VIDEOS AND ANIMATIONS FOR VOCABULARY LEARNING:
A STUDY ON DIFFICULT WORDS

Chih-cheng Lin
National Taiwan Normal University
cclin@ntnu.edu.tw

Yi-fang Tseng
National Taiwan Normal University
yifangzeng@hotmail.com

ABSTRACT
Studies on using still images and dynamic videos in multimedia annotations produced inconclusive results. A further examination, however, showed that the principle of using videos to explain complex concepts was not observed in the previous studies. This study was intended to investigate whether videos, compared with pictures, better assist English learners to learn difficult words. It adopted a three-group immediate posttest and delayed posttest quasi-experimental design. Ten target words were selected and embedded in a reading text, each of which was annotated by three annotation types: text-only, text and picture, and text and video. Three intact classes, a total of 88 students, were recruited in a junior high school in northern Taiwan, each of which was randomly assigned to one of the three groups. All participants took the pretest two weeks before the experiment, the immediate posttest after reading the text, and the delayed posttest two weeks after the experiment. The result revealed significant differences between the three groups, in which the video group outperformed the other two groups. Pedagogical implications and suggestions for future research are also given.
Keywords: multimedia annotations, videos, vocabulary learning

1. INTRODUCTION

Annotations, as an instructional design to facilitate reading comprehension, have been employed in multimedia learning for more than a decade. Since their pioneer studies in 1996, Chun and Plass have successfully brought attention to investigate how multimedia annotations could assist language learners in acquiring unknown vocabulary words. A major concern of studies on multimedia annotations was the competitions among media, including texts, pictures, videos and audios, for vocabulary learning. That is, how each annotation type, for example, textual definitions, pictures, animations, films and sound clips, or how combinations of the aforementioned annotation types, could help language learners learn unknown vocabulary words. The combination of textual definitions and pictures has thus far been believed to be more effective for vocabulary learning than single annotations as textual definitions or pictures; it is also more effective than combinations of textual definitions and other types of media (Shahrokni, 2009; Yeh & Wang, 2003; Yoshii, 2006; Yoshii & Flaitz, 2002). One of the inconclusive results is the comparisons between results of textual definitions and still images and those of textual definitions and dynamic videos (Akbulut, 2007; Al-Seghayer, 2001; Chun & Plass, 1996a, 1996b). The reason making animations and films inferior to pictures, based on a further examination done in this study, is that the concepts conveyed by the target words in the previous comparative studies may not have needed such dynamic videos as animations and films. Simple pictures together with textual definitions could have served the purpose of defining the target words. Thus, the purpose of this study is to investigate whether or not dynamic videos, such as animations and films, assist language learners in their learning vocabulary words that entail video presentations, in particular, those that convey meanings difficult for learners to comprehend. The sole research question of the study is: Is there any difference among effectiveness of Taiwanese high school students viewing pictures and animations/films on learning difficult English words?

2. LITERATURE REVIEW
2.1 Vocabulary learning and annotations

Annotations embedded in or appended to reading texts were considered beneficial for reading comprehension and vocabulary acquisition. Marginal annotations using short definitions served as bridges connecting learners' previous knowledge to new information; during their processing of texts, annotated vocabulary words were learned and acquired. Advantages of using annotations in reading tasks were that annotations attracted learners' attention to reading and learning; their proximity to the reading texts provided minimal interruption of reading flow; their appearances also avoided wrong guessing or improper inferences; and, most importantly, annotations made it easier for learners to comprehend difficult texts, which eventually made them more independent and autonomous learners. (Nation, 2001). Researchers of annotations focused their efforts on three major areas, and they were marginal text annotations, multiple-choice text annotations and annotations in target languages, to which we now turn.

Marginal text annotations were the focus of study when annotations started to play a role in foreign language classrooms in 1980s. Text annotations appended to reading texts were implemented for such affective reasons as drawing attention, lowering anxiety, promoting motivation, and securing understanding (Davis, 1989; Johnson, 1982). Later, their effectiveness became a research interest in reading comprehension (Jacobs, 1994), in incidental vocabulary acquisition (Hulstijn, Hollander, & Greidanus, 1996) and in both reading comprehension and vocabulary learning (Jacobs, Dufon & Hong, 1994). It is generally agreed that annotations help enhance learners' later performance on reading comprehension and vocabulary learning.

