Semantic Clustering in EFL Secondary School Students' Vocabulary Learning

Jun-Hyeok Choi[*]
(International Graduate School of English)
Hyun-Sook Chung
(International Graduate School of English)

Choi, Jun-Hyeok, & Chung, Hyun-Sook. (2016). Semantic clustering in EFL secondary school students' vocabulary learning. *English Teaching, 71*(3), 121-143.

This study aims to investigate the effects of semantic clustering on Korean secondary school students' English vocabulary learning. The present study was conducted with 56 secondary school students in South Korea. In order to determine if and how semantic clustering affects second language learners, the participants were divided into two groups: a semantically related (SR) word group and a semantically unrelated (SU) word group. The two comparison groups were taught the same 25 vocabulary items in different clusters (SR vs. SU). In the experiment, vocabulary test scores were gathered and analyzed three times: on a pre-, post-, and delayed-test. In the results, scores from the SU word group significantly outperformed the SR word group on the post-test, but on the delayed-test. Findings suggest that presenting new L2 words in SU sets is more effective than SR sets in the short term with Korean EFL secondary school students.

Key words: vocabulary presentation methods, semantic clustering, interference theory, distinctiveness hypothesis

1. INTRODUCTION

As Wilkins (1972) noted, "without grammar very little can be conveyed, without vocabulary nothing can be conveyed" (p. 111). This argument not only applies to first language (L1) acquisition but also to second language (L2) acquisition. However, in spite of this awareness of the importance of vocabulary in communication, vocabulary learning in second language teaching was neglected up until the 1980s (Meara, 1980) because

---

[*] Jun-Hyeok Choi: First author; Hyun-Sook Chung: Corresponding author

scholars and teachers believed that L2 learners would naturally acquire the target vocabulary in much the same way that they learned their first languages (Coady, 1997). Vocabulary learning, however, has become a topic of interest in the field of English language teaching (ELT) and learning (Carter & McCarthy, 1988; Coady, 1997; Folse, 2004; Read, 2000; Schmitt, 2000). Due to the growing body of supportive studies, researchers now generally agree that not only incidental vocabulary learning (as described above) but also intentional learning is crucial for vocabulary growth in a second language. Morin and Geobel (2001), for example, maintained that more robust vocabulary knowledge can be attained through explicit instruction. Incidental and intentional vocabulary learning are now considered complementary (Hulstijn, 2001; Schmitt, 2000). It is, however, true that many learners struggle in this area (Gu & Johnson, 1996; Jones, 1995; Lawson & Hogden, 1996; Porte, 1988; Sanaoui, 1995). In response, the main focus of L2 vocabulary teaching presently is on uncovering effective and efficient means of teaching vocabulary in L2 contexts (Folse, 2004).

Since a number of studies (Ramachandran & Rahim, 2004; Rodriquez & Sadoski, 2000; Webb, 2007) have focused on the effects of various vocabulary presentation techniques for intentional learning such as providing visual aids, example sentences, and L1 translations with target words, a long-standing question in vocabulary teaching is whether L2 vocabulary should be taught in semantic sets (Finkbeiner & Nicol, 2003). Some researchers have maintained that if words that fall in the same semantic field, for example "fruits" (e.g., *apple*, *orange*, *banana*) or "clothes" (e.g., *shirt*, *jacket*, *pants*), are taught simultaneously, learning will be impeded by interference (e.g., Erten & Tekin, 2008; Finkbeiner & Nicol, 2003; Papathanasiou, 2009; Tinkham, 1993, 1997; Waring, 1997; Wilcox & Medina, 2013). However, other researchers have argued that teaching new words in semantically related sets is more effective (e.g., Baleghizadeh & Naeim, 2011; Hashemi & Gowdasiaei, 2005; Hoshino, 2010; Morin & Goebel, 2001).

In South Korea, the results of previous research have been inconsistent (e.g., Bak, 2012; Chung, 2013; Jang, 2014). Furthermore, few studies regarding this issue have been conducted, and the range of participants has been limited to elementary school students (e.g., Bak, 2012; Chung, 2013; Jang, 2014). Therefore, further study is needed to clarify which manner of vocabulary teaching is more beneficial. Thus, the present study aims to investigate the effects of semantic clustering on Korean EFL secondary school students in an attempt to provide more empirical evidence related to EFL vocabulary learning.

2. LITERATURE REVIEW

2.1. Semantic Clustering in L2 Vocabulary Learning

Because the importance of vocabulary learning has been well established, the main focus of L2 vocabulary learning is now centered on uncovering the most effective means of undertaking it (Folse, 2004). Among the various methods of presenting vocabulary, semantic clustering is commonplace, yet there is little consensus on the effects of this presentation technique.

When several words that share a similar concept are grouped, some researchers label this a lexical set (henceforth, LS) (e.g., Gairns & Redman, 1986), while others call it a semantic cluster (e.g., Marzano & Marzano, 1988). Tinkham (1997) defined it as "a group of words with semantic and syntactic similarities, such as *apricot, peach, plum, nectarine, pear, apple*, which fall under the superordinate concept *fruit* and come from a single syntactic word class" (as cited in Wilcox & Medina, 2013, p. 1057).

For decades, teachers and materials developers have presented vocabulary to learners in semantically related sets. In South Korea, examples of this technique can easily be found in the textbooks of elementary school students under topics such as occupation, family members, and weather (Bak, 2012). Unfortunately, the effectiveness of this technique is insufficiently supported by empirical evidence (Tinkham, 1993; Waring, 1997).

2.1.1. Research findings supporting semantic clustering

The idea that presenting vocabulary in semantic clusters will improve L2 acquisition and retention originated in L1 research. Among the implications from L1 research, many L1 researchers acknowledged that vocabulary learning can take place only when learners relate new lexical items to existing background knowledge (schemas) which was previously learned (Stoller & Grabe, 1993). It was, therefore, believed that vocabulary in semantically related groups would activate schema and facilitate vocabulary learning more easily.

