Project Title	**:**	Multimedia annotations for vocabulary acquisition: Modes, modalities, application, effects and pedagogical implications
Grantee	**:**	The Education University of Hong Kong
Principal Investigator	**:**	ZOU Di Department of English Language Education The Education University of Hong Kong
Co-investigators	**:**	HUANG Gwo-jen Graduate Institute of Digital Learning and Education National Taiwan University of Science and Technology Paul STAPLETON Department of English Language Education The Education University of Hong Kong WANG Fu-lee School of Science and Technology The Open University of Hong Kong XIE Haoran Department of Computing and Decision Sciences Lingnan University

Final Report

by

Principal Investigator

Final report for the project - *Multimedia annotations for vocabulary acquisition: modes, modalities, application, effects and pedagogical implications*

Abstract

This project examined multimedia annotation-enhanced vocabulary learning. A mobile application was developed, which included 12 tasks (four reading comprehension tasks, four cloze exercises, and four sentence-writing tasks). The four reading comprehension tasks were basically the same except that different annotations were provided for the target vocabulary, which were textual, pictural, graphics interchange format (GIF), and video annotations. Similarly, the four cloze exercises and the four sentence-writing tasks were only different in terms of the annotation types. A total of 500 working adults in Hong Kong participated in the study, among whom, 20 explored using the application for language learning and reported their learning experiences and perceptions afterwards. Another 360 participants were post-tested immediately and one week later to measure their immediate learning and retention of the target words. The remaining 120 participants were interviewed to investigate their thinking processes. The results showed statistically significant interaction effects between the task type and the annotation type. Concerning the main effects of the task type, sentence writing was significantly more effective than cloze, which was significantly more effective than reading comprehension. For the main effects of the annotation type, pictural annotations were similarly effective as GIF annotations, which were significantly more effective than video annotations. Textual annotations were significantly less effective than the three multimedia annotation types. Most participants showed positive attitudes towards multimedia annotation-enhanced

vocabulary learning, considering it interesting and effective. They also suggested that provision of diversified multimedia annotations and word-focused learning tasks for learner selection was conducive to active learning.

Keywords

multimedia annotation; vocabulary learning; mobile application; technique feature analysis; involvement load hypothesis

1. Introduction

The advent of computer- and mobile-assisted language learning has created great opportunities for adopting multimedia learning in support of vocabulary acquisition (Mohsen & Balakumar, 2011). As an essential part of word learning in multimodality, multimedia annotations, also known as multimedia glosses, have been examined in several studies (Chun, 2006). Multimedia annotations explain words with short definitions or notes with different modalities and modes such as "text, picture, video, and sound" (Chun & Plass, 1996, p. 183). They are considered superior to traditional annotations in enhancing comprehension and learning of target words, meeting the needs and preferences of learners, and making better use of authentic materials (Abraham, 2008). Multimedia annotations are effective in promoting word learning because, according to the principle of visual memory, images are more likely to be remembered than words, and consequently words that are strongly associated with images can be learned better (Akbulut, 2007). Also, words with multimedia annotations are significantly better noticed and recognized (Yanguas, 2009); and multimedia annotations enhance retention as they provide learners with multiple access routes to the word and

strengthen a deep memory trace (Al-Seghayer, 2005; Yoshii, 2006).

However, the main foci of most previous studies on multimedia annotations are static pictorial annotations, as they can be easily integrated into paper-based language instructional materials; whereas very little research has been conducted to investigate dynamic annotation formats such as GIF and videos. This is perhaps because dynamic annotation formats were not widely applied in common language instructional materials (i.e., paper-based materials) as they can only be displayed on digital devices (e.g., computers, tablets, and smart phones). Nevertheless, with the development of educational technologies and the popular use of digital devices in language education, this situation has been changed now. Dynamic multimedia annotations are commonly integrated into online learning resources, e-learning systems, and mobile word-learning applications (apps) nowadays. This project therefore conducted a comprehensive study on diversified types of annotations and the use of them in different word-focused vocabulary learning tasks. The main objectives are listed as follows:

1) To investigate the effects of textual, pictural, GIF, and video annotations on immediate learning and retention of target vocabulary.
2) To examine the effectiveness of multimedia annotation-enhanced tasks in promoting word learning.
3) To study the interaction effects between the type of annotations and the type of tasks on vocabulary learning.
4) To inspect the perceptions and preferences of Hong Kong working adults for different multimedia annotations.
5) To discuss appropriate pedagogies to be used with multimedia instructional materials for

vocabulary acquisition.

6) To suggest methods for material writers and teachers in developing multimedia annotations and integrating them into instructional materials, and for language learners in selecting appropriate materials and activities.

2. Review of literature of the project

2.1 Multimedia annotations

Mayer and Moreno (2003) defined multimedia learning as learning from words (either printed or spoken) and pictures (either static as charts, graphs, diagrams, and illustrations, or animated as interactive animations and videos). Numerous studies have been conducted to investigate how implementation of different multimedia principles can affect abilities of learners in developing meaningful learning and what principles are most effective in guiding the design of multimedia-enhanced instructional materials. Bull (2013) found that instructional multimedia pieces with words and pictures were more effective than using words alone. Sorden (2012) suggested that multimedia pieces are better designed for learners when words and pictures are placed near each other. Höffler and Leutner (2007) noted an advantage in using animations over static pictures, particularly when the animation is directly connected to the topic or lesson being addressed. Jamet, Gavota, and Quaireau (2008) discovered that color changes and flashing for cueing are facilitative for retention, transfer, and text-image matching tasks. Similarly, Scheiter, Schüler, Gerjets, Huk, and Hesse (2014) observed that adding animation to verbal explanations helped learners recall immediate information but did not aid in the transfer part of the experiment. Moreover, a static picture can be more effective in certain situations. Pictures must be easy to understand, have limited text, and relate directly to the main

objective being taught. Several researchers (e.g., Oud, 2009; Sage, Bonacorsi, Izzo, & Quirk, 2015) also found that learning is more meaningful and effective, and cognitive overload can be prevented when learners have control over the pace of multimedia. Grouping lessons into smaller segments provides learners the control to focus their attention on the aspects of the lesson that they prefer (Tabbers & de Koeijer, 2010).