Multiple-choice text annotations, as opposed to single text annotations, referred to learners' choosing a proper definition for an unknown word from a given list of options. It is argued that the invitation to choosing the correct definition from a list induces deep processing of the target vocabulary words, which further helps learners retain the target vocabulary words (Hulstijn, 1992; Nagata, 1999; Rott, 2005; Watanabe, 1997). However, wrong choices or improper inferences from a given list of annotations debilitated vocabulary learning (Hulstijn, 1992; Watanabe, 1997). Immediate feedback after learners' making choices rendered multiple-choice annotations effective (Nagata, 1999).

Annotations in target languages instead of in learners' mother tongue were also believed to enhance vocabulary acquisition (Jacobs, Dufon & Hong, 1994; Miyasako, 2002; Myong, 2005). While annotations in learners' mother tongue provided direct access in their lexicon, those in target languages required learners to make sense of texts by forming meanings of target vocabulary words. This instructional design for vocabulary learning and reading comprehension was either competitive to their counterparts in learners' mother tongue (Jacobs, Dufon & Hong, 1994; Miyasako, 2002) or superior (Myong, 2005).

Based on the previous findings, this study concerns the use of marginal single text annotations in learners' native language for the following reasons. Supported by the noticing hypothesis (Schmidt, 1990), the bold-faced design, with its highlighting effect, was employed to attract learners' attention while reading. Accurate meanings were provided to avoid wrong guessing. And, because of our participants' proficiency level, their native language was chosen.

2.2 Multimedia annotations for vocabulary learning
The learning theory that studies on multimedia-enhanced vocabulary learning are based upon is a generative theory of multimedia learning, whose major concern is learners' cognitive load during learning (Mayer, 2005; Plass & Jones, 2005). The theory first categorizes all input information into two types, verbal and non-verbal; then, it maintains that new information prompt our brain to perceive, comprehend, subsume, and merge it into old, existing system. When the content of learning is well controlled (intrinsic cognitive load) and the presentation is properly designed (extraneous cognitive load), effective learning is likely to take place (germane cognitive load). Ineffective learning occurs usually when the presentation of the learning content is carelessly designed, which causes learners' attention to split (Sweller, 2005). Take multimedia annotations for example. Learning vocabulary words from textual definitions alone creates insufficient links for retrieval of meanings whereas learning vocabulary words with textual definitions and some visual aids constructs stronger meaning representations for future retrieval. Proponents of multimedia-based vocabulary learning, taking advantages of this instructional principle, concentrate their efforts on providing most helpful non-verbal aids so as to enhance vocabulary learning and retention. With the innovative elements multimedia offers, the process of language learning can be more entertaining and supportive by activating students' visual and auditory senses (Kayaoglu, Akbas & Ozturk, 2011) and enhance learners' autonomy and motivation by providing them with a greater variety of effective learning strategies (Kilickaya & Karjka, 2010).

Although texts, pictures, animations, films and sounds could all be candidates for multimedia annotations, practitioners and researchers used pictures most frequently as the non-verbal aid for vocabulary learning. Starting from Kost, Foss, and Lenzini (1999), pictures were used together with textual definitions in studies on annotations and found effective. They were later implemented on computers (Yoshii & Flaitz, 2002) and on the Internet (Shahrokni, 2009) for reconfirmation of their superiority. In many following studies, pictures were regarded as a reference so that other factors were investigated, such as languages of definitions (Yoshii, 2006), learners' age (Acha, 2009), parts of speech (Shahrokni, 2009), and principles of vocabulary learning (Yanguas, 2009). The implementations of pictorial annotations were supported by their efficiency in vocabulary learning, more efficient than other non-verbal media as videos (Chun & Plass, 1996a, 1996b) and audios (Yeh & Wang, 2003). From the past 15 years of research on multimedia annotations, language learners seemed to have a preference for textual definitions accompanied by pictorial annotations when asked to comprehend reading texts with unknown vocabulary words.

Textual definitions along with pictorial annotations may have better learning results than other kinds. However, the results of learning vocabulary words with textual definitions and pictures and those of learning vocabulary words with textual definitions and animations or films have not yet reached a consensus. In Chun and Plass' studies (1996a, 1996b), college learners of German favored learning vocabulary words with pictorial annotations, whereas English learners in Al-Seghayer's study (2001) preferred learning vocabulary with video clips. Others (Akbulut, 2007) found the two modalities equal. A further examination of the word list given in the video related studies, Chun and Plass' being the only one, found that a crucial feature of using animations, films, or video clips is not observed, that is, its capacity of simplifying complex concepts as defined in Weiss, Knoelton & Morrison (2002).

