Arab World English Journal
مجلة اللغة الانكليزية في العالم العربي
Refereed e-Journal. ISSN: 2229-9327

AWEJ Volume 2, number 2, April 2011 pp. 91- 132

YouTube in an EFL Composition Class

Wenhua Hsu, PhD
I-Shou University, Taiwan

Abstract

This study investigated the extent to which the gap between the passive/receptive and active/productive vocabulary of EFL learners was narrowed in a college freshman composition class with the YouTube video clips incorporated and examined its effect on advanced vocabulary use in writing. To measure the vocabulary used in the videos and the compositions as well as 101 Taiwanese students' lexical proficiency, the RANGE software and the Vocabulary Size Test were utilized. The results show that there was a higher percentage of more sophisticated vocabulary use in timed paragraph writing after viewing the YouTube video clips. The average increasing proportion was 6.02% in turning higher-level vocabulary in the reservoir of passive vocabulary into free active vocabulary. Higher proficiency brings about a greater extension of productive vocabulary as a consequence of better uptake from English exposure. The increased productivity of advanced vocabulary ranged from 2.86% to 7.99% for learners of low- to high-intermediate English proficiency. The research implies that there is a need for activating less frequent words in students' passive vocabulary but not yet fully part of their active vocabulary. The positive effect of YouTube confirms that multiple exposures to English before writing may boost productive vocabulary development.

Keywords: Passive/receptive vocabulary; active/productive vocabulary; vocabulary levels; lexical frequency profile

YouTube in an EFL Composition Class

ESL learners' productive vocabulary is richer than EFL learners' because they are exposed to a larger number of lexical items in their milieu, which propels a rapid growth of productive vocabulary. In contrast, for EFL learners, immersion in a full-English context outside of the classroom may be difficult. The limited hours in the classroom may therefore be the only opportunities for language practice for them. This study was motivated by the concern about how to help EFL learners to access their stock of vocabulary more actively and thus stretch productive vocabulary.

Writing provides a greater opportunity for experimentation with productive use of receptive vocabulary than speaking, as the speed of reaction is not so pressing and students have more time to formulate their ideas. However, teaching English writing has never been easy. In addition to providing feedback on each student's drafts, the ambiance of integrating multimedia into instruction has also given writing teachers no choice but to build up computer-based teaching materials. Multimedia appears to

provide additional channels for exposure to English. As such, reading-to-write may no longer be the exclusive task in the composition classroom. Teaching composition without computer-assisted instruction seems to be outmoded. The ever-increasing presence of multimedia technology provides writing classes with a new tool to improve writing and can thus place audiovisual-to-writing activities on a similar footing.

In recent years, YouTube has become enormously popular, especially with young adults. This video-sharing site (http://www.youtube.com), made accessible to anybody around the world, supplies a wide variety of video content. Surfing on YouTube by typing some keywords and an 'enter' hit on the computer keyboard brings about a wide variety of video content in different languages. Once appropriate videos for teaching and learning have been found, the URL (web page address) displayed at the top of the computer screen along with the HTML code can be saved in the browser for class use. In case of Internet disconnection in class, one may turn to the Keepvid site (http://keepvid.com) to download the YouTube video clips to the computer beforehand.

Despite being enormously popular, research on the effect of YouTube on language learning and teaching is still in its infancy, and even less has been explored on the effect of YouTube on writing. From a language learning point of view, the video clips offer a huge multimedia library of real language use by real people, a potentially rich

resource for language learning or corpus collection. While more and more ESL/EFL teachers are beginning to tap into this source for their listening, speaking or listening-to-speak class, YouTube videos, as linguistic input to catalyze the use of the more sophisticated vocabulary, may also be useful for the listening-to-write class. For learners in an EFL environment, the degree of active English language use seems to depend greatly on any contact with the language they might have within the classroom. Consequently, this research sought to explore the possibility of maximizing higher-level vocabulary use in writing with multimedia incorporated, more specifically, the greater use of less frequent vocabulary as a result of the effect of multiple exposures to English.

Deriving from the speculation that YouTube could play a valuable role in activating less frequent words which are in students' passive vocabulary but not yet fully part of their active vocabulary (Corson, 1997), the purpose of this study is to examine the extent to which the gap between the passive and active vocabulary of EFL learners was narrowed with the incorporation of YouTube video clips into a college freshman composition class. The following questions guided this study:

1. In terms of vocabulary levels, what is EFL learners' everyday range of free active vocabulary?

2. In terms of advanced vocabulary use, to what extent does passive vocabulary become active vocabulary after viewing the YouTube video clips?
3. Does English proficiency make a significant difference in the increase of advanced vocabulary production after viewing the YouTube video clips?

Literature Review

Lexical competence

English learners' vocabulary knowledge and competence have been classified in different ways. For example, Faerch et al. (1984) defined lexical knowledge as a continuum from a vague familiarity with a word form (knowing it exists in the target language) and ending with the ability to use the word correctly in free production. Nation (1990) described lexical knowledge as taxonomy of components. Knowing a word means knowing its form, its position, its function and its meaning. Henriksen (1999) split lexical competence into three parts—partial to precise knowledge, depth of knowledge, and receptive to productive use ability. He noted the fact that if learners cannot use a word correctly nor access it freely for production, this does not mean that they do not know the word, but that they have not yet achieved adequate control over word access.

In a dichotomous way, Nation (1983) distinguished receptive/recognition vocabulary (understood in reading) and productive vocabulary (used in writing or speech). Laufer and Paribakht (1998) instead named receptive vocabulary as passive vocabulary and productive vocabulary as active. They further divided active vocabulary into two types: controlled/elicited active vocabulary ("words learners can use if required") and free active vocabulary ("words learners voluntarily choose to use") (Ibid., p.369). According to Laufer and Paribakht (1998, pp.370-371), passive knowledge of a word is defined as "understanding its most frequent meaning". Controlled active knowledge means "a cued recall of the word". Free active knowledge refers to "spontaneous use of a word in a context generated by the user, in response to a writing assignment". In their research, they found out that the three dimensions of vocabulary knowledge (passive, controlled active and free active vocabulary) developed at different rates. Free active vocabulary developed more slowly and less predictably than did passive vocabulary. Their findings also showed that the gap between passive and active vocabulary widened somewhat as EFL learners acquired more vocabulary, but for ESL students, the gap lessened for learners with advanced levels of passive knowledge. It is generally agreed that passive/receptive knowledge normally precedes active knowledge of a word (Laufer, 1998; Meara, 1996; Nation,

1990). Laufer and Paribakht (1998) suggested that a moderate increase in passive vocabulary within a proficiency level would not necessarily result in a more sophisticated free active vocabulary. Learners may need to learn a lot of passive vocabulary (e.g. to move from an intermediate to advanced proficiency level) before growth in free active vocabulary becomes apparent.

