The Effect of using Images and their Memorability on L2 Vocabulary Acquisition

Niek Warringa

Department of Psychology, University of Groningen

PSEMACT-10: Master's Thesis Applied Cognitive Neuroscience

Supervisor: Dr. Mark Nieuwenstein

Second evaluator: <kan ik even niet terugvinden>

January 8, 2024

Abstract

The present study investigated the impact of image presence and memorability on second language (L2) vocabulary retention for abstract words. Ninety-nine Dutch-speaking psychology bachelor students participated in a 2x2x2 mixed design study, with image presence as the between-subjects factor and memorability and retention interval as within-subjects factors. The experiment, conducted through OpenSesame on the MindProbe server, employed a vocabulary list of 40 abstract Finnish words paired with Dutch translations. Words were selected based on their concreteness level and the possibility to link them to a memorable or non-memorable picture that represented the same concept. Participants were assigned to either the image absent or the image present condition. Based on various theories including multimedia learning theory and dual coding theory, it was expected that pictures would enhance the encoding phase of an abstract vocabulary retention test. Contrary to expectations and previous research that did find positive effects, the results revealed no significant effects of image presence or memorability on vocabulary retention accuracy in both short-term and long-term retention phases. Future research should be directed in further exploration of boundary conditions in which visual aids do and do not enhance L2 vocabulary learning.

Keywords: Second Language Learning, Image Memorability, Multimedia Learning, Visual Aids

The Effect of using Images and their Memorability on L2 Vocabulary Acquisition

Acquiring a second language (L2) is a process that many individuals undertake at some stage in their lives, due to numerous factors such as educational and societal requirements or personal interest. Second language learning is a rewarding but also challenging endeavour for many people (Spence, 2022) and one of its key aspects is acquiring new vocabulary, which are the building blocks of every language and facilitate communication and comprehension. Memorizing L2 vocabulary can be especially difficult for abstract words (Altarriba & Bauer, 2004; de Groot, 1992; de Groot et al., 1994; Schwanenflugel et al., 1992), although it is different from L1 vocabulary learning, because an individual has already developed a conceptual and semantic system through their native language (Wu, 2012). However, the advantage of this system is cancelled out by the effects of ageing. Research shows that second language proficiency equal to the level of native speakers can only be achieved if the learning process starts before the age of ten years (Hartshorne et al., 2018).

Decades of research have made clear that images are better remembered than words, (e.g. Bevan & Steger, 1971; Lieberman & Culpepper, 1965; Kirkpatrick, 1894) and therefore, combining words with images may be an interesting approach in making vocabulary learning more efficient. In the present study, an experiment was conducted that investigated how the presence of an image influences creating associations between foreign and native vocabulary. In addition, the nature of the used words (abstract or concrete) and images (degree of memorability) were taken into consideration. The rationale behind the experiment will be explained in more detail in the following paragraphs, where we will review various theories that shed light on the influence of using images during the encoding of L1-L2 word pairs.

Using Images During Encoding Predicts Facilitation of L2 Vocabulary Learning

Multimedia Learning Theory

The best way to learn L2 vocabulary, i.e. having the biggest chance of remaining in the long-term memory, is deep processing (Reisberg, 2018). Deep processing occurs when elaborative retrieval is taking place, which can be achieved by multiple aspects of a retrieval task. For example, if a task lets you think about the meaning of the word or somehow makes you link it to existing knowledge. In Mayer's multimedia learning theory (MMLT; 1997), it is described how the combination of images and text can be used to enhance study practices, which is called the multimedia principle (Fletcher & Tobias, 2005; Mayer, 2009). MMLT is based on cognitivism and assumes three core concepts. The first one is the dual channel assumption, which implies that information can be processed through both verbal and visual channels and will be clarified later. The second assumption states both the visual and verbal channels have limited capacity for simultaneously processing information. Third, meaningful learning is an act of making connections between the channels and appropriate cognitive processing by attending to the relevant information, organizing the information and integrating it logically with long term knowledge.

Multimedia Learning Principles

Alongside the multimedia principle, there are other principles proposed by Mayer (2009) that also need to be taken into account in the attempt of optimizing foreign language vocabulary learning. There are twelve principles that originate from the cognitive load theory by Sweller (1988) and aim to manage the extraneous load (caused by factors that are not related to the learning goal), intrinsic load (the difficulty and complexity of the material itself) and germane load (acquiring and automating schemas in long-term memory). Conceptualization of the germane principle resulted from observations that an increase in cognitive load does not decrease learning per se but could increase learning as well (Sweller et al., 1998). Table 1 gives an overview of the relevant principles. Not all principles were included, because not all twelve principles are applicable to the context of vocabulary learning (but are to other learning environments, e.g. presentations). All of the principles were tested in different learning contexts and showed predominantly positive effects on learning outcomes (for an overview of studies, see Mayer, 2009). The ratio of the tests that reported positive effects and their median effect sizes (Cohen's *d*) are included in Table 1.

Table 1

Mayer's Principles of Multimedia Learning

Principle	Implication	Effect on Cognitive Load	Empirical Evidence
Multimedia	Use both pictures and words	Optimize germane load	11/11, *d* = 1.39
Coherence	Exclude irrelevant words, pictures and noises	Reduce extraneous load	13/14, *d* = 0.97
Signaling	Emphasize essential words and pictures	Reduce extraneous load	5/6, *d* = 0.52
Spatial Contiguity	Place words near the corresponding graphic	Reduce extraneous load	5/5, *d* = 1.09
Temporal Contiguity	Present words and the corresponding graphic at the same time	Reduce extraneous load	8/8, *d* = 1.31
Segmenting	Learners should learn in chunks, in which they themselves control the pace and timing	Manage intrinsic load	3/3, *d* = 0.98
Modality	People learn better from spoken words and pictures, than from printed words and pictures	Manage intrinsic load	17/17, *d* = 1.02
Personalization	People learn better with familiar, informal language (e.g. when target vocabulary is presented within context sentences)	Optimize germane load	11/11, *d* = 1.11
Voice	Auditorily presented vocabulary by human voices are processed better than computed voices	Optimize germane load	3/3, *d* = 0.78

Picture Superiority Effect

To better understand why MMLT suggests that people learn better form words together with pictures than from pictures alone, first the phenomenon that pictures and images are more likely to be remembered than words needs to be clarified, which is called the picture superiority effect (PSE; Paivio et al., 1968). The PSE has been robust across various experimental methods, such as free recall-, signal- and drawing tasks. In free recall tasks, participants receive a list of words and pictures and are then asked to recall as many as they can without importance of order (Paivio & Csapo, 1973; Mintzer & Snodgrass, 1999). Signal tasks are similar to free recall tests but with a time limit included (Defeyter et al., 2009) and in drawing tasks, participants get the task to draw or write down the words and objects that have been presented (Paivio & Csapo, 1973). Moreover, it is more likely for studied stimuli presented as pictures to be discriminated from unstudied items, than it is for their verbal counterparts (Shepard, 1967; Standing et al., 1970).

One theory behind this picture superiority effect is the dual-coding theory (DCT), which claims that stimuli can be encoded in image codes and words in verbal codes (Paivio, 1986, 1991, 2013). When either a word or a picture is encoded, it is processed through the respective visual or verbal coding pathway. However, these systems are interconnected, which means that a picture or word may also evoke the opposite channel. These connections are more likely to happen from image codes to verbal codes than vice versa, because images have more distinct features which can more easily evoke both visual and verbal representations.. According to the code-additivity hypothesis (Paivio, 1991), encoding through multiple channels simultaneously causes better (long-term) storage of the input, which is in line with the previously discussed germane principle (increasing cognitive load does not necessarily decrease learning). Thus, in this context pictures are more memorable than words because it is more likely they evoke multiple coding channels. For example, this was supported by the study of Snodgrass et al. (1974) in which subjects were more likely to recognize words corresponding to the pictures they had seen than to identify the picture that was a visual representation of the word presented. Moreover, not only are pictures more likely to be dually coded, but their distinctive and salient features also make them stand out in memory (Rust & Mehrpour, 2020), causing the memorability of an image code to be stronger than the verbal code, which gives pictures another advantage over words (Paivio & Csapo, 1973). Therefore, in DCT the PSE is best explained by the combination of dual coding and dominance of the image code.

These mechanisms behind the PSE predict various potential benefits of presenting a picture together with an L1-L2 word pair. When pictures are presented together with L1-L2 word pairs during encoding, learners can form associations between the L2 word and the picture, as well as between the L2 word and the L1 word, leading to deeper processing because it enables encoding through both verbal and visual channels, which might not occur with words alone. Therefore, pictures can enhance encoding and retrieval of L2 words by providing an additional pathway for memory access. In addition, the memorability of the picture associated with the L2 word is higher than its L1 translation. Consequently, even if the initial link between the two words is forgotten, the learner might still recall the translation of the L2 word through the verbal code of the picture that is remembered, which would naturally occur in the first language.