In their studies (1996a, 1996b), Chun and Plass investigated whether multimedia annotations facilitated reading comprehension and vocabulary learning. They selected 36 target words and evenly divided them into each of the three annotation types, that is, text only, text and picture, and text and video. Their German learners were asked to read a reading text in which the target words were embedded. The results showed that text and picture were more effective than text only and text and video in vocabulary learning. The English equivalents of the target words annotated by video were: to sit up, shaking of one's head, sign language, to stretch, lighter, white caps, helicopter, mouths, irritated, sad, missed, hostile. And, those annotated by picture were: to explode, to doze, to thaw, to threaten, catch, schools of fish, cutter, fisherman's cap, poor, measurable, anxious, pensive. According to the video principle, the necessity for video presentations for the video list is not clear. In other words, the concepts conveyed by the target words were familiar to their college participants. The ease with the concepts conveyed by the target words may have rendered dynamic visual aids distracting, which in turn made the learning results of video presentations inferior to those of pictures. This paper argued that dynamic visual aids as opposed to static ones are considered facilitative when used to explain difficult words to learners. Difficult words in this study are defined as those that may cause problems for learners to learn because of their unfamiliar spellings and/or concepts.

3. METHODS

The study recruited 88 seventh graders from three intact classes in a junior high school in northern Taiwan. They were beginners of English when the study took place; and they had four hours of English classes per week and each class lasted for 50 minutes. All participants had never had experiences learning vocabulary with multimedia annotations online.

The reading text, 417 words (See Appendix A), was first adapted from several passages on world festivals by the English teacher of the three classes and later revised by an experienced English teacher and writer. The Flesch Reading Ease of the text was 71.1 and its Flesch-Kincaid Grade Level was 6.0, both showing that the readability level of the text was moderate and equal to a reading text for 6th graders. Ten nouns in the text were selected and they were: boa, bazaar, derby, falconry, gasp, jester, raffle, revelry, ringtoss, and smoothie. The nouns were deemed to be challenging because their forms were unfamiliar or their meanings conveyed ideas difficult to comprehend to our young learners of English. None of the ten target nouns was included in the basic vocabulary for a junior high graduate in Taiwan; thus, they were all beyond the level of the participants.

The digital version of the reading text was duplicated onto a Moodle site. When reading the text, participants could click on the words that were highlighted. After the target words were clicked on, its annotation appeared in the display area on the right-hand side of the screen. The text group read Chinese definitions only, the picture group had Chinese definitions and pictures, and the video group had Chinese definitions and animations or films. The snapshot of target word derby prepared for the picture group is shown in Figure 1.
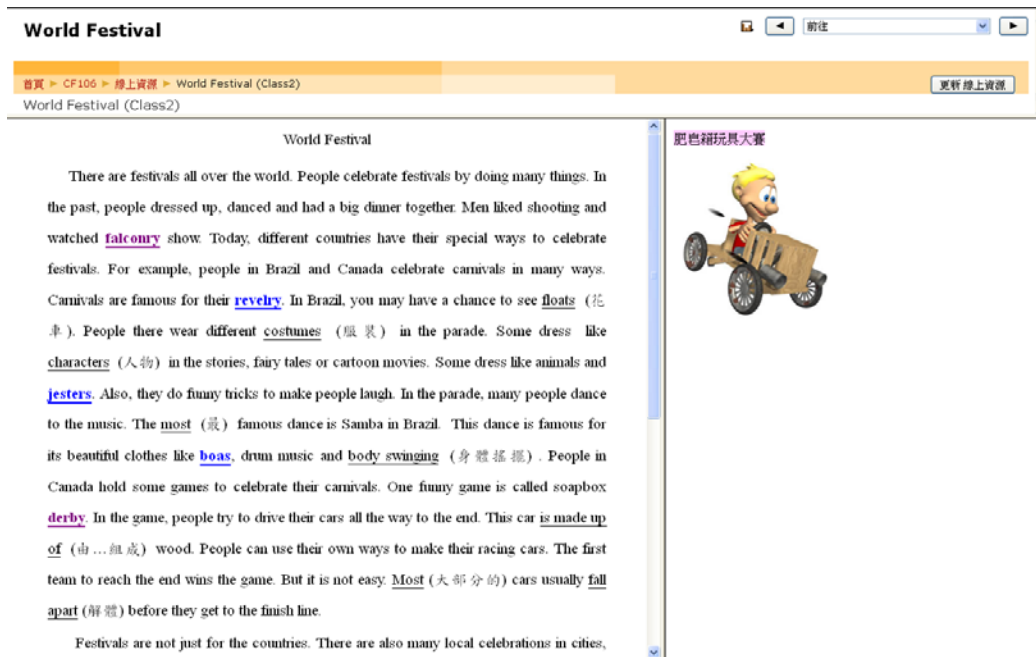
Figure 1. Screen design of the reading text and multimedia annotations