Measures of vocabulary size

The Vocabulary Levels Test (Beglar & Hunt, 1999; Nation,1983, 1990, 2001; Schmitt, Schmitt & Clapham, 2001) for measuring passive vocabulary size and controlled active vocabulary (Laufer & Nation, 1999) is currently a widely-used diagnostic test that looks at five separate word-frequency levels of a learner's vocabulary (the 2nd, 3rd, 5th, 10th 1000, and the Academic Word List containing 570 words or the University Word List containing 836 words) (Each level and version of the test can be obtained at http://www.lextutor.ca/tests/). If a learner's vocabulary reaches the level of the 4th 1000 or the 6th to 9th 1000 or above the 10th 1000, the estimate may be made by extrapolating from the scores of the 3rd 1000, 5th 1000 and 10th 1000.

To complement the insufficiency of the Vocabulary Levels Test (namely, lacking the vocabulary tests between the 3rd – 5th, the 6th – 9th and the 10th – 14th 1000

word-frequency levels), the Vocabulary Size Test was hence developed by Nation and Beglar (2007), which provides a more comprehensive measure of a learner's receptive vocabulary from the 1st 1000 to the 14th 1000 word families of English. Different from the Vocabulary Levels Test that includes productive versions for elicited active vocabulary, the Vocabulary Size Test is merely a proficiency measure used to determine how much vocabulary learners know.

Measures of lexical richness

Vocabulary proficiency is perhaps one of the indicators of overall composition quality. The quality of learners' vocabulary use in writing, for example, the diversity of lexical choice and the correctness of lexical form, may affect reader judgments of the writing (Engber, 1995). A well-written composition makes effective use of vocabulary.

Several measures of lexical richness have been applied to quantify the degree to which a writer uses a varied and large vocabulary. The most widely used measures of lexical richness in the literature include: (Engber, 1995; Laufer & Nation, 1995)

(1) Lexical variation (LV): A type/token ratio of words used by learners, expressed as the ratio of the number of different lexical items to the total number of lexical items in the writing.

(2) Lexical originality (LO): The percentage of words in a given piece of writing that is

used by one particular writer and no one else in the group.

(3) Lexical density (LD): The percentage of lexical items, i.e. nouns, verbs, adjectives and adverbs, in the total number of words in a text.

(4) Lexical sophistication (LS): The ratio of advanced words to the total number of lexical words.

The four popular lexical richness measures have their own weaknesses. The indices may not be stable across groups of different proficiency levels, and additionally subject to topics, sentence structure involving function words, the definition of sophisticated vocabulary and text length.

In an attempt to find a measure that shows the relationship between learners' vocabulary size and the richness of vocabulary in their language production, Laufer and Nation (1995) proposed the measure *Lexical Frequency Profile* (LFP), which examines the relative percentage of words learners use from different vocabulary frequency levels in their writing. The profile embraces the ratios of the total number of word types of the 1st 1000 most frequent words, the 2nd 1000 most frequent words, the academic words (university word list, UWL, Xue & Nation, 1984) and the less frequent words in none of the three lists. The four categories of the 1st, 2nd 1000 words, the UWL and the words not in any lists constitute 100 per cent of the count. A learner's

LFP is measured by the number of words that come from the four levels of vocabulary.

Regarding how learners' vocabulary size is reflected in their vocabulary use and distinguishing what frequency levels of words are used, the notion of the LFP seems to serve the purpose of the current research among the five measures for lexical richness (LV, LO, LD, LS and LFP). This study is more concerned about whether a measure can discriminate between subjects who use frequent and less frequent vocabulary, not just between those who can or cannot use many derived forms of a few words, as shown by lexical variation.

Research Method

The course background and participants

The course *English Composition* was a 2-credit hour required course under the General English program for freshmen at a university in southern Taiwan. The researcher-teacher was lucky enough to have two classes of approximately equal size for the course, totaling 101 students. All of them were freshmen and had been learning English for about 8 to 9 years, at the low- to high-intermediate proficiency levels. They majored in a range of academic areas: electronic and electric engineering, business and medical sciences. It was assumed that two classes could be compared if one of them had the pedagogical treatment (i.e. YouTube served as additional channel for more

exposure to English) while the other simply had not.

The class lectures were given in a classroom with a teacher's console commanding a networked computer, a projector, a big screen and an electronic whiteboard. The textbook *Great Paragraphs* (Folse, Muchmore-Vokoun & Solomon, 2009) was used. The writing class was devoted to four rhetorical styles of paragraphs: description, narration, comparison & contrast, and cause-and-effect, each spanning four 2-hour weeks over an 18-week semester.

In the first 2-hour meeting for a writing theme, the whiteboard lectures of the content in the textbook proceeded in a way so that methods, patterns of organizations and various rhetorical styles were presented and discussed, such as chronological order in narrative paragraph writing, point-by-point or block method in comparison & contrast, focus-on-causes and focus-on-effects methods. Class exercises for each writing unit revolved around formulating a topic sentence, developing supporting ideas, building better sentences and bringing the paragraph to a logical conclusion. In the second meeting for a writing unit, the instructional objective was to make students aware of cohesion and keeping their writing focused. Cohesive devices such as connectors, content prompts and sample paragraphs were provided by PowerPoint slides for reference. At the writing stage (i.e. in the third meeting for a writing unit), for

instance, for a writing theme *Cause-and-Effect*, several writing topics were given to the students for choice. They were *the stresses of being a celebrity*, *the effects of the lottery*, *what would happen if we used credit cards only instead of paper money* and so on. To avoid plagiarism, students were required to complete their drafts in class for a topic from the writing topic options. Students received general evaluative remarks (no grade) and indications of errors by underlining in red ink in the following week.

