenges, and joys of classroom-based quasi-experimental research. In R. M. DeKeyser & G. P. Botana (Eds.), Doing SLA research with implications for the classroom: Reconciling methodological demands and pedagogical applicability. *Language Learning & Language Teaching* 52, 31–54. John Benjamins. https://doi.org/10.1075/lllt.52.03sat

Schmidt, R. W. 1990. The role of consciousness in second language learning. *Applied Linguistics*, 11, 129–158.

Schmitt, Norbert. 2014. Size and Depth of Vocabulary Knowledge: What the Research Shows. *Language Learning,* 64:4, 913–951. https://doi.org/10.1111/lang.12077

Spada, Nina. 2005. Conditions and challenges in developing school-based SLA research programs. *The Modern Language Journal*, 89:3, 328–338. https://doi.org/10.1111/j.1540-4781.2005.00308.x

Stangroom, Jeremy. 2018. T-test Calculator for 2 dependent means. https://www.socscistatistics.com/tests/ttestdependent/default2.aspx Accessed: 13.04.2024

Swain, Merrill. 1985. Communicative competence: some roles of comprehen-sible input and comprehensible output in its development. Rrowley, M: New bury House.

Sweller, John. 1994. Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*. 4:4, 295–312. doi:10.1016/0959-4752(94)90003-5

Sweller, John. 2005. Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19–30). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.003

Sweller, John. 2010. Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review* 22, 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, John; Ayres, Paul; Kalyuga, Slava. 2011. Cognitive load theory. Springer: New York.

Sweller, John. 2019. Cognitive load theory and educational technology. *Education Tech Research Dev* 68, 1–16. https://doi.org/10.1007/s11423-019-09701-3

Sydorenko, Tetyana. 2010. Modality of input and vocabulary acquisition. *Language Learning and Technology*, 14:2, 50–73. http://dx.doi.org/10125/44214

VanPatten Bill. 1990. Attending to Form and Content in the Input: An Experiment in Consciousness. *Studies in Second Language Acquisition*, 12:3, 287-301. doi:10.1017/S0272263100009177

Vermeer, Anne. 2001. Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, 22, 217–234.

Webb, Stuart. 2008. Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30:79–95.

Zhang, Ruofei & Zou, Di. 2021. A state-of-the-art review of the modes and effectiveness of multimedia input for second and foreign language learning. *Computer Assisted Language Learning, 35:9, 2790–2816.* https://doi.org/10.1080/09588221.2021.1896555

APPENDICES

APPENDIX 1 - School's consent letter

Hr. Direktor

04.03.2024

Avaldus

Palun Teie nõusolekut, et viia Saku Gümnaasiumi A, B, D, E klassi õpilaste seas 2024.a läbi minu magistritööks vajalik uuring. Uuringu tulemusena valmib magistritöö, mis võrdleb uue sõnavara omandamise abivahenditena kasutatavate piltide ja videoklippide efektiivsust mnemotehnilist mõju eel- ja järeltestide kaudu. Igalt õpilastelt ja vähemalt ühelt lapsevanemalt on küsitud kirjalik nõusolek.
Magistritöö juhendaja on Tartu Ülikooli Maailma keelte ja kultuuride instituudi anglistika osakonna kaasprofessor Jane Klavan. Uuringu läbiviimisel saadud tulemused on pseudonüümitud tuvastamist mittevõimaldava koodiga.

Lugupidamisega,
Jüri Roots

Kooli direktor

(allkiri tähistab koolipoolse esindaja nõusolekut)

APPENDIX 2 - Student & parent consent form

Nõusolek uurimustöös osalemiseks

Tere hea õpilane ja lapsevanem!

Olen Saku Gümnaasiumi inglise keele õpetaja Jüri Roots. Palun teie nõusolekut enda magistritööga seotud uurimustöö läbiviimiseks inglise keele tundide raames Saku Gümnaasiumis.

Uurimustöö liik: teoreetilis-rakenduslik uurimus
Uurimustöö valdkond: sõnavara õppimise abivahendid, meeldejätmise efektiivsus kahe meediumi võrdluses (staatiline pilt vs video). Kokku 3 eksperimenti (koos mälutestidega)..
Uurimustöös kasutatavad vahendid: videoprojektor, paber, pastakas.
Uurimustöö läbiviimise koht: Saku Gümnaasium, õpilasele tuttavas keskkonnas
Uurimustöö läbiviimise ligikaudne aeg: veebruar-märts 2024
Isikuandmete kaitse: kõik uurimustöös avaldatavad andmed on anonümiseeritud. Seosesd konkreetsete isikutega on teada ainult uurimustöö läbiviijale (st. mulle)

Olen nõus osalema (õpilane) / lubama lapse osalemise (lapsevanem)

Lapsevanema nimi: _____ Lapsevanema allkiri:_____

Õpilase nimi: _____ Õpilase allkiri: _____

Kuupäev: _____/02/2024

Teil on igal ajal õigus oma nõusolek tagasi võtta. Selleks kirjutage palun minu meiliaadressile @saku.edu.ee, mille leiate kooli koduleheküljelt.

