The Effects of using Synonyms Annotation for English Video Subtitles on Content Understanding, Vocabulary Learning, and Cognitive Loading

Wang, T. I.[1]
Professor
Department of Engineering Science, National Cheng Kung University, Taiwan
Tainan, Taiwan

Wu, Y. H.
Student
Department of Engineering Science, National Cheng Kung University, Taiwan
Tainan, Taiwan

Conference Topic: Technology in Learning

INTRODUCTION

Cognitive load theory, developed by John Sweller in late 1980s [1], uses fundamental entities, like the working memory, the long-term memory, the schema construction, and the schema automation. Working memory, a.k.a. short-term memory has a limited capacity, which can hardly process simultaneously more than 7 elements, like digits, letters, and words, and last only for a few seconds [2]. After repeatedly processing, some elements may go into long-term memory and these elements, by the schema construction process initiated by the schema automation, may combine with other ones already there to become systematic knowledge. Working memory plays an important role in this process. One of the factors closely related to the effectiveness of the working memory process is the cognitive load. A learner's working memory capacity usually directly determines the learner's cognitive load capacity. Overloading the working memory increase the total cognitive load of a learner and, once it goes beyond the limited capacity, will decrease the learner's perceiving, understanding, and problem solving performance [1].

Sweller divided cognitive loads into three types: the intrinsic, the extraneous, and the germane cognitive loads. The intrinsic cognitive load has to do with a learner's prior knowledge of a domain and the degree of difficulty of course materials. A too difficult course material requires learners to put more efforts to understand it will certainly create heavier intrinsic cognitive load. On the other hand, the design and the presentation of course materials affect the extraneous cognitive load. A course material that delivers too many messages or informations simultaneously, even it is easy, will cause learners to swing among the perceptions, trying to catch and organize all the information; this would certainly create heavier intrinsic load too. Some germane cognitive load can help learners to remember the course contents and promote their learning performance.

With the widespread of Internet technology, new ways of teaching and learning have also emerged. Video learning is now one of the new effective ways of teaching and learning for many disciplines. Many researches have proved learners may achieve better performance when watching proper videos than taking traditional classroom lecturing [3] [4] [5] [6]. One of the major advantages of learning by videos is that, by creating a semi-realistic environment, videos can give learners multisensory stimulations, which can enhance their motivation [4] [6] [7] [8] [9].

A lot of institutes or organizations now open their multimedia, including videos, learning materials to the public. The most famous one is the Open Course Ware (OCW), originated from Massachusetts Institute of Technology (MIT) in 2001. The OCW now has shared all their course videos and other

---

[1] Corresponding Author: WANG, T. I.
[2] The National Science Council (Now, Ministry of Science and Technology), Taiwan, has been supported this study under the contracts NSC 101-2511-S-006-011-MY2 and NSC 102-2511-S-006-017-MY3.

materials online free to all the students, self-learner, and educators. It is easy for learners to access those online videos and learn the knowledge of various new technologies. However, for English as Foreign Language - EFL learners, their limited volume of English vocabulary might pose as a barrier to understand videos with difficult contents and subtitles [10]. This is especially true for EFL self-learners. To ease the problem, some researchers have been using dictionary style definitions for the vocabulary of subtitles for learners while they are watching videos [11]. However, such method, making learners to stop the video and click on the subtitle to read the full explanations of the vocabulary and then go back to the scenarios, would bring too much messages and lead to the increase of learners' cognitive load [1]. It has been recognized that increased information in the course materials could increase the cognitive load of a learner and, when it goes beyond the learner's cognitive load capacity, would lead to the learner's reduced performance [1] [12]. Some researches show that replaces the difficult words with simpler synonyms in an article can help learners to understand the content of the article more easily [13] [14]. Synonyms known to a learner express the meanings of unfamiliar words and reduce the complexity of the content. For example dictionary style definition for the word "glad", in one case, would have "Eagerly disposed to act or to be of service etc." totally over 43 words, but one of the synonyms would be "Happy" only. Webb [15] shows learners can learn the unfamiliar words in articles effectively through synonyms.

