Modern English Education, Vol. 16, No. 4, Winter 2015
http://dx.doi.org/10.18095/meeso.2015.16.4.04

Appropriate Numbers of Citations for Korean High School Students' Data-Driven Learning*

Sooin Chun
Dankook University

Chun, Sooin. (2015). Appropriate numbers of citations for Korean high school students' Data-Driven Learning. *Modern English Education, 16(*4), 65-90.**

As a part of basic research for developing concordance materials for young EFL learners, this study explores appropriate numbers of concordance citations contained in the materials, given that secondary EFL students successfully undertake Data-Driven Learning (DDL) in the classroom. In past decades, there have been increasing demands regarding *deductive DDL* for language learners at low proficiency levels. However, few practical suggestions have been made how to develop the materials. This current study focuses on the importance of suitable materials for young learners and suggests guidelines for materials development in terms of citations limited to classroom activity. It followed three Korean teachers of English and 78 high school students. Students' learning outcomes were qualitatively analyzed using DDL materials developed with different numbers of citations, and interview transcripts. Based on students' findings and preferences, it was concluded that 12 citations were most appropriate for the classroom activity studied. Three participant teachers stated that the numbers were dependent on learning settings. The study provides practical information on appropriate numbers of citations at least required for DDL activity in the secondary classroom.

**[DDL/concordance/materials development/inductive learning/COCA
데이터추론학습/콘코던스/학습자료개발/귀납적 학습/현대미국영어코퍼스]**

I. INTRODUCTION

Over the past decades, corpora have had a strong influence in diverse areas of language education such as classroom concordancing, syllabus design, materials development,

* The present research was conducted by the research fund of Dankook University in 2015.

language testing, and interlanguage research (Xiao & McEnery, 2005). Regarding corpus-based approaches, a number of experimental studies have been conducted on classroom concordancing, which Johns (1991, 1994) suggested be named Data-Driven Learning (DDL). Classroom concordancing activities are associated with learners' direct use of corpus data, either undertaking hands-on concordancing or employing ready-made concordance materials. Technically, pedagogical effects of DDL can be evaluated for two aspects of learners' improvement: 1) enhancement of language knowledge, and 2) development of cognitive skills. In terms of advantage, S. Chun's study (2011) concluded that the latter approach is more helpful within a short testing period. This approach seems to be critical for young learners who are expected to grow in intellectual balance, i.e., in cognitive abilities as well as language skills.

DDL cannot be viewed solely as a language learning activity or task for young learners. It is also important to develop balanced intelligence through language learning, using diverse learning techniques. DDL can be used as an individual work activity that develops learners' autonomy, and as a group task which can provide educational benefits of cooperative learning. Since improving critical thinking skills has recently been an essential teaching goal of schools, DDL can also be adopted into the students' syllabus or curriculum.

However, despite positive results shown in a number of previous studies for adults, DDL has not been used much as a learning method for young learners in EFL classrooms, and therefore cognate research has been rare. Lack of use can in part be attributed to burdens that language learners face during concordance tasks. First, corpus texts are too difficult for young ESL or EFL learners (Chamber, 2007); second, the learning process takes too much time and energy for them (Farr, 2008); last, it requires full knowledge of concordancing tools (Bernardini, 2002; Flowerdew, 2012; Gavioli, 2001). Accordingly, some studies (Boulton, 2008, 2009a, 2009b, 2010, 2012) suggest "deductive DDL" (Cresswell, 2007, p. 270) using paper-based materials, which is indirect in comparison with hands-on concordancing, for language learners at low proficiency levels. S. Chun (2014a, 2014b), however, insists that suitable concordance materials should be made available to undertake this DDL activity successfully. Despite the importance of guidelines for teachers and educators to develop suitable materials, there has been no further research that is theoretically supported on this issue.

In line with her previous studies (S. Chun, 2014a, 2014b), this research explores basic needs to construct concrete guidelines concerning how to develop DDL materials for young learners. Whereas the previous studies suggested how content of materials should be within texts, the current study focuses only on structure and numbers of concordance citations using target words, as the external conditions of DDL materials. As a pilot study, this research developed some materials using *Frequency List* first, and then analyzed appropriate numbers of citations for students' effective learning in the classroom. It also

investigated basic needs from both teachers and students toward the developed materials. This study examined the issues noted below.

1) When students undertook DDL effectively, how many citations did they use?
2) What were students' preferences for numbers of citations during the activity?
3) What were teachers' perspectives regarding different numbers of citations?