The study had two instruments, a pretest and a posttest, to collect data. The pretest administered two weeks before the treatment consisted of 10 target nouns and 10 distracters, in which the participants were asked to write down the Chinese meanings of the English words. For those that they didn't know, they could put down a check under the "I don't know" column. The posttest consisted of 10 questions of production and 10 questions of recognition (See Appendix B). Participants needed to write down the English words according to the Chinese meanings in the production test and to choose the correct Chinese meaning of the English word from three alternatives in the recognition questions. A question weighed a point, and 20 was the maximum and 0 the minimum in the posttest. The posttest was given to the participants twice, one immediately after they finished reading the passage and the other two weeks after the treatment.

The scores of the pretest were analyzed by one-way ANOVA to inform whether or not the three groups had an equal start and whether or not the participants had previous encounters of the target nouns. A three-way mixed ANOVA, three types of annotations (text, text and picture, and text and video), two times of measurement (immediate and delayed posttests), and two types of tests (production and recognition tests), with times of measurement and types of tests as the repeated measures, were used to analyze the scores of the two posttests.

4. RESULTS AND DISCUSSIONS

The results of the one-way ANOVA of the pretest showed no significant differences ($F(2,85) = 0.084$, $n.s.$), suggesting an equivalence among the three annotation groups. Their mean scores, 0.20 for the text group, 0.19 for the picture group, and 0.26 for the video group, showed that they hardly knew the target words before the treatment.

In both posttests, the video group scored the highest among the three annotation groups. Across the groups, the participants scored higher in the recognition tests than in the production tests; also, they performed better in the immediate posttest than in the delayed posttest. The descriptive statistics of the two posttests in general and of the production and recognition tests in particular are shown in Table 1.

Table 1. Descriptive Statistics of the Two Posttests

Groups	N	Immediate Posttest M (SD)		Delayed Posttest M (SD)	
		Production M (SD)	Recognition M (SD)	Production M (SD)	Recognition M (SD)
Text	30	11.03 (3.634)		10.90 (3.407)	
		2.13 (2.193)	8.90 (2.023)	2.10 (2.339)	8.80 (1.627)

Picture	27	11.19 (3.981)		10.33 (4.252)	
		2.37 (3.027)	8.81 (1.520)	1.96 (2.624)	8.37 (2.372)
Video	31	13.35 (3.527)		12.71 (3.368)	
		3.65 (3.241)	9.71 (0.739)	2.94 (3.065)	9.77 (0.956)
Total	88	11.90 (3.821)		11.36 (3.773)	
		2.74 (2.903)	9.16 (1.553)	2.35 (2.704)	9.01 (1.797)

A three-way repeated measures ANOVA was conducted to examine whether the vocabulary learning varied by group (of annotation), time (of measurement), and type (of test). The results of which are shown in Table 2.

Table 2. Repeated Measures ANOVA Summary for Group, Time, and Type

Source	SS	df	MS	F
Between-subjects effects				
Group	94.628	2	47.314	3.824*
Error	1051.620	85	12.372	
Within-subjects effects				
Time	6.475	1	6.475	5.192*
Time × Group	1.982	2	0.991	0.795
Error	105.993	85	1.247	
Type	3747.366	1	3747.366	628.386*
Type × Group	1.715	2	0.857	0.144
Error	506.896	85	5.963	
Time × Type	1.095	1	1.095	0.995
Time × Type × Group	3.435	2	1.717	1.560
Error	93.562	85	1.101	

The three-way ANOVA revealed no significant interaction effect, including time-by-group interaction effect ($F_{(2,85)} = 1.247$, $n.s.$), time-by-type interaction effect ($F_{(2,85)} = 0.857$, $n.s.$), and the interaction effect among all three factors ($F_{(2,85)} = 1.717$, $n.s.$); however, the main effect of each of the three factors was found significant. The analyses revealed significant differences among the annotation groups ($F_{(2,85)} = 3.824$, $p < .05$), between immediate and delayed posttests ($F_{(1,85)} = 5.192$, $p < .05$), and between production and recognition tests ($F_{(1,85)} = 628.386$, $p < .05$). Of the differences among the three different types of annotation, post hoc comparisons showed that the video group (M = 13.032) significantly outperformed the text group (M = 10.967) and the picture group (M = 10.759). Concerning the significant differences between the two posttests, marginal means showed higher scores of the immediate posttest (M = 11.858) than those of the delayed posttest (M = 11.314), suggesting a memory loss of the target words over two weeks of time. Finally, in terms of the significant differences between the two types of tests, marginal means showed that the scores of the recognition tests (M = 9.062) are higher than those of the production ones (M= 2.065), indicating that for our participants, recognizing target words was an easier task than producing them.