This assumption relates to Ausubel's (1963) "advance organizer theory." Ausubel defined an advance organizer as a superordinate word or concept in the mind that can activate existing schemas. When learners encounter new words, the words can be understood via the previously learned concept. Although he claimed that superordinate words or concepts should be presented to learners before they are exposed to new information, not simultaneously, this notion of connecting vocabulary with an existing schema is consistent with L1 studies.

Another theoretical concept supporting semantic clustering is semantic field theory

(Lehrer, 1974), which is based on the assumption that vocabulary is cognitively organized by interrelationships and networks between words rather than being organized in lists of random words. The mind classifies vocabulary by making connections in meaning, and these connections are semantic fields (Channell, 1981). In Channell's (1981) interpretation, words that are close in meaning are located closer to each other in the mental lexicon. If true, when words are presented with semantic similarity, they will be established in the mind more effectively.

Similarly, Aitchison (2003) stated that native speakers appear to have a word-web that has a logically similar organization, including "coordination, collocation, superordination, synonym" (p. 86). This word-web, termed a mental lexicon, helps native speakers recall and retain vocabulary better because they notice connections among the words (Aitchison, 2003; Schmitt, 1997).

Because many L1 vocabulary teaching strategies have been used in L2 vocabulary learning (Wilcox & Media, 2013), there was an expectation that the presentation of vocabulary in semantic sets would also facilitate L2 vocabulary learning (Baleghizadeh & Naeim, 2011; Erten & Tekin, 2008; Schmitt, 2010). Several experimental research findings have supported semantic clustering.

Morin and Goebel (2001) investigated the effectiveness of semantic mapping as a vocabulary learning strategy. Four classes of English speaking learners in a college received Spanish vocabulary instruction through either a communicative-activity-only condition or a semantic mapping plus communicative activity condition. After a semester-long observation, researchers concluded semantic mapping to be an effective strategy. However, this experiment only revealed the effectiveness of semantic mapping. It did not show a difference between semantically related (henceforth, SR) words and semantically unrelated (henceforth, SU) words.

Hashemi and Gowdasiaei (2005) examined sixty intermediate-level EFL adults from Iran, ranging from 20 to 30 in age. Divided into two groups, the participants were presented with new vocabulary either in LS or a semantically unrelated set (SU). Results showed that post-instruction, the LS group was better than the SU group in measures of vocabulary knowledge. They found that semantic sets led to a significant improvement over unrelated sets in both breadth and depth of vocabulary knowledge (i.e., how many and how well the target words are known). Hashemi and Gowdasiaei also found that upper-level intermediate students had significantly greater gains than their lower-level intermediate peers. These findings highlight the fact that learners' L2 level is indeed important in the debate on semantic presentation.

A similar experiment was conducted by Baleghizadeh and Naeim (2011) with a single adult learner. The learner learned two sets of forty-two words, the first set following a semantic clustering method and the other set following randomly grouped words. This

experiment also showed the positive effect of semantic clustering on a meaning recall test and interview.

Similarly, Hoshino (2010) tested Japanese university EFL students in a classroom setting. Hoshino investigated which of five types of word lists (synonyms, antonyms, categorical, thematic, and arbitrary) facilitated L2 vocabulary learning. Hoshino found that the most effective type of word list was a categorical list, and there was no significant difference among the other groupings. Although these results yielded empirical support for semantic clustering, they are limited in generalizability because the number of target words per set was limited to two (e.g., Set 1: *moth, wasp*; Set 2: *asthma, diabetes*, Set 3: *calf, chick*) (Wilcox & Medina, 2013). This number is lower than in many other studies.

Recently in South Korea, Chung (2013) conducted a similar experiment to investigate the effectiveness of semantic clustering with 6th grade elementary school students using unknown English words and Korean equivalents. On the immediate test, there was no significant difference between the two groups; however, one month later on the retention test, scores of the semantically related group were significantly higher than the semantically unrelated group. Chung concluded that semantic clustering has more positive effects in learning vocabulary than presenting semantically unrelated words.

2.1.2. Research findings against semantic clustering

Two powerful theories argue against semantic clustering: interference theory and the distinctiveness hypothesis. In interference theory (Baddeley, 1990; Higa, 1963), the main argument is that "when words are being learned at the same time, but are too 'similar' or share too many common elements, then these words will interfere with each other thus impairing retention of them" (Waring, 1997, pp. 261-262) because traces in memory often compete with one another.

Higa (1963) categorized word pairs into seven kinds of meaning relationships (e.g., near synonyms, free associates, opposites, unrelated, connotation, partial response identity, and coordinates) and differentiated the strength of effect to explain the interference. Higa's research, however, does not fully support theories against semantic clustering because he maintains that coordinate items can still be helpful (Nation, 2000). The coordinate item is identical semantic clustering under one superordinate term.

The distinctiveness hypothesis, on the other hand, focuses on differences rather than similarities, yet the conclusion is the same. The main argument is that as learners learn vocabulary, distinctiveness helps to organize concepts in the mental lexicon. Researchers (Hunt & Elliot, 1980; Hunt & Mitchell, 1982) hypothesized that increasing the non-similarity of information would increase the possibility of learning; hence, vocabulary should be presented in a nonrelated fashion.

In line with these theories, there have been several influential studies. Tinkham (1993), in his study of advanced English speakers learning artificial words, found evidence against semantic clustering. Tinkham observed how easily participants could learn English or artificial words in two ways: semantically related and semantically unrelated. Tinkham's participants were able to learn unrelated words statistically faster than related words.

Warring (1997) replicated Tinkams's (1993) study with L1 Japanese plus L2 artificial word pairs. Warring noted that the group using semantically unrelated words made only 5% of the errors on the immediate test, whereas the corresponding group using semantic clustering made 25% of the errors. These studies suggest that there is an interfering effect of grouping words semantically.