II. LITERATURE REVIEW

1. Effectiveness of Data-Driven Learning (DDL) for Young Learners

Cognitive enhancement in teaching and learning has now become mainstream. Dewey (1933) initially provided diverse perspectives on human thinking, such as the stream of consciousness, imagination and mindfulness, which are usually restricted to things or individual belief and can be reflective as a chain of thought leading to a conclusion. These perspectives influenced diverse fields of education over the past 50 years. Since Bloom's study (1956), many others (Adey & Shayer, 1994; Ashman & Conway, 1997; Corno & Randi, 1999; de Corte, 2000; Elliott, 2000; Hamers & Overtoom, 1997; Leat & Higgins, 2002; Lipman, 2003; Newton, 2000; Wood, Bruner & Ross, 1976; Zimmerman, 2000) have focused on learners' cognitive process for the improvement of teaching, learning, and assessment. Many educators and teachers (Adams & Wallace, 1991; de Corte, Verschaffel & van de Ven, 2001; Fuchs et al., 2003; Leat, 1998; Meyer, Young & Bartlett, 1989; Shayer & Adey, 2002; Taba, 1966; Wallace, 2001) have also suggested new curriculum and educational methods that develop learners' cognitive abilities in different areas of school subjects. Those methods employing critical thinking skills have also been highlighted in language education. They focus on overall areas of language learning: listening, speaking, reading, and writing as language skills, and grammar and vocabulary as language knowledge. In line with the shift in instructional goals and the development of information and technology, DDL has become a new language learning method. Some studies (Bernardini, 2002, 2004; Gavioli, 2001; O'Sullivan, 2007; Sun, 2003; Wolff, 1997) have also been conducted regarding how to use it to develop cognitive abilities.

The aspect of DDL, related to cognitive skills use, deserves more attention from educators and teachers in the secondary classroom. Since young students are expanding cognitive abilities as well as language knowledge, DDL can assume an important role in their learning. It can usefully be adopted within a classroom activity either as individual or group work depending on the design of materials. The critical role of DDL for young learners is also that it requires both problem-solving skills and critical thinking. Problem-

solving skills include the process of finding and recognizing problems, hypothesizing rules and testing them (Beyer, 1988), while critical thinking abilities are related to solving problems and making decisions (Stiggins, 1987). In line with these classifications, DDL is also understood as problem-solving tasks (Bernardini, 2001) or reasoning-gap tasks (Prabhu, 1987). In Marzano and Kendall's (2007) classification, it can also achieve all six learning objectives that are from cognitive, meta-cognitive, and self-learning. Classroom use of DDL can therefore be important to achieve educational objectives for overall growth in intelligence in young learners.

Considerations regarding learners' thinking abilities by some prominent educationists (Ashman & Conway, 1997; Beyer, 1988; Bloom, 1956; Lipman, 2003; Stiggins, 1987) have become more important in educational fields. DDL, which focuses on inductive thinking skills, is therefore assuming a critical role as a new and effective language learning method. DDL can assist younger learners to achieve many extensive goals of education. However, despite such benefits of DDL for young learners, it still has not gained much attention from educators or researchers. Only a few studies (Cheng, Warren & Xun-feng, 2003; S. Chun, 2011; Kennedy & Miceli, 2001; Sun, 2003) focus on learners' cognitive skills and their learning strategies during concordancing activities.

Based on these studies noted above, DDL should be highlighted, particularly for young learners who need to enhance their abilities yet keep them in balance. This is because DDL can assume an important role to assist them in diverse areas: cognitive and meta-cognitive abilities, information and application knowledge. However, it is necessary that DDL be effectively implemented as a syllabus into classrooms for novice learners. To do this well, appropriate methods by educators and researchers must first be developed and later introduced.

2. DDL Materials for Low-proficiency Language Learners

In principle, DDL is inductive learning, but it can be viewed by two different approaches. Deductive or indirect DDL means that a teacher mediates learners' concordancing using paper-based materials, whereas hands-on concordancing, conducted in the learner's direct use of concordance software is viewed as inductive DDL (Boulton, 2009a, 2009b; Cresswell, 2007; Johns & King, 1991). Since inductive DDL comprises classroom activities and tasks through which learners identify language rules by directly observing the concordance output, some studies (e.g., Cobb, 1997, 1999) note that it is more effective for learners to take a role as a language researcher. On the other hand, some (Breyer, 2006; S. Chun, 2012; Granger & Meunier, 2008: Krishnamurthy & Kosem, 2007; Osborne, 2004; Romer, 2006) mention the impediment of DDL and emphasize the necessity of ready-made and user-friendly tools for teachers and language learners. In keeping with their notion,

other studies (Boulton, 2008, 2009a, 2009b, 2010, 2012; Cresswell, 2007; Johns & King, 1991) suggest paper-based DDL materials for the low-proficiency language learners. This type of DDL can also be viewed as "a guided approach or a combination of an inductive and deductive approach where the elements of explanation and corpus use are tailored to the needs of the student" (Johansson, 2009, p. 41).