The differences between the two posttests and those between the two test types are expected. The significant memory loss of target words in the delayed posttest is considered normal and in agreement with previous studies. After a span of two weeks the teenagers in the study demonstrated a degree of forgetting the difficult words that they had learned. One point that deserves attention among the three groups, though, is the forgetting rate, the differences between the two posttests over the immediate posttest. The one-way ANOVA revealed no significant differences among the three annotation groups ($F_{(2,85)} = 1.437$, n.s.) with the text group showed the lowest forgetting rate (M = 0.01, SD = 0.246) followed by the video group (M = 0.03, SD = 0.198) and by the picture group (M = 0.08, SD = 0.246). This phenomenon shown in the text group concurred with the average belief of learning vocabulary words with textual definitions: once a word is learned by textual definition, it is part of the lexicon. The effectiveness of learning difficult words with textual definitions alone requires further investigation. Similarly, the drastic differences found between the performances in recognition tests and those in production tests are common, too. The word-supply tests present more problems to our teenagers than the multiple-choice tests which correct answers can be found among other two alternatives. The appearances of possible correct answers helped our adolescents retrieve the meanings of the target words. All tests, including immediate and delayed posttests and recognition and production tests, showed a decrease in scores except for the recognition test of the video group in their delayed posttest. Although the increase from 9.71 in the immediate posttest to 9.77 in the delayed posttest was found, it was not statistically significant. An explanation was wild guessing encouraged by the multiple-choice test type in the recognition test.

Correct responses of each target words in the production and recognition tests of the two posttests were tabulated

in Table 3. From the frequency counts and percentage of correct responses in the table, it showed that the shortest noun boa was the easiest to learn of all.

Table 3. Frequency Counts and Percentage of Correct Responses of Target Words

Words	IMp	%	IMr	%	IM	%	DEp	%	DEr	%	DE	%	Both	%
boa	59	67.05	85	96.59	144	81.82	65	73.86	84	95.45	149	84.66	293	83.24
bazaar	14	15.91	79	89.77	93	52.84	14	15.91	76	86.36	90	51.14	183	51.99
derby	18	20.45	77	87.50	95	53.98	10	11.36	83	94.32	93	52.84	188	53.41
falconry	18	20.45	79	89.77	97	55.11	14	15.91	83	94.32	97	55.11	194	55.11
gasp	35	39.77	86	97.73	121	68.75	26	29.55	79	89.77	105	59.66	226	64.20
jester	36	40.91	83	94.32	119	67.61	23	26.14	83	94.32	106	60.23	225	63.92
raffle	12	13.64	75	85.23	87	49.43	9	10.23	78	88.64	87	49.43	174	49.43
revelry	9	10.23	81	92.05	90	51.14	6	6.82	70	79.55	76	43.18	166	47.16
ringtoss	20	22.73	78	88.64	98	55.68	22	25.00	78	88.64	100	56.82	198	56.25
smoothie	21	23.86	83	94.32	104	59.09	17	19.32	79	89.77	96	54.55	200	56.82

Note. IM stands for immediate posttest, DE delayed posttest, p production test, and r recognition test. The total number of each production test is 88 and so is that of each recognition test. That total number of the immediate posttest (IM) is 176 and so is that of the delayed posttest. Finally, the number of the both posttests is 352.

Except for gasp in the recognition test of the immediate posttest (86, 97.73%), boa received the highest scores of all words in all other tests; in particular, the scores in the two production tests, 59 (67.05%) of the immediate posttest and 65 (73.86%) of the delayed posttest, left those of other target words far behind. The three-letter word boa with its meaning easy to comprehend became the most popular word of all (293, 83.24%). The percentage of correct responses of other target words fell between 64 for gasp and jester and 47 for revelry. The most difficult words for our teenage participants to learn were revelry and raffle, receiving 166 (47.16%) and 174 (49.43%) correct responses in both posttests, respectively. Following the patterns in scoring, both revelry and raffle scored the least in two production tests, with revelry's 9 (10.23%) and 6 (6.82) and raffle's 12 (13.64%) and 9 (10.23%), respectively. As expected, our teenagers' unfamiliarity to the spellings as well as their form-meaning associations both contributed to their low scores. The scoring patterns of each annotation group agree to those discussed earlier, the details of which are included in Appendix C.