The previous studies generally argued that multimedia input helped L2 learners achieve deep processing of information (Winke, Gass, & Syodorenko, 2010), elongated retention of knowledge (Mayer, Lee, & Peebles, 2014), promoted motivation (Chen, Wang, Zou, Lin, & Xie, 2019), and enhanced capability of applying knowledge in authentic contexts (Syodorenko, 2010). However, conflicting results regarding the effectiveness of different multimedia input modes have been reported as well (Mohsen & Balakumar, 2011; Xu, 2010). Some researchers found that pure text was more effective than text plus picture in promoting vocabulary learning (e.g., Acha, 2009; Boers, Warren, He, & Deconinck, 2017), while others identified the opposite results (e.g., Bisson, van Heuven, Conklin, & Tunney, 2015; Warren, Boers, Grimshaw, & Siyanova-Chanturia, 2018). Specifically, based on the data collected through eye-tracking technology about the participants' attention to multimodal stimuli, Bisson et al. (2015) found that the participants spent remarkably more time looking at pictures than texts. They also paid more attention to word knowledge presented in text plus picture than that in text and audio, which resulted in better learning outcomes. Conflicting results as such inundated the field of multimedia-enhanced language learning and may have resulted in educators' and researchers' confusion and hesitation in selecting and using multimedia in L2 learning, thus it is necessary to conduct a comprehensive research on diversified multimedia annotations and investigate the

interaction effects between the annotation type and the task type.

2.2 Word-focused tasks

Among various word-focused tasks, reading comprehension, cloze, and sentence writing are three of the most practiced exercises in and out of language classrooms (Zou, 2017). The current literature generally agreed that sentence writing is more effective than cloze, which is more effective than reading comprehension in promoting vocabulary learning (Zou, 2017). The involvement load hypothesis argues that tasks with higher involvement load lead to more effective word learning, and the involvement load of a task consists of the need, search, and evaluation (Laufer & Hulstijn, 2001). Need is the drive to learn. It is moderate when it is externally imposed and strong when it is self-driven. Search is the attempt to find the meaning or form of a word; it has only one degree of prominence. Evaluation is another cognitive dimension. Moderate evaluation involves comparison of meanings or forms of words, and strong evaluation involves generation of original contexts using target words (Laufer & Hulstijn, 2001). According to the involvement load hypothesis, reading comprehension with marginal textual annotations induces moderate need, no search, and no evaluation. Cloze with marginal textual annotations involves moderate need, no search, and moderate evaluation. Sentence writing with marginal textual annotations involves moderate need, no search, and strong evaluation. Thus, the involvement load hypothesis anticipates that sentence writing is most effective, followed by cloze, and reading comprehension is the least effective among the three. The results of previous empirical studies on task-based vocabulary learning basically support the assumptions of the involvement load hypothesis.

However, most previous studies on task-based language learning with annotations apply

textual annotations or do not provide annotations, so it is uncertain whether the effectiveness of these tasks is influenced by the type of annotations that are involved in them. This project therefore aims to investigate the interaction effects between the task type and the annotation type. Specifically, it examines whether textual annotation-enhanced tasks with higher involvement load lead to more effective word learning than multimedia annotation-enhanced tasks with lower involvement load.

3. Theoretical framework of the project

The cognitive theory of multimedia learning (Mayer, 2014; Moreno & Mayer, 2002) is one of the most notable theories in the field of multimedia-enhanced language learning (Liu, Jang, & Roy-Campbell, 2018). It was based on three assumptions. First, the dual-coding theory (Paivio, 1991) argued that language learners process different multimedia input through two channels of sensory memory, auditory input (i.e., audio) through an auditory channel and visual input (i.e., text, picture, animation, and captions/subtitles) through a visual channel. Moreover, the generative theory argued that the process of learning consists of active selection of auditory and visual information from the sensory memory, coding of the information into auditory and visual representations, organization of the representations into verbal and pictorial modes in the working memory, and integration of the newly-learned knowledge with the prior one (Mayer, 2001). Meaningful learning occurs when learners obtain a deep understanding by selecting, organizing, and integrating newly presented information with prior knowledge (Tempelman-Kluit, 2006). Meaningful and deep learning consequently depends on the cognitive capacities of learners in selecting relevant information, organizing it into coherent representations, and integrating it with other knowledge (Mayer, 2008). However, the cognitive

load theory argued that the working memory and cognitive channels are limited in capacity and could be overloaded when there is redundant input (Sweller, 1988). Cognitive overload occurs when the intended cognitive processing of a learners exceeds the learner's available cognitive capacity (Mayer & Moreno, 2003). Therefore, how to introduce new, engaging information without causing cognitive overload is an important topic for multimedia designers and educators.

Based on those three assumptions, the cognitive theory hypothesised a series of principles, among which three were frequently cited (e.g., in Boers et al., 2017; Chen et al., 2019b; Gruhn, Segers, & Verhoeven, 2019; Lee & Mayer, 2018b; Liu et al., 2018). First, the modality principle argued that when the instructional content is presented through visual and auditory input simultaneously, students have both auditory and visual channels stimulated, so they can establish both auditory and visual representations of target knowledge, construct cognitive connections between them, integrate the connected representations into their long-term memory, and achieve high learning efficiency (Moreno & Mayer, 2002). However, the redundancy principle argued that when redundant multimedia input is imposed on a single channel, learners would have cognitive overload and split attention between different input, so they may have difficulty in fully processing information and thus reduced learning efficiency (Moreno & Mayer, 2002). Third, the coherence principle argued that when there is extraneous, irrelevant, or unneeded multimedia input, learners would have distraction from learning, disrupted knowledge processing, and thus reduced learning efficiency (Mayer, 2001; Mayer & Moreno, 2003).