Ette tänades,

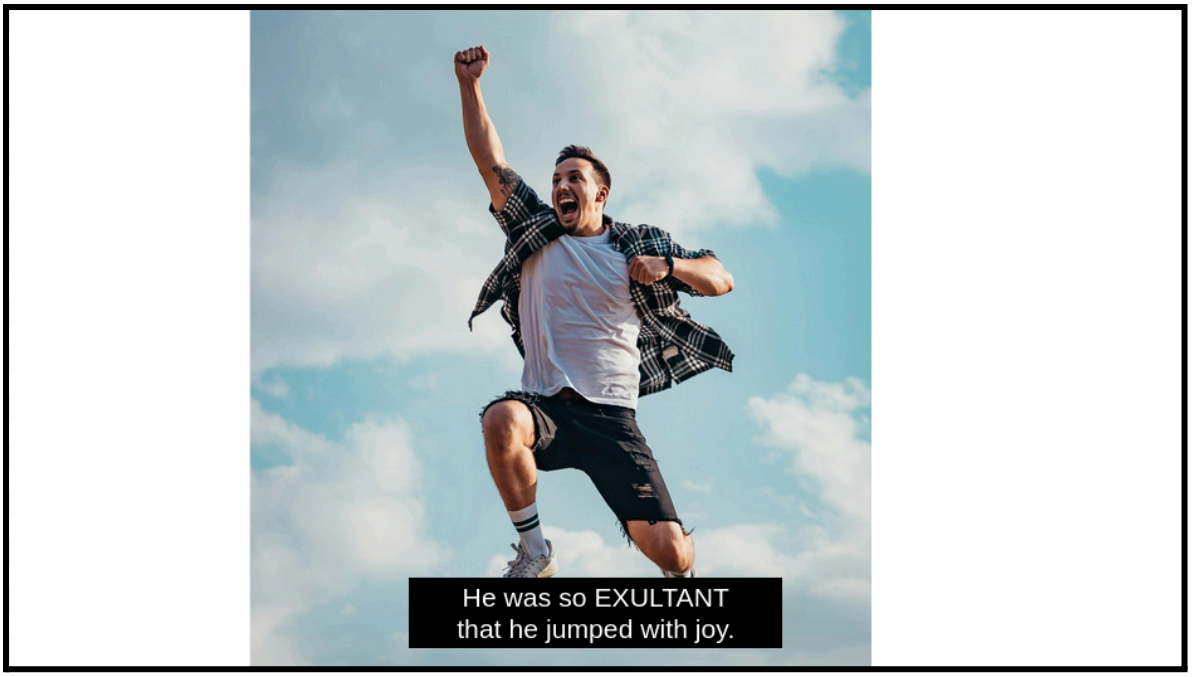
Jüri Roots

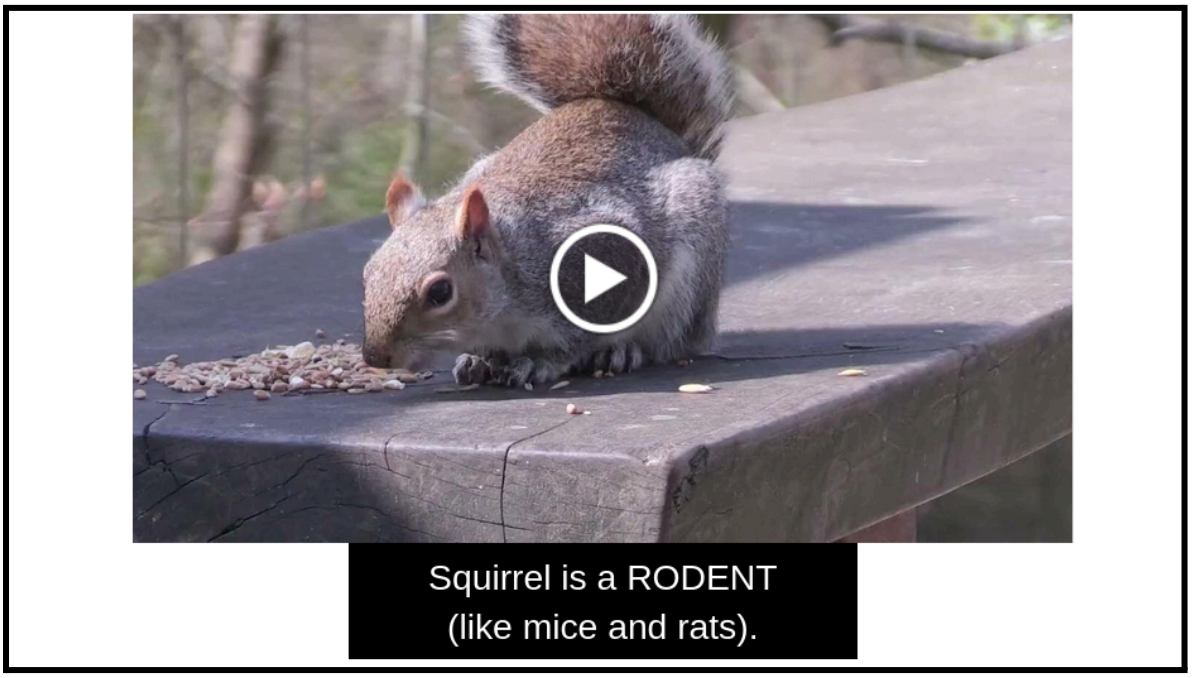
APPENDIX 3 - Immediate pre- and post-test answer sheet