Several years later, Tinkham (1997) examined both semantic and thematic clusters, introducing thematic clusters as an alternative to reduce interference between words. He stressed that semantic clustering serves as a hindrance, "while thematic clustering serves as a facilitator of new language vocabulary learning" (p. 138). In response to the researcher's questionnaire, participants noted that words in the semantically related set, "too similar" or "all the same" (p. 160) were more difficult to remember.

Similarly, Finkbeiner and Nicol (2003) examined 47 undergraduate monolingual English speakers of who learned 32 new artificial words from four different semantic categories in either a related or unrelated condition. They also concluded that recall of words learned in semantically related sets was slower than in semantically unrelated sets in both directions of translation (L1-L2 and L2-L1).

Among these studies, some only observed immediate posttest results, without a delayed test, and some used artificial words as equivalents of target words. As a result, it is difficult to generalize these research results.

To overcome this weakness, other experiments were conducted. Papathanasiou (2009) assessed participants immediately after their final vocabulary lesson and two weeks later. Papathanasiou conducted an experiment in an authentic L2 classroom setting. Although the researcher concluded that semantically related clustering impedes L2 vocabulary learning (using an immediate and delayed test), the results were only accounted for on participants at the beginner level.

Recently, studies supporting this conclusion have returned similar results. Erten and Tekin (2008) investigated a group of 60 fourth graders learning under different conditions: semantic sets versus semantically unrelated sets. Wilcox and Medina (2013) conducted a similar experiment with 32 novice-level learners of L2 Spanish at a university. In both cases, they reached similar conclusions: Learning semantically unrelated words yields better results than learning vocabulary in semantically related sets.

In South Korea, Bak (2012) and Jang (2014) reported similar conclusions after investigating elementary school students. Bak (2012) conducted research with 5th grade

students. After dividing them into two groups, 40 unknown English words with Korean equivalents were taught in one of two ways (SR vs. SU). The SU group was found to yield better results on both short- and long-term retention. Similarly, Jang (2014) examined 3rd grade students with the same research questions. Forty target words were taught in two different ways (SR vs. SU), and the participants were asked to take a post-test. The results showed that the SU group significantly outperformed the SR group. As reported, there have been many inconsistent experimental results on both sides. One side supports semantic clustering, while the other side is against it with respect to new vocabulary learning.

Moreover, there is little consistency in terms of participant attributes such as age, level of proficiency and in data collection methods. For age, some research was conducted with elementary students (e,g., Bak, 2012; Chung, 2013; Erten & Tekin, 2008; Jang, 2014), some with adults (Baleghizadeh & Naeim, 2011; Finkbeiner & Nicol, 2003; Hashemi & Gowdasiaei, 2005; Hoshino, 2010; Morin & Goebel, 2001; Tinkham,1993; Warring, 1997; Wilcox & Media; 2013), and some with both (e.g., Papathanasiou, 2009). Among the research conducted with relatively young learners as participants, disconfirmed semantic clustering, but Chung's (2013) results were supportive. Among the research conducted with adult participants, Morin and Gobel (2001), Hashemi and Gowdasiaei (2005), Baleghizadeh and Naeim (2011), and Hoshino (2010) supported for semantic clustering, but Tinkham (1973), Warring (1997), Finkbeiner and Nicol (2003), and Wilcox and Media (2013) disconfirmed it. In Papathanaiou's (2009) research, only with the adult participants was there a clear result against semantic clustering. Consequently, no consistency exists between ages and the research results. Proficiency levels also varied. Some research indicated that the participants' proficiency levels were at a higher than pre-intermediate level (e.g., Baleghizadeh & Naeim, 2011; Hashemi & Gowdasiaei, 2005), while those in other studies were at a low level (e.g., Erten & Tekin, 2008; Wilcox & Media, 2013). On the other hand, some research did not indicate the learners' proficiency levels in the target language clearly (e.g., Bak, 2012; Chung, 2013; Finkbeiner & Nicol, 2003; Hoshino, 2010; Morin & Gobel, 2001; Tinkham, 1993; Warring, 1997). The participants, however, were considered to be novice learners because they were taught with either artificial words or unknown words. To observe the effect on proficiency, some researchers divided the groups into levels (e.g., Hashemi & Gowdasiaei, 2005; Papathanasiou, 2009). Hashemi and Gowdasiaei (2005) concluded that semantic clustering had a positive effect and the upper level group presented greater gains. In Papathanasiou's (2009) research, the beginner group yielded a clear result, running counter to semantic clustering, while the upper level group indicated a neutral result. In sum, regardless of the participants' proficiency levels, the results did not show a clear consensus. In the data collection, different skills were considered in calculating the results. Some research used productive

skills (e.g., Baleghizadeh & Naeim, 2011: Tinkham, 1993; Warring, 1997), while others used receptive skills (e.g., Bak, 2012; Chung, 2013; Erten & Tekin, 2008; Finkbeiner & Nicol, 2003; Hashemi & Gowdasiaei, 2005; Hoshino, 2010; Jang, 2014; Morin & Gobel, 2001; Papathanisiou, 2009; Wilcox & Media, 2013). The results, however, did not indicate any meaningful pattern according to the type of data collection.

Because of these circumstances, it is more difficult to conclude one way or the other between the SR way and the SU way (Ishii, 2014). In the same vein, although several research studies have been carried out on semantic clustering in South Korea (e.g., Bak, 2012; Chung, 2013; Jang, 2014), the results are conflicting, and the range of participants has been limited to elementary school students. Even if it is not possible to generalize the conclusion widely under different experimental situations, a narrower conclusion considering the specific conditions in South Korea (EFL situation and classroom environment) will also be helpful. Therefore, further study is needed to clarify which manner of vocabulary teaching is more beneficial in South Korea.