Based on the aforementioned theories and review of relevant literature, the basic

hypothesis of this project include that (1) when the multimedia annotations are well designed and integrated into the learning tasks, multimedia annotation-enhanced tasks with low involvement load are not necessarily less effective than textual annotation-enhanced tasks with high involvement load; and (2) interaction effects exist between the type of multimedia annotations and the types of word-focused tasks in promoting vocabulary learning.

4. Methodology

4.1 Research design

The general design of the research is illustrated in Figure 1. A total of 500 working adults in Hong Kong participated in the project, 20 of whom were interviewed to examine their thoughts of multimedia annotation-enhanced word learning and the vocabulary learning app. The remaining 480 subjects were randomly assigned to complete 12 tasks (see Table 1). The 12 tasks involved three types of exercises (i.e., reading comprehension, cloze, and sentence writing) and four types of annotations (i.e., textual, pictorial, GIF, and video annotations). The three exercises were selected as they were among the most frequently practiced word learning activates; and almost all English learners had experience practicing them. The four types of annotations were different in that pictorial annotations involved images that depicted meanings of words to assist textual definitions in explaining word meanings, GIF annotations involved animations, and video annotations involved short videos. To complete these tasks, the participants needed approximately 30 minutes on average.

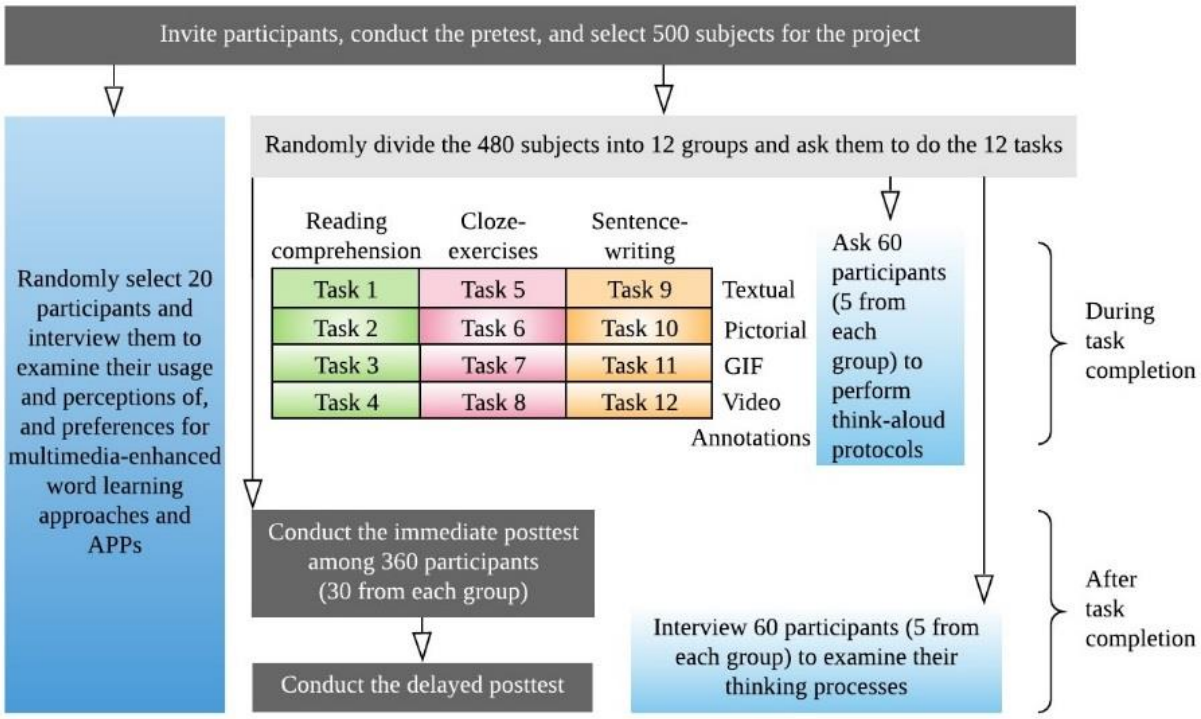

Figure 1: Research design

Table 1. Allocation of participants to tasks

	Think-aloud	Interviews	Posttests
T1: Reading comprehension with textual annotations	5	5	30
T2: Reading comprehension with pictorial annotations	5	5	30
T3: Reading comprehension with GIF annotations	5	5	30
T4: Reading comprehension with video annotations	5	5	30
T5: Cloze exercises with textual annotations	5	5	30
T6: Cloze exercises with pictorial annotations	5	5	30
T7: Cloze exercises with GIF annotations	5	5	30
T8: Cloze exercises with video annotations	5	5	30
T9: Sentence writing with textual annotations	5	5	30
T10: Sentence writing with pictorial annotations	5	5	30
T11: Sentence writing with GIF annotations	5	5	30
T12: Sentence writing with video annotations	5	5	30

4.2 Participants

Working adults in Hong Kong were invited to participate in the project, and the following criteria were used while inviting the participants: (1) the numbers of female and male

participants were balanced; (2) their ages ranged from 20 to 50; (3) they were EFL learners; and (4) their self-reported English proficiency levels were intermediate. This sample represented a sizable proportion of EFL learners in Hong Kong.

A pretest was also conducted among the invited participants one month before the treatment. Only those who were evaluated as intermediate learners and knew very little about the target words were selected as the subjects of the research. The revised version of Nation's Vocabulary Levels Test by Schmitt, Schmitt and Claphan (2001) was used as the measurement tool. This test had been widely regarded as a very reliable assessment tool that can effectively measure learners' vocabulary sizes and further reflect their English proficiency levels (Nation, 2001). Meara and Milton (2003) also found that learners' vocabulary sizes were closely associated with their CEF (Common European Framework) levels as defined by the British Council, and a learner who knew around 3000 words was regarded as an intermediate learner.