YouTube as explanatory variable

To increase students' exposure to English and to add variety to class activities, YouTube served as an add-on component that afforded students background knowledge for a writing theme. Before the writing phase began, the researcher surfed through YouTube for short videos for the writing topics. Typing a keyword (e.g. entering a coke fountain and Chinese ghost festival) may return hundreds of results. Considering some factors such as video length, sound quality and content, the optimum video clip (based on the teacher's intuition and experience) was finally selected for later class use. In the meantime, a difficult vocabulary guide was created and listed on a class handout if necessary.

Subsequently, to observe the effect of the YouTube video clips on writing, the

same class routines proceeded for both classes except for one viewing the YouTube videos for seven minutes before writing and the other not. To fill in the 7-minute time slot in the YouTube class, the non-YouTube class was given the same length of time to brainstorm the vocabulary to be possibly used on a writing topic.

A series of comparisons using t tests and ANOVA on the SPSS 17.0 would therefore be straightforward, with the expectation that multiple exposures to English would help in converting passive/receptive vocabulary into spontaneous use and thereby free active vocabulary.

Data collection procedures

To gauge non-English majoring students' proficiency level in terms of receptive/passive/recognition vocabulary knowledge, the Vocabulary Size Test was administered to 101 students at the beginning of the course. Nation and Beglar (2007) made this test by sampling from the most frequent 14,000 word families of English from the British National Corpus. It consists of 140 items (ten from each 1000 word level of the fourteen 1000 word lists). Here are two sample items from the 6th and the 14th 1000 word levels (For a complete version, the Vocabulary Size Test can be freely downloaded from http://www.victoria.ac.nz/lals/staff/paul-nation.aspx).

The 6th 1000 word level: Cavalier	The 14th 1000 word level: Cordillera
He treated her in a **cavalier** manner.	They were stopped by the **cordillera.**

a. without care.
b. politely.
c. awkwardly.
d. as a brother would.

a. a special law.
b. an armed ship.
c. a line of mountains.
d. a member of the whale family.

Because there are ten items at each 1000 word level, each item in the test represents 100 word families. According to Nation and Beglar (2007), a test-taker's score needs to be multiplied by 100 to get a total vocabulary size up to the 14th 1000 word family level.

After the Vocabulary Size Test was given to 101 students, the students' compositions were sampled, according to their correspondingly similar vocabulary scores across two classes. For example, if five students from one class had test scores at the range of 3500-3700 (＝35－37 out of 140 test items being correct * 100) and the other class happened to have four students with the scores within this range, their compositions from both classes would be chosen as a research focus. By doing so, both classes could remain intact and comparisons would be more easily made between any two groups of students from both classes having approximately identical receptive vocabulary size. Any difference arising from the explanatory variable (with and without YouTube video viewing) would be more easily detected. Finally, totaling 26 pairs of students across two classes (26 students from the YouTube class as well as 26 from the non-YouTube class having the correspondingly similar vocabulary size) were

selected.

Data processing

To ease out the topic factor that may constrain vocabulary use and lexical diversity due to the subject matter, 52 students' four paragraph-compositions across four writing units (150-200 words in length for each type of paragraph writing) were entered into the RANGE program for the analysis of vocabulary levels

The RANGE program created by Nation & Heatley (2002) (a freely downloadable zip file is also accessible at the above website) was conducted to analyze the students' compositions with the Lexical Frequency Profile (LFP) measure. The LFP shows the percentage of words in the writing texts that come from different vocabulary-frequency levels. The RANGE is equipped with the British National Corpus 14,000 high frequency word families. Based on English words' occurring frequency, range and dispersion figures, they are divided into fourteen base word lists, each containing exactly 1000 word families. Because words are ranked in terms of how frequently/commonly they are used, for instance, the 3rd 1000 words on a frequency list mean they are less frequent than the 2nd 1000 words and more frequent than the 4th 1000 words. This implies that the likelihood of the 2nd 1000 words being known is higher than that of the 3rd 1000 and the 4th 1000 words. Through this software, the

vocabulary levels of a text can be worked out by comparing the word lists made from the target text with the fourteen base word lists.

When it was applied to the students' compositions, the calculation was as follows. If a 200-word composition contained 120 words belonging to the 1st 1000 most frequent words, 40 belonging to the 2nd 1000, 20 from the 3rd 1000, 15 from the 4th 1000 and 5 from the 5th 1000, this composition's LFP ratios would be 60%-20%-10%-7.5%-2.5%, as opposed to the LFP in Laufer and Nation's (1995) version which shows four divisions: the percentage of the words from the 1st 1000 most frequent words, the 2nd 1000, the University Word List and the less frequent words. Due to the limited vocabulary of EFL non-English majoring students, the above two types of LFP were thus modified into a more condensed version, consisting of the percentage of the basic 2000 words and the percentage of vocabulary use beyond 2000, i.e. 80%-20% against the above 60%-20%-10%-7.5%-2.5%. The percentage score beyond the 2000 most frequent words represents the learner's lexical richness in free expression (use of less frequent vocabulary at will), which is the concern of the present research. The increase of advanced/less frequent vocabulary use was defined as growth in the free active vocabulary.

The students' compositions were typed into text files for computing analysis.

Proper nouns were deleted from the samples. When a word was clearly used incorrectly, it was omitted, as it could not be considered as part of the student's productive lexicon. Since this research was concerned about the use of the higher level and more varied vocabulary, if there were errors in spelling, part of speech, verb tense, subject-verb agreement and other grammatical mistakes that did not affect the meaning of the vocabulary item, the word was corrected, regarded as an attempt for vocabulary use and counted as free active vocabulary. The YouTube video clips were transcribed and entered into the computer for the vocabulary level test on the RANGE.

Results and Discussion

Research Question 1: In terms of vocabulary levels, what is EFL learners' everyday range of free active vocabulary?