The findings in the present study lent support to Weiss et al's definition of videos (2002). That is, videos in this study provide a concrete reference and a visual context for complex and difficult target words. The dynamic presentations in the form of animations and films, rather than the static ones, facilitated our participants' schema construction of the difficult vocabulary words, which can be ascribed to the following reasons. First and foremost, because videos have the capabilities of representing form-meaning connections and of providing a gestalt, dynamic stimuli are easily remembered, making effective building mental images of the target words. Furthermore, because videos also provide rich contexts and cultural authenticity embedded in the target words, the learning of them is made meaningful. And, because videos are made of a series of images, our participants pay attention to and stay focused on changes of images illustrating the meanings of the target words. All these reasons of form-meaning connections, contextual richness and attention getting help our teenage learners learn the target words effectively with videos (Al-Seghayer, 2001, p. 225).

The potential difficulties of learning unknown words in the present study lie in the insufficient information given by textual definitions and still images. Based on the scores in two posttests, some target words, for example, falconry, derby, revelry, raffle, and bazaar selected in the study, are difficult for our participants to comprehend from reading textual definitions alone. The existence of still images helps little, too. When verbal and visual information are both available, according to Chun and Plass (1997), the learner's attention may be directed to the type of information deemed more important or more interesting and taken away from the other mode which may in fact contain more important information. Accordingly, the participants may have focused more on the pictures instead of inferring meanings expressed through textual definitions. In other words, pictures can neither illustrate the complicated meanings embedded in the target words nor serve as a retrieval channel. Instead, they caused a split in the participants' attention to the processing of the target words. The process of integrating the two different sources of information, textual definitions and still images, to some extent, resulted in additional cognitive load for participants learning with pictures (Sweller, 2005). This can also account for the absence of

significant differences between the picture group and the text group in both posttests.

5. CONCLUSIONS

This paper has presented a study on multimedia annotations based on the argument that videos are designed to present complex concepts (Weiss et al., 2002), complex concepts being those difficult to learners. The finding of this study is that learning difficult words with textual definitions and videos is more effective than learning them with textual definitions and pictures and with textual definitions alone. Not only do our teenage learners enjoy the contents of the target words presented in animations and films, they also focus their efforts on learning them. The rich contexts in the dynamic video clips later help our adolescent learners recall the meanings of the difficult words. It is, therefore, our pedagogical suggestions that language practitioners and classroom teachers take advantage of videos when learners find textual definitions and pictorial aids in vain for comprehending the meanings of vocabulary items. As free online video sources have gained their popularity, making a link to play animations and films to explain vocabulary should be regarded essential in modern classrooms.

The limitations of the present study are two-fold: the target words of nouns and the teenage participants. Animations and films must not be employed to explain nouns only. Verbs and adjectives should be incorporated in multimedia-based vocabulary learning and teaching as well. For example, to sauté in cuisines and to fumble in sports may be a couple examples for video annotations, in particular, in English classrooms for teenagers. The primary reason is that to some learners video presentations of some verbs better depict the meanings than other media. Participant-wise, adult or advanced language learners should be recruited in learning difficult words with videos. For adult language learners, who have already had enough worldly knowledge and have few unfamiliar concepts, their requests for multimedia annotations may vary. Similarly, advanced language learners, who have acquired most basic vocabulary and have developed their own habits of learning vocabulary, may demand different combinations of multimedia annotations. Both groups are worth further investigations.

REFERENCES

Acha, J. (2009). The effectiveness of multimedia programmes in children's vocabulary learning. *British Journal of Educational Technology, 40*(1), 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. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, *5*(1), 202-232.

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

Chun, D.M., & Plass, J.L. (1996b) Facilitating reading comprehension with multimedia. *System*, *24*(4), 503-519.

Chun, D.M. & Plass, J.L. (1997). Research on text comprehension in multimedia environments. *Language Learning & Technology, 1*(1), 60-81.

Davis, J.N. (1989). Facilitating effects of marginal glosses on foreign language reading. *The Modern Language Journal, (1)*, 41-48.

Hulstijn, J.H. (1992). Retention of inferred and given word meanings: experiments in incidental vocabulary learning. In P.J. Arnaud & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 113-125). London: Macmillan.

Hulstijn, J.H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal, 80*(3), 327-339.

Jacobs, G.M. (1994). What lurks in the margin: use of vocabulary glosses as a strategy in second language learning. *Issues in Applied Linguistics, 5*, 115-137.

Jacobs, G.M., Dufon, P., & Hong, F.C. (1994). L1 and L2 vocabulary glosses in L2 reading passages: their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading, 17*, 19-28.