Other theories assume an inherent distinctiveness of pictures, which helps explain the picture superiority effect (Ensor et al., 2019). One model that has received attention is the sensory semantic model of encoding (Nelson, 1979; Nelson et al., 1977). This model is similar to the DCT in the way it assumes two types of coding, which both can be activated through words and pictures. However, instead of visual and verbal codes the model assumes semantic codes which give meaning, and sensory codes which are about the perception of the physical organization of a word or picture. It assumes memory representations can be active for both the meaningful and the observable aspect of the stimulus. Differences at these levels can explain the picture superiority effect. The sensory and semantic code configuration of a picture may be stronger than the codes for its (verbal) label, because pictures are more distinct through their variety of visual features and characteristics, compared to words (Van Den Broek et al., 2021). In other words, the lines and curves forming a picture are more unique than those in words and provide more information and meaning than a word can. Whereas the dual coding theory assumes a combination of a higher probability of dual coding and dominance of the image code, the semantic sensory theory argues that the greater strength of image codes is primarily the reason for the picture superiority effect.

According to the sensory semantic model of encoding, pictures can facilitate encoding and retrieval of L2 words by having stronger sensory and semantic codes than words. When pictures are presented together with L1-L2 word pairs during encoding, learners can encode the L2 word more deeply and distinctively by attending to both the sensory and semantic codes of the picture, as well as the semantic code of the L1 word. Subsequently during recall, when only the L2 word is shown, the associated picture can activate both the sensory and semantic codes of the L2 word. This activation, in turn, can stimulate the semantic code of the L1 word (i.e. the corresponding translation). Therefore, pictures can enhance encoding and retrieval of L2 words by providing more information and meaning than words.

Abstract and Concrete Words

While the picture superiority effect shows an effect for pictures compared to words, there is also such an effect within language called the concreteness effect (Paivio, 1991). Prior studies have shown concrete words are processed more thoroughly than abstract words. In contrast to concrete words, abstract words are words that refer to ideas, feelings, qualities or states that are not directly observable or tangible, such as charity, love, justice or confidence.

By means of the DCT this is explained by the claim that concrete nouns, comparable to pictures, evoke a second imagery processing system in the right hemisphere. Meanwhile, abstract nouns, because of their poor imaginability rely fully on the verbal path in the left hemisphere only (Jessen et al., 2000). This had been demonstrated in an earlier study that examined participants with brain damage in the right hemisphere, as they recalled concrete words significantly worse than the control group, while the abstract word recollection was similar in both groups (Villardita et al., 1988).

The concreteness effect can also be explained by the context availability theory, which claims that concrete nouns have more context available than abstract nouns. All additional information that helps to understand and remember a word is context information, such as examples, situations and synonyms. Because concrete words are more familiar and meaningful, they inherently have more context information and thus are more easily processed and remembered than abstract words (Schwanenflugel & Shoben, 1983; Schwanenflugel et al., 1988). The fMRI study by Jessen et al. (2000), which was mentioned earlier, also supports this because they saw greater activation outside of the primary language processing areas for concrete words, both linked to visual and contextual processing (Jessen et al., 2000). Therefore, they suggest support for both context availability and dual coding theory in explaining the concreteness effect.

These findings imply that presenting an image together with an L1-L2 word pair would facilitate efficient vocabulary learning mainly for abstract words, as they are less likely to already have a mental image associated with them than concrete words. Moreover, pictures could play a significant role in providing the context which helps deeper processing and better memory storage for abstract words.

Results from Previous Studies

Multiple studies reported beneficial effects for using images during encoding of concrete L2 vocabulary (e.g., Akbulut, 2007; Hald et al., 2016; Shahrokni, 2009; Tonzar et al., 2009; Yeh & Wang, 2003) and abstract L2 vocabulary (Andrä et al., 2020; Farley et al., 2012; 2014; Kim & Gilman, 2008; Yeh & Wang, 2003). However, there are also studies that did not report multimedia benefits (Al-Seghayer, 2001; Boers et al., 2009; Cohen & Johnson, 2011; Dubois & Vial, 2001) or negative effects (Acha, 2009). Remarkable, but not surprising in light of the theories discussed, is that all of the studies that reported no or negative effects used concrete words in their experiments. This included a bachelor thesis project (also at RUG) supervised by Nieuwenstein (2023), which also reported no effects in their study for image-assisted vocabulary learning. Presumably, the nature of concrete words is such that mental imagery takes place inherently and the addition of a picture would be superfluous (Shen, 2010). However, the distinction between abstract and concrete words does not suffice in explaining the absence of beneficial effects, because there are also studies that did report benefits for concrete vocabulary. Therefore, it seems more likely that there is a combination of factors that together cause there to be no effect. One of these factors could be that some of the research designs did not or only partially follow the multimedia principles. For example, in the aforementioned studies there seems to be an advantage for presenting pictures at the same time as words rather than presenting words and pictures separately (e.g. a second before a word pair appears), which is in line with the temporal contiguity principle. Moreover, not only the nature of the used vocabulary, but also the nature of the used images in terms of how memorable they are could have been a factor in the mixed results of the studies.

Image Memorability

How well a picture is remembered, can be defined as image memorability and is the degree to which an image is remembered by different viewers. Previous studies have shown that image memorability is a highly consistent and intrinsic property of images, independent of other image features (Isola et al., 2011; Bainbridge et al., 2013). For example, Isola et al.

(2011) measured the memorability of various images from different semantic categories using a recognition task with 665 participants. They found that some images were consistently better remembered than others, regardless of the participants' individual differences. Similarly, Bainbridge et al. (2013) demonstrated that the memorability of face images was also stable across different viewers, using a comparable recognition task. These results indicate that memorability is an objective and inherent characteristic of images, rather than a subjective and extrinsic one. In addition, the study by Kramer et al. (2023) offers an extensive overview on the subject by analysing 26.107 images of 1854 different object concepts, which were presented to 13.946 participants. In their study, memorability was quantified as the proportion of correct recognitions, minus the proportion of false alarms for that particular item. They too found that participants tended to remember and forget mainly the same images, regardless of order and context, which suggest memorability is an intrinsic property of an image. Moreover, it is suggested that semantic properties, which are the meaningful properties we use to categorize objects in (e.g. animacy, utility), play a more important role than visual properties (e.g. shape, brightness) in determining how intrinsically memorable a picture is. They obtained these results by using a combination of human ratings and computational methods to categorize each object concept into 49 semantic and visual dimensions (e.g. metal/tools & long/thin), which could then be used as predictors of memorability. As for prototypicality, the degree to which an object conforms to the standard representation of its category or concept, its relationship with memorability is more equivocal than assumed by past studies. Often was assumed that the most memorable images included either the most prototypical (Bainbridge et al., 2017) or the most atypical (Mohsenzadeh et al., 2019) version of an object. Although there tends to be a positive relationship between prototypicality and memorability for most objects, there is also a significant amount of variability across concepts and categories. More recent work suggest that a combination of high prototypicality for semantic features and high atypicality for visual features tend to contribute most to memorability (Koch et al., 2020). What is most relevant for application in the current study, is that semantic features explain almost forty percent of the variance in memorability, compared to almost ten percent for visual features. The authors' main claim in this study is that semantic features have a higher memorability than visual features due to the deeper processing that they entail.

Associative Memory

Linking theories of picture superiority to enhance abstract L2 vocabulary lies in deeper processing as a result of encoding through multiple codes (the code-additivity hypothesis). However, this also depends on the assumption that information is formed in an associative structure. This implies that memorizing different elements of the same thing is related. Madan et al. (2010) found in their study on word pairs that higher imageability, being the property of a word defined as subjective evaluations of contributing to mental imagery, improved the recall of the association between these words. Therefore it is suggested that pictures can compensate for the lack of imageability of (mainly) abstract word pair associations. In addition, Horner & Burgess (2013) found that the retrieval of one element of an event is depended on the retrieval of the other elements, suggesting that events are stored as coherent associative structures, which takes place within the hippocampus. These studies form an argument for a holistic view on learning, in which word pairs and multi-element events as are learned coherent and integrated units, rather than as separate and isolated items, i.e. different elements involved in encoding and retrieval of the associations interact and influence each other in both directions. From this follows that if the memorability of one of these elements is positively affected, the memorability of the whole association is increased. Therefore, it is suggested that by adding a picture to abstract L1-L2 word pairs, the strength of its association can be increased by improving the imageability of the word pairs, but also by adding an element to the association that intrinsically has a higher memorability, in accordance with the picture superiority effect. Moreover, it is expected that this assumed effect can be most optimally utilized if the added picture is the most memorable version of pictures available depicting the word pair.

The Present Study

To summarize the goals of the present study on second language learning, we aimed to make the process of vocabulary acquisition more efficient by looking into the effect of pictures during the encoding phase of a vocabulary retrieval test. In the present study Finnish was used as the foreign language because of its lexical dissimilarity with Dutch (Bella et al., 2021). The previously discussed bachelor thesis project, which found no effects for the use of images paired with concrete words, formed the starting point of conducting a similar study design within the same university, but now with a focus on abstract words. In light of the context of the discussed theories and previous studies we expected to find a concreteness effect in abstract L2 vocabulary learning through the addition of pictures during encoding.