Table 1 presents the means of passive vocabulary size and their standard deviations for the total 26 pairs of EFL students selected from two classes based on their correspondingly similar scores on the Vocabulary Size Test (26 students from the YouTube class and 26 from the non-YouTube class). Because the composition classes included the students from different departments, as expected, the dispersion from the average test score of the whole class was large, ranging from 2500 to 6900. For easy

observation, the students' vocabulary scores were further categorized into three proficiency levels: the vocabulary size above 5000 words, of 3000-5000 words and below 3000 words.

Table 1: EFL learners' passive vocabulary size measured by the Vocabulary Size Test

Passive vocabulary size	No. of EFL students	Mean	SD
above 5000	12	5,592 min=5,200; max=6,900	596
3000-5000	30	4,007 min=3,400; max=4,700	429
below 3000	10	2,630 min=2,500; max=2,800	95
	N=52		

Note: The number of test items being correct (n=140 vocabulary items) multiplied by 100 = the passive/receptive vocabulary size

Most of the students' (30 out of 52) vocabulary size was at the range of 3000-5000 words (mean=4,007, SD=429). None of the 10 low-proficiency students' vocabulary scores was below 2000 words. According to Nation (1990), the 2000-word level contains the high-frequency words that all learners need to know to read unsimplified texts. Concerning this, these 10 students had met the threshold level for the minimum vocabulary requirement. Among the 12 students who had knowledge of words above the 5000-word level, only one from one class reached 6900 words and one in the other class had 6800 words, with ten converging at 5200-5500 words.

The following is an example for one paragraph writing on the topic entitled *Chinese Ghost Festival* from a student in the non-YouTube class, whose

receptive/passive vocabulary was 6900 words.

The ghost festival is an important holiday for Taiwanese people. The Gate of Hell is open in the seventh lunar month. Many spirits get the chance to visit the world. Ghosts are like human beings. Some are good and some are bad. The good ghosts do not hurt people. Instead, they help people to do good things. The bad ones are awful and they make people suffer and fear. A long time ago, many people were killed every seventh month of the lunar calendar. Consequently, people prepare a lot of food and things for the spirits and their ancestors, such as meat, wine, fruits and ghost money. They burn ghost money and daily necessities made of paper, for example, paper car, paper house, paper shoes and paper clothes for the dead people. In this way, at the end of lunar July, the ghosts would return to their own places and leave people in peace again.

Table 2: The vocabulary levels of the composition entitled *Chinese ghost festival*

BNC	Tokens/ %	Types/ %	# of word families	Headword
1st 1000	127 81.94%	68 76.40%	62	
2nd 1000	13 8.39%	12 13.48%	12	
3rd 1000	9 5.81%	5 5.62%	4	calendar, consequent, ghost, necessity
4th 1000	1 0.65%	1 1.12%	1	festival
5th 1000	0 0.00%	0 0.00%	0	
6th 1000	1 0.65%	1 1.12%	1	ancestor
7th 1000	0 0.00%	0 0.00%	0	
8th 1000	1 0.65%	1 1.12%	1	Taiwan
9th 1000	3 1.94%	1 1.12%	1	lunar
Total	155	89	82	

Regardless of rhetoric, idioms or grammar, Table 2 displays the vocabulary levels of the above paragraph along the scale of the BNC 14,000 high-frequency word lists. The student's free active vocabulary in this 155-word composition involved using 82 word families, one of which was up to the 9th 1000 word frequency level (lunar), one in the 8th 1000 level (Taiwan), one in the 6th 1000 (ancestor), one in the 4th (festival) and four in the 3rd 1000 level (ghost, necessity, calendar and consequent). In this research, the top 2000 words along the BNC scale were viewed as frequent vocabulary while the words above 2000 were regarded as less frequent or advanced vocabulary. The increase of advanced/less frequent vocabulary use signals productive/active vocabulary development. Therefore, this student's lexical frequency profile (LFP) in a simplified version (as reported in the *Data Processing* section) was 90.24%—9.76%. That is, 90.24% of the total word families in her paragraph writing belonged to the basic 2000 word families. (90.24% was equal to 62 plus 12 word families in the 1st and 2nd 1000 most frequent words divided by the total 82 word families.), against 9.76% for the vocabulary use beyond the 2000 frequent words. Although this student occasionally used higher-level lexical items, most of her vocabulary use centered upon the basic 2000 words (90.24% in word families; 90.33% in tokens). Comparing with her receptive/passive vocabulary size, i.e. 6900 words, the LFP ratios imply that this

student's productive proficiency had not reached a plateau yet in terms of advanced/less frequent vocabulary use, and her great store of receptive vocabulary was waiting for activation. To improve the vocabulary profile of her compositions, a case in point of this was to increase the proportion of advanced vocabulary use.

Using the student's LFP as an index (i.e. having a 6900-word receptive vocabulary versus 9.76%—the usage rate of the advanced/less frequent vocabulary beyond 2000 in a production/writing task), Table 3 addresses the first research question pertaining to EFL learners' normal range of vocabulary use and Table 4 shows the relationship between passive vocabulary size and free active vocabulary. The usage rate of the advanced/less frequent vocabulary in free written expression (measured by the ratio of the word families beyond 2000 to the total word families used) reflects one's extent of free active vocabulary.