Despite the importance of the materials for successful DDL, there have been few studies that discuss what the materials should be and how they should be developed. For young language learners, S. Chun (2014a, 2014b) has recently suggested some concrete guidelines for producing effective DDL materials. Her study (S. Chun, 2014a) notes that teachers, who can be the materials developers, must first understand the characteristics of corpora and then collect concordance examples for target words. She also notes that citations should consist of similar characteristics within a general corpus, which together emphasize the representativeness and balance of collected texts. Particularly for novice learners to learn general language rules or knowledge, appropriate materials should contain citations from diverse genre and include different part of speech (POS) of the target word. They should also contain numbers of them retaining similar ratios of frequency in each genre as the source corpus. If the same numbers of concordances from each genre were collected, students might not know how differently in frequency the target word is used in each genre. Based on these suggestions, S. Chun outlines overall procedures to develop *deductive DDL* materials (Figure 1).

However, her studies above suggest the content of concordance citations using a target word/chunk in the material, but not its external condition, which is the number of citations. It is important to know what amounts of text are appropriate so that learners undertake it effectively and successfully within limited learning conditions. In principle, sufficient concordance samples are useful for inferring language rules, but constraints of time and potential fatigue exist when both teachers and learners conduct DDL in the classroom. Such issues on numbers of citations are related to DDL usage rather than theoretical considerations, and are limited to effectiveness of teaching and learning time management. Therefore, this study approaches the issue regarding appropriate numbers of citations, based on needs of both teachers and learners who are directly involved in DDL.

3. Online Tool for Developing DDL Materials

Easy and user-friendly tools are essential for teachers and DDL material developers.

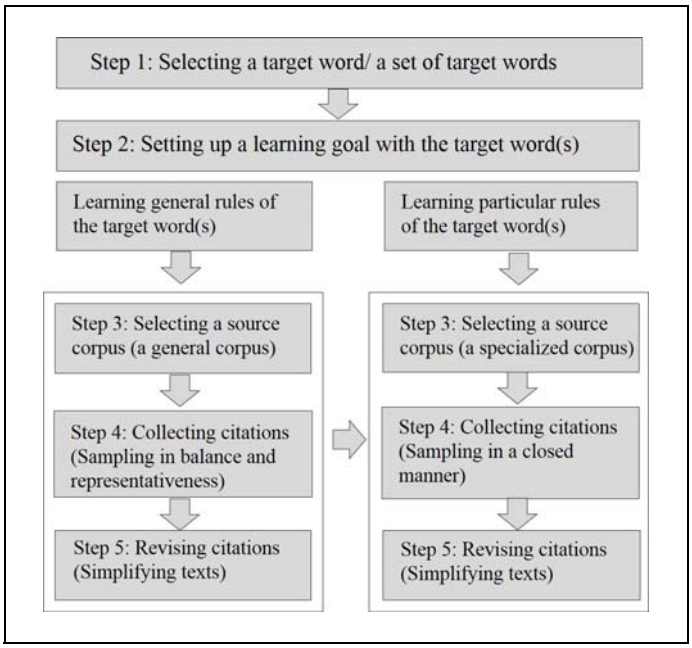

FIGURE 1 How to Develop DDL Materials

Frequency List (FL)[1] is an online tool that profiles vocabulary frequency levels. The site allows users to see detailed information on the top 60,000 words (lemmas) of English, based on data from Corpus of Contemporary American English (COCA)[2]. It shows the overall frequency for each word, as well as the frequency of words in different kinds of English—spoken, fiction, magazines, newspapers, and academic writing. For each word the user can also find the 20-30 most frequent collocates and see 200 or more concordance lines. It also displays a list of synonyms and words with more specific and more general meanings, and only by clicking the related area, the user can see the entries for those related words as well (Figure 2). FL features are beneficial for developing DDL materials since they display essential types of vocabulary knowledge of a target word in different forms and meanings. These include essential information on the targets: 1) all types of POS, 2) frequencies by genre, and 3) meaning groups, in a single screen (Figure 2). These features are easy for teachers and beginner corpus users to apply, and are helpful even for vocabulary learners. According to S. Chun (2014a, 2014b), DDL materials should include characteristics of the source corpus, i.e., a general corpus, so that they learn diverse types of vocabulary knowledge of the target word. In other words, the material for a target word

[1] Frequency List: http://www.wordandphrase.info/frequencyList.asp
[2] Corpus of Contemporary American English: http://corpus.byu.edu/coca/

can be understood as a micro-sized general corpus.

Teachers might not be knowledgeable regarding corpora. Even if they are, they are busy with classroom management. The convenient, user-friendly features of FL can help them, not only reduce their workload but also produce decent materials which satisfy the requirements noted by S. Chun. Accordingly, beneficial functions of FL were used to develop DDL materials for this study.