This study is designed to examine to what extent semantic relatedness affects L2 vocabulary retention for secondary school students in South Korea. For this purpose, the effects of two types of vocabulary presentation, SR sets and SU sets, are compared.

The research questions are as follows:

1. To what extent do different types of vocabulary presentation (SR vs. SU) affect L2 vocabulary retention in the short term?
2. To what extent do different types of vocabulary presentation (SR vs. SU) affect L2 vocabulary retention in the long term?

3. METHOD

3.1. Participants

This study was conducted at a middle school in Gyeongsang province, South Korea. The school runs 37 classes in total, and each class has an average of 30 students. The participants are 63 first-grade students from two classes. The participants take four English classes per week during the regular school semester following the requirements of the National English curriculum: three classes led by a native Korean speaking teacher and one by a native English speaking instructor. All the participants are native Korean speakers. In the classes with the Korean teacher, both Korean and English were used.

For the treatment, two classes with the same native English speaking instructor were selected. One class was randomly assigned to the group learning with SR sets; the second

class was assigned to the group learning with SU sets.

Due to the absence of three participants at one of the vocabulary tests, their scores were excluded from the data analysis. In addition, since few test scores increased on the delayed test, all the participants were asked whether they had studied additionally for the delayed test. With this result, the scores of another four participants who studied additionally for the delayed test, two participants from the SR group and two from the SU group, were eliminated. The scores of a total 56 participants, 29 participants from the SR group and 27 participants from the SU group, 21 boys and 35 girls, were analyzed in the present study.

3.2. Materials

3.2.1. Selection of L2 words

To select the target vocabulary items, subordinate semantic categories were needed. For this purpose, Jang's (2014) five categories of animals, movements, personality, human body parts, and descriptions of people, were used as a basis. Several additional concepts were extracted from three other sources: *Collins Junior Thesaurus* (Bloomfield, 1989), *The Heinle Picture Dictionary* (O'Sullivan, 2005), and *Oxford Picture Dictionary* (Anderson-Goldstein & Shapiro, 2008). At the end of the selection process, eight semantic concepts were chosen: animals, movements, human body parts, housework, plants, jobs, cooking, and personality. In the next step, the target items were drawn from the dictionaries listed above. To provide different stimuli with the target items, three different parts of speech (noun, adjective, and verb) were assigned to each semantic category (e.g., nouns for animals and body parts, verbs for movements). For the current study, the criteria used in Jang's (2014) study were modified as follows:

- Each word has to fall under one of the eight superordinate concepts.
- All the words have to be concrete rather than abstract.
- All the words have to be easily understood through L1 equivalents.
- All words have to be clearly paired with a corresponding picture without confusion.
- All words are likely to be familiar to the subjects in their L1, but may not be familiar in the L2.

Unlike in Jang's (2014) study, several words with more than three syllables were used (e.g. *accountant*, *manicurist*, *receptionist*). These words are familiar to the participants of the present study because of their ages. With these criteria, 10 words from each category, for a total 80 words, were selected for the current experiment.

To exclude known words, a pilot test was conducted with participants from another

school in the same province not among the participants of the present experiment. As Jang (2014) noted, it is impossible to extract unknown words in advance without revealing the target words. For this reason, piloting was carried out with students not participating in the study. The participants in the piloting were deemed to have had a similar ability in English.

In the piloting test, the students were given 80 words chosen by the criteria noted above and asked to write the meaning in Korean. Those words left as unknown were also assumed to be unknown to the target participants. The pilot group of 30 students completed a survey. All words recognized as known by more than two out of the 30 students were excluded.

To choose semantic subordinates, the five semantic groups with more than five words left were selected: animals, movements, human body parts, cooking, and personality. From this list, the five words that were not chosen by the participants were retained in each semantic category. Finally, 25 words were chosen for the treatment. Table 1 presents the final word list.

TABLE 1

Vocabulary List Organized in Semantic Sets

Category	Vocabulary Items
Animals	antlers, hoof, whisker, trunk, mane
Movements	crawl, leap, crouch, stroll, tumble
Jobs	accountant, butcher, manicurist, welder, receptionist
Cooking	sauté, grate, dice, chop, stir
Personality	daring, timid, caring, arrogant, cruel

To compare the effect of semantic clustering on vocabulary learning, the 25 words were rearranged into two different sets: an SR set and an SU set. Two comparison groups learned the same vocabulary items, but the order and the composition of the words varied. Table 2 presents the word list organized by group.

TABLE 2

Vocabulary List Organized by Group

Group	Set No	Vocabulary Items
	1	antler, hoof, whisker, trunk, mane
	2	crawl, leap, stroll, crouch, tumble
SR	3	accountant, butcher, manicurist, welder, receptionist
	4	sauté, grate, dice, chop, stir
	5	daring, timid, caring, arrogant, cruel
	1	antler, crawl, accountant, sauté, daring
	2	hoof, leap, butcher, grate, timid
SU	3	whisker, stroll, manicurist, dice, caring
	4	trunk, crouch, welder, chop, arrogant
	5	mane, tumble, receptionist, stir, cruel

3.2.2. Instructional materials

In the experiment, the same vocabulary items were taught to the two groups via two different semantic clustering types (SR vs. SU). For each group, the instructional materials were presented via two methods: whole class learning material and individual learning material.

The whole class learning materials were made into PowerPoint files and presented through PowerPoint to prevent any possible teacher effect. Each set of material for the whole class was designed to run automatically, displaying five new words with their pronunciations, corresponding pictures, and Korean equivalents. Various information was provided at the same time during teaching to maximize the participants' recall. The teacher's role was limited to managing the presentations and leading the lessons during the planned time.