Johnson, P. (1982). Effects on reading comprehension of building background knowledge. *TESOL Quarterly, 16*(4), 503-516.

Kayaoglu, M.N., Akbas, R.D., & Ozturk, Z. (2011). A small scale experimental study : Using animations to learn vocabulary. *The Turkish Online Journal of Educational Technology*, *10*(2), 24-30.

Kilickaya, F., & Krajka, J. (2010). Comparative usefulness of on-line and traditional vocabulary learning. *The Turkish Online Journal of Educational Technology*, *9*(2), 55-63.

Kost, C.R., Foss, P., & Lenzini, J.J. (1999). Textual and pictorial glosses: Effectiveness of incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals, 32*(1), 89-113.

Mayer, R.E. (2005). Cognitive theory of multimedia learning. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). New York: Cambridge University Press.

Mayer, R.E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology, 11*(4), 256-265.

Miyasako, N. (2002). Does text-glossing have any effects on incidental vocabulary learning through reading for Japanese senior high school students? *Learning Education & Technology, 39*, 1-20.

Myong, H. K. (2005). Glosses, comprehension, and strategy use. *Reading in a Foreign Language, 17* (2).

Nagata, N. (1999). The effectiveness of computer-assisted interactive glosses. *Foreign Language Annals, 32*(4), 469-479.

Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Plass, J.L. & Jones, L.C. (2005). Multimedia learning in second language acquisition. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 467-488). New York: Cambridge University Press.

Rott, S (2005). Processing glosses: A qualitative exploration of how form-meaning connections are established and strengthened. *Reading in a Foreign Language, 17*(2), 95-124.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11,* 129-158.

Shahrokni, S. A. (2009). Second language incidental vocabulary learning: The effect of online textual, pictorial, and textual pictorial glosses. *TESL-EJ, 13*(3). Retrieved June 1, 2010, from http://www.tesl-ej.org/wordpress/issues/volume13/ ej51/ej51a3/

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp.19-29). New York, NY: Cambridge University Press.

Watanabe, Y. (1997). Input, intake, and retention: Effects if increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition, 19*, 287-307.

Weiss, R, E., Knowlton, D.S. & Morisson, G.R. (2002). Principles for using animation in computer-based instruction: theoretical heuristic for effective design. *Computer in Human Behavior, 18*(4), 465-477.

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), 48-67.

Yeh, Y., & Wang, C.W. (2003). Effects of multimedia vocabulary annotations and learning styles on vocabulary learning. *CALICO Journal, 21*(1), 131-144.

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

Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of picture and annotation types. *CALICO Journal*, *20*(1), 33-58.

Appendix A
Reading Text and Vocabulary Tests

World Festival

There are festivals all over the world. People celebrate festivals by doing many things. In the past, people dressed up, danced and had a big dinner together. Men liked shooting and watched falconry show. Today, different countries have their special ways to celebrate festivals. For example, people in Brazil and Canada celebrate carnivals in many ways. Carnivals are famous for their revelry. In Brazil, you may have a chance to see floats. People there wear different costumes in the parade. Some dress like characters in the stories, fairy tales or cartoon movies. Some dress like animals and jesters. Also, they do funny tricks to make people laugh. In the parade, many people dance to the music. The most famous dance is Samba in Brazil. This dance is famous for its beautiful clothes like boas, drum music and body swinging. People in Canada hold some games to celebrate their carnivals. One funny game is called soapbox derby. In the game, people try to drive their cars all the way to the end. This car is made up of wood. People can use their own ways to make their racing cars. The first team to reach the end wins the game. But it is not easy. Most cars usually fall apart before they get to the finish line.

Festivals are not just for the countries. There are also many local celebrations in cities, towns, schools or neighborhood. In these places, it is common to see bazaars. People set up different stands and sell many things here. Some stands sell toys, clothes, or arts; while some stands sell food like snacks, candy or juice. If you feel thirsty, you can have a try to drink the smoothie. It is good to have an icy drink with much fruit when you feel hot. For outdoor activities, you can watch shows like dancing or singing, or play games like ring toss. If you are good at this game, you have a good chance to take away the biggest prize or whatever you like. If you are not good at this game, you can still try your luck in the raffle to win other prizes. In some places, the festivals also have some stunt shows like juggling with balls or riding on the monowheel. It is very exciting to watch these shows. People usually give a gasp of surprise and admire the performance. No matter what activity the festivals have, its purpose is to make people feel relaxed and have fun in it.