4.3 Materials

The main research materials included a reading text, 10 target words, and associated multimedia annotations. The text was about a story of "A scary night", which was adapted from Yoshii's (2006) research material. It was selected as its length and difficulty level were appropriate for intermediate EFL learners, and the topic seemed interesting. The 10 words were "burglarize," "dash," "grin," "inflammation," "rake," "scribble," "shatter," "shiver," "tumble," and "wrath". They were selected as they were tangible to the senses and could be easily imagined. Moreover, according to Davis's (2012) Corpus of Contemporary American English (COCA), they were out of the 6000 most frequently used words, thus were unlikely to be unknown to the participants who were intermediate English learners and knew approximately

3000 words.

The textual annotations of the target words were created by presenting the basic and essential dictionary definitions of the words. The pictorial, GIF, and video annotations were developed by the researchers through three stages. At the first stage, the researchers consulted the target words in dictionaries and read the text to understand the meanings of the words in the context. Then the researchers created short videos to present the meanings; based on the videos, pictures were selected, and GIFs were generated. The main differences between the videos and GIFs are that the videos were longer, and the GIFs included no audio explanation of the target words. A pilot study had been conducted to examine the appropriateness of the pictorial, GIF, and video annotations, positive results of which had been identified.

The same scoring system of Zou (2016, 2017) was used in this project to evaluate the learning of the target words of the participants in the post-tests. "A meaning was graded zero if it is completely incorrect, a half score if it is a semantically acceptable equivalent of the target word, and a full score if it is a comparable meaning to that of the target word. A sentence was graded zero if it has a completely inappropriate semantic context for the target word, a half score if it has an appropriate semantic context but the target word is used ungrammatically, and a full score if it has an appropriate semantic context and the target word is used grammatically" (Zou, 2017, p. 59). Blind scoring was employed using two trained raters who scored the answers separately.

4.4 The multimedia annotation-enhanced vocabulary learning app

A mobile app was developed to promote multimedia annotation-enhanced vocabulary learning for this project. The interfaces of the app are presented in the following figures. All 12

tasks can be easily accessed using this app, and learners can select different types of tasks and annotations freely. The pictural, GIF, and video annotations of this project also include text definitions of the target words.

Figure 2. System interfaces

Figure 3. Task interfaces

Figure 4. Annotation interfaces

5. Data collection and analysis

To measure the effectiveness of the 12 tasks, 30 participants from each group (see Table 1) were tested immediately after task completion and one week later. The two post-tests were the same; the participants were asked to provide English synonyms or definitions of the target words and generate original sentences using them. Blind scoring was employed using two trained raters who scored the answers separately. After marking the post-tests, all scores were entered into SPSS. Statistical tests such as one-way ANCOVAs and two-way ANCOVAs were conducted to investigate the effectiveness of different tasks in promoting initial learning and retention of the target words and examine whether interaction effects exist between the annotation type and the task type.

To investigate the thinking processes induced by the 12 tasks and triangulate the collected data, 60 participants (five from each group) were trained to perform think-aloud protocols while completing the respective tasks, and another 60 (also five from each group) were interviewed afterward. These subjects may self-report in either English (the target language) or Chinese (their first language), as desired. Their initial learning of the target words was measured by the immediate post-test, and they were also asked with questions concerning why they were (not) able to recall the meanings of the words and generate sentences using them, and what they thought about the learning tasks and annotations. They did not participate in the delayed post-test, and their test data in the immediate post-test were discarded from the statistical analyses to reduce possible effects of engagement in self-reporting on the following test performance.

The audio data obtained from self-reporting were transcribed into Microsoft Word. The

transcriptions were in both English and Chinese because, as explained previously, the subjects performed self-reporting in either of the language that they prefer. After obtaining a holistic view of all the data through the transcription stage, the investigators coded the data following Zou (2016, 2017). All data were initially skimmed to note certain distinct facilitative elements for word learning, the relevant literature of which were then reviewed to identify theories for explanations. Subsequently, re-examination of the data was conducted, focusing on the participants' thinking processes while doing the assigned tasks with different annotations. The thinking patterns of the subjects were identified by analyzing their thinking processes and noting repetitive expressions with similar contexts (Zou, 2017).

6. Results and Discussion

6.1 Effectiveness of the tasks in promoting vocabulary learning

The participants' scores of the immediate and delayed posttests are presented in Table 2. The results showed that the twelve tasks were effective in promoting immediate learning and retention of the target vocabulary, considering that the participants had almost zero prior knowledge of the target words before the treatment. For better illustration, Figure 5 presents the estimated marginal means of the scores in the immediate and delayed posttests with the participants' scores in the pretest as the covariate.

Two one-way ANCOVAs were conducted to test whether statistically significant differences existed among the 12 tasks in promoting immediate learning and retention of the target vocabulary, after controlling for the participants' pretest scores. One investigated the participants' scores in the immediate posttest, the other their delayed posttest. The data could be analyzed using one-way ANCOVAs as they passed the basic assumptions, including normal

distribution, homogeneity of regression slopes, and homogeneity of variance.

The results of the univariate tests indicated significantly different effectiveness of the 12 tasks, specifically, $F(11, 347) = 84.56$, $p < .001$, *partial* $\eta^2 = .72$ for the immediate posttest scores, and $F(11, 347) = 27.20$, $p < .001$, *partial* $\eta^2 = .46$ for the delayed posttest scores.