While each set of whole class materials ran, each new word was presented three times. When each target word was presented, the pronunciation was provided twice. Consequently, the participants were exposed to the same word three times, and the pronunciation of the target words was played a total of six times. Each presentation in one session ran approximately four minutes. For the 25 target words, five presentation material files (including five target words each) were made, and they were provided twice during the 10 days for each group, respectively.

The material for individual learning was made into worksheets with the corresponding items presented through the PowerPoint file. The format was identical to the whole class material except that pronunciation was provided through two types of written form in a phonetic alphabet and Korean instead of audio files (e.g., [hu:f] and [후우프]). The participants were allowed to study the individual material for five minutes.

3.2.3. Test materials

Pre-tests, post-tests, and delayed tests were administered. The pre-test was administered before the treatment, the post-test was administered the day after the final treatment, and the delayed test was administered exactly four weeks after the post-test. The format of each test was based on "the matching format of the Vocabulary Levels Test" (Nation, 2013). Although some researchers (e.g., Campion & Elley; Stewart & White) noted disadvantages with this type of test (as cited in Nation, 2013), the level test was selected due to the accuracy of the estimate of words known at each level.

All three tests are exactly the same in format except for the composition of words. That is, the words for test items were selected randomly and placed in a different order on each of the three tests. Despite concern for the practice effect, as the condition would affect

both groups evenly, and the main purpose of this study was only to differentiate the effects of two different presentation ways (SR vs. SR), the same format was used for all three tests (pre-, post-, and delayed test). All the tests lasted approximately 10 minutes. Twenty Korean equivalents out of the 25 total equivalents were randomly selected to make the question items, and they were divided into five groups. The five groups each with four equivalent items were written on the right side of the test sheet. As a counterpart to each group of Korean equivalents on the left side, ten target words were extracted randomly, including six distracters. These items were written on the left side. The question required participants to match the question items on the left side with their target words on the right side. Each test sheet consists of 20 question items. Clear instructions and an example of how to answer were provided on the first page of each test (see Appendix).

One point was allocated for each correct answer, so the maximum score was 20 points for each test. Answers were not provided to prevent the possibility of affecting the test results.

3.3. Procedures

The experiment lasted six weeks, from November 5th to December 18th, 2015, until the day of the one-month delayed test. To minimize any possible influence of forgetting the target words due to the long treatment period, the treatments were designed to follow in a row. Each treatment, including the level-tests, was administered during the self-study time in the mornings under the control of the Korean teachers.

On the first day, the pre-test was administered. From the second day, instructional materials, including 5 new target words each, were provided as a treatment ten times for two weeks. During the treatment, each group (SR vs. SU) studied these five target words a day through two types of instructional materials: whole class and individual. Each whole class set of materials was made using PowerPoint and was presented for four minutes. As soon as the whole class study ended, the participants were given five minutes to review what they had learned through the individual material, rendered in worksheet form. When each group finished the fifth treatment, all of the words were again taught, because the number of target words was only 25.

When all 10 treatments had concluded, a post-test was administered to both groups the following day. To examine the long-term retention of the target words relative to the different treatments, exactly four weeks later, a delayed test was administered.

In order to minimize the effects of any unexpected variables, the participants were asked not to self-study the individual materials following each instructional treatment. The teachers who participated in the study did not provide any information that might have affected the participants' test results. For the pre-, post-, and delayed-test, ten minutes was

allocated to complete each test, respectively.

3.4. Data Analysis

The purpose of the data analysis is to examine the effects of two different types of vocabulary presentation (SR vs. SU) on second language vocabulary learning. For this purpose, three test scores (pre, post, delayed) were analyzed quantitatively by SPSS.

To begin the process, scores on the three tests were computed descriptively, using means and standard deviations to account for the effect of semantic clustering. Secondly, a repeated measures analysis of variance (ANOVA) was employed to observe any changes within the test scores. The within-subjects variable was Time (pre, post, delayed) and the between-subjects variable was Group (SR vs. SU). The dependent variable was vocabulary test scores.

4. RESULTS

The main purpose of this study is to investigate the effect of semantic clustering on L2 vocabulary learning. For this purpose, vocabulary test scores collected from two different experimental groups were analyzed. Table 3 presents the descriptive statistics for the vocabulary test scores.

TABLE 3
Descriptive Statistics for Vocabulary Test Scores

| Time	Group	N	M	SD
Pre-test	SR	29	3.14	1.80
	SU	27	4.26	2.87
	Total	56	3.68	2.42
Post-test	SR	29	13.59	5.42
	SU	27	17.22	4.10
	Total	56	15.34	5.12
Delayed-test	SR	29	10.79	5.94
	SU	27	14.26	5.45
	Total	56	12.46	5.92

To determine whether this result indicates a statistically significant difference between the two groups, a repeated measures ANOVA was performed to examine the effects of Group (SR vs. SU) and Time (pre-, post-, delayed-test) on vocabulary scores. For an alpha level of .05, the results show a highly significant main effect for Group. The mean number of accurately answered items in SU group was statistically greater than the mean number of accurately answered items in SR group. A similar significant result was found for Time.

The complete source table (Table 4) shows that there was a significant Time x Group interaction ($F(2, 108) = 3.21$, $p < .05$). In addition, as Figure 1 shows, the SU group outperformed the SR group.