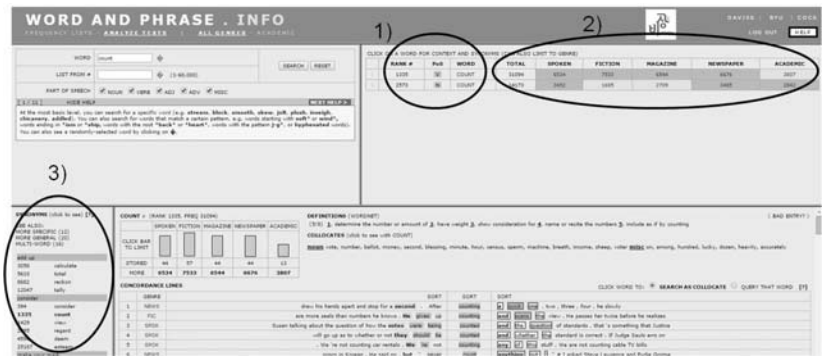

FIGURE 2 The Feature of *Frequency List*

III. METHOD

1. Participants

The study followed three high school English teachers and their 78 high school students in Gyeonggi Province in Korea. Participant teachers, who were interested in novel techniques for vocabulary teaching but had no particular knowledge regarding DDL, volunteered for the study. They were also expected to have had at least three years of teaching experience, so they could understand the activities related to this study and handle them competently in class. Teachers 1, 2 and 3 have five, seven, and ten years of teaching experience respectively. They conducted a DDL activity with their students, a total of 78 male high school juniors, who had won a grade of Average in the *National Level Test*, administered by Korea Institute for Curriculum and Evaluation. These volunteer students were classified into three groups of 26, A, B, and C, and undertook the activity in after-classes. Both teachers and students were also interviewed on their preferences toward the DDL materials developed for this study.

2. DDL Materials Development

Three different types of DDL materials were used. Based on S. Chun's (2014a, 2014b) guidelines, the materials were developed using different numbers of concordance citations for a single target word. According to her studies, the citations within the learning materials for each target word could not be collected in the same numbers from each genre and for each different form in order to retain "representativeness" (Biber, 1993, p. 243) of the source corpus. However, for this study, the numbers of citations were controlled using the fixed numbers, 7, 12, and 17, respectively. However, for this study, based on the maximum number suggested by teachers as feasible to manage in class, three were randomly chosen, the fixed numbers 7, 12 and 17. Since the current study initially attempts to explore appropriate numbers of citations for the materials, they were done within a narrow range of choice, given teachers' and students' burden of conducting DDL within the classroom. The materials were then developed using the procedure outlined below.

1) Selecting Target Words

For the DDL materials development, *count* was chosen as a target word. This was selected based on data from an earlier survey on vocabulary teaching with 30 secondary school teachers. In the survey, they were asked questions about target words to be taught and 22 answered that words frequently shown in textbooks should first be taught intensively, rather than in other lexicons. They also recommended 18 different words for their students including single words and synonymous word pairs. It was chosen based on their suggestions and the researcher's consideration for ESL student's learning difficulty. When the vocabulary level of *count* is checked using *Frequency List*, it is in diverse frequency ranges in COCA and has different part of speech (POS), i.e., verb and noun in 1K and 2K respectively. This selection was lastly confirmed by three participant teachers as a teachable word in their classroom.

2) Collecting Concordance Samples

Seventeen concordances for the target word were collected from COCA, to enable learners to infer general and diverse language rules as much as possible. When sampling concordances from the source corpus, *Frequency List* (FL) was used. As shown in Figure 3, the table in FL for the numbers of sampled COCA concordances was utilized to develop a set of different numbers of citations for a target word. The 17 citations collected from FL is a reduced number to the extent that it contains similar characteristics of samples in COCA, as shown in Table 1. Selected citations were then re-arranged as reduced numbers—to 12

and 7, respectively, in order to develop a new set.