Table 2. Summary of descriptive statistics of the participants' learning performance

		N	Immediate posttest		Delayed posttest	
Task type	Annotation type		M	SD	M	SD
Reading comprehension	Text	30	11.23	1.22	8.90	2.35
	Picture	30	14.23	1.30	11.80	2.83
	GIF	30	14.23	1.25	12.70	2.10
	Video	30	13.66	1.24	11.90	2.27
	Total	120	13.34	1.75	11.32	2.78
Cloze	Text	30	13.96	1.29	11.86	2.78
	Picture	30	17.53	1.27	15.53	1.59
	GIF	30	17.26	1.33	15.20	1.95
	Video	30	15.66	1.24	14.03	2.07
	Total	120	16.10	1.91	14.15	2.56
Sentence writing	Text	30	17.20	1.27	14.80	2.09
	Picture	30	17.60	1.24	15.50	2.02
	GIF	30	17.73	1.25	15.73	2.27
	Video	30	17.36	1.27	14.83	2.10
	Total	120	17.47	1.26	15.21	2.13
Total	Text	90	14.13	2.75	11.85	3.41
	Picture	90	16.45	2.02	14.27	2.80
	GIF	90	16.41	2.01	14.54	2.47
	Video	90	15.56	1.96	13.58	2.46
	Total	360	15.64	2.39	13.56	2.99

Table 3. Univariate tests

		SS	*df*	MS	*F*	*Sig.*	η^2
Immediate posttest	Contrast	1498.87	11	136.26	84.56	.00	.72
	Error	559.10	347	1.61			
Delayed posttest	Contrast	1490.39	11	135.49	27.20	.00	.46
	Error	559.10	347	1.61			

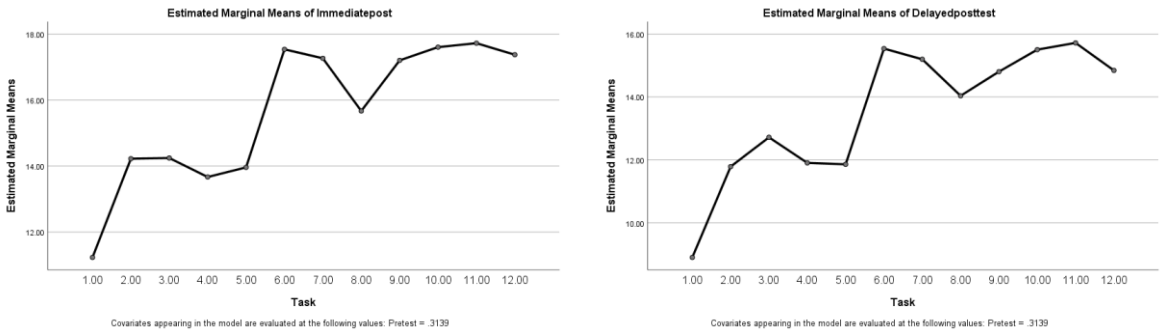

Figure 5. Estimated marginal means of the scores in the immediate and delayed posttests

The results of the pairwise comparisons also showed that, concerning the effectiveness of promoting immediate learning of the target vocabulary,

(1) Task 1 (Reading comprehension with textual annotations) was significantly less effective than other tasks;

(2) Task 2 (Reading comprehension with pictorial annotations), Task 3 (Reading comprehension with GIF annotations), Task 4 (Reading comprehension with video annotations), and Task 5 (Cloze exercises with textual annotations) were similarly effective;

(3) Task 8 (Cloze exercises with video annotations) was more effective than Task 2, 3, 4, and 5;

(4) Task 6 (Cloze exercises with pictorial annotations), Task 7 (Cloze exercises with GIF annotations), Task 9 (Sentence writing with textual annotations), Task 10 (Sentence writing with pictorial annotations), Task 11 (Sentence writing with GIF annotations), and Task 12 (Sentence writing with video annotations) were similarly effective and significantly more effective than other tasks.

Concerning the effectiveness of these 12 tasks in promoting retention of the target vocabulary, the results of the pairwise comparisons suggested that

(1) Task 1 was significantly less effective than other tasks;

(2) Task 2, 3, 4, and 5 were similarly effective;

(3) Task 6, 7, 8, 9, 10, 11, and 12 were similarly effective and significantly more effective than other tasks;

(4) Task 3 and Task 8 were not significantly different.

6.2 Interaction effect between task types and annotation types in the immediate posttest

Two two-way ANCOVAs were conducted to test whether there was an interaction effect between the task type and the annotation type, after controlling for the participants' pretest scores. The data could be analyzed using two-way ANCOVAs as they passed the basic assumptions, including normal distribution, homogeneity of regression slopes, and homogeneity of variance.

The data of the participants' scores in the immediate posttest were analyzed first. The results of the tests of between-subjects effects indicated a statistically significant interaction between the task type and the annotation type, whilst controlling for their prior knowledge of the target vocabulary, $F(6, 347) = 12.04$, $p < .001$, *partial* $\eta^2 = .17$ (see Table 4).

The main effects of the two independent variables were then examined. As shown in Table 5, the main effects of the task type were significant, $F = 330.22$, $p < .001$, *partial* $\eta^2 = .65$. Also, the main effects of the annotation type were significant, $F = 65.84$, $p < .001$, *partial* $\eta^2 = .36$ (see Table 6). Specifically, as demonstrated in Table 7, the mean differences among the three types of tasks were all statistically significant. Such results indicated that sentence writing was significantly more effective than cloze, which was then significantly more effective than reading comprehension. Table 8 presents the pairwise comparisons of the annotation types, the

results of which indicated that picture and GIF were similarly effective, and they were significantly more effective than video, which was then significantly more effective than text.

Table 4. Results of the tests of between-subjects effects (the immediate posttest)

Source	Type III SS	*df*	MS	*F*	*Sig.*	η^2
Corrected Model	1499.66[a]	12	124.97	77.56	.00	.72
Intercept	63639.81	1	63639.81	39496.82	.00	.99
Pretest	1.45	1	1.45	.90	.34	.00
Task type	1064.14	2	532.07	330.22	.00	.65
Annotation type	318.28	3	106.09	65.84	.00	.36
Task type * Annotation type	116.44	6	19.40	12.04	.00	.17
Error	559.10	347	1.61			
Total	90137.00	360				
Corrected Total	2058.77	359				

[a]. R Squared = .72 (Adjusted R Squared = .71)

Table 5. Main effects of the *task type* on learning performance (the immediate posttest)

	SS	*df*	MS	*F*	Sig.	η^2
Contrast	1064.14	2	532.07	330.22	.00	.65
Error	559.10	347	1.61			

Table 6. Main effects of the *annotation type* on learning performance (the immediate posttest)

	SS	*df*	MS	*F*	Sig.	η^2
Contrast	318.28	3	106.09	65.84	.00	.36
Error	559.10	347	1.61			

Table 7. Results of the pairwise comparisons of the *task type* (the immediate posttest)

(I) Task type	(J) Task type	MD (I-J)	SE	Sig.[b]
Reading comprehension	Cloze	-2.76*	.16	.00
	Sentence writing	-4.13*	.16	.00
Cloze	Reading comprehension	2.76*	.16	.00
	Sentence writing	-1.36*	.16	.00
Sentence writing	Reading comprehension	4.13*	.16	.00
	Cloze	1.36*	.16	.00

*. The mean difference is significant at the .05 level.