Moreover, image memorability and its potential (interaction) effect on L2 vocabulary learning was investigated. We expected that the assumed positive effect of adding pictures would be stronger for memorable pictures as opposed to non-memorable pictures. Because of the assumed deeper processing in the image present and the memorable conditions than in the absent and nonmemorable conditions, it is expected that this results in comparatively higher scores, in particular on the long-term retention (Craik & Lockhart, 1972).

Finally, we assessed relationships between time spent encoding and accuracy scores on retention, as engagement and achievement are positively associated (Nagy and Ulitzsch, 2021). In the context of a vocabulary test, engagement involves enhanced effort and involvement, which could be expressed in longer study times. Moreover, although deep processing does require more attention, deep processing does not necessarily take more time than shallow processing (Craik, 2002), therefore we did not expect an effect of image presence and memorability on encoding times.

Materials and Methods

Participants

Before participant recruitment took place, the study was approved by the Ethics Committee of the Faculty of Behavioural and Social Sciences (EC-BSS) at the University of Groningen. An a priori power calculation (Nieuwenstein, 2023) determined the estimated sample size that was required, estimated through the studies of Farley et al. (2012;2014) an effect size of .71 was expected for the difference between the groups with and without pictures. To have 80% power to find this effect at $p < .05$ we would need 50 participants, 25 per condition (with/without picture). For the effect of memorability there was no estimate of the effect size. Therefore, we adopted Brybaert's (2019) recommendation of assuming an effect size of .4, because this is a reasonable estimate of the smallest effect size that would be relevant and reliable (i.e. replicable) for our research question. To establish this effect we would need (at least) 41 participants in the condition with pictures and 41 participants in the condition without pictures, resulting in an estimated sample size of $N = 82$.

Participants were recruited from a pool of first-year psychology bachelor students in return for partial course credit. All participants were Dutch and/or spoke the Dutch language fluently. In the pre-screening questions it was asked whether they spoke Finnish, Estonian, Danish, Hungarian, Icelandic, Norwegian or Swedish, in order to make sure we only used the data of participants with no prior L2 knowledge. Among the final 99 participants, 74 were female, 24 male and 1 non-binary. Ages ranged from 18-31 (M = 20.5, SD = 3.5). Participants were assigned to either the image absent or present version of the experiment, after they had signed up for one of the versions of the experiment.

Materials

Software

The experiment was built in OpenSesame version 3.3.14 (Mathôth et al., 2012) and was run on MindProbe, a free online server powered by Jatos (Lange et al., 2015). A link to the server was placed within the SONA platform, so that the experiment was executable via the participant's own computer, regardless of location.

Vocabulary List and Pictures

For the experiment a vocabulary list of 40 abstract Finnish words with Dutch translations was composed. The process of selecting the abstract words with matching pictures was not as straightforward as it might seem. It involved keeping consistency in the abstractness of the words by keeping the concreteness scores as low as possible, which were rated by participants in Brysbaert et al. (2014) on a scale from 1 to 5. For each abstract word two pictures from the THINGS database (Kramer et al., 2023) were selected with as much difference as possible in their memorability scores, which were rated on a scale from 0 to 1. In all cases, the memorable and the non-memorable images portrayed the same concept. Because choosing a suitable picture for an abstract word was challenging due to their intangibility, familiar symbols and metaphors that would be associative to the word for most participants were used (see Figure 1 for an example). This has mostly been done using AI to keep matching as objective as possible, but some degree of subjectiveness was inevitable. Moreover, the selection of words and pictures was thus interdependent and finding suitable pictures for some abstract words was sometimes unsuccessful, which then resulted in selecting a slightly less abstract to be able to link two suitable pictures to it.

Figure 1

Example item of a metaphorically used image for target word loyal (as loyal as a dog)

The final set of selected words had a concreteness score between 1.07 and 2.93 (*M* = 2.05, *SD* = 0.42) and the average difference in memorability for the memorable and non-memorable pictures was 0.25 (*SD* = 0.04). The vocabulary list with all concreteness scores can be found in table A1, along with a measure of word frequency based on film and television subtitles (SUBTLEX) and the age of acquisition (AoA) to indicate the familiarity and accessibility of the words used (Brysbaert et al., 2014). A list of the pictures that were used along with their memorability scores can be found in table A2.

Procedure

The present study used a 2x2x2 mixed between-within subjects design with three independent variables: image presence (present vs. absent), retention interval (short-term vs. long-term) and image memorability (low vs. high). Image presence was the between-subjects factor in this design, with 49 participants in the absent group and 50 participants in the present group. Memorability and retention interval both were within-subjects factors. The dependent variable was vocabulary acquisition, measured by the proportion of correct answers per participant (in the remainder of this paper referred to as *accuracy*). The correctness of the answers was checked with some leniency, i.e., words that were misspelled for up to two letters were counted as correct. To control for order and sequence effects, we used a counterbalanced design with two image present and two image absent versions[1] of the experiment. In each version, the presentation of the words was reversed, so that each word was paired with a memorable or a non-memorable picture in the image present condition, and with a memorable or a non-memorable dummy condition in the image absent condition. This ensured that the means for the memorable and non-memorable conditions were computed based on the same subsets of items across the participants in both conditions. Participants were assigned to one of the four versions of the experiment by clicking on a version in the SONA portal (without

[1] even though non-memorable and memorable items are practically speaking not meaningful in the two image absent versions of the experiment.

knowing the details of the experiment they just signed up for). Participation in one study version automatically meant exclusion from the other three versions.

Prior to the start of the experiment, participants filled in an informed consent and answered a number of demographical questions. In the next screen participants were informed with the structure and duration of the experiment, which would take approximately thirty minutes and was built up out of five encoding and short-term test blocks of eight words followed by one final test block testing all 40 words in a random order. The sequence and items presented per each block were fixed, but the order of the items within each block was randomized. The average word length per block was kept approximately the same. In the encoding phase, participants saw a Finnish word, its Dutch translation, and a non-memorable image, a memorable image or no image, depending on the experimental condition they signed up for (Figure 2). To facilitate learning, they subsequently pressed spacebar to hide the Dutch word and the image (if applicable), after which they had to type in the Dutch translation for the Finnish word that would remain visible on screen along with the answer that was typed in by the participant. Once this procedure was done for all words in a block, a short-term testing phase without images started in which participants were asked to type the translations of all eight Finnish words from the respective block in random order by heart, followed by feedback regardless of the correctness of an answer given (short-term retention). This procedure was repeated for all blocks, after which participants took a final test that consisted of all 40 Finnish words without any image or feedback (long-term retention). The experiment ended with a debriefing session, in which participants were informed about the purpose of the study.

Figure 2

Non-memorable Condition (a) Memorable Condition (b) and No Image Condition (c)

Results

The present study investigated the impact of images and their memorability on second language (L2) learning for abstract words. Specifically, we hypothesized a positive effect of the presence of images during the encoding phase of a vocabulary retention task. Further, we hypothesized a larger positive effect for images with higher memorability scores, compared to images with lower memorability scores. The hypotheses were evaluated through statistical analyses including mixed ANOVAs, paired samples t-test and Spearman's rank correlation coefficients.

Image Presence, Memorability and Retention Interval Effects on Accuracy Scores

To investigate the potential influence of image presence and memorability on the average vocabulary retention scores across two distinct task phases (the retention interval), a three-way mixed ANOVA was conducted. Independence of the between-subjects groups was guaranteed, because each mean score reflected a distinct participant. A normal Q-Q Plot, as shown in Figure B1, reflects the distribution of the model residuals, which was right-skewed. Taken into account that we had a large sample size with practically equal groups and ANOVA is fairly robust against the normality assumption under those prerequisites (Blanca et al., 2023), we decided to follow through with the model. Conducting a Levene's test indicated the homoscedasticity assumption was met, $F(1, 97) = 1.88$, $p = .17$. The sphericity assumption was met, because the within-subject factors had only two levels.

Table 2 shows the descriptive statistics for the mean accuracy scores across all possible conditions. Participants scored slightly better in the image present condition (M = .40, SD = .17) than in the image absent condition (M = .38, SD = .21). However, this difference appeared to be numerically small and, as Table 3, indicates did not reach statistical significance. The interaction effect of memorability and image presence was non-significant. A paired samples t-test comparing the memorable image condition (M = .40, SD = .19) and the non-memorable image condition (*M* = .40, *SD* = .16) using only participants in the image present condition showed no significant effect of memorability on the mean accuracy scores *t*(49) = -.04, *p* = .971, Cohen's *d* = -.005). The main effect of retention interval yielded an F-ratio of *F*(1, 97) = 143.52, *p* <.001, *ηp2* = .6, indicating that the mean correct score was significantly higher in the short-term test (*M* = .46, *SD* = .21) than in the long-term test (*M* = .32, *SD* = .18). As can be seen in Table 3, there were no significant interaction effects between any of the factors on the mean accuracy scores and therefore, no simple effects tests were performed.