To help learners who use videos for learning to grasp the meaning of unfamiliar words in the subtitles and understand the contents easily, this study propose a new way of annotating the vocabularies in the subtitles by their synonyms. The study conducts two major experiments to investigate how well this method helps learners in understanding the contents of videos and, in the meantime, in learning the vocabulary and the knowledge in the videos effectively. Analysis are also performed on the data collected from experiments and on the answers to a post-experiment questionnaire to see what the cognitive load may vary, in comparing to the traditional dictionary style annotation method.

1 THE PROPOSED METHOD AND SYSTEM DESIGN

This study proposes a new method for e-learning systems using English videos as course materials. The method uses synonyms annotation, instead of dictionary-style full annotation, on the subtitles of English videos to reduce the extraneous cognitive load of learners that would otherwise decrease their performance. A prototype system is built specifically as the learning and experimenting platform. *Fig.* 1 shows some core modules of the system.

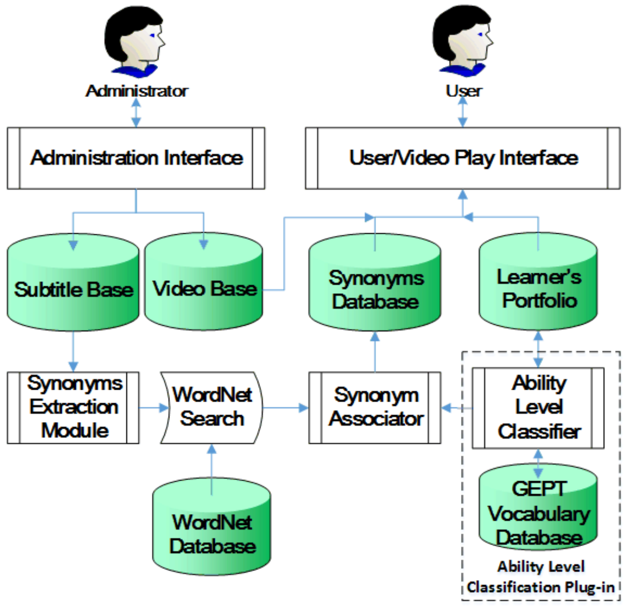

Fig. 1. Core modules of the prototype system

1.1 English ability of a learner

To dispatch synonyms adaptively, the proposed method needs to know the current English ability of learners. For better flexibility, the prototype system is designed to identify a learner's current English ability level by a plug-in module, which can be easily adapted to different evaluation systems. This study presumes that the learners are from Taiwan for subsequent experiments sake. In this case, to estimate a learner's initial English ability, tests similar to the General English Proficiency Test, GEPT, in Taiwan is used and carried out when a learner first registers. GEPT has been used by millions of people since 2000. Its test results has been used by many enterprises and universities as one of the criteria for employment or graduation. This study uses the same proficiency levels in GEPT, including Elementary, Intermediate, and High-Intermediate, Advanced, and Superior, for classifying a learner's English ability level. Each level has its own referred vocabulary, recommended by the GEPT. The prototype system keeps them in a hash table for quick indexing. A learner's ability level may advance after passing certain tests. The ability level of a learner will determine the displaying orders of a synonyms list. Synonyms appearing in the GEPT vocabulary referred for the same proficiency level as a learner's level will get higher priority and appear in the front of the list. Appearing in the vocabulary of other levels get lower priority. Synonyms not appearing in the GEPT vocabulary database use the order found in the WordNet database.