[b]. Adjustment for multiple comparisons: Bonferroni.

Table 8. Results of the pairwise comparisons of the *annotation type* (the immediate posttest)

(I) Annotation type	(J) Annotation type	MD (I-J)	SE	Sig.[b]
Text	Picture	-2.32*	.18	.00
	GIF	-2.27*	.18	.00
	Video	-1.43*	.18	.00
Picture	Text	2.32*	.18	.00
	GIF	.04	.18	1.00
	Video	.88*	.18	.00
GIF	Text	2.27*	.18	.00
	Picture	-.04	.18	1.00
	Video	.84*	.18	.00

*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

6.3 Interaction effect between task types and annotation types in the delayed posttest

The data of the participants' scores in the delayed posttest were then analyzed. Table 9 demonstrates the results of the tests of between-subjects effects, indicating a statistically significant interaction effect between the task type and the annotation type on the participants' retention, whilst controlling for their prior knowledge of the target vocabulary, $F(6, 347) = 4.13$, $p < .001$, *partial* $\eta^2 = .06$.

Table 9. Results of the tests of between-subjects effects (the delayed posttest)

Source	Type III SS	*df*	MS	*F*	*Sig.*	η^2
Corrected Model	1491.97[a]	12	124.33	24.96	.00	.46
Intercept	48041.47	1	48041.47	9644.86	.00	.96
Pretest	3.11	1	3.11	.62	.43	.00
Task type	971.49	2	485.74	97.51	.00	.36
Annotation type	395.49	3	131.83	26.46	.00	.18
Task type * Annotation type	123.43	6	20.57	4.13	.00	.06
Error	1728.42	347	4.98			
Total	69480.00	360				
Corrected Total	3220.40	359				

[a]. R Squared = .46 (Adjusted R Squared = .44)

Table 10. Main effects of *task type* on learning performance (the delayed posttest)

	SS	df	MS	F	Sig.	*η2*
Contrast	971.49	2	485.74	97.51	.00	.36
Error	1728.42	347	4.98			

Table 11. Main effects of *annotation type* on learning performance (the delayed posttest)

	SS	df	MS	F	Sig.	*η2*
Contrast	395.49	3	131.83	26.46	.00	.18
Error	1728.42	347	4.98			

Table 12. Results of the pairwise comparisons of *task type* (the delayed posttest)

(I) Task type	(J) Task type	MD (I-J)	SE	Sig.[b]
Reading comprehension	Cloze	-2.83*	.28	.00
	Sentence writing	-3.89*	.28	.00
Cloze	Reading comprehension	2.83*	.28	.00
	Sentence writing	-1.06*	.28	.00
Sentence writing	Reading comprehension	3.89*	.28	.00
	Cloze	1.06*	.28	.00

*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

Table 13. Results of the pairwise comparisons of *annotation type* (the delayed posttest)

(I) Annotation type	(J) Annotation type	MD (I-J)	SE	Sig.[b]
Text	Picture	-2.42*	.33	.00
	GIF	-2.69*	.33	.00
	Video	-1.74*	.33	.00
Picture	Text	2.42*	.33	.00
	GIF	-.26	.33	1.00
	Video	.68	.33	.25
GIF	Text	2.69*	.33	.00
	Picture	.26	.33	1.00
	Video	.94*	.33	.02

*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

The main effects of the two independent variables were then examined. As shown in Table 10, the main effects of the task type were significant, $F = 97.51$, $p < .001$, *partial η2* = .36. Also, the main effects of the annotation type were significant, $F = 26.46$, $p < .001$, *partial η2* = .18

(see Table 11).

Specifically, as demonstrated in Table 12, the mean differences among the three types of tasks were all statistically significant. Such results were the same as those of the immediate posttest. Table 13 presents the results of the pairwise comparisons of the annotation types, which were different from those of the immediate posttest. Three findings concerning the main effects of annotation types on learners' retention of target vocabulary were indicated, (1) text annotations were significantly less effective than other multimedia annotations; (2) picture annotations were similarly effective as GIF and video annotations; and (3) GIF annotations were significantly more effective than video annotations.

6.4 Interview and think-aloud results

The interview and think-aloud results of this project showed that the participants had overall positive attitudes towards multimedia annotation-enhanced vocabulary learning. They also considered the app that was developed for this project easy-to-use and helpful for word learning. Concerning the different types of annotations, most participants felt that pictural and GIF annotations were more effective than text and video annotations. Text annotations were less vivid and more boring; while video annotations were to some extent tedious as the videos lasted several seconds longer than what was necessary for learners to understand the vocabulary meanings. Many participants also commented that GIF annotations were more efficient than video annotations as they could express the same amount of information in a shorter period. Picture annotations had similar advantages in this respect, but picture annotations were not as effective as GIF annotations in presenting meanings of verbs. Animations could demonstrate the process of doing something, so they were conducive to comprehension and memorization

of meanings of verbs. This is not so important for cloze and sentence writing tasks, as these two types of tasks to some extent involve generative use of target words. The creation of contexts for the target words plays a crucial role in promoting vocabulary learning, thus the advantages of multimedia annotations were somewhat weakened in sentence writing tasks.