Table 2

Descriptive Statistics For Accuracy Scores by Image Presence, Memorability and Retention Interval

Image Presence	Memorability	Retention Interval	*N*	*M*	*SD*
Absent	Memorable	Long-Term	49	.36	.21
Absent	Memorable	Short-Term	49	.42	.21
Absent	Non-Memorable	Long-Term	49	.34	.21
Absent	Non-Memorable	Short-Term	49	.40	.22
Present	Memorable	Long-Term	50	.36	.17
Present	Memorable	Short-Term	50	.44	.19
Present	Non-Memorable	Long-Term	50	.36	.15
Present	Non-Memorable	Short-Term	50	.44	.17

Table 3

Three-Way Mixed ANOVA Results

Predictor	Sum of Squares	*Dfn*	Error Sum of Squares	*Dfd*	*F*	*p*	η_p^2
Image Presence	0.039	1	13.859	97	0.27	.60	.003
Memorability	0.006	1	0.245	97	2.30	.13	.02
Image Presence X Memorability	0.005	1	0.245	97	2.15	.15	.02
Retention Interval	0.436	1	0.295	97	143.52	<.001	.60
Image Presence X Retention Interval	0.003	1	0.295	97	1.13	.29	.01
Memorability X Retention Interval	0.000	1	0.000	97	0.00	>.999	<.001
Image Presence X Memorability X Retention Interval	0.000	1	0.000	97	0.00	>.999	<.001

Encoding Time and Accuracy Scores

Next, the relationship between study time (in milliseconds) in the encoding phase and the proportion of correct answers was investigated. Scatterplots, which are presented in Figure B2 and B3, show a monotonic relationship between response time and accuracy. However, the variables of interest were not normally distributed and there was one clear outlier identified. Therefore, Spearman's rank correlation coefficient between the mean accuracy scores and encoding time was calculated across participants for both task phases rather than Pearson's correlation coefficient. There was a positive correlation between response time and accuracy in both the short-term retention task phase, $r(97) = .60$, $p < .001$, and the long-term retention task phase, $r(97) = .47$, $p < .001$, indicating higher vocabulary retention scores for participants who took more time during the encoding phase of the experiment.

Image Presence and Memorability Effects on Encoding Time

To investigate how image presence and memorability might play a role in the amount of study time during the encoding phase, a two-way mixed ANOVA was conducted. With image presence or absence determining the between-subjects group again, independence was guaranteed. A QQ-plot, as shown in Figure B4, reflects the distribution of the model residuals, which was right-skewed. However, as mentioned earlier, the sample size is large with practically equal groups and thus, we decided to follow through with the model. Moreover, Levene's test indicated the homoscedasticity assumption was met, $F(1, 97) = 0.37$. Using the same IV's as in the previous ANOVA (both two levels), the sphericity assumption was met.

Table 3 shows the descriptive statistics for the mean study times cross all possible conditions. Participants took slightly longer during the encoding phase in the image present condition ($M = 4607.47$, $SD = 2922.84$) than in the image absent condition ($M = 4486.56$, $SD = 3475.20$). However this difference appeared to be numerically small and, as Table 4, indicates did not reach statistical significance. Study times corresponding to the memorable pictured items (*M* = 4676.08, *SD* = 2861.51) were slightly higher than they were for the nonmemorable pictured items (*M* = 4538.86, *SD* = 2984.17), yet this effect was not found to be statistically significant (Table 4). The interaction effect between memorability and image presence on encoding time. was non-significant. Therefore, no simple effects tests were performed. These results suggest that that study time was not influenced by image presence and memorability.

Table 3

Descriptive Statistics For Encoding Times by Image Presence and Memorability

Image Presence	Memorability	*N*	*M*	*SD*
Absent	Memorable	49	4143.51	2455.89
Absent	Non-Memorable	49	4829.60	4494.50
Present	Memorable	50	4676.08	2861.51
Present	Non-Memorable	50	4538.86	2984.17

Table 4

Two-Way Mixed ANOVA Results

Predictor	Sum of Squares	*Dfn*	Error Sum of Squares	*Dfd*	*F*	*p*	η_p^2
Image Presence	723679	1	1566705535	97	0.04	.83	<.001
Memorability	3727566	1	530011496	97	0.68	.41	.007
Image Presence X Memorability	8387544	1	530011496	97	1.54	.22	.02

Note. *Dfn* indicates degrees of freedom numerator. *Dfd* indicates degrees of freedom denominator.

Exploratory Analysis: Word Difficulty

We did not find any statistically significant effects for our hypotheses. Therefore we considered the possibility that the difficulty of the Finnish words could affect the outcomes. As previously discussed, the intrinsic cognitive load may be increased by the complexity and difficulty of the material, meaning that more working memory resources are needed to process and store the words, which may reduce the efficiency and effectiveness of learning combined with images. For example, some word pairs may be more difficult than others, including words that are longer, have irregular spelling and/or have multiple meanings. Moreover, some of the abstract words might have been more difficult than others, because they had less cross-linguistic similarity, less contextual support, and less emotional salience (Altarriba & Basnight-Brown, 2011).

To test this, we calculated the mean scores for each word across participants, rather than for each participant across words. This created a continuous scale of difficulty, where lower mean scores indicated harder words and higher mean scores indicated easier words. We then calculated the difference scores for the image presence variable by subtracting the mean scores for each word in the image absent condition from the mean scores for each word in the image present condition (i.e. a positive difference score meaning higher mean accuracy for the image present condition and vice versa). We first plotted these difference scores against the mean scores for each word (Figure B5), which shows an unclear trend in terms of monotonicity and linearity. Therefore, we ran a Spearman's rank correlation to examine the relationship between word difficulty and the effect of image presence. There was a weak, positive correlation between the scores, $r(40) = .247$. However, two groups could be identified: a larger group ($N = 33$) of relatively difficult words and a smaller group ($N = 7$) of relatively easy words. Therefore we ran the same analysis again, including only the larger group of relatively difficult words, which found the correlation coefficient to be .141.

We also plotted the difference scores for the memorability variable against the mean scores for each word, using only the data from the image present condition. Figure B6 shows a similar unclear trend between the difference scores and the mean scores for each word. Again, the items could be separated into two groups of relatively difficult words ($N = 33$) and relatively easy words ($N = 7$). We ran a Spearman's rank correlation including all words, which found a (very) weak positive relationship between the scores, $r = .075$. After which we ran the same analysis again, including only the relatively difficult items, which found a weak positive relationship between the difference scores and mean accuracy scores of r = .232.

Discussion

Formatted: English (United Kingdom)

The present study aimed to broaden our understanding of the way people acquire a second language (L2), specifically abstract L2 vocabulary. Two main hypotheses were tested. First, we expected a positive effect for using visual aids in a vocabulary retention test spread over two distinct task phases. Second, we expected a positive effect for memorability, which would result in higher retention scores for words accompanied with a memorable image compared to words accompanied with a non-memorable image of the same concept. Neither of the hypotheses were supported, as no significant effects were found in our analysis.

Image Presence

We expected that a picture during encoding could help retrieve the L1 translation when only a L2 word was presented in a retention task. Not only because the abstract vocabulary was now more likely to be dually coded and processed deeper, but also by providing an extra memory pathway through the verbal code of the picture. However, the connections from image codes to verbal codes are stronger than the connections from verbal codes to image codes, as was discussed earlier (Paivio, 1986; 1991; 2013). Therefore, it is likely that when only the L2 word was presented, the reverse connection from the verbal code to the image code was not (always) properly made and could not serve as memory pathway to the verbal representation of the picture (which would occur in the native language). Furthermore, even if the verbal representation of a picture could be retrieved this would not result in a correct L1 translation of the L2 word per se, as the verbal representation of the picture was not identical to the abstract word.

However, this does not explain the findings of previous studies (Andrä et al., 2020; Farley et al., 2012; 2014; Kim & Gilman, 2008; Yeh & Wang, 2003) that did report beneficial effects for image assisted abstract L2 vocabulary learning, which we failed to replicate. Although not all studies mentioned earlier in the present research reported positive effects regarding the usage of visual aids in learning L2 vocabulary, the present study is now the only one among them concerning abstract vocabulary. Therefore, our findings contribute to the literature by offering a highly powered test of the effect. We aimed to find possible explanations by creating an overview of the studies, their methods and the circumstances that might influence the effects of visual aids. To do this in a theoretically substantiated way, the comparison of the studies will also incorporate the principles of MMLT that have been examined. These and other study characteristics were revised and are schematically displayed in Table 5. Since the present study is directed at the examining of the relationship between abstract L2 vocabulary learning and image usage, only the studies that also focused on abstract words or made a clear distinction between abstract and concrete words were included in the overview. For the sake of conciseness, only the most effective learning techniques are listed under *visual aids* and *memoranda*, e.g., if a study reported positive effects for both the visual aid, spoken text and written text conditions as well as the visual aid and written text condition, the greater effect would be selected for further comparison of the principles.