TABLE 4

Repeated Measures ANOVA for Vocabulary Test Scores by Time and Group

Source	SS	df	MS	F	Sig.
Within Subjects					
Time	4161.661	2	2080.831	242.30	.000
Time ⨯ Group	55.233	2	27.616	3.21	.044
Error	927.470	108	8.588		
Between Subjects					
Intercept	18650.185	1	18650.185	419.13	.000
Group	315.185	1	315.185	7.08	.010
Error	2402.809	54	44.496		

FIGURE 1

Estimated Marginal Means for Vocabulary Test Scores
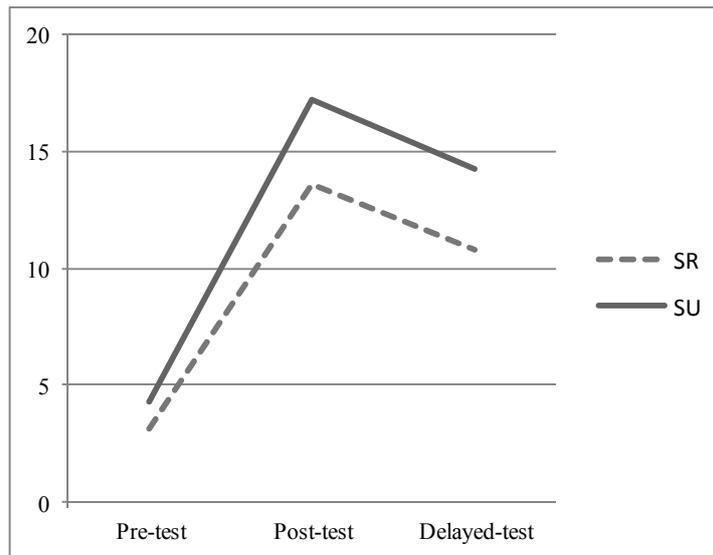
To identify where significant changes occurred, a post-hoc test was conducted. However, since there were three pairwise comparisons by Time and Group, the significance level was adjusted to .017 (= .05/3). Table 5 shows the results of the comparisons by Time, and Table 6 shows the results of the comparisons by Group.

TABLE 5

Post Hoc Test: Pairwise Comparisons by Time

Group	(I) Time	(J) Time	MD (I-J)	Std. Error	Sig.
SR	Pre	Post	-10.448	.791	.000
	Pre	Delayed	-7.655	.906	.000
	Post	Delayed	2.793	.574	.000
SU	Pre	Post	-12.963	.820	.000
	Pre	Delayed	-10.000	.939	.000
	Post	Delayed	2.963	.595	.000

TABLE 6

Post Hoc Test: Pairwise Comparisons by Group and Time

Time	(I) Group	(J) Group	MD (I-J)	Std. Error	Sig.
Pre	SR	SU	-1.121	.637	.084
Post	SR	SU	-3.636	1.293	.007
Delayed	SR	SU	-3.466	1.528	.027

Both post-test scores were significantly higher than the pre-test scores, and both post-test scores decreased significantly on the delayed-test. However, the delayed scores remained significantly higher than the pre-test scores. This result shows that both presentation techniques were effective in teaching new vocabulary.

Table 6 provides more important information about the significant interaction between "Group" and "Time." First, there was no significant difference between the two groups on the pre-test. This means that the two groups were at the same level before the treatment. Second, on the post-test, the SU group noticeably outperformed the SR group. Lastly, on the delayed test, the SU group demonstrated higher results. Nevertheless the scores were not significantly higher than for the SR group. The test results in the current study suggest that the SU technique is more effective in presenting new vocabulary for short-term retention, but not for long-term retention.

5. DISCUSSION

The research question investigated in the present study sought to observe the effects of semantic clustering on EFL secondary school students' new vocabulary learning in South Korea. The SU group scored significantly higher than the SR group on the post-test. These results suggest that presenting new words in semantically unrelated sets is more beneficial in terms of short-term retention. These findings match those observed earlier in studies critical of semantic clustering in L2 vocabulary learning to some degree (e.g., Erten & Tekin, 2008; Finkbeiner & Nicol, 2003; Higa, 1963; Papathanasiou, 2009; Tinkham, 1993, 1997; Waring, 1997; Wilcox & Medina, 2013).

Two important theories provide a theoretical justification for this view: The first is interference theory (Baddeley, 1990; Higa, 1963) and the second is the distinctive hypothesis (Hunt & Elliot, 1980; Hunt & Mitchell, 1982).

The notion of interference aroused the interest of psychologists in the 1940s and 1950s as the major cause of forgetting (Nation, 2000). In addition, McGeoch and McDonald (1931) found that too similar words can interfere with first language learning. These findings were formulated as interference theory (Waring, 1997). The main concept of this theory is that when words are too similar, or when they share too many commonalities, they will interfere with each other (Waring, 1997). This theory points to the similarities between words as the main cause of interference.

The distinctiveness hypothesis is another cited theoretical perspective against semantic clustering. Its arguments are consistent with those of interference theory. One difference, however, is that this theory focuses on the distinctions rather than the similarities between words. The main argument is that as a learner acquires vocabulary, distinctiveness helps organize new concepts in the mental lexicon. Researchers (Hunt & Elliot, 1980; Hunt & Mitchell, 1982) hypothesized that increasing the non-similarity of information increases the possibility of learning.

Since the earliest research by Higa (1963), who brought the concept of interference into foreign language learning, many experimental studies have supported interference theory in the field of SLA (e.g., Erten & Tekin, 2008; Finkbeiner & Nicol, 2003; Higa, 1963; Papathanasiou, 2009; Tinkham, 1993, 1997; Waring, 1997; Wilcox & Medina, 2013). Although recently a handful of results appear to contradict the research listed above, a large proportion of studies run counter to semantic clustering in presenting new words. Ishii (2014) even pointed out that "following such research, the negative impact of semantic clustering is sometimes treated almost as if it were an established fact" (p. 28). The results of the present study are little different from the outcomes of these studies. Both theoretical bases also appear to support the interpretation of the results of the present study. The research questions of the current study, therefore, appear to be resolved.

From a different point of view, there are several possible explanations for this result. Firstly, this experiment, as a replication study, intended to show whether there was a difference between previous studies conducted with primary school students (e.g., Bak, 2012; Chung, 2013; Jang, 2014) and with secondary school students. In other words, it was designed to determine the effect of age toward semantic clustering because there was a certain expectation that more mature students could use semantic clustering as a strategy. However, the result indicates that semantic clustering impedes EFL students' vocabulary learning generally regardless of the age of participants in South Korea.