Table 3: Passive vocabulary size and the usage rate of the advanced vocabulary

Passive/receptive vocabulary size	Class w/o YouTube No. of students	% of the advanced vocabulary (above 2000) utilized in four compositions as a whole	Post Hoc Tests: Sheffe	
			vocabulary size	Mean Diff. Sig.
min 5200—6900 max mean=5592	6	Mean=8.64%, SD=0.938% min=7.2%; max=9.76%	5200-6900 3400-4700	1.98% Sig.=0.000
min 3400—4700 max mean=4007	15	Mean=6.66%, SD=0.503% min=5.6%; max=7.4%	3400-4700 2500-2800	3.64% Sig.=0.000
min 2500—2800 max mean=2630	5	Mean=3.02%, SD=0.526% min=2.2%; max=3.6%	5200-6900 2500-2800	5.62% Sig.=0.000

	N=26	F(df 25)=112; Sig.=0.000	

Note: The average running words in four compositions as a whole=703, SD=122, N=26;
The mean of the total word families used in four compositions=218, SD=34, N=26

Table 4: Correlation between passive vocabulary size and advanced vocabulary use

	Passive/receptive vocabulary size
% of advanced vocabulary use (beyond the top 2000 frequent words)	Pearson r=0.931 Sig. (2-tailed)=0.000; N=26

Note: N=26, the number of the students from the non-YouTube Class

Table 3 tells us the EFL students' everyday range of vocabulary use in a series of production tasks (4 compositions). The students at higher proficiency levels (having a vocabulary size between 5200 and 6900 words) used a significantly greater proportion of advanced/less frequent vocabulary (mean=8.64%, $p<0.01$) than the other two groups of students at lower proficiency levels (the average usage rates: 6.66% and 3.02%). More specifically, the EFL learners from low- to high-proficiency levels respectively used an average of 3 words, 6－7 words and 8－9 words above the 2000-word level for every 100 words generated while formulating their ideas, with a correspondingly overwhelming majority of 97 words, 93－94 words and 91－92 words below the 2000-word level.

Table 4 confirms such a tendency—the larger one's vocabulary size; the higher percentage of more sophisticated vocabulary use, which shows growth in free active vocabulary (Pearson correlation=0.931, $p<0.01$). In spite of the post hoc tests

indicating significant differences in the usage rates of less frequent vocabulary among the three groups of students, the increase of free active vocabulary from 3.02% or 6.66% to 8.64% (mean difference=+1.98%, +3.64%, +5.62%; all $p<0.01$) was not proportionate to the rise of receptive/passive vocabulary from the level of 2500－2800 or 3400－4700 words to the level of 5200－6900 words. More precisely, an increase of every 1500 receptive words (see Table 3, vocabulary size for three groups, mean=5592, 4007, 2630, roughly at an interval of 1500 words) resulted in an increase of at most 3.64% in advanced vocabulary use. With the advancement of proficiency levels (from having a vocabulary of 3000－5000 words to having a more than 5000-word receptive vocabulary), the increase of the advanced vocabulary ratio decreased progressively to 1.98%. Alternatively, for the students with a vocabulary size between 2000-3000 words, the gap between passive vocabulary and free active vocabulary was perceivably the smallest. With the development of passive/receptive vocabulary, the passive-active vocabulary gap significantly widened.

The result sustains Laufer and Paribakht's(1998) suggestion that learners need to learn a lot of passive vocabulary before growth in free active vocabulary becomes apparent. A moderate increase in passive vocabulary would not necessarily result in a more sophisticated free active vocabulary. As Table 3 has shown, even though most of

the students' vocabulary has reached far beyond 2000 (21 students out of 26 with a vocabulary size above 3000 words), the scanty use of a more advanced vocabulary beyond the top 2000 most frequent words was quite disappointing (6.66%—8.64%). Notwithstanding a vocabulary size of over 5000 words, high-intermediate EFL learners' free active vocabulary was still confined to the top 2000 frequent words. The students seemed to fail to use timed writing as a great opportunity to experiment with and extend their own vocabulary range.

Research Question 2: In terms of advanced vocabulary use, to what extent does passive vocabulary become active vocabulary after viewing the YouTube video clips?

The following is another example for one paragraph writing on the topic entitled *A Coke Fountain* from a student in the non-YouTube class, whose receptive/passive vocabulary was 4000 words. Her free active vocabulary in this 151-word composition was up to the 5th 1000 word level along the scale of the BNC fourteen 1000 word lists (i.e. *fountain*, see Table 5). Nevertheless, the lexical item 'coke fountain' had already been given in the composition title. Rigorously speaking, the student's free active vocabulary only spread to the 3rd 1000 word level, lower than her receptive vocabulary size of 4000 words. Her LFP ratio (95.77%—4.23%) also indicates a low percentage of higher-level vocabulary use.

A Coke fountain is a very interesting game. You can play it when you are bored. It is very easy to make a Coke fountain. It will not cost much money and time. But you can not do it in the house, or it will be dirty. To do it, you need to prepare something: one bottle of Coke, a big drink tube, a thick paper, twenty Mentos and some friends you want them to have a surprise. Following is the method to play. First, put these Mentos into the tube and put the paper under the tube to stop Mentos dropping. Second, put the Coke under the paper. When all are done, open the Coke and take the paper away as soon as possible. When the Mentos fall to Coke, then you will see a Coke fountain run as high as four meters. You will see your friends get wet.

Table 5: The vocabulary levels of the composition entitled *A Coke Fountain* without viewing YouTube

BNC	Tokens/ %	Types/ %	# of word families	Headword	LFP ratio
1st—2nd 1000	134 88.74%	73 94.8%	68		95.77%—4.23% (% below 2000 —% above 2000 in word families)
3rd 1000	3 1.98%	1 1.3%	1	tube	
4th 1000	7 4.63%	1 1.3%	1	Coke	
5th 1000	3 1.99%	1 1.3%	1	fountain	
not in any list	4 2.65%	1 1.3%	X	Mentos*	
Total	151 100%	77 100%	71		

Note: Mentos is the brand name of some sweets/candy sold in Taiwan.

As a contrast, below is an example from a student who was in the YouTube class and also had a receptive vocabulary of 4000 words on the Vocabulary Size Test.

What can Coke be used for? In addition to being a refreshing beverage, it can also be

made an amazing experiment—a Coke fountain. All you need to do is to follow the four steps. First of all, purchase all the items you need at a convenience store: a roll of Mentos, a bottle of 2-liter diet Coke, a jumbo straw and a piece of cardboard. Second, bring all the items to a place outside like a park or garden, never do this indoors. Third, stuff the Mentos into the straw and aim it at the opening of the Coke with the cardboard between them to hold Mentos. Finally, remove the cardboard quickly and let the Mentos drop into the Coke. It is wise for you to prepare an umbrella or wear goggles in advance because a Coke fountain has just been created. It can erupt 4 meters high.