Name: _____ Class: _____

Exultant	
Aghast	
Crevice	
Rodent	
To accede	
To vault	
Despondent	
Spry	
To dart	
To hunker	
Singlet	
Aglet	
Leery	
Dilapidated	
Grapnel	
Agraffe	
To clamber	
To avert	
Repugnant	
Livid	
Shingle	
Apron	
To hurl	
To throng	

APPENDIX 4 - Slide examples





APPENDIX 5 - Vocabulary list with definitions and captions

MW = Merriam Webster Online Dictionary

VERBS:

1. To dart: to thrust or move with sudden speed (MW)

   Caption: We saw a man DART towards us (Author).

2. To hurl: When you hurl something, you throw it hard. (vocabulary.com),

   Caption: He HURLED the javelin very far with great force (Author)

3. To vault: to execute a leap using the hands or a pole (MW)

   Caption: He VAULTED over the barricade (Author).

4. To avert: to see coming and ward off : avoid (MW)

   To avert is to turn away or to prevent. You might avert your gaze or avert a disaster — either way, you are avoiding something. (vocabulary.com).

   Caption: Luckily, the cameraman AVERTED being hit by the car! (Author).

5. To throng: to crowd together in great numbers (MW)

   Caption: The fans THRONGED to the stadium to listen to their idol (Author).

6. To hunker: crouch, squat - usually used with *down* (MW).

   To crouch or squat on one's heels: He hunkered to be at eye level with his dog.

   I can’t hunker with this bad knee. (dictionary.com)

   Caption: The rapper HUNKERED and hid his face in a dramatic pose during the photoshoot (Author)

7. To accede: to express approval or give consent **:** to agree to a request or demand (MW)

   to give consent, approval, or adherence; agree; assent; to accede to a request; to accede to the terms of a contract (dictionary.com).

   Caption: She acceded to the terms of the new contract (Author).

8. To clamber: to climb awkwardly or with effort especially by using both the hands and the feet (MW). To climb, using both feet and hands; climb with effort or difficulty. (dictionary.com)

Caption: The hiker clambered to the top of the wet and slippery slope (Author).

ADJECTIVES**:**

1. Exultant: filled with or expressing great joy or triumph (MW)

Caption: He was so EXULTANT that he jumped with joy (Author).

2. Livid: very angry, enraged (MW)

Caption: This man is very angry - he is LIVID! (Author).

3. Dilapidated: decayed, deteriorated, or fallen into partial ruin especially through neglect or misuse (MW).

Caption: This old manor looks abandoned and dilapidated (Author).

4. Aghast: struck with terror, amazement, or horror : shocked and upset (MW).

Caption: The man is AGHAST - shocked and terrified (Author).

5. Despondent: feeling or showing extreme discouragement, dejection, or depression (MW)

Caption: The man looks very despondent about something (Author).

6. Leery: suspicious, wary (MW). Being leery is being distrustful. If you're leery, it will probably show on your face and in your posture. You may squint and wrinkle your brows or take a few steps back. If you're leery that someone won't keep a secret, you won't confide in them, and if you're leery about whether they'll keep a promise to you, you won't count on them. Sometimes people are leery because of some proof, but often, just having a bad feeling or instinct is enough to make a person leery (vocabulary.com).

Caption: Being LEERY is being distrustful, to suspect (Author).

7. Repugnant: a repugnant thing is a thing offensive, detestable, or obscene. It can be repugnant to your mind or your morals (vocabulary.com).

   The images on her social media feed were just disgusting and REPUGNANT! (Author).

8. Spry: able to move quickly, easily, and lightly (MW).

   He is a SPRY (active, nimble, agile) dancer (Author).

NOUNS

1. Crevice: a narrow opening resulting from a split or crack (MW).

   There was a crevice in the asphalt (Author).