After narrowing down the discussed studies regarding L2 vocabulary acquisition to abstract L2 vocabulary, we can see that five out of the remaining six studies reported beneficial effects and only the present study reported no significant effects. This makes it all the more interesting to investigate the differences in the current study as opposed to the studies that did find an effect. What immediately stands out, is that both Farley et al. (2012; 2014) studies are similar to the present study in the way they follow MMLT principles, which indicates it is likely that the differences in effects these studies found compared to the present study need to be sought elsewhere. However, the three other studies that reported beneficial effects, followed various principles that the present study did not, which might have affected learning outcomes. In particular the modality principle, followed by Andrä et al. (2020) and Kim & Gilman (2008), which suggests that people learn more deeply from spoken words and pictures than from printed words and pictures. More specifically, in these studies the words enter the cognitive system through the ears and the pictures through the eyes, while in the

Table 5

Overview of studies including abstract L2 vocabulary learning

Study	**Participants**	**L1**	**L2**	**Study Procedure**	**Visual aids and Memoranda**	**Effect**	**Extraneous load principles**	**Intrinsic cognitive load principles**	**Germane load principle**
Andrä et al. (2020)	54 primary school children (8 y/o)	German	English	Delayed free recall and translation test after 5 days of training	Pictures and gestures during training, auditorily presented L2 words	Positive	Coherence, temporal contiguity	Modality	Multimedia, voice
Kim & Gilman (2008)	172 middle school children (14 y/o)	South-Korean	English*	Pretest, Instruction, Posttest, week-delayed retention test through 'degree of certainty' scores (1-100)	Graphics, visually and auditorily presented L2 target words within context sentences	Positive	Signaling principle, spatial contiguity, temporal contiguity	Modality	Multimedia, personalization, voice, image
Yeh & Wang (2003)	82 freshman university students	Taiwanese	English*	Pretest, digital 'Thanksgiving' courseware, posttest containing MC, word association and a cloze test questions (after 4 – 14 days)	Pictures, visually presented L2 words, alone and in context sentences	Positive, especially for concrete objects	Coherence, signaling, spatial contiguity	Segmenting	Multimedia, personalization
Farley et al. (2012)	87 university students learning Spanish	Various (American)	Spanish*	Pretest, 3 in class PowerPoint instructional phases, immediate and delayed production recall posttest	Pictures during encoding, visually presented L2 words	Positive	Coherence, spatial contiguity, temporal contiguity	Segmenting	Multimedia
Farley et al. (2014)	97 university students learning Spanish	Various (American)	Spanish*	Similar to Farley et al. (2012)	Pictures during encoding, visually presented L2 words	Positive, only for immediate posttest	Coherence, spatial contiguity, temporal contiguity	Segmenting	Multimedia
Present study (2023)	99 first year university students	Dutch	Finnish	5 blocks of 8 words in the encoding phase, followed by immediate production tests and a 40 word posttest after all five blocks	Pictures during encoding, visually presented L2 words	No significant effects	Coherence, spatial contiguity, temporal contiguity	Segmenting	Multimedia

* Already have some L2 command, which was controlled for in the pretests

Commented [NW1]: Heb je comment wel gezien maar wanneer ik hier twee kolommen van maakte, kon ik op geen mogelijkheid de tabel nog op 1 pagina krijgen...

present study both printed words and pictures entered the cognitive system through the eyes, possibly resulting in an overload in the visual system (Mayer, 2009). However, this does not mean that we lacked the rationale behind our choice for printed words with pictures, as Mayer (2009) also described some boundary conditions, in which printed words might be more effective. One of these conditions is whether it concerns non-native speakers, which we logically deemed to be of utmost importance in the context of L2 vocabulary learning.

Another striking difference is that our study is the only one that actually uses a completely new language as the L2 component, with a large lexical dissimilarity to the L1 component of the study. The other studies involved a language that was in a beginner stage of language acquisition (Farley et al, 2012; 2014), was already at a basic, secondary educational, level (Yeh & Wang, 2003; Kim & Gilman 2008) or was not as divergent to the first language in terms of lexical dissimilarity as Finnish is to Dutch (Andrä et al., 2020), in combination with young participants (8 years old) of whom it is known that they learn a second language more easily (Hartshorne et al., 2018). The studies did report some form of controlling for the pre-existing knowledge, but it remains a possible explanation for better retention of the vocabulary items because these participants already possessed a (basic) conceptual and semantic system for this foreign language. Besides having a basic understanding of the language in which the target vocabulary was presented, participants in both studies conducted by Farley et al. (2012; 2014) had a personal interest in acquiring and retaining the vocabulary, being students that were enrolled in a Spanish course. This is in contrast to a considerable limitation of the present study, in which participants had no intrinsic motivation of learning Finnish and presumably their main reason for participating in the study was the mandatory amount of participation credits to be obtained in the first year of their Psychology bachelor.

These suppositions of differences in intrinsic motivation and prior exposure get supported by comparing average accuracy scores for each study in the comparable test phases, which may reflect the difficulty of and/or the engagement with the vocabulary to be learned. Andrä et al. (2020) report a mean accuracy score on their three-day delayed post-test of 0.38, compared to 0.36 in the present study. A difference that is bigger than it seems, bearing in mind that the long-term posttest of the current study still fell within tens of minutes after the learning phase. On the other hand, last named study did test the L2-L1 word pairs, right after having tested L1-L2 word pairs, which probably had an increasing effect on the mean L2-L1 accuracy scores. Other studies also reported higher mean accuracy scores (Farley et al., 2012: 0.53 & Farley et al., 2014: 0.46), in line with the assumed higher personal interest and engagement. Moreover, Yeh & Wang (2003) and Kim & Gilman (2008) recorded no comparable accuracy scores, but they did report high scores on their respective ways (see Table 5) of post-testing, suggesting less difficult material than in the present study.

Furthermore, another source of variation between the studies is how they selected illustrative images for the abstract vocabulary items. After all, for dual coding to take place the words and pictures should be related to the same concept, and they should complement each other, not contradict or confuse each other (Peterson et al., 2010). In the present study, it might have been so that the used images, being mostly proverbial in meaning, were themselves too abstract/different to elicit enough semantic processing or association to create a lasting memory trace to the L1 translation. As a consequence, the use of image was superfluous and merely resulted in an increased cognitive load rather than dual coding. On the other hand, one could argue that a lesser match between image and word in terms of concept representation would actually elicit deeper processing, as it requires more thought to think of the possible link between the two, and thus facilitate deeper memory encoding. As the mean scores in the image present and image absent condition showed no significant difference, the two explanations may both be valid and have offset each other's effects.

We did try to match the words to an image as objectively as possible so that for most participants there was a certain association between word and picture, by making use of artificial intelligence. However, a limitation of the present study is that only one researcher made the final selection of images, albeit with a supervisor's check. Future research could adopt a more systematic method of image selection, such as using multiple assessors who agree on the chosen images. Unfortunately, other studies were not as concise in describing their method of image selection, therefore it is assumed they used some form of subjective selection, but the details remain unclear (e.g. they did not report any usage of AI models). Moreover, Andrä et al. (2020) did note that their visual aids existed of black and white line drawings, created by a professional cartoon artist, which is an interesting approach in light of reducing cognitive load.

Memorability

The second research question of this study was whether the memorability of the images used in the abstract L2 vocabulary learning task had an effect on the recall of the words. The hypothesis was that using images during encoding that were intrinsically more memorable would enhance the retention of the words. However, our results did not support this hypothesis as there was no significant difference in the mean accuracy scores between the memorable and non-memorable image conditions in both task phases, which suggests that the memorability of the images did not have a strong impact on the learning of the words.

Memorability scores of the pictures (acquired from the THINGS database), ranged between 0.53 and 0.95, with a mean difference between the memorable and non-memorable picture of the same concept of 0.25 (*SD* = 0.04), which seems reasonably distinctive. However, the lack of a significant effect suggests that this difference might not have been substantial enough to influence word retention significantly and memorability effects on vocabulary learning might only come into play when the difference is larger still.

Furthermore, it seems logical that the effect of image memorability relies strongly on the presence of the main effect for image presence and possible explanations can be found in continuation of the explanations discussed in the previous section. We explained earlier that if the memorability of one element of an association is increased, in this case the image memorability, the memorability of the whole association would be increased as a result. However, we did not find effects for the presence of an image in increasing the association between the L1-L2 vocabulary, which suggests that image memorability is no part of the association and would be an irrelevant factor in abstract word learning and thus, increasing the memorability of this element would have no effect.