Table 6: The vocabulary levels of the composition entitled *A Coke Fountain* after viewing YouTube

BNC	Tokens/ %	Types/ %	# of word families	Headword	LFP ratio
1st—2nd 1000	125 83.34%	73 84.9%	69		85.19%—14.81% (% below 2000—% above 2000 in word families)
3rd 1000	8 5.33%	5 5.8%	5	cardboard, liter, convenience, refresh, straw	
4th 1000	7 4.67%	2 2.33%	2	Coke, umbrella	
5th 1000	2 1.33%	1 1.16%	1	fountain	
6th 1000	2 1.33%	2 2.33%	2	erupt, goggles	
7th 1000	1 0.67%	1 1.16%	1	jumbo	
12th 1000	1 0.67%	1 1.16%	1	beverage	
not in any list	4 2.67%	1 1.16%	X	Mentos	
Total	150 100%	86 100%	81		

After viewing the Coke geyser video clips from YouTube, the student with a 4000-word vocabulary indeed performed better in using more sophisticated vocabulary (see Table 6, one word family up to the 12th 1000 word level and 11 word families scattering around the 3rd—7th 1000 word levels; the percentage of advanced vocabulary

use＝14.81%), compared with the student of equal proficiency from the other class without viewing YouTube (altogether three words being sporadically between the 3rd—5th word levels; the usage rate of advanced vocabulary＝4.23%).

The reader may be interested to see what English words the student in the YouTube class may have been exposed to before she wrote this timed paragraph-composition. Below is part of the transcript for the YouTube video clip about how to make a Coke geyser.

…A marvelous Coke fountain can be created by mixing Mentos, mint candy and the best refreshing beverage, Coke. This activity is probably best done outside, in a wide open space or on a huge lawn. Prepare a roll of Mentos, a PVC tube or a jumbo straw that is big enough to hold Mentos, a piece of cardboard or a flashcard and a two-liter bottle of Coke. Either diet or regular Coke will work for this experiment, but Coke Zero works better than anything else. In other words, diet Coke erupts higher. Position the bottle on the ground so that it will not tip over. Put the flashcard under the straw to stop Mentos from dropping. Unwrap the whole roll of Mentos and fill the straw with the mint candy. Open the cap of the Coke. Put the straw with the flashcard under it on top of the two-liter bottle and align the straw with the opening of the bottle. Warn the spectators to stand back. Quickly remove the flashcard so that all of the Mentos drop into the bottle at once, and then move out of the way as fast as you can. A Coke fountain forms, gushing four meters high, probably as high as four meters.

(from http://www.youtube.com/watch?v=5taG_-sCAtQ&feature=fvsr)

Table 7 demonstrates that the video used a more sophisticated vocabulary, totaling 19 above the top 2000 frequent words out of 101 word families uttered and taking up to 18.81%. (Purely out of interest, the text of this paper was also conducted on the RANGE program, which resulted in the LFP ratio of 70.52%－29.48%.)

Table 7: The vocabulary levels of the transcript for the YouTube video clip regarding a coke geyser

BNC	Tokens/ %	Types/ %	# of word families	Headword	LFP ratio
1st—2nd 1000	166 79.81%	90 81.8%	82		81.19%—18.81% (% below 2000—% above 2000 in word families)
3rd 1000	16 7.69%	8 7.27%	8	cap, flash, lawn, cardboard, liter, refresh, straw, tube	
4th 1000	9 4.33%	2 1.82%	2	Coke, mint	
5th 1000	2 0.96%	1 0.91%	1	fountain	
6th 1000	5 2.40%	4 3.64%	4	align, candy, erupt, spectator	
7th 1000	1 0.48%	1 0.91%	1	jumbo	
11th 1000	2 0.96%	2 1.82%	2	gush, PVC	
12th 1000	1 0.48%	1 0.91%	1	beverage	
not in any list	6 2.88%	1 0.91%	X	Mentos	
Total	208 100%	110 100%	101		

Table 8, derived from Table 6 and Table 7, compares the advanced word families used and not used in the paragraph-composition after viewing YouTube. When exposed to English for around 7 minutes through the Coke geyser video clip, this student with the capability of 4000 words picked up eight advanced words from what she had heard (i.e. cardboard, liter, refresh, straw, fountain, jumbo, erupt and beverage, excluding Coke given in the title) and used them in her paragraph writing. In addition to listening, the student took advantage of what she saw on the screen and experimented with the

words she had learned before (i.e. umbrella, goggles and 'convenience' at a convenience store) to describe what she observed and associated with. For instance, the student may have caught sight of the spectators with raincoats, umbrellas and goggles standing aside to watch the experiment of mixing Coke and Mentos. The funny scene may have impressed her so that she put forward her suggestion to the reader in her writing about being wise to prepare an umbrella or wear goggles in advance before doing a Coke fountain.

When we complain about our students' writing lacking content, multimedia like YouTube video clips may prove a good prompt in terms of more stimulation for inspiration and more exposure to chunks of the English lexicon. It was very encouraging to see that even so short as 7 minutes' English exposure could also help the student to expand her habitual domain in vocabulary use from the top 2000 frequent words to the advanced words beyond 2000. By inference, the repetition of vocabulary, lexical diversity and variation in sustained exposure would help in strengthening the retention and extension of active vocabulary.

Table 8: Advanced vocabulary use in the student composition and the video transcript

BNC	YouTube transcript	Student composition after viewing YouTube	The student's vocabulary uptake	The student's vocabulary extension
3rd 1000	cap, flash, lawn, cardboard, liter, refresh, straw, tube	cardboard, liter, convenience, refresh, straw	cardboard, liter, refresh, straw	convenience

4th 1000	Coke, mint	Coke, umbrella	Coke	umbrella
5th 1000	fountain	fountain	fountain	
6th 1000	align, candy, erupt, spectator	erupt, goggles	erupt	goggles
7th 1000	jumbo	jumbo	jumbo	X
11th 1000	gush, PVC	X	X	X
12th 1000	beverage	beverage	beverage	X

Note: The word *Coke* was not counted because it was given in the composition title.