2. Singlet: a sleeveless athletic jersey (MW, modified)

   This wrestler is wearing a wrestling singlet (Author).

3. Grapnel: a small anchor with usually four or five flukes used especially to recover a sunken object or to anchor a small boat (MW)

   The grapnel has several hooks and is hanging on the rope (Author).

4. Shingle: coarse rounded detritus or alluvial material especially on the seashore that differs from ordinary gravel only in the larger size of the stones (MW).

   Coarse beach gravel of small waterworn stones and pebbles (vocabulary.com).

APPENDIX 6 - Delayed post-test example (flexiquiz.com)

Write as many words as you remember from the slideshow. **Small letters only.**

These words were shown using **STILL IMAGES** (you were shown the images without descriptions):

1. He was so _____ (+ed) that he jumped with joy (elated, overjoyed).
2. There was a _____ in the asphalt (a deep crack that occurs in natural stone etc)
3. She _____ (+ed) to the terms of the new contract (to agree, accept).
4. The man looks very _____ about something (depressed, weighed down).
5. We saw a man _____ towards us (to accelerate quickly, run fast).
6. The rapper _____ (+ed) in a dramatic pose (to crouch, squat)
7. This wrestler is wearing a wrestling _____ (a sleeveless one piece clothing).
8. Being _____ is being distrustful, to suspect (others)
9. The _____ has several hooks (a small anchor)
10. The hiker _____ (+ed) up from the slippery slope (climb awkwardly by using hands and legs)
11. The images on her social media feed were just disgusting and _____! (very unpleasant and disturbing)
12. There is no sand here, only large pebbles - this is a _____ beach.

RESÜMEE

TARTU ÜLIKOOL
ANGLISTIKA OSAKOND
Jüri Roots

ENG: Effectiveness of still images vs. short video clips as visual aids in vocabulary learning and retention - a comparative study among 8-9th grade EFL students
EST: Piltide ja videote efektiivsus mnemotehniliste abivahenditena uue sõnavara õppimisel Eesti põhikooli 8-9klassi õpilaste seas: võrdlev uurimus
Magistritöö
2024
Lehekülgede arv: 61

Uurimustöö võrdleb inglise keelt võõrkeelena õppivate 8-9 kl õpilaste uue sõnavara meeldejätmise efektiivsust kõrge kognitiivse koormuse tingimustes piltide ja lühikeste videoklippide ning märksõnade / selgitavate tekstide kaasabil. Uurimuse läbiviimiseks korraldati kolmeosaline pseudoeksperiment, mille käigus 50-le õpilasele näidati slaidiesitlustes segamini pilte ja lühikesi videoklippe märksõnade ja lühidefinitsioonidega. Õpilased märkisid ankeet-küsitluses (nii paber- kui digitaalversioonis), kui paljud sõnakujud või sõnade tähendused neile meelde olid jäänud. Pseudoeksperimendi eesmärgiks oli võrrelda piltide ja videote efektiivsust meeldejätmise abivahenditena.
Magistritöö koosneb kahest peatükist. Esimeses peatükis antakse ülevaade olulisematest teooriatest ja seisukohtadest teise keele omandamise ning multimeedia kasutamise valdkonnas. Samuti osutatakse teoreetilistele vastuargumentidele, mida erinevad uuringud on tõstatanud. Teoreetiline osa loob aluse mitmete empiirilises osas tõstatuvate küsimuste võimalikuks tõlgendamiseks.
Empiirilises osas kirjeldatakse pseudoeksperimendi läbiviimise tingimusi ning lähtekohti selle planeerimisel, samuti saadud tulemusi.
Tulemustest võib reservatsioonidega välja lugeda lühivideo võimaliku eelise staatilise pildi kui mnemotehnilise abivahendi ees võõrkeelse sõnavara õppimise kontekstis ühe Eesti kooli näite varal.

Märksõnad: inglise keel, sõnavara, multimeedia, video, pilt, mälu

Lihtlitsents lõputöö reprodutseerimiseks

ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Jüri Roots,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose Effectiveness of still images vs. short video clips as visual aids in vocabulary learning and retention - a comparative study among 8-9th grade EFL students,
mille juhendaja on Jane Klavan, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commonsi litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Jüri Roots
12.05.2024

Autorsuse kinnitus

Kinnitan, et olen koostanud käesoleva magistritöö ise ning toonud korrektselt välja teiste autorite panuse. Töö on koostatud lähtudes Tartu Ülikooli maailma keelte ja kultuuride kolledži anglistika osakonna magistritöö nõuetest ning on kooskõlas heade akadeemiliste tavadega.

Jüri Roots
14.05.2023

Lõputöö on lubatud kaitsmisele.
[Juhendaja allkiri]

Jane Klavan