1.2 Video subtitle vocabulary synonyms extraction

The synonyms extraction module performs several tasks in sequence. First, it scans the subtitle file and extracts all the sentences from it. Secondly, the part of speech (POS) of the vocabulary must be known to precisely identify its synonyms, which have the same meaning in a sentence. This module uses the Stanford parser to identify the POS tags of all the words in a sentence. For example, the subtitle sentence “Thank you for joining us for today” has POS tags created by the Stanford Parser as shown in *Table* 1. The module can know the word “joining” here is *verb* and *gerund* instead of a *noun* and find its appropriate synonyms. In general, the module has to find the stem of a word before finding its synonyms. Synonyms of a word are identified and collected by submitting its stem and POS tag to the WordNet, a lexical database that clusters English words into synonyms sets, called synsets, and provides short, general definitions, and various semantic relations between these synonym sets. Collected synonyms are associated with the original word in the subtitle and the GEPT proficiency level, if any, before they go into the synonyms base.

Table 1. POS tags created by Stanford Parser

Sentence	Thank you for joining us for today
POS tags	(ROOT (S (VP (VB Thank) (NP (PRP you)) (PP (IN for) (S (VP (VBG joining) (NP (PRP us)) (PP (IN for) (NP (NN today)))))))))

1.3 System Operation modes

For experiments, the prototype system has two video playing modes and a simple web-based user/ video playing interface to minimize the distraction to learners. *Fig.* 2 shows the synonyms annotation mode on the left, where the learner just click on the word “joining” and two synonyms, “fall_in” and “get_together” appear to the lower left corner of the display. When a learner click on a word in the subtitle, the video stops playing and the system retrieves synonyms of the word from the database, the current English ability level is used to set the priority of these synonyms, and the system displays the synonyms accordingly. The learner can push the forward button to resume video playing. *Fig.* 2, on the right, shows the dictionary-style annotation mode, where, instead of the synonyms, the full dictionary style definitions for the word clicked by the learner appear on the screen.

2 THE EXPERIMENTS AND THE RESULTS

To see whether the proposed method is feasible and effective, this study conducts two experiments to evaluate how well learners can learn the contents and the vocabulary from the videos and how their extraneous cognitive loads may vary under different subtitle annotation methods. Over forty college students, who took the course "Introduction to Computer Science," volunteered to take part in the experiments. Forty-two students participated in the first experiment, in which they watched easy videos of various contents on technologies. Forty-four students participated in the second experiment, in which they watched more difficult videos. Before each experiment, students joined a pre-test to evaluate their volume of vocabulary and some prior knowledge on the technologies in the contents of the videos. Students are then S-shuffled to form two groups, the experimental group and the control group, based on their scores of pre-tests. The experimental group watched videos with subtitles annotated by synonyms in some chosen vocabulary, while the control group watched videos with subtitles annotated by complete dictionary-style definition in the same chosen vocabulary.

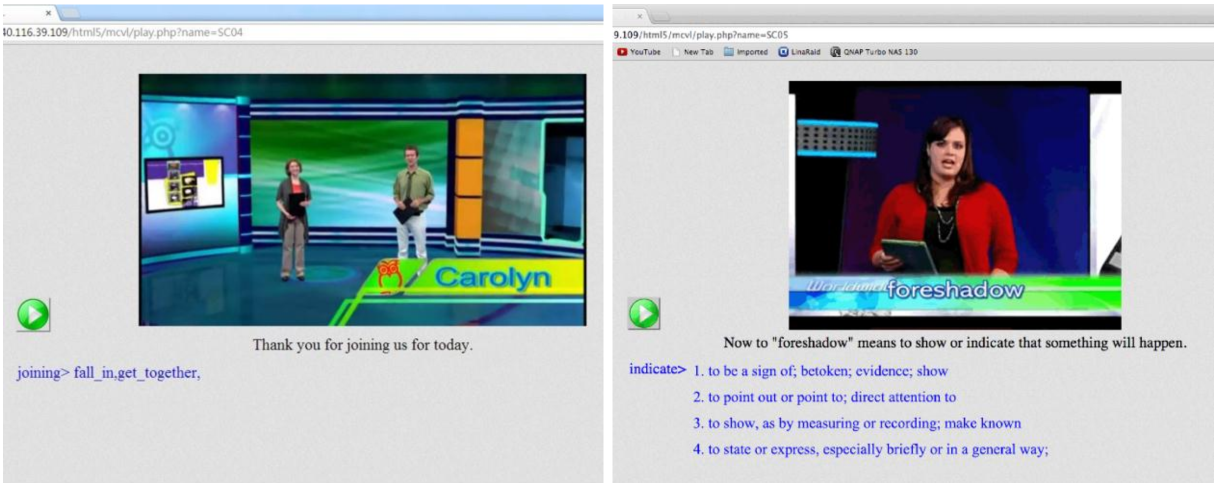

Fig. 2. The prototype system: synonyms annotation mode (left) and dictionary-style full annotation mode (right)