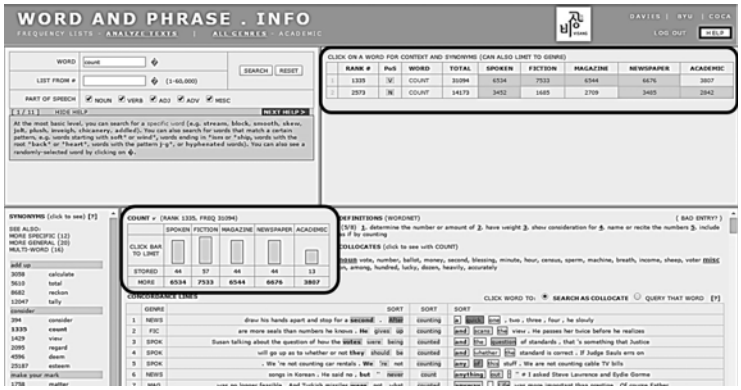

FIGURE 3 FL Displaying Reduced Numbers of Concordance Samples from COCA

TABLE 1

Reduced Numbers of Citations from COCA for *count*

Genre	Stored in COCA (verb)	Stored in COCA (noun)	Stored in FL (verb)	Stored in FL (noun)	Collected in the material (verb)	Collected in the material (noun)	Total in the material
Spoken	6534	3452	44	55	3	1	**4**
Fiction	7533	1685	57	24	3	0	**3**
Magazine	6544	2709	44	34	2	1	**3**
Newspaper	6676	2842	44	55	2	2	**5**
Academic	3807	3485	13	34	2	1	**2**
Total	31094	14173	202	202	12	5	**17**

3) Simplifying Texts

Once the citations were collected, they were simplified into structures to help students understand the texts easier. Based on Nation's (1990) suggestion regarding how to simplify the language, the texts were revised subject to the limits of keeping the authenticity of the original texts. All spoken citations were given double quotation marks so that students recognized a spoken mode (Example 1). Unnecessary modifiers were eliminated (Example 2); difficult proper nouns were replaced by easier or more common ones, or by pronouns (Example 3), and vocabulary at the low frequency level was replaced by words at a higher level (Example 4). Examples are below:

(1) Original: 2000: SPOK: CNN_SatMorn (source): …*are we going to apply? How*

are we going to count ? In their pleadings, the Gore campaign has said…

Revised: "How are we going to count these ballots"

(2) Original: 2008: MAG: Nation Parks (source): …*After several doses of chemotherapy, his white-blood-cell count dropped precipitately*. …

Revised: The cell count dropped quickly.

(3) Original: 1954: MAG: Outdoor Life (source):...*like Beardsley, he counts on his binoculars. We hunt a lot of black timber* …

Revised: He counts on his binoculars to find the trees.

(4) Original: 2001: NEWS: San Francisco Chronicle (source): … *Pakistan is counting on substantial economic aid* ….

Revised: The country is counting on major financial help from others.

3. Procedures

Data collection was conducted in three steps. In the first step, each volunteer teacher undertook a DDL activity for each group of students in after-school classes. The students inferred language rules based on concordance examples given in the material, and then wrote a report regarding their findings or conclusions. Group A used the material, A-1 (7 citations); Group B used A-2 (12 citations); Group C used A-3 (17 citations). The teachers provided students only with brief directions given in the material regarding how to undertake the activity for 30 minutes. In the second step, they were then given a whole set of A (Appendix 1) and were surveyed regarding their preferences concerning different amounts of citations for the target word. They were asked to mark the number of citations that they thought would be best for the activity, and then briefly describe reasons for their choices. Last, the three teachers who had conducted the classroom activity and six students—two sampled from each group—were individually interviewed in depth on their opinions over the materials and the activity. Data collections were obtained following consent from the students, their teachers, and the school administration, after acknowledgment and reassurance that any information obtained would be used anonymously and restricted to this study.

4. Data Analysis

Students' reports were qualitatively analyzed to investigate their DDL learning. There were 78 reports in total containing 1220 words; an average of 15 words per report were coded by the researcher and by two other trained coders. They worked independently and, then discussed discrepancies among their results together. The finalized coded data were classified and summarized into tables in Appendix 2. The rationale for adopting a

qualitative approach in this study was that both the students' and the teachers' individual responses as the material users could be investigated in depth. Furthermore, the qualitative analysis of the data provides detailed information on the reasons of the participants' failure in the recognition of all the items and their preference of the particular numbers of citations, which would be useful for the development of DDL materials in the future.

Students' descriptions on findings or conclusions in the reports were coded and classified into 1) *forms*, including POS and affix, 2) *meanings*, including definitional meanings, register, and associations, and 3) *application*, including lexical chunks and frequency. The classifications were based on the framework of vocabulary knowledge of Nation (2001) and Schmitt (2000). Cases of faulty reasoning were also counted as findings since they recognized those as types of word knowledge from the concordances given. However, since the students' descriptions regarding language rules were simple and general, their findings were roughly grouped and placed under the categories of form, meaning, and application. For example, in Figure 4, student's descriptions were coded in four items under the categories: two for different POS, and two for meanings of each POS. The numbers of students' findings were then counted and summarized using absolute scales as shown in Tables A, B and C (see Appendix 2 for more details). Statistical data in these tables were used only for supplementary purposes, in order to show an overview of their performance.