6.5 Discussion

The results of this project provided empirical support to the hypothesis that multimedia annotation-enhanced tasks with low involvement load are not necessarily less effective than textual annotation-enhanced tasks with high involvement load. For example, reading comprehension induces lower involvement load than cloze, but the results of this research showed that reading comprehension tasks with picture, GIF, or video annotations were similarly effective as cloze exercises with text annotations. Similarly, although sentence writing tasks induced higher involvement load than cloze exercises, but the integration of picture and GIF annotations into cloze exercises could significantly increase their effectiveness and lead to similarly effective vocabulary learning as sentence writing tasks.

The results also provided empirical support to the other hypothesis that interaction effects exist between the type of multimedia annotations and the type of word-focused tasks in promoting vocabulary learning. Moreover, the main effects of the task type and the annotation type were statistically significant.

However, it is noteworthy that the addition of multimedia to sentence writing tasks did not lead to significantly better learning outcomes, compared to textual annotation-enhanced vocabulary learning. This is perhaps because sentence writing induced generation of original contexts for target vocabulary, and the effects of multimedia were weakened by learners' active

contextualization of target vocabulary. In other words, learner-generated contexts played a dominant role in promoting vocabulary learning, which to some extent diminished the facilitative effects of multimedia annotations.

Moreover, a possible reason for the superiority of picture and GIF annotations over video annotations is that video annotations involved more multimedia input (e.g., voice over) than picture and GIF annotations, and consequently caused redundancy effects.

7. Conclusions and Recommendations

7.1 Recommendations for teachers

Concerning appropriate pedagogies to be used with multimedia instructional materials for vocabulary acquisition, the data of this project indicated three suggestions for teachers.

1) Multimedia annotations, especially picture and GIF annotations, are effective in promoting vocabulary learning, so teachers are advised to make more use of multimedia annotations in language education.
2) Multimedia annotations are effective supplements to reading-based tasks such as reading comprehension and cloze exercises.
3) It is not necessary to supplement writing tasks with multimedia annotations.

7.2 Recommendations for material developers

Concerning methods for material developers in designing multimedia annotations and integrating them into instructional materials, the data of this project indicated five suggestions.

1) Material developers are advised to develop annotations of different types (i.e., textual annotations and annotations that involve both text and images/GIFs/videos) and provide learners with personalized learning environments where they can freely decide and select

what types of annotations they want.

2) It is also suggested that material developers and teachers provide learners with diverse learning tasks so that they can select appropriate approaches to learning that best meet their needs and preferences. The effectiveness of multimedia annotations and personalized learning has been widely acknowledged in the literature (Zou, Xie, & Wang, 2018).
3) The content of the annotations and associated learning materials ought to be accurate. Textual annotations should be developed with reference to the standardized dictionaries; and images/GIFs/videos should be able to precisely depict and express the meanings of the target words. Imprecise information will lead to misunderstanding and should be avoided.
4) The annotations and associated learning materials should present the target words in contexts, as de-contextualization tends to lead to ineffective learning (Chen, Wang, Zou, Lin, & Xie, 2019).
5) The videos should be within several seconds. Our data indicated that learners feel long videos time-consuming and useless, and it is important to present the key messages in a precise and concise way.

7.3 Recommendations for language learners

Concerning methods for language learners in selecting appropriate materials and activities, the data of this project indicated five suggestions.

1) Learners are advised to select annotations that involve both text and images/GIFs/videos, if they are provided with such personalized learning opportunities, as multimedia annotations are significantly more effective than single annotations that involve only text. The literature also supports this suggestion (e.g., Chun & Plass, 1996; Boers et al., 2017;

Jones & Plass, 2002; Turk & Ercetin, 2014, etc.).

2) Learners of verbal learning style tend to learn better with textual annotations, those of aural learning style tend to learn better with video annotations, those of visual learning style tend to learn better with image/GIF annotations; and it is suggested that learners select the annotations based on their learning styles if they are provided with such personalized learning opportunities.

3) Learners of lower language proficiency are likely benefit more from learning with annotations that involve dynamic animations and/or sound, as such annotations can help them better understand the meanings; while learners of higher language proficiency may not need the additional animations and can understand the meanings well without them, so they may find annotations that involve static pictures more conducive to efficiency.

4) Learners are advised to select learning activities that induce higher involvement load (i.e., writing and close exercises), rather than activities with low involvement load (i.e., reading comprehension), as they promote more effective learning. This is also supported by Laufer and Hulstijn's involvement load hypothesis (2001).

5) Learners are advised to consider their cognitive capacities while selecting annotations and learning activities, as cognitive overload leads to ineffective learning although higher involvement load induces more effective learning. That is, it is suggested that learners select simpler annotations and learning activities with lower involvement load, if they feel that they have difficulty processing too much information (i.e., the dynamic animations, the sound, etc.). Our data generally indicated that learners with greater cognitive capacities tended to benefit more from writing exercises with GIF annotations; and reading

comprehension and cloze exercises with image/text annotations are more beneficial for learners with smaller cognitive capacities.

7.4 Limitations of the project and directions for follow-up research

This project is limited in five aspects, and follow-up research of it can be conducted from these perspectives. Firstly, only 10 target words were investigated in this project, and future research may consider examining more target words. Secondly, this project examined four types of annotations, and future studies may include audio annotations and annotations presented in the formats of augmented reality and hyperlinks in the investigation as well. Thirdly, this project focused on three types of tasks only, and future research may investigate the interaction effects between multimedia annotations and other common word-focused learning tasks such as composition writing, matching exercises, and flashcards. Fourthly, several videos of this research are longer than ten seconds, which may be a key factor that influences the effectiveness of video annotations. Future research may investigate videos that are around five seconds and check whether the lengths of videos play an important role in promoting vocabulary learning. Fifthly, this project conducted one delayed posttest one week later, and future research may consider conducting two delayed posttests, with one two weeks later to examine the long-term effects of multimedia annotation-enhanced vocabulary learning.

8. Bibliography

Abraham, L. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, 12(3), 199-226.