2.1 The contents of the videos and the pre-/post- tests

The experiments uses videos authorized to the University by Studio Classroom, a company devoted to publish multimedia magazines for English learning for Taiwanese since 1962. Contents of published videos are mostly on the hot issues of people, technologies, cultures, and social phenomenon. Every video is classified into low, medium, and high level of difficulty by some experts. Owing to the students recruited are all from the college of engineering, 3 consultant experts from the Department of Foreign Language and Literature of the University picks the videos for the experiments. Six videos, each about ten minutes in length, are all about technologies. Three videos in the low difficulty level (easy), are for the first experiment and three in the high difficulty level (difficult), are for the second experiment. When the videos go into the prototype system, the synonym extraction module starts the process to find all the synonyms for the vocabulary in their subtitles. Consultant experts then filter the vocabulary and create the pre-tests and the post-tests on both the vocabulary and the knowledge of the technologies in the videos. Before each experiment, all the students took the pre-test to reveal their current vocabulary capacity and their prior knowledge on the technologies in the videos, and after each experiment, students took the post-test to disclose their learning performance on two aspects: the vocabulary improvement and the video content (knowledge) understanding.

2.2 The questionnaire

An after experiment questionnaire is designed for measuring the participants' cognitive load, which is based on the NASA-TLX (National Aeronautics and Space Administration - Task Load Index) [16]. There are 6 subscales, including Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration in a standard NASA-TLX questionnaire. In this study, because the length of the videos are fixed and all the learners can finish watching the videos without any pressure, the subscale Temporal Demand is omitted from the questionnaire. Each subscale includes three questions with their answers rated in a 7-point Likert Scale; average of the three questions' scores is the score of that subscale. The lower the score the lower the cognitive load is.

2.3 The structure of the experiments

The experiments aim to investigate how well learners can learn the contents and the vocabulary of videos and how their extraneous cognitive loads may vary under different subtitle annotation methods. The analysis on the experiment results follows the organization as show in *Fig*. 3. Analysis is on the results of pre-tests and the post-tests on both the vocabulary capacity and on the knowledge in the videos to reveal the learning performance of learners. Analysis is also on the learners' answers to a post-experiment questionnaire to see how their cognitive loads may vary. Thus, the two Independent variables in *Fig. 3* are the annotation methods, one is for the experimental group, which receives synonym annotation, and the other is for the control group, which receives dictionary-style complete definition annotation. The dependent variables are the learning performance and the cognitive load. The learning performance variable includes analysis on Vocabulary learning and the video content understanding. The variable cognitive load has five factors as describe in the above section.