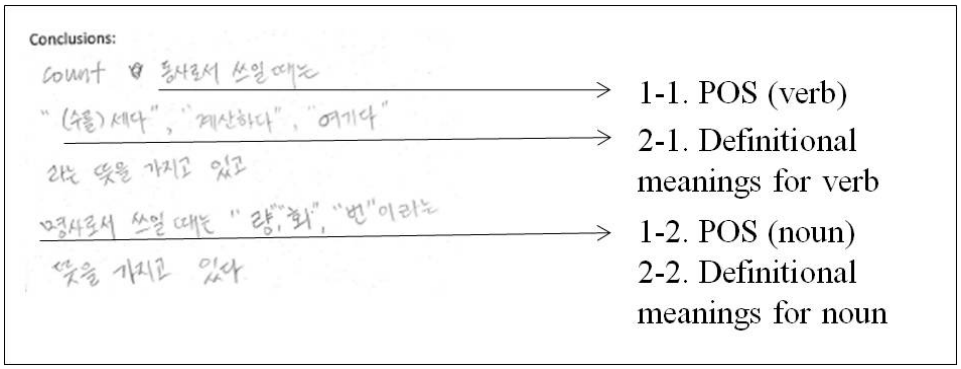

FIGURE 4 An Example of Coding for Students' Findings

After completing the activity, students were surveyed on their preferences for material types, using a whole set of material including three different numbers of citations. Their responses were observed and summarized using a table (Table D in Appendix 2). They were also used as supplementary data to examine their needs for the materials. Lastly, each teacher was interviewed for 15 minutes regarding appropriate numbers of citations and difficulties they had faced during the activity.

IV. RESULTS

1. Students' Findings by Types of Materials

Table A shows what types of vocabulary knowledge students recognized when using the developed DDL materials. According to the data, there is a slight difference among the numbers of findings for each group. Group 1 described a total of 85 items; Group 2 had a total of 99; Group 3 had a total of 101. The averages of the findings are 3.2, 3.8 and 3.9 items respectively. The numbers of findings per student vary from two to five items, and not all items shown in each material were recognized and described. There is no correlation between the length of descriptions and the numbers of findings. For example, a student described his findings with 35 words but mentioned only the different meanings of verb and noun of the target word, while another had about five items with 15 words. Throughout three groups, items which students recognized most were forms and meanings of the target word.

Table B shows types of word knowledge and the numbers of students in each group who recognized each type. According to the coded data, POS and definitional meanings of the verb were recognized most, while phrasal verbs and frequency were least. In effect, form knowledge was most easily recognized and application knowledge was the least. In addition, participle, phrasal verb, and frequency were described by only a few students throughout three groups. Unlike Groups 2 and 3, Group 1 recognized the smallest numbers of each item except those related to verb.

Table C shows the distributions of students who recognized different numbers of items in each group. The numbers of students who recognized five items from Groups 1, 2, and 3 are four, six, and seven respectively; the numbers of students who had four items are six, 11, and 12; the numbers for three items are nine, six, and four; the numbers for two items were seven, three, and three. Given the data, the numbers of students who recognized the least and the most numbers of items were not affected by different numbers of citation, whereas those who recognized four or three items were dependent on them.

Following the above overview of students' performance, their reports were qualitatively examined in depth with respect to how the materials affected the students' performance as below:

1) Performance of Group 1

Group 1 using A-1 (7 citations) recognized the least numbers of the items. All 26 students recognized the verb form and meaning, whereas many failed to do items related to the corresponding noun. Among a total of 17 students who described five items, only four

recognized all the POS, their definitional meanings, and either participle or phrasal verb (Figure 5). Items regarding the phrasal verb, *count on*, and the participle, *counting* or *counted*, however, were not shown in the reports that described less than four items. In addition, these students were the least among the three groups. The material they used included only two citations showing the participle, whereas A-2 and A-3 had three, and included two citations showing the phrasal verb, whereas the other materials had three and five respectively. Those numbers of citations were even smaller than those for the other items. In the interviews with two students who failed to recognize the phrasal verb, one answered that he had not even found those examples in the material. The other who succeeded in inferring the rule reported that he almost missed it, but finally recognized it since he had learned it. Based on the above, it is understandable that the numbers of citations in A-1 were not enough for the students to recognize the rules easily.