Encoding Time

A positive relationship between study time and the mean accuracy scores was found to be significant, similar to the results in a bachelor thesis project supervised by Nieuwenstein (2023) regarding concrete words. However, this result is anything but groundbreaking, as it has antecedents in research on learning and memory dating all the way back to nineteenth century research by founding figures as Ebbinghaus and Bartlett (Sternberg & Pickren, 2019). More relevant to the current research is that this bachelor thesis also found a significant effect of image presence on study time compared to the no picture condition (i.e. image presence decreased study time), suggesting lower mean accuracy scores in their image presence group as an indirect result. They argued that participants may overestimate how well they encode the vocabulary when images are present by thinking images made the learning easier, based on earlier research by Carpenter & Olsen (2012). In their study, Carpenter & Olsen (2012) also investigated the role pictures had in vocabulary learning and found that images can facilitate learning, but only after eliminating the overconfidence caused by these images, either through warning them to not be overconfident or through retrieval practice. The exploration of the relationship between the addition of pictures and the response time during the encoding phase was also included in present study's analyses. Contradictory to the study by Carpenter & Olsen (2012) and the mentioned bachelor thesis (2023), the positive relationship between response time and mean accuracy scores was not found to be moderated by the presence or absence of images (and their memorability) during the encoding phase, suggesting various other possible explanations for differences in the time a participant took for the encoding of abstract vocabulary items, such as different learning strategies and the previously mentioned assumed lack of intrinsic motivation among some of the participants.

Word Difficulty

An exploratory analysis involving word difficulty as possible moderator for the effect that images have on L2 abstract vocabulary acquisition was conducted. The results suggested no differences in how image presence and memorability influence vocabulary learning across different word difficulties, given the current operationalization of word difficulty. Alternatively, researchers may include word difficulty as variable by using a multi-faceted features method representing word difficulty, which considers various aspects of a word, such as its length, syllables, domain specificity, semantic relatedness, morphological relatedness and syntactic complexity (Zhang et al., 2020).

Conclusion

The present study investigated the effects of image presence and image memorability on the learning of abstract L2 vocabulary. The results showed that neither image presence nor image memorability had a significant effect on the retention of words in both task phases. These findings are in contrast with previous studies that reported positive effects of visual aids on abstract L2 vocabulary learning. Possible reasons for this discrepancy include the use of Finnish as the L2 component, which is a complex language that differs greatly from the first language of the participants, the selection of images and the way they did (not) represent the same concept as the abstract vocabulary, the superfluity (distraction) of images, visual (instead of auditorily) presentation of words, and the lack of intrinsic motivation and interest of the participants. The present study suggests that using visual aids for abstract L2 vocabulary learning might not be as effective as expected. Therefore, teachers and learners should be cautious when selecting and using visual aids for L2 vocabulary learning. Future research should be directed in further elaboration of the circumstances and boundary conditions under which visual aids can serve as a useful tool in vocabulary acquisition, for example by using auditorily presented vocabulary and an intrinsically motivated participant pool.

References

Acha, J. (2009). The effectiveness of multimedia programmes in children's vocabulary learning. *British Journal of Educational Technology*, 40(1), 23–31. https://doi.org/10.1111/j.1467-8535.2007.00800.x

Akbulut, Y. (2007). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language. *Instructional Science*, 35(6), 499–517. https://doi.org/10.1007/s11251-007-9016-7

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, 5(1), 202–232.

Altarriba, J., & Basnight-Brown, D. M. (2011). The acquisition of concrete, abstract, and emotion words in a second language. *International Journal of Bilingualism*, *16*(4), 446–452. https://doi.org/10.1177/1367006911429511

Altarriba, J., & Bauer, L. (2004). The Distinctiveness of Emotion Concepts: A Comparison between Emotion, Abstract, and Concrete Words. *American Journal of Psychology*, *117*(3), 389. https://doi.org/10.2307/4149007

Andrä, C., Mathias, B., Schwager, A., Macedonia, M., & von Kriegstein, K. (2020). Learning foreign language vocabulary with gestures and pictures enhances vocabulary memory for several months post-learning in eight-year-old school children. *Educational Psychology Review*, 32(3), 815–850. https://doi.org/10.1007/s10648-020-09527-z

Bainbridge, W. A., Dilks, D. D., & Oliva, A. (2017). Memorability: A stimulus-driven perceptual neural signature distinctive from memory. *NeuroImage*, *149*, 141–152. https://doi.org/10.1016/j.neuroimage.2017.01.063

Bella, G., Batsuren, K., & Giunchiglia, F. (2021). A Database and Visualization of the Similarity of Contemporary Lexicons. In *Lecture Notes in Computer Science* (pp. 95–

Field Code Changed

104). Springer Science+Business Media. https://doi.org/10.1007/978-3-030-83527-9_8

Bevan, W., & Steger, J. L. (1971). Free Recall and Abstractness of Stimuli. *Science*, *172*(3983), 597–599. https://doi.org/10.1126/science.172.3983.597

Blanca, M. J., Arnau, J., García-Castro, F. J., Alarcón, R., & Bono, R. (2023). Non-normal data in repeated measures ANOVA: impact on Type I error and power. *PubMed*, *35*(1), 21–29. https://doi.org/10.7334/psicothema2022.292

Boers, F., Píriz, A. M. P., Stengers, H., & Eyckmans, J. (2009). Does pictorial elucidation foster recollection of idioms? *Language Teaching Research*, 13(4), 367–382. https://doi.org/10.1177/1362168809341505

Brysbaert, M., Warriner, A.B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46, 904-911.

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), 92–101. https://doi.org/10.1037/a0024828

Craik, F. I. M. (2002). Levels of processing: Past, present. . . and future? *Memory*, *10*(5–6), 305–318. https://doi.org/10.1080/09658210244000135

Craik, F. I. M., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/s0022-5371(72)80001-x

Cohen, M. T., & Johnson, H. L. (2011). Improving the acquisition of novel vocabulary through the use of imagery interventions. Early Childhood Education Journal, 38(5), 357–366. https://doi.org/10.1007/s10643-010 -0408-y

Field Code Changed

Field Code Changed

Defeyter, M. A., Russo, R., & McPartlin, P. L. (2009). The picture superiority effect in recognition memory: A developmental study using the response signal procedure. *Cognitive Development*, *24*(3), 265–273. https://doi.org/10.1016/j.cogdev.2009.05.002

De Groot, A., Dannenburg, L., & Vanhell, J. (1994). Forward and Backward Word Translation by Bilinguals. *Journal of Memory and Language*, *33*(5), 600–629. https://doi.org/10.1006/jmla.1994.1029

De Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*(5), 1001–1018. https://doi.org/10.1037/0278-7393.18.5.1001

Dubois, M., & Vial, I. (2001). Multimedia design: The effects of relating multimodal information. Journal of Computer Assisted Learning, 16(2), 157–165. https://doi.org/10.1046/j.1365-2729.2000.00127.x

Ensor, T. M., Bancroft, T. D., & Hockley, W. E. (2019). Listening to the Picture-Superiority Effect. *Experimental Psychology*, *66*(2), 134–153. https://doi.org/10.1027/1618-3169/a000437

Fletcher, J. P., & Tobias, S. (2005). The Multimedia Principle. In *Cambridge University Press eBooks* (pp. 117–134). https://doi.org/10.1017/cbo9780511816819.008

Hald, L. A., de Nooijer, J., van Gog, T., & Bekkering, H. (2016). Optimizing word learning via links to perceptual and motoric experience. Educational Psychology Review, 28(3), 495–522. https://doi.org/10.1007/ s10648-015-9334-2

Hartshorne, J. K., & Tenenbaum, J. B. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *PMC*. https://dspace.mit.edu/handle/1721.1/126641

Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multi-element

Field Code Changed

Field Code Changed

Field Code Changed

events. *Journal of Experimental Psychology: General, 142*(4), 1370–1383. https://doi.org/10.1037/a0033626

Jessen, F., Heun, R., Erb, M., Granath, D., Klose, U., Papassotiropoulos, A., & Grodd, W. (2000). The Concreteness Effect: Evidence for Dual Coding and Context Availability. *Brain and Language*, *74*(1), 103–112. https://doi.org/10.1006/brln.2000.2340

Kim, D., & Gilman, D. A. (2008). Effects of text, audio, and graphic aids in multimedia instruction n for vocabulary learning. Journal of Educational Technology & Society, 11(3), 114–126.

Kirkpatrick, E. A. An experimental study of memory. *Psychological Review*, 1894, 1, 602-609.

Koch, G. E., Akpan, E., & Coutanche, M. N. (2020). Image memorability is predicted by discriminability and similarity in different stages of a convolutional neural network. *Learning & Memory*, *27*(12), 503–509. https://doi.org/10.1101/lm.051649.120

Kramer, M. A., Hebart, M. N., Baker, C. I., & Bainbridge, W. A. (2023). The features underlying the memorability of objects. *Science Advances*, *9*(17). https://doi.org/10.1126/sciadv.add2981

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): an easy solution for setup and management of web servers supporting online studies. *PLOS ONE*, *10*(6), e0130834. https://doi.org/10.1371/journal.pone.0130834

Lieberman, L. R., & Culpepper, J. D. (1965). Words versus Objects: Comparison of Free Verbal Recall. *Psychological Reports*, *17*(3), 983–988. https://doi.org/10.2466/pr0.1965.17.3.983

Nelson, D. L., Reed, V. S., & McEvoy, C. L. (1977). Learning to order pictures and words: A model of sensory and semantic encoding. Journal of Experimental Psychology:

Field Code Changed

Human Learning and Memory, 3, 485–497. https://doi.org/10.1037/0278-7393.3.5.485

Nelson, D. L. (1979). Remembering pictures and words: Appearance, significance, and name. In L. S. Cermak & F. I. M. Craik (Eds.), Levels of processing in human memory (pp. 45–76). Hillsdale, NJ: Erlbaum.