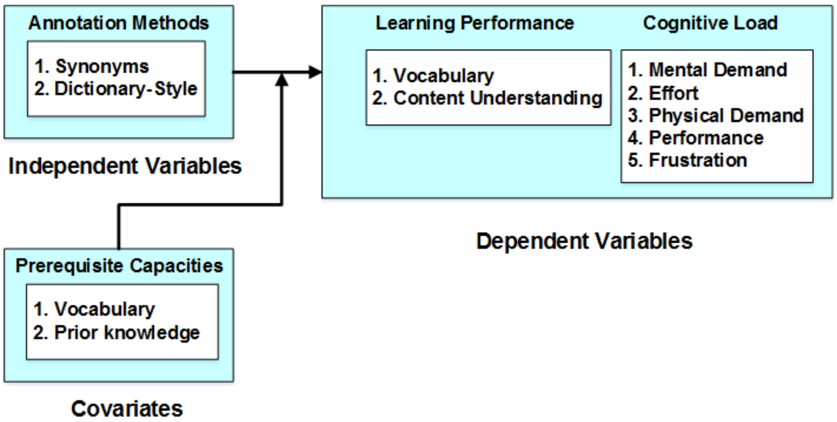

Fig. 3. The structure of the experiments

3 RESULTS AND DISCUSSION

This study uses ANCOVA and MANCOVA on the data collected in the experiments to investigate learners' performance on the vocabulary improvement, understanding of the video contents, and the perception of cognitive load. Owing to the limited paragraph length, the following description will omit some details of the analysis.

3.1 Learning performance on vocabulary improvement

For both experiments, which uses easy and difficult videos respectively, the independent samples t tests on the prerequisite capacities yield (t = 0.097；p = 0.923 > 0.05) and (t = 0.295；p = 0.770 > 0.05) to show that there are no significant differences between the two groups. The distribution of the two groups are quite uniform. For both experiments, learners of both groups are seen to improvement on the vocabulary capacity, as shown in *Table* 3. Yet the ANCOVA results show, as in *Table* 2, there is no significant difference between the experimental group and the control group regarding to the English vocabulary improvement. The possible reason for this might be because the system does not provide a mechanism for practicing vocabulary repeatedly to learners.

Table 2. Learning performance on vocabulary improvement

Group		Easy Videos[1]		Difficult Videos[2]	
		Prerequisite Capacities	Learning Performance	Prerequisite Capacities	Learning Performance
Experimental group	Average	72.3810	84.1270	70.1515	82.1212
	Standard deviation	8.63731	7.44646	9.83877	11.97641
Control group	Average	71.2698	83.4921	68.7879	77.7273
	Standard deviation	6.27837	7.56279	8.39157	9.10761

ANCOVA **1**: (F = 0.001；p = 0.976 > 0.05) **2**: (F = 3.535；p = 0.067 > 0.05)

3.2 Learning performance on the understanding of video contents

Regarding to the video content understanding, as shown in *Table 3*, there is no significant difference between the experimental group and the control group in the first experiment that uses easy videos. While there is indeed significant difference between the two groups in the second experiment that uses difficult videos. This might be because learners, when watching difficult videos, will have higher intrinsic cognitive load and the use of synonym annotations on the video subtitles can reduce their extraneous cognitive load and keep their total cognitive load within their capacity. Therefore, learners can understand the contents in the video more easily.

Table 3. Learning performance on video content understanding

Group		Easy Videos[1]		Difficult Videos[2]	
		Prerequisite Capacities	**Learning Performance**	**Prerequisite Capacities**	**Learning Performance**
Experimental group	Average	57.4603	86.3492	43.6364	73.6364
	Standard deviation	12.37851	12.73249	14.36151	15.86681
Control group	Average	58.0952	83.1746	43.3333	60.6061
	Standard deviation	11.38085	11.85450	14.80026	13.15903

ANCOVA **1**: ($F = 0.708$；$p = 0.406 > 0.05$)　　**2**: ($F = 20.905$；$p =$ ***0.000 < 0.05)

3.3 The effect on the cognitive load

To see if the different annotation methods provide the same assistants for the learners of both groups, the number of using annotations are retrieved from their learning records, and are found to be 26.9524 and 25.2381 clicks in average in experiment 1 and 31.7727 and 29.8182 clicks in average in experiment 2. Running an independent sample t test on these data reveals ($t = 0.451$；$p = 0.654 > 0.05$) and ($t = 0.451$；$p = 0.654 > 0.05$), indicating that there are no significant difference between the two groups in both experiments. This also support that the learner distribution of the groups is even.