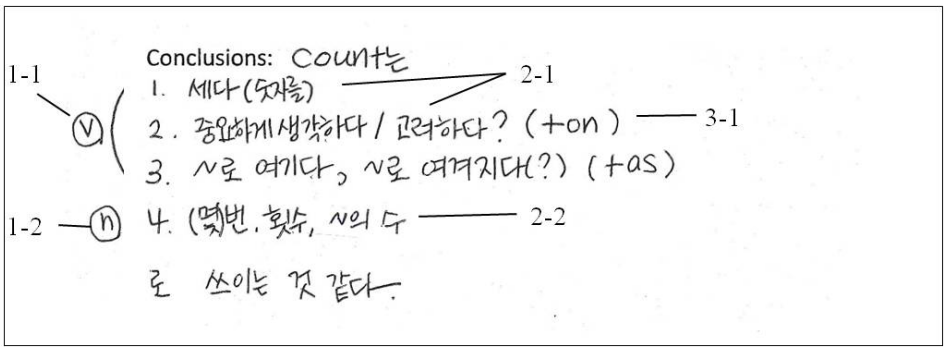

FIGURE 5 A Sample of Students' Findings

2) Performance of Group 2

Group 2 performed better on the activity than Group 1 did, and recognized more items from the material, A-2. According to data in Table D, they did better on the participle, phrasal verb, and frequency. There was an observable change in the numbers of students who recognized overall items when compared with Group 1. In particular, students' recognition on noun form and meanings was noticeably improved. It is assumed that this meaningful change occurred by adding one more citation, No.12. The concordance shows the plural of *count* as well as a number that is quantifying the noun. Three students in this group described the plural, while none did in Group 1. This is observed from the students' reports as shown in Figure 6.

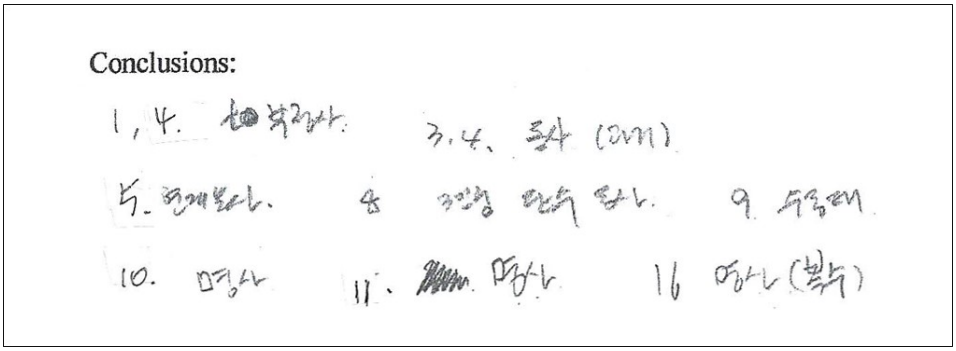

FIGURE 6 A Sample of Students' Findings on Forms

In the interviews with two students in this group, they both stated that they figured out meaning of the noun, *count*, from the citation. It is thereby assumed that the example was useful for students to infer its noun form and meaning easily, so it affected the increase of the total numbers of student who succeeded in reasoning the rules.

3) Performance of Group 3

Performances of Groups 2 and 3 were similar in the total numbers of findings, and the numbers of items that students recognized. Interestingly, even though A-3 contained the most numbers of citations among three materials, students' outcomes did not differ much from Group 2's. However, data for the phrasal verb, *count on,* were unique and outstanding. The most students, five, in Group 3 recognized the phrasal verb form and its meaning. The numbers of citations showing the phrasal verb were the most in A-3. Therefore, it is assumed that these sufficient examples helped students focus on the form easily and inferred the rule including its meaning. In the interviews with two students in this group, they both reported that they could easily discover the pattern in which the phrasal verb was used since "there were many sentences".

Given the summarized data from Tables A, B, and C, the performances of Group 2 and 3 were similar with respect to total numbers and patterns of their findings, whereas Group 1 performed noticeably different. In other words, the biggest numbers, 17 citations, did not improve students' performance, nor did the smallest number, seven citations. Regardless of the types of materials, all students, except one, throughout three groups, recognized forms and meanings of the verb and noun. However, only 20 out of 78 recognized the phrasal verbs. From these results, it is concluded that students did not use all the citations in the material effectively as evidence.

2. Students' Preferences on Numbers of Citations

After completing the activity, students were surveyed to examine their preferences for differential materials. Students in each group were given a whole set of material, that is A-1, A-2, and A-3 and were then asked questions: 1) Among the materials, which do they think is best for their learning them, and give a mark next to the number; 2) why do they think so, and write their other opinions on the material or activity.

Summarized data are presented in Table D to show students' preferences for different materials. Three groups of students gave similar choices for numbers of citations. Those who selected A-1 and A-2 were similar in number within the different groups. Interestingly, Group 3, who had used A-3, also preferred A-1 and A-2 more, though the students who chose A-3 outnumbered those in other groups. The three groups tended to prefer the materials they used for the classroom activity, but in general, preferred smaller numbers of citations regardless of results of their performance. Their descriptions in the reports and individual interviews revealed that they made choices based on their fatigue from reading the texts than on their learning results. It can not be thought that their choices were made based on professional knowledge that teachers have—that students are able to discriminate characteristics of the differential materials in the same way that teachers have.