Appendix B
Vocabulary Posttest
Part A:　　Production Test

1.　f _ _ _ _ _ _ y ：放鷹打獵，獵鷹訓練

2.　r _ _ _ _ _ y ：狂歡，尋歡作樂

3.　j _ _ _ _ r ：弄臣，小丑

4.　d _ _ _ y ：肥皂箱玩具大賽

5.　b _ a ：圍巾

6.　b _ _ _ _ r ：市集

7.　s _ _ _ _ _ _ e ：冰沙

8.　r _ _ _ _ _ _ s ：套圈圈遊戲

9.　r _ _ _ _ e ：抽獎活動

10.　g _ _ p ：(因驚訝或驚喜) 倒抽一口氣

Part B:　　Recognition Test

11. (　　) falconry：　（1）放鷹打獵訓練　　（2）馬術競賽會　　（3）騎士長槍比武活動

12. (　　) jester：　　（1）魔術師　　　　（2）小丑　　　　　（3）街頭藝人

13. (　　) derby：　　（1）肥皂箱玩具大賽　（2）抽獎比賽　　　（3）套圈圈遊戲

14. (　　) boa：　　　（1）手套　　　　　（2）圍巾　　　　　（3）緊身褲

15. (　　) revelry：　（1）打架　　　　　（2）祈福　　　　　（3）狂歡

16. (　　) bazaar：　（1）教會　　　　　（2）市集　　　　　（3）遊行

17. (　　) smoothie：（1）氣泡水　　　　（2）冰沙　　　　　（3）彈珠飲料

18. （　　）ringtoss：　　（１）抽獎比賽　　　　　　（２）肥皂箱玩具大賽　　　　（３）套圈圈遊戲

19. （　　）raffle：　　　（１）肥皂箱玩具大賽　　（２）抽獎活動　　　　（３）套圈圈遊戲

20. （　　）gasp：　　　　（１）喘氣　　　　　　　　（２）點頭　　　　　　（３）倒抽一口氣

Appendix C
Scores of Each Target Words by Three Annotation Groups

Table 4: Frequency Counts and Percentage of Correct Responses in Immediate Posttest

Words	Production Test						Recognition Test					
	Text		Picture		Video		Text		Picture		Video	
	N	%	N	%	N	%	N	%	N	%	N	%
boa	17	56.67	18	66.67	24	77.42	29	96.67	25	92.59	31	100.00
bazaar	2	6.67	4	14.81	8	25.81	29	96.67	21	77.78	29	93.55
derby	4	13.33	4	14.81	10	32.26	23	76.67	24	88.89	30	96.77
falconry	2	6.67	4	14.81	12	38.71	27	90.00	23	85.19	29	93.55
gasp	13	43.33	9	33.33	13	41.94	28	93.33	27	100.00	31	100.00
jester	11	36.67	8	29.63	17	54.84	25	83.33	27	100.00	31	100.00
raffle	2	6.67	4	14.81	6	19.35	26	86.67	22	81.48	27	87.10
revelry	1	3.33	2	7.41	6	19.35	27	90.00	23	85.19	31	100.00
ringtoss	5	16.67	6	22.22	9	29.03	26	86.67	21	77.78	31	100.00
smoothie	7	23.33	5	18.52	9	29.03	27	90.00	25	92.59	31	100.00

Note. The maximum for the text group is 30, that for the picture group is 27, and that for the video group is 31.

Table 5: Frequency Counts and Percentage of Correct Responses in Delayed Posttest

Words	Production Test						Recognition Test					
	Text		Picture		Video		Text		Picture		Video	
	N	%	N	%	N	%	N	%	N	%	N	%
boa	21	70.00	18	66.67	26	83.87	28	93.33	26	96.30	30	96.77
bazaar	3	10.00	6	22.22	5	16.13	25	83.33	21	77.78	30	96.77
derby	1	3.33	2	7.41	7	22.58	28	93.33	25	92.59	30	96.77
falconry	5	16.67	2	7.41	7	22.58	29	96.67	24	88.89	30	96.77
gasp	10	33.33	5	18.52	11	35.48	27	90.00	22	81.48	30	96.77
jester	6	20.00	4	14.81	13	41.94	28	93.33	24	88.89	31	100.00
raffle	3	10.00	3	11.11	3	9.68	25	83.33	22	81.48	31	100.00
revelry	0	0.00	3	11.11	3	9.68	22	73.33	18	66.67	30	96.77
ringtoss	9	30.00	5	18.52	8	25.81	24	80.00	23	85.19	31	100.00
smoothie	5	16.67	4	14.81	8	25.81	28	93.33	21	77.78	30	96.77

Note. The maximum for the text group is 30, that for the picture group is 27, and that for the video group is 31.