Table 4 shows the average of the learners' scores and the standard deviation of each subscale in the first experiment. The value in the Global column is the average of all the subscales' values. Except for the Performance subscale, the experimental group has lower scores than the control group in all the subscales. The Global score of the experimental group is lower than that of the control group by 0.42. Using the ANCOVA on the Global score yields ($F = 5.943$；$p = 0.020 < 0.05$), indicating a significant difference between the two groups. Using the MANCOVA on all the subscales' scores of both groups yields (*F, p*) value pairs of (4.836, *0.034*), (4.974, *0.032*), (5.402, *0.026*), (11.614, **0.002**), and (2.589, 0.116), indicating there are significant differences between the two groups in subscales of Mental Demand, Effort, Physical Demand, and Performance. The control group, however, is better than the experimental group in the Performance subscale. The reason for this might be that learners of control group have to spend more time to understand the meaning of the full definitions, and they feel they perform well. In the Frustration subscale, there is no significant difference between the two groups. The possible reason for this might be that the easy videos do not induce enough cognitive loads that go over the capacities of all the learners in both groups. This can be indirectly proved from the fact that the learning performance on the content understanding for the easy videos has no significant difference between the two groups.

Table 5 shows the average of the learners' scores and the standard deviation of each subscale in the second experiment. The same situation happens here that, except for the Performance subscale, the experimental group has lower scores than the control group in all the subscales. The Global score of the experimental group is still lower than that of the control group by 0.38. Using the ANCOVA on the Global score yields ($F = 6.534$；$p = 0.014 < 0.05$), indicating a significant difference between the two groups. Using the MANCOVA on all the subscales' scores of both groups yields (*F, p*) value pairs of (4.219, *0.047*), (5.520 *0.024*), (1.715 0.198), (10.917 **0.002**), and (12.016, *0.001*), indicating there are significant differences between the two groups in subscales of Mental Demand, Effort, Frustration, and Performance. The control group, still, is better than the experimental group in the Performance subscale. In the Physical Demand subscale, there is no significant difference between the two groups in the second experiments.

The analysis of the answers to the questionnaire concludes that there is a significant difference in the learners' extraneous cognitive load between the experimental group and the control group; watching videos with synonym annotated subtitles induces lower cognitive load than watching dictionary-style annotated ones. This happens in both watching the easy and the difficult video cases. When watching easy videos, learners' reduced extraneous cognitive load does not further make significant differences in both their English vocabulary improvement and video content understanding. The reason might be that the total cognitive load of learners is still within their cognitive load capacity [17] [18]. While, learners' reduced extraneous cognitive load, when watching difficult videos, does induce a significant difference in their understanding of video content.

Table 4. Scores of all the subscales of the questionnaire in easy videos

Group		Mental Demand	Effort	Physical Demand	Performance	Frustration	Global
Experimental group	Average	3.0952	3.6190	3.0476	4.7143	2.4286	3.3810
	Standard deviation	1.26114	1.11697	1.35927	1.05560	1.28730	0.57238
Control group	Average	3.9524	4.4286	3.9524	3.5714	3.0952	3.8000
	Standard deviation	1.16087	1.20712	1.11697	1.07571	1.22085	0.49800
(F, p)		(4.836, *0.034*)	(4.974, *0.032*)	(5.402, *0.026*)	(11.614, **0.002**)	(2.589, 0.116)	(5.943, *0.020*)

Table 5. Scores of all the subscales of the questionnaire in difficult videos

Group		Mental Demand	Effort	Physical Demand	Performance	Frustration	Global
Experimental group	Average	4.0000	4.2727	3.7273	3.8182	3.4545	3.8545
	Standard deviation	1.34519	0.88273	0.93513	0.73266	0.91168	0.44156
Control group	Average	4.7727	4.9545	4.1364	2.9091	4.5455	4.2295
	Standard deviation	1.06600	1.04550	1.03719	1.01929	1.18431	0.50772
(F, p)		(4.219, *0.047*),	(5.520, *0.024)*	(1.715, 0.198)	(10.917, **0.002**)	(12.016, *0.001*)	(6.534, *0.014*)