Table 9 offers a general picture of the LFP ratios for both classes with and without the treatment of viewing the YouTube videos on the four compositions over one semester. The results reveal a clear-cut difference in the percentage of frequent and less frequent vocabulary use between the two classes (mean LFP ratio for class without YouTube=93.58%—6.42% versus mean LFP ratio for class with YouTube=87.56%—12.44%). The inferiority of the non-YouTube class to the YouTube class in producing more sophisticated words was statistically noticeable (mean=6.42% for class without YouTube versus mean=12.44% for the YouTube class; t=7.114, sig.=0.000<0.01). This verified that although the learners might not automatically put their recognition vocabulary into productive use, they were able to stretch their active vocabulary after exposure to English.

Table 9: Two classes' mean LFP ratios for the four compositions as a whole

	Class without YouTube (n=26)	Class with YouTube (n=26)
Mean LFP ratios (% of the word families used below the 2000-word level—% above 2000)	mean=93.58%—**6.42%** SD=1.97% 97.8%—**2.2%** min. 90.24%—**9.76%** max.	mean=87.56%—**12.44%** SD=3.84% 95.1%—**4.9%** min. 82.28%—**17.72%** max.

Independent-samples t test	t (df 50)=7.114, Sig. (2-tailed)=0.000<0.01
Levene's test for equality of variances	F=10.184; Sig.=0.002; equal variances not assumed, t (df 37.332)=7.114, p=0.000<0.01

In reply to Research Question 2, 'In terms of advanced vocabulary use, to what extent does passive vocabulary become active vocabulary after viewing the YouTube video clips?', a direct deduction of 6.42% from 12.44% was carried out. The positive outcome hints an increased productivity of recognition vocabulary, equivalent to an expansion of 6.02% in turning higher-level vocabulary in the reservoir of passive vocabulary into free active vocabulary. Through multiple exposures to English, one may enlarge his/her normal range of free active vocabulary by voluntary attempts to use higher-level vocabulary. The increased attempts to use higher-level vocabulary also indicated higher awareness of vocabulary use in writing, which may contribute to overall improvement in writing quality as a consequence of richer advanced vocabulary.

Research Question 3: Does English proficiency make a significant difference in the increase of advanced vocabulary production after viewing the YouTube video clips?

Research question 3 arose from Research question 2, with a particular interest in the impact of proficiency on the uptake from English audio-visual materials and therefore on the growth in activating the prior-learned advanced vocabulary. A vertical

look at Table 10 illustrates the relationship between passive vocabulary size and free active vocabulary measured by the usage rate of the advanced vocabulary in written production for both classes separately. The one-way ANOVA results across three proficiency levels for both classes demonstrate that no matter whether the YouTube video clips were supplied, there was a significant difference in the percentage of advanced vocabulary use between the students of different proficiency levels (F=112, Sig.=0.000 for the non-YouTube class, and F= 83.93, Sig.=0.000 for the YouTube class). The students with higher proficiency levels (i.e. a command of above 5000 words) made apparent progress in the production of higher-level vocabulary, compared with those at lower-proficiency levels (steady progression, 3.02%<6.66%<8.64% for the non-YouTube class versus a rapid ascent 5.88%<12.96<16.63% for the YouTube class).

Table 10: The relationship between passive vocabulary size and free active vocabulary, measured by % of advanced vocabulary for both classes

Passive vocabulary size and sample size	Non-YouTube Class % of advanced vocabulary (above 2000) in four compositions	YouTube Class % of advanced vocabulary (above 2000) in four compositions altogether	Independent samples t tests ----------Mean diff.
above 5000 mean=5592 N=6	Mean=8.64%, SD=0.938% min=7.2%; max=9.76%	Mean=16.63%, SD=0.825% min=15.76%; max=17.72%	Mean diff.=7.99% t (df 10)=15.66 Sig.=0.000
3000 - 5000 mean=4007 N=15	Mean=6.66%, SD=0.503% min=5.6%; max=7.4%	Mean=12.96%, SD=1.68% min=9.1%; max=14.95%	Mean diff.=6.3% t (df 28)=13.927 Sig.=0.000
below 3000 mean=2630 N=5	Mean=3.02%, SD=0.526% min=2.2%; max=3.6%	Mean=5.88%, SD=0.646% min=4.9%; max=6.6%	Mean diff.=2.86% t (df 8)=7.677 Sig.=0.000

N=26	F(df 25)=112; Sig.=0.000	F (df 25)=83.93; Sig.0.000	F(df25)=39.44; all p<0.01 on post hoc

Concerning the varying effect of YouTube on advanced vocabulary use in writing across different proficiency levels, Research Question 3 can be answered from a horizontal view of Table 10. As shown in the right column, there was a significant difference between the students of different proficiency levels in the increase of free active vocabulary after viewing YouTube (from top to bottom in turn, $t = 15.66$, sig.=0.000; $t = 13.927$, sig.=0.000; $t = 7.677$, sig.=0.000). The YouTube videos motivated and inspired the students to make attempts to produce more sophisticated words than the other class doing without (the mean difference between two classes from the high- to the low- intermediate level respectively=7.99%, 6.3%, 2.86%). In particular, the YouTube treatment seemed to be more effective amongst the groups of high-intermediate and intermediate students (7.99%>6.3%>>2.86%, F=39.44, p<0.01 and all p<0.01 on post hoc tests). This may be partially ascribed to better uptake from English exposure for those with a bigger passive vocabulary. Moreover, from the low-intermediate level below 3000 words (mean=2630 words) to the intermediate level of around 3000-5000 words (mean=4007 words), the progress in free active vocabulary became more prominent than that in the promotion from the intermediate to the high-intermediate level (6.3%>>2.86% versus 7.99%>6.3%). This hints that the

students with an additional 1500 words from a low-intermediate to an intermediate level may have been on a fast learning curve and had a greater potential for extending productive vocabulary.