Secondly, as the purpose of this study was to find other explanations for the effects reported, the participants' proficiency level was examined. Proficiency level did not

greatly affect the result of the present study because the result was consistent with previous studies in classroom settings (e.g., Bak, 2012; Jang, 2014). Normally, South Korean classrooms consist of students at different levels. This outcome is contrary to a previous study (e.g., Papathanasiou, 2009) which suggested that semantic clustering impeded vocabulary learning only at the beginning-level.

Up to presently, several other studies have attended to other variables (e.g., Ishii, 2014; Wilcox & Media, 2013). In Ishii's (2014) study, the visual similarity of target words was considered a variable for semantic relatedness, and Wilcox and Media (2013) considered the phonological similarly of the target words as a variable because only the beginner-level group improved on the test. Ishii (2014) suggested a visual similarity as a variable, although she admitted that semantically related words were harder to learn than unrelated words. In addition, arguments that semantic clustering can be useful depending on learners' proficiency or when the target words are known have been put forward (Folse, 2004; Nation, 2000). These arguments have expanded the range of possibilities that can influence the results of semantic clustering.

Thirdly, in the same vein, "awareness" of what they are doing was expected to be a variable in the present study. For this purpose, the SR group was notified that they would learn semantically related words during the class session. Despite expectations that semantic clustering could be used as a strategy by mature learners, it appears that "awareness" is not a variable because the test results were consistent with those of previous studies (e.g., Bak, 2012; Jang, 2014).

6. CONCLUSION

6.1. Implications

According to the results of the present study, presenting words in semantic sets for second language vocabulary learning appears to be less beneficial than presenting them in semantically unrelated sets in terms of short term retention. First, this finding suggests that the widespread practice of using semantic clustering in English language teaching materials should be reconsidered. Teaching materials should follow the results of research findings, yet many in-use L2 course books and textbooks present new words under a set of shared common semantic characteristics (Erten & Tekin, 2008; Tinkham, 1993). It is important that second language teachers, material developers, and curriculum designers know which direction to follow in presenting words to maximize the potential of learning target words. Accepting this principle, materials writers should reflect the principle of providing new words in semantically unrelated sets.

In the same vein, to minimize the effects of interference, previous studies introduced several alternatives. First, thematic clustering was proposed instead of semantic clustering (Bolger & Zapata, 2011; Erten & Tekin, 2008; Folse, 2004; Tinkham, 1997; Waring, 1997). According to Tinkham (1997), in thematic clustering words share a common theme (e.g., *frog, pond, hop, swim, green,* and *slippery*), and the associative strength between the words is weaker than with semantic clustering because the words come from different parts of speech. This type of clustering is better suited to the majority of current ELT programs.

Another proposed alternative to minimize interference is to teach semantically grouped words at different times (Nation, 2000). Nation assumed that one cause of the difficulty of learning semantically related words is due to the different frequencies of the words. He suggested that if words can be learned according to frequency, frequent words first and less frequent words later, learners will be able to learn more easily and without confusion. The role of teachers, material developers, curriculum designers, and even the materials per se, is crucial in minimizing interference.

6.2. Limitations and Suggestions for Further Study

It is necessary to address some limitations in order to generalize the findings of the present study. Firstly, this study tested only receptive knowledge of the target vocabulary. As Nation (2013) pointed out, what it means to know a word includes knowing the receptive vocabulary use and productive vocabulary use. In addition, in previous studies, there have been different results with different data collection methods. Therefore, testing both areas of word knowledge may result in a more comprehensive understanding of the issues. Secondly, the participants were quite young. Initially, this experiment, as a replication study, was designed to identify the effect of semantic clustering on secondary school students. From this perspective, the representability of secondary school students went unfulfilled because the participants were in the first year of middle school. As a result, more varied techniques to measure the participants' proficiency and a wider range of participants are needed in future studies. In addition, to collect more data about the participants' impressions toward the different presentation techniques, in-depth interviews are needed.

REFERENCES

Adelson-Goldstein, J., & Shapiro, N. (2008). *Oxford picture dictionary.* New York: Oxford University Press.

Aitchison, J. (2003). *Words in the mind: An introduction to the mental lexicon* (3rd ed.). Oxford: Blackwell.

Ausubel, D. (1963). *The psychology of meaningful verbal learning*. New York, NY: Grune & Stratton.

Baddeley, A. D. (1990). *Human memory theory and practice*. Hillsdale, NJ: Erlbaum.

Bak, S. Y. (2012). The effects of semantic relatedness on EFL vocabulary recall and retention. *English Teaching, 67*(3), 281-306.

Baleghizadeh, S., & Naeim, M. Y. (2011). Enhancing vocabulary retention through semantic mapping: A single-subject study. *Language Society and Culture, 32*, 11-16.

Bloomfield, L. (1989). *Collins junior thesaurus*. London, England: Collins Educational.

Bolger, P., & Zapata, G. (2011). Semantic categories and context in L2 vocabulary learning. *Language Learning*, *61*, 614-646.

Campion, M. E., & Elley, W. B. (1971). *An academic vocabulary list*. Wellington: NZCER.

Carter, R., & McCarthy, M. (1988). Lexis and structure. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 18-38). Harlow, UK: Longman.

Channel, J. (1981). Applying semantic theory to vocabulary teaching. *ELT Journal, 35*(6), 115-122.

Chung, S. R. (2013). *Effects of presenting semantic/non-semantic category word sets on elementary English learners' vocabulary acquisition*. Unpublished master's thesis, Seoul National University of Education, Seoul.

Coady, J. (1997). L2 vocabulary acquisition through extensive reading. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 225-237). Cambridge: Cambridge University Press.

Erten, I. H., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically-unrelated sets. *System, 36*(3), 407-422.

Finkbiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics*, *24*, 369–383.

Folse, K. S. (2004). *Vocabulary myths*. Ann Arbor, MI: University of Michigan Press.

Gairns, R. & Redman, S. (1986). *Working with words*. Cambridge: Cambridge University Press.

Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning, 46*, 643-679.

Hashemi, M. R., & Gowdasiaei, F. (2005). An attribute-treatment interaction study: Lexical-set versus semantically-unrelated vocabulary instruction. *RELC Journal, 36*(3), 341-361.

Higa, M. (1963). Interference effects of intralist and word relationships in verbal learning. *Journal of Verbal Learning and Verbal Behavior, 2*, 170-175.

Hoshino, Y. (2010). The categorical facilitation effects on L2 vocabulary learning in a

classroom setting. *RELC Journal, 41*(3), 301-312.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.

Hunt, R. R., & Elliot, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: Learning, Memory and Cognition, 109*, 49-74.

Hunt, R. R., & Mitchell, D. B. (1982). Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8*, 81-87.

Ishii, T. (2014). Re-examining semantic clustering: Insight from memory models. *The Language Teacher, 38*(5), 27-30.

Jang, H. J. (2014). The effects of semantic clustering on EFL young learners' vocabulary learning. *English Teaching*, *69*(3), 25-37.

Jones, F. R. (1995). Learning an alien lexicon: A teach-yourself case study. *Second Language Research, 11*, 95-111.

Lawson, M. J., & Hogben, D. (1996). The vocabulary-learning strategies of foreign-language students. *Language learning*, *46*(1), 101-135.

Lehrer, A. (1974). *Semantic fields and lexical structures*. Amsterdam: North-Holland.

Marzano, R. J., & Marzano, J. S. (1988). *A cluster approach to elementary vocabulary instruction*. DE: International Reading Association.

McGeoch, J. A., & McDonald, W. T. (1931). Meaningful relation and retroactive inhibition. *American Journal of Psychology, 43*, 579-588.

Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching and Linguistics, 13*, 221-246.

Morin, R., & Goebel, J. (2001). Basic vocabulary instruction: Teaching strategies or teaching words? *Foreign Language Annals, 34*(1), 8-17.

Nation, I. S. P. (2000). Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal, 9*, 6-10.

Nation, I. S. P. (2013). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

O'Sullivan, J. C. (2005). *The Heinle picture dictionary.* Boston, MA: Heinle ELT.

Papathanasiou, E. (2009). An investigation of ways of presenting vocabulary. *ELT Journal, 63*(2), 313-322.

Porte, G. (1988). Poor language learners and their strategies for dealing with new vocabulary. *ELT Journal*, *42*(3), 167-172.

Ramachandran, S. D., & Rahim, H. A. (2004). Meaning recall and retention: The impact of

the translation method on elementary level learners' vocabulary learning. *RELC Journal*, *35*(2), 161-178.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Rodríguez, M., & Sadowki, M. (2000). Effects of rote, context, keyword, and context/keyword methods on retention of vocabulary in EFL classrooms. *Language Learning*, *50*(2), 385-412.

Sanaoui, R. (1995). Adult learners' approaches to learning vocabulary in second languages. *The Modern Language Journal*, *79*(1), 15-28.

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 199-227). Cambridge: Cambridge University Press.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, England: Palgrave Macmillan.

Stewart, J., & White, D. A. (2011). Estimating guessing effect on the vocabulary levels test for differing degrees of word knowledge. *TESOL Quarterly, 45*(2), 370-380.

Stoller, F., & Grabe, W. (1993). Implications for L2 vocabulary acquisition and instruction from L1 vocabulary research. In T. Huckin, M. Haynes & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 25-45). Norwood, NJ: Ablex Publishing Corporation.

Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System, 21*(3), 371-380.

Tinkham, T. (1997). The effects of semantic and thematic clustering of second language vocabulary. *Second Language Research, 13*(2), 138-163.

Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System, 25,* 261-274.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*, 48–65.

Wilcox, A., & Medina, A. (2013). Effects of semantic and phonological clustering on L2 vocabulary acquisition among novice learners. *System, 41*, 1056-1069.

Wilkins, D. A. (1972). *Linguistics in language teaching.* London: Arnold.

APPENDIX
Vocabulary Test Sample

진단평가

( )학년 ( )반 ( )번 이름:_____

이 평가지는 이제부터 시작할 단어 공부를 위해 여러분이 아는 단어를 제외하기 위한 것입니다. 그러므로 모르는 단어가 많이 나오는 것이 정상입니다. 편안한 마음으로 꼼꼼하게 한 문제씩 풀어 봅시다.

☞ 주어진 단어를 보고, 알맞은 영어단어를 <보기>에서 찾아 번호를 쓰세요.

보기		
1. business		
2. clock	____	벽
3. horse	____	말
4. pencil	____	연필
5. shoe		
6. wall		

☞ 다음과 같이 답하세요.

보기		
1. business		
2. clock	6	벽
3. horse	3	말
4. pencil	4	연필
5. shoe		
6. wall		

<보기>에 있는 단어 중 몇 개는 오른쪽에 해당하는 뜻이 없습니다. 이 문제에서는 business, clock, shoe 가 뜻이 주어지지 않은 단어입니다.

☞ 문제는 다음 페이지부터 시작됩니다.

Applicable levels: Secondary

Jun-Hyeok Choi
Department of English Teaching
International Graduate School of English
81 Yangjae-daero 17-gil, Gangdong-gu
Seoul 05408, Korea
Phone: 02-6477-5114
Email: choi3427@igse.ac.kr

Hyun-Sook Chung
Department of English Teaching
International Graduate School of English
81 Yangjae-daero 17-gil, Gangdong-gu
Seoul 05408, Korea
Phone: 02-6477-5114
Email: sook@igse.ac.kr