Madan, C. R., Glaholt, M. G., & Caplan, J. B. (2010). The influence of item properties on association-memory. *Journal of Memory and Language*, *63*(1), 46–63. https://doi.org/10.1016/j.jml.2010.03.001

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32*(1), 1–19. https://doi.org/10.1207/s15326985ep3201_1

Mayer, R. E. (2009). Multimedia learning (2nd ed). New York: Cambridge University Press.

Mayer, R. E. (2013). Multimedia Instruction. In *Springer eBooks* (pp. 385–399). https://doi.org/10.1007/978-1-4614-3185-5_31

Mintzer, M. Z., & Snodgrass, J. G. (1999). The Picture Superiority Effect: Support for the Distinctiveness Model. *American Journal of Psychology*, *112*(1), 113. https://doi.org/10.2307/1423627

Mohsenzadeh, Y., Mullin, C., Oliva, A., & Pantazis, D. (2019). The perceptual neural trace of memorable unseen scenes. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-42429-x

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. https://ci.nii.ac.jp/ncid/BA19613767

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*(3), 255–287. https://doi.org/10.1037/h0084295

Paivio, A. (2013). Imagery and Verbal Processes. In *Psychology Press eBooks*. https://doi.org/10.4324/9781315798868

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, *5*(2), 176–206. https://doi.org/10.1016/0010-0285(73)90032-7

Paivio, A., Rogers, T. M., & Smythe, P. R. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, *11*(4), 137–138. https://doi.org/10.3758/bf03331011

Peterson, P. L., Baker, E. L., & McGaw, B. (2010). *International encyclopedia of education*.

Reisberg, D. (2018). *Cognition: Seventh International Student Edition*. W.W. Norton & Company. 215-219

Rust, N. C., & Mehrpour, V. (2020). Understanding image memorability. *Trends in Cognitive Sciences*, *24*(7), 557–568. https://doi.org/10.1016/j.tics.2020.04.001

Schwanenflugel, P. J., Akin, C. E., & Luh, W. (1992). Context availability and the recall of abstract and concrete words. *Memory & Cognition*, *20*(1), 96–104. https://doi.org/10.3758/bf03208259

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *9*(1), 82–102. https://doi.org/10.1037/0278-7393.9.1.82

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*(5), 499–520. https://doi.org/10.1016/0749-596x(88)90022-8

Shahrokni, S. A. (2009). Second language incidental vocabulary learning: The effect of online textual, pictorial, and textual pictorial glosses. The Electronic Journal for English as a Second Language, 13(3), 1–17.

Shen, H. (2010). Imagery and verbal coding approaches in Chinese vocabulary instruction. *Language Teaching Research*, *14*(4), 485–499. https://doi.org/10.1177/1362168810375370

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

Spence, C. (2022, April 29). *How learning a new language changes your brain | Cambridge English*. World of Better Learning | Cambridge University Press. https://www.cambridge.org/elt/blog/2022/04/29/learning-language-changes-your-brain/

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. https://doi.org/10.3758/bf03337426

Field Code Changed

Sternberg, R. J., & Pickren, W. E. (2019). The Cambridge Handbook of the Intellectual History of Psychology. In *Cambridge University Press eBooks*. https://doi.org/10.1017/9781108290876

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J., van Merrienboer, J.J.G. & Paas, F.G.W.C. Cognitive Architecture and Instructional Design. *Educational Psychology Review* 10, 251–296 (1998). https://doi.org/10.1023/A:1022193728205

Field Code Changed

Tonzar, C., Lotto, L., & Job, R. (2009). L2 vocabulary acquisition in children: Effects of learning method and cognate status. Language Learning, 59(3), 623–646. https://doi.org/10.1111/j.1467-9922.2009.00519.x

Van Den Broek, G. S. E., Van Gog, T., Jansen, E., Pleijsant, M., & Kester, L. (2021). Multimedia effects during retrieval practice: Images that reveal the answer reduce vocabulary learning. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000499

Villardita, C., Grioli, S. M., & Quattropani, M. C. (1988). Concreteness/Abstractness of Stimulus-Words and Semantic Clustering in Right Brain-Damaged Patients. *Cortex*, *24*(4), 563–571. https://doi.org/10.1016/s0010-9452(88)80050-9

Wasser, B., Finkelstein, M., & Goldberg, L. S. (1974). On the fate of visual and verbal memory codes for pictures and words: Evidence for a dual coding mechanism in recognition memory. *Journal of Verbal Learning and Verbal Behavior*. https://doi.org/10.1016/s0022-5371(74)80027-7

Wu, X. (2012). Vocabulary Learning in a Second Language. In *Springer eBooks* (pp. 3422–3424). https://doi.org/10.1007/978-1-4419-1428-6_943

Yeh, Y., & Wang, C.-W. (2003). Effects of multimedia vocabulary annotations and learning styles on vocabulary learning. CALICO Journal, 21(1), 131–144. https://doi.org/10.1558/cj.v21i1.131-144

Yoshii, M. (2006). L1 and l2 glosses: Their effects on incidental vocabulary learning. Language Learning & Technology, 10(3), 85–101.

Zhang, S., Jia, Q., Shen, L., & Zhao, Y. (2020). Automatic classification and comparison of words by difficulty. *Communications in Computer and Information Science*, 635–642. https://doi.org/10.1007/978-3-030-63820-7_72

Field Code Changed

Formatted: English (United Kingdom)

Appendix A

Table A1

Used Words and Corresponding Scores

Dutch Word	Finnish Word	SUBTLEX	AoA	Concreteness score
gezond	terveitä	1392	6,53	2.13
futurisme	futurismi	1	13,17	1.07
loyaal	uskollinen	235	11,78	1.87
rijkdom	varallisuus	436	7,46	1.87
magie	taika	1442	7,64	1.87
eer	kunnia	3261	8,96	1.4
bevolkingsgroei	väestönkasvu	4	12,00	2.33
koudbloedig	lisko	16	10,35	2.07
diepgang	syvyyttä	42	10,53	1.73
geheim	salaisuus	4450	6,46	2.33
spannend	jännittävä	908	7,84	2.47
ijdel	turhamainen	122	10,84	1.8
evenwicht	saldo	305	8,39	2.33
tijd	aika	39289	5,72	1.33
idee	idea	21121	6,71	1.27
competitief	kilpailukykyinen	20	10,46	1.93
geslaagd	ohitettu	614	9,60	1.73
digitaliseren	digitoida	10	12,44	2.47
akoestiek	akustiikka	22	11,44	2.4
opslag	varastointi	369	9,75	2.53
kunst	taide	1622	8,02	2.33
forensisch	oikeuslääketiede	185	13,83	1.73
trots	ylpeys	4874	7,34	1.8
schattig	söpö	1254	6,59	2.47
waarschuwing	varoitus	955	8,22	2.47
verdediging	puolustus	1009	8,44	2.67
straf	rangaistus	1550	4,84	2.93
formeel	muodollinen	190	11,72	1.73
sterk	vahva	4164	5,03	2.73
wetenschap	tiede	1019	10,39	1.73
samenhang	johdonmukaisuus	37	10,45	1.87
waarde	arvo	1121	7,96	1.73
klassiek	klassiko	192	9,39	1.8
controle	tarkistaa	3145	7,96	2.33
onderzoek	tutkimusta	4805	8,78	2.33
veilig	turvallinen	7649	6,71	2.07
vrij	vapaa	10185	5,53	1.73
onderhoud	huolto	299	9,50	2.4
verraden	pettää	1739	8,35	2.33
enorm	valtava	1948	7,46	1.8