Conclusion

Findings and pedagogical implications

As the sample size of student writing was small, the results can only be regarded as indicative rather than conclusive. Drawing upon the notion of the ranking of BNC frequent word lists and lexical frequency profile (LFP), the present study presented the function the YouTube video clips may serve in an EFL writing course. The principal concern was threefold:

(1) Although all of the students' recognition vocabulary was over 2000 words, ranging from 2500 to 6900 words, the extent of their productive vocabulary was still within the 2000 most frequent words, with an occasional use of 3 - 9 words over the 2000-word level as opposed to 91 - 97 words used below 2000 for every 100 words produced (LFP=96.98%-3.02%~91.36%-8.64%). Obviously, the majority of the passive/receptive words did not enter the active realm. The EFL students with larger passive/receptive vocabulary also had larger free active vocabulary in writing.

However, their productive/active vocabulary did not develop with their receptive/passive vocabulary proportionately, namely, a growing gap between the passive-active vocabularies. The situation was even worse amongst the cohort of high-intermediate students.

This could have been improved by encouraging the students to use a more sophisticated lexicon than they would otherwise use. The production task could be approached from a vocabulary focus. In the early stages of writing instruction, teachers may need to place emphasis on vocabulary so that learners become accustomed to thinking of vocabulary as part of writing preparation.

(2) Moving from the non-YouTube class to the YouTube class, an overall tendency could be seen regarding the increasing proportion of higher-level vocabulary use (6.42%→12.44%). Their productive vocabulary performance appeared to be affected by the YouTube videos. It can be a dynamic means that fosters language production. Part of the students' deeply dormant passive vocabulary seemed to be aroused. For this reason, any possible exposure to English in the writing class is highly advocated. Through audio-visual materials, the EFL students were unconsciously reminded to use their non-basic recognition vocabulary that was contextually appropriate for the topic in an immediate writing task.

This has some pedagogical implications for writing teachers, who may need to choose or prepare the audiovisual content relating to the writing topic, allowing students to make attempts at contextually-related words, especially less frequent vocabulary.

(3) This study confirms that exposure to English before writing may boost productive vocabulary development. The data results also add to a better understanding of the extent to which free active vocabulary can be expanded depending on the proficiency level. Higher proficiency brings about a greater extension of productive vocabulary as a consequence of better uptake from English exposure. (7.99%>6.3%>2.86%, $F=39.44$, $p<0.01$ and all $p<0.01$ on post hoc tests)

This suggests that through vocabulary profile and vocabulary size tests, teachers need to recognize the levels of their students and find suitable YouTube video clips (and multimedia materials alike) which are lexically challenging and manageable. For low-proficiency students, whose vocabulary store is small and basic, teachers may need to act as a vocabulary resource by introducing new vocabulary, which helps to convert recognition vocabulary or newly-learned vocabulary into productive vocabulary and helps retention. For high-intermediate or advanced students, a little more audio-visual arousal/prompt induces them to use their passive vocabulary more

actively.

In sum, it may not be surprising to see that free active vocabulary lagged behind and did not grow at the same rate as the learners' passive/receptive vocabulary. Frequent words are more likely to pass from passive to active lexicon, because they are indispensable for daily communication and constant use reinforces their retention. On the other hand, less frequent words are encountered less frequently and used less in communication. Therefore, in an EFL context, limited exposure and lack of practice hinder successful passage of words from receptive to productive vocabulary.

Accordingly, the researcher would like to conclude that the fulfillment of the goal for extending productive vocabulary requires no radical new teaching approach but rather the integration of easily accessible YouTube videos and writing tasks such that they complement each other. It is urged that writing courses be exploited flexibly by supplying audio-visual materials to consolidate less frequent vocabulary use.

Limitations and recommendations

The proposed incorporation of YouTube into a composition course, grounded in the assessment of vocabulary levels and meticulous selection of YouTube content does not necessarily operate in all types of writing classes, nor does it signify a panacea for

improving writing. However, the proposal does offer an alternative way to the systematic instruction of writing cycles.

Although the current survey contributes to the literature of passive/receptive and active/productive vocabulary in an EFL writing setting, it has worked within a narrow focus on vocabulary. When viewing YouTube videos, students may pay little attention to the rule-governed aspect of language form and syntax. It is not sufficient merely to teach and learn vocabulary in isolation. Phrases and collocations in writing are also worthwhile exploring.

Taken together, without deviating too much from the standard approach, which focuses on organization and the communication of ideas, it should be possible to introduce activities which can stimulate the use of more advanced vocabulary right the way through the writing cycle. The aim of this research has been to generate that awareness to prepare students for enhancing their English productive abilities.

References

Beglar, D., & Hunt, A. (1999). Revising and validating the 2000 word level and the university word level vocabulary tests. *Language Testing, 16*(2), 131-162.

Corson, D. (1997). The learning and use of academic English words. *Language Learning, 47*(4), 671-718.

Engber, C.A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing, 4*(2), 139-155.

Faerch, C., Haastrup, K., & Phillipson, R. (1984). *Learner language and language learning*. Clevedon, UK: Multilingual Matters.

Folse, K.S., Muchmore-Vokoun, A., & Solomon, E.V. (2009). Great paragraphs: An introduction to writing paragraphs. Boston: Heinle Cengage Learning.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*(2), 303-317.

Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics, 19*(2), 255-271.

Laufer, B., & Paribakht, T.S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning, 48*(3), 365-391.

Laufer, B., & Nation, I.S.P. (1995). Lexical richness in L2 written production: Can it be

measured? *Applied Linguistics, 16*(3), 307-322.

Laufer, B., & Nation, I.S.P. (1999). A vocabulary size test of controlled productive ability. *Language Testing, 16*(1), 33-51.

Meara, P. (1996). The dimension of lexical competence. In Brown, G., Malmkjaer Williams, J. (Eds.), *Performance and competence in second language acquisition* (pp.35-53). Cambridge: New York: Cambridge University Press.

Nation, I.S.P. (1983). Testing and learning vocabulary. *Guidelin*es, 5(1), 12-25.

Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

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

Nation, I.S.P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Nation, I.S.P., & Heatley, A. (2002). RANGE [Computer software]. Retrieved from June 30, 2010 from http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-88.

Xue, G., & Nation, I.S.P. (1984). A university word list. *Language Learning and*

Communication, 3(2), 215-229.