Acha, J. (2009). The effectiveness of multimedia programmes in children's vocabulary learning. *British Journal of Educational Technology, 40*, 23–31.

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.

Al-Seghayer, K. (2005). The effect of multimedia annotation modes on L2 vocabulary acquisition. *Research in technology and second language education: Developments and directions*, 3, 133.

Bisson, M. J., van Heuven, W. J., Conklin, K., & Tunney, R. J. (2015). The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. *The Quarterly Journal of Experimental Psychology, 68*(7), 1306-1326.

Boers, F., Warren, P., Grimshaw, G., & Siyanova-Chanturia, A. (2017a). On the benefits of multimodal annotations for vocabulary uptake from reading. *Computer Assisted Language Learning, 30*(7), 709-725.

Bull, P. (2013). Cognitive Constructivist Theory of Multimedia: Designing Teacher-Made Interactive Digital. *Creative Education*, 4, 614-619.

Chen, M. P., Wang, L. C., Zou, D., Lin, S. Y., & Xie, H. (2019). Effects of caption and gender on junior high students' EFL learning from iMap-enhanced contextualized learning. *Computers & Education, 140*, 103602.

Chun, D. (2006). CALL technologies for L2 reading. In: Ducate, L. and Arnold, N. (eds.), *Calling on CALL: From theory and research to new directions in foreign language teaching*. San Marcos, TX: CALICO, 69-98.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, 80(2), 183-198.

Davies, M. (2012). Corpus of Contemporary American English (1990–2012), available at http://corpus.byu.edu/coca/

Gruhn, S., Segers, E., & Verhoeven, L. (2019). Moderating role of reading comprehension in children's word learning with context versus pictures. *Journal of Computer Assisted Learning, 36,* 29-45.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A metaanalysis. *Learning and instruction*, 17(6), 722-738.

Jamet, E., Gavota, M., & Quaireau, C. (2008). Attention guiding in multimedia learning. *Learning and instruction*, 18(2), 135-145.

Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The modern language journal, 86*(4), 546-561.

Laufer, B., & Hulstijn, H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1-26.

Lee, H., & Mayer, R. E. (2015a). Visual aids to learning in a second language: Adding redundant video to an audio lecture. *Applied Cognitive Psychology, 29*(3), 445-454.

Liu, Y., Jang, B. G., & Roy-Campbell, Z. (2018). Optimum input mode in the modality and redundancy principles for university ESL students' multimedia learning. *Computers & Education, 127*, 190-200.

Mayer, R. E. (2008). Applying the science of learning: evidence-based principles for the design

of multimedia instruction. *American Psychologist*, 63(8), 760.

Mayer, R. E. (2014). Multimedia instruction. In *Handbook of research on educational communications and technology* (pp. 375-399). Springer, New York, NY.

Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia learning in a second language: A cognitive load perspective. *Applied Cognitive Psychology, 28*(5), 653-660.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia-enhanced learning. *Educational Psychologist, 37*(1), 43-52.

Meara, P. M., & Milton, J. L. (2003). *X_Lex. The Swansea vocabulary levels test*. Newbury: Express.

Mohsen, M. A., & Balakumar, M. (2011). A review of multimedia glosses and their effects on L2 vocabulary acquisition in CALL literature. *ReCALL*, 23(02), 135-159.

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia-enhanced learning: When reading helps listening. *Journal of Educational Psychology, 94*, 156–163.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. New York: Cambridge University Press.

Oud, J. (2009). Guidelines for effective online instruction using multimedia screen casts. *Reference Services Review*, 37(2), 164-177.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie, 45*(3), 255.

Sage, K., Bonacorsi, N., Izzo, S., & Quirk, A. (2015). Controlling the slides: Does clicking help adults learn? *Computers & Education*, 81179-190. doi:10.1016/j.compedu.2014.10.007.

Scheiter, K., Schüler, A., Gerjets, P., Huk, T., & Hesse, F. W. (2014). Extending multimedia research: How do prerequisite knowledge and reading comprehension affect learning from text and pictures. *Computers in Human Behavior*, 3173-84. doi:10.1016/j.chb.2013.09.022.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18, 55–89.

Sorden, S. (2012). The cognitive theory of multimedia learning. *Handbook of educational theories.* Charlotte, NC: Information Age Publishing.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

Syodorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology, 14*(2), 50-73.

Tabbers, H. K., & de Koeijer, B. (2010). Learner control in animated multimedia instructions. *Instructional Science*, 38(5), 441-453.

Tempelman-Kluit, N. (2006). Multimedia learning theories and online instruction. *College & Research Libraries*, 67(4), 364-369.

Türk, E., & Erçetin, G. (2014). Effects of interactive versus simultaneous display of multimedia glosses on L2 reading comprehension and incidental vocabulary learning. *Computer Assisted Language Learning, 27*(1), 1-25.

Warren, P., Boers, F., Grimshaw, G., & Siyanova-Chanturia, A. (2018). The effect of gloss type on learners' intake of new words during reading: evidence form eye-tracking. *Studies in Second Language Acquisition, 40*(4), 883-906.

Winke, P., Gass, S., & Syodorenko, T. (2010). The effects of captioning videos used for foreign

language listening activities. *Language Learning & Technology, 14*(1), 65-86.

Xu, J. (2010). Using multimedia vocabulary annotations in L2 reading and listening activities. *CALICO Journal, 27*(2), 311-327.

Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning & Technology*, 13(2), 48-76.

Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, 10(3), 85-101.

Zou, D. (2016). Comparing dictionary-induced vocabulary learning and inferencing in the context of reading. *Lexikos*, 26(1), 372-390.

Zou, D. (2017). Vocabulary acquisition through cloze exercises, sentence-writing and composition-writing: Extending the evaluation component of the involvement load hypothesis. *Language Teaching Research*, 21(1), 54-75.

Zou, D., Xie, H.R. & Wang, F.L. (2018). Future trends and research issues of technology-enhanced language learning: A technological perspective. *Knowledge Management & E-Learning, 10*(4), 426-440.