4 CONCLUSION

This study proposes to use simple synonym annotations on subtitles of videos for learning to reduce information complexity and to reduce learners' extraneous cognitive loads so that they can learn the vocabulary and the knowledge in the videos effectively. A platform is established and two experiments are conducted to investigate how the proposed method will affect learners' understanding of videos' content, learners' vocabulary improvement, and their cognitive load of learning. Over forty college students joined the experiments that uses videos of different difficulty levels, to evaluates on their effectiveness of learning, and measure their learning cognitive load. The result can show that: (1) Learners can still learn vocabulary from synonym annotated subtitles. (2) Using synonym annotations learners can make a significant improvement in understanding the content of difficult videos. (3) Synonym annotations can reduce learners' extraneous cognitive load effectively when learning with difficult videos. This confirms the hypothesis that the simple way of annotating subtitle's vocabulary with synonyms can not only reduce learners' cognitive load but also increase their understanding of video contents and volume of vocabulary effectively.

REFERENCES

[1] Sweller, J. (1988), Cognitive load during problem solving: Effects on learning. *Cognitive science*, Vol. 12, No. 2, pp. 257-285.

[2] Miller, G. A. (1956), The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, Vol. 63, No. 2, pp. 81.

[3] King, J. (2002), Using DVD feature films in the EFL classroom. *Computer Assisted Language Learning*, Vol. 15, No. 5, pp. 509-523.

[4] Secules, T., Herron, C., & Tomasello, M. (1992), The effect of video context on foreign language learning. *The Modern Language Journal*, Vol. 76, No. 4, pp. 480-490.

[5] Tschirner, E. (2001), Language acquisition in the classroom: The role of digital video. Computer. *Assisted Language Learning*, Vol. 14, No. 3-4, pp. 305-319.

[6] Weyers, J. R. (1999), The effect of authentic video on communicative competence. *The Modern Language Journal*, Vol. 83, No. 3, pp. 339-349.

[7] Greenfield, R. (2003), Collaborative e-mail exchange for teaching secondary ESL: A case study in Hong Kong. *Language Learning and Technology*, Vol. 7, No. 1, pp. 46-70.

[8] Stepp-Greany, J. (2002), Student perceptions on language learning in a technological environment: Implications for the new millennium. *Language Learning & Technology*, Vol. 6, No. 1, pp. 165-180.

[9] Brett, P. (1998), Using multimedia: A descriptive investigation of incidental language learning. *Computer Assisted Language Learning*, Vol. 11 No. 2, pp. 179-200.

[10] Hazenberg, S., Hulstun, J. (1996), Defining a Minimal Receptive Second-language Vocabulary for Non-native University Students: An Empirical Investigation, *Applied Linguistics*, Vol. 17, No. 2, pp. 145-163

[11] Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. Language Learning & Technology Vol. 5, No. 1, pp. 202-232

[12] Cerpa, N., Chandler, P., & Sweller, J. (1996), Some conditions under which integrated computer-based training software can facilitate learning. *Journal of Educational Computing Research*, Vol. 15, No. 4, pp. 345-367.

[13] Kameenui, E. J., Carnine, D. W., & Freschi, R. (1982), Effects of text construction and instructional procedures for teaching word meanings on comprehension and recall. *Reading research quarterly*, 1992, pp. 367-388.

[14] Stahl, S. A., Jacobson, M. G., Davis, C. E., & Davis, R. L. (1989), Prior knowledge and difficult vocabulary in the comprehension of unfamiliar text. *Reading research quarterly*, Winter1989, Vol. 24, No. 1, pp. 27.

[15] Webb, S. (2007), The effects of synonymy on second-language vocabulary learning. *Reading in a Foreign Language*, Vol. 19, No. 2, pp. 120 - 136

[16] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), Human mental workload. Amsterdam: North Holland, pp. 139-183.

[17] Paas, F., Renkl, A., & Sweller, J. (2003), Cognitive load theory and instructional design: Recent developments. *Educational psychologist*, Vol. 38, No. 1, pp. 1-4.

[18] Sweller, J., & Chandler, P. (1994), Why some material is difficult to learn. *Cognition and instruction*, Vol. 12, No. 3, pp. 185-233.