Table A2

Used Pictures and Corresponding Scores

Mem	HitRate_Mem	NonMem	HitRate_NonMem	NLWoord	FINWoord	Diff_Mem
apple_01b.jpg	0.95	apple_07s.jpg	0.66	gezond	terveitä	0.29
space_shuttle_10s.jpg	0.88	space_shuttle_12s.jpg	0.68	futurisme	futurismi	0.2
dog_10s.jpg	0.9	dog_01b.jpg	0.68	loyaal	uskollinen	0.22
yacht_04s.jpg	0.9	yacht_09s.jpg	0.6	rijkdom	varallisuus	0.3
wand_08s.jpg	0.89	wand_01s.jpg	0.66	magie	taika	0.23
medal_08s.jpg	0.9	medal_02s.jpg	0.71	eer	kunnia	0.19
baby_19s.jpg	0.95	baby_06s.jpg	0.73	bevolkingsgroei	väestönkasvu	0.22
lizard_08n.jpg	0.93	lizard_04n.jpg	0.68	koudbloedig	lisko	0.25
scuba_05s.jpg	0.87	scuba_04s.jpg	0.58	diepgang	syvyyttä	0.3
padlock_04n.jpg	0.92	padlock_02n.jpg	0.68	geheim	salaisuus	0.24
roller_coaster_07s.jpg	0.93	roller_coaster_06s.jpg	0.71	spannend	jännittävä	0.22
mirror_02s.jpg	0.85	mirror_01b.jpg	0.63	ijdel	turhamainen	0.22
balance_beam_11s.jpg	0.9	balance_beam_05s.jpg	0.69	evenwicht	saldo	0.21
clock_02n.jpg	0.8	clock_06n.jpg	0.53	tijd	aika	0.27
lightbulb_04s.jpg	0.88	lightbulb_08s.jpg	0.62	idee	idea	0.26
chessboard_11s.jpg	0.88	chessboard_09s.jpg	0.63	competitief	kilpailukykyinen	0.25
champagne_10s.jpg	0.9	champagne_05s.jpg	0.55	geslaagd	ohitettu	0.35
computer_03n.jpg	0.93	computer_10n.jpg	0.7	digitaliseren	digitoida	0.23
guitar_10s.jpg	0.9	guitar_02s.jpg	0.63	akoestiek	akustiikka	0.27
hard_disk_08s.jpg	0.88	hard_disk_06s.jpg	0.67	opslag	varastointi	0.21
painting_04s.jpg	0.89	painting_09s.jpg	0.62	kunst	taide	0.27
magnifying_glass_11s.jpg	0.85	magnifying_glass_07s.jpg	0.6	forensisch	oikeuslääketiede	0.26
peacock_12s.jpg	0.93	peacock_03s.jpg	0.71	trots	ylpeys	0.22
puppy_14s.jpg	0.95	puppy_09s.jpg	0.71	schattig	söpö	0.25
smoke_alarm_09s.jpg	0.89	smoke_alarm_05s.jpg	0.65	waarschuwing	varoitus	0.24
shield_09s.jpg	0.92	shield_03n.jpg	0.67	verdediging	puolustus	0.26
handcuff_12s.jpg	0.93	handcuff_08s.jpg	0.69	straf	rangaistus	0.23
tuxedo_02s.jpg	0.9	tuxedo_12s.jpg	0.65	formeel	muodollinen	0.25
dumbbell_12s.jpg	0.95	dumbbell_13s.jpg	0.73	sterk	vahva	0.22
lab_coat_11s.jpg	0.92	lab_coat_15s.jpg	0.71	wetenschap	tiede	0.21
jigsaw_puzzle_02s.jpg	0.93	jigsaw_puzzle_14s.jpg	0.7	samenhang	johdonmukaisuus	0.23
gold_04s.jpg	0.98	gold_11s.jpg	0.66	waarde	arvo	0.32
violin_01b.jpg	0.95	violin_07n.jpg	0.63	klassiek	klassiko	0.32
gate_04s.jpg	0.88	gate_12s.jpg	0.65	controle	tarkistaa	0.23
stethoscope_02s.jpg	0.9	stethoscope_12s.jpg	0.64	onderzoek	tutkimusta	0.26
seatbelt_08s.jpg	0.93	seatbelt_09s.jpg	0.69	veilig	turvallinen	0.23
bird_20s.jpg	0.85	bird_04s.jpg	0.63	vrij	vapaa	0.23
screwdriver_09s.jpg	0.88	screwdriver_02s.jpg	0.69	onderhoud	huolto	0.18
mole_07s.jpg	0.93	mole_15s.jpg	0.68	verraden	pettää	0.24
whale_08s.jpg	0.92	whale_13s.jpg	0.64	enorm	valtava	0.28

Appendix B

Figure B1

Distribution of the Model Residuals (Image Presence, Memorability, Retention Time on Accuracy)

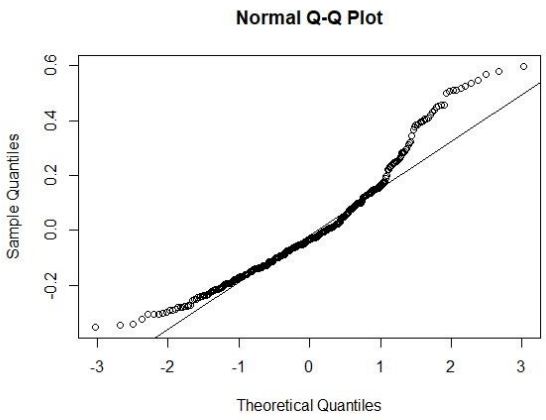

Figure B2

Scatterplot of Relation between Encoding Time and Short-Term Accuracy

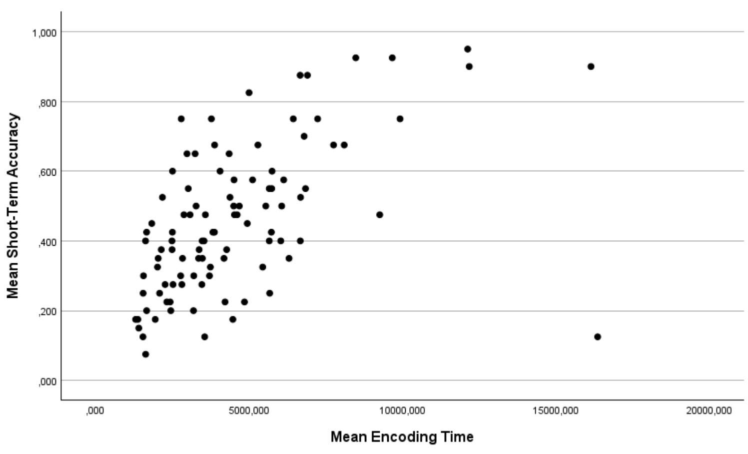

Figure B3

Scatterplot of Relation between Encoding Time and Long-Term Accuracy

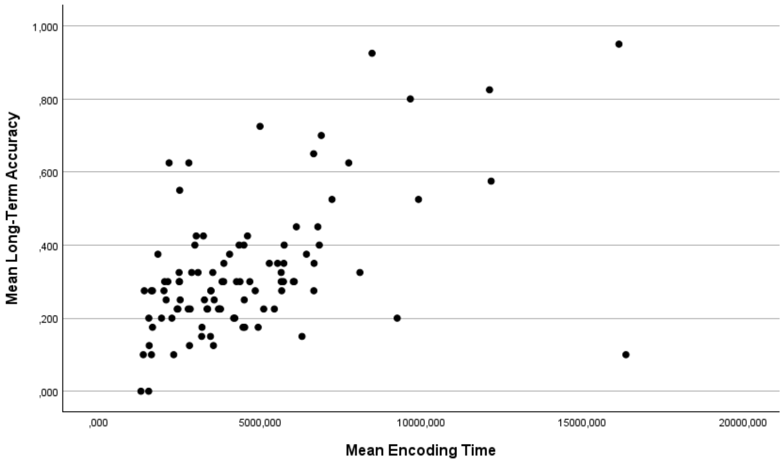

Figure B4

Distribution of the Model Residuals (Image Presence, Memorability on Encoding Time)

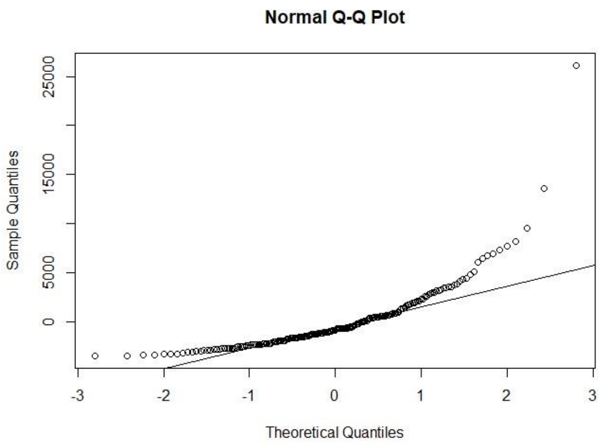

Figure B5

Scatterplot of the Relation between Word Difficulty and Image Presence

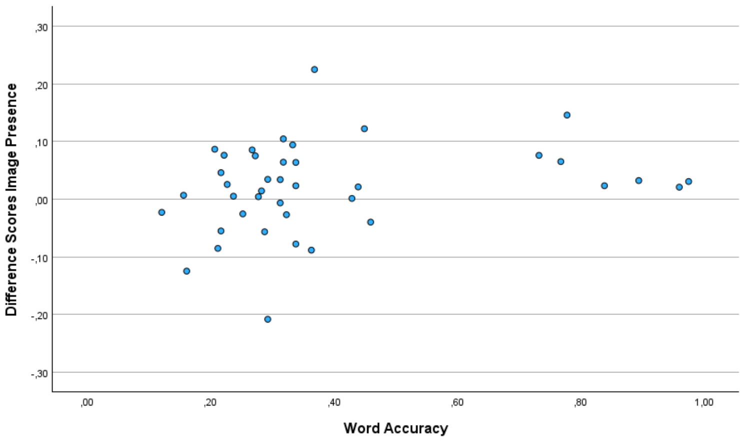

Figure B6

Scatterplot of the Relation between Word Difficulty and Memorability