In their responses, students who chose A-3 stated that they understood the purpose of DDL activity and rationally selected it as the best one (Figure 7). Those who chose A-2 also commented on their rationale for their choice, as shown in Figure 8. Students who chose A-3 also explained the rationale for choosing, briefly, e.g., "they are too many," "it takes too long to read," "it is too hard to read them all," and "they are confusing."

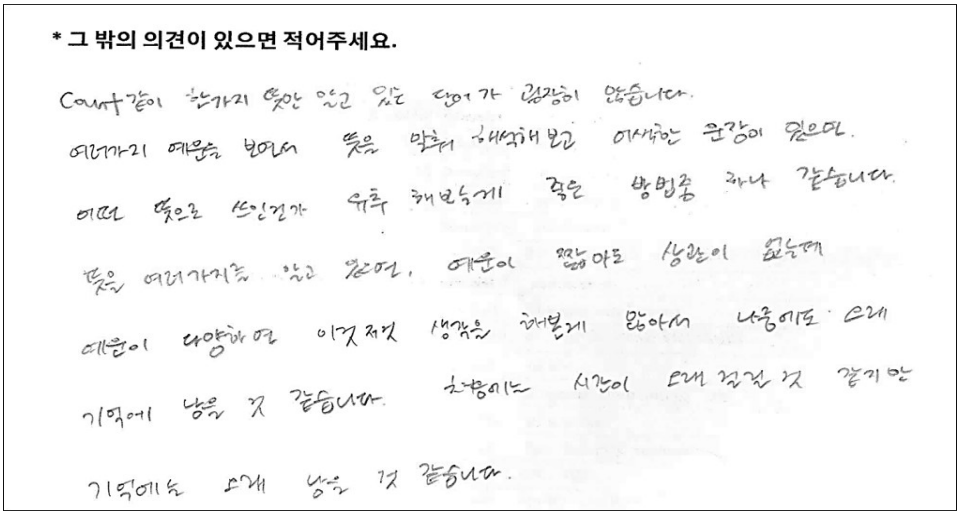

FIGURE 7 A Sample of Student' Comment on the Materials

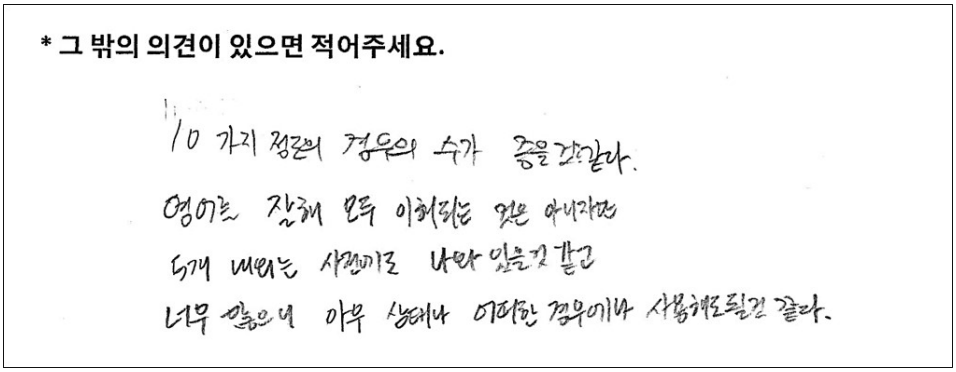

FIGURE 8 A Sample of Student' Comment on the Materials

Given the data from students' reports and survey responses, the material— including the most citations, A-3—was not preferable for overall groups of students. Between Groups 1 and 2, the numbers of students' findings were distinctively different but more students preferred A-2; though students' findings from Groups 2 and 3 were similar, more in Group 3 chose it. It cannot be deduced from this result only that these two materials can be used equally for effective learning. However, the students' comments on A-2 should be regarded as important opinions, since students are the core and the center of main agents of education. Their learning fatigue should be carefully considered when developing the materials.

In the individual interviews after the survey with six students sampled from three groups: three who chose A-1 the best and the other three who chose A-2 the best, students who chose A-1stated as below:

Student 1-1: "I know many texts are good,… but they are too many.… (she laughed)."

Student 2-1: "First time, I didn't understand what we were supposed to do, but later I knew we had to read the texts to find something…they were difficult to read,…. so I wouldn't like to read too many."

Student 3-1: "When I saw many sentences, I thought I hate the activity, but later, it was useful. I think my friends think so too."

The above descriptions imply what students considered about the material when choosing the best, which were the amounts and difficulty of the texts.

3. Teachers' Perspective toward Different Types of Materials

The participant teachers were also given the whole set of the materials—A-1, A-2, and

A-3—to review. In the 15 minute interview, they were then asked questions: 1) What number of material among A-1, A-2, and A-3, do they think are best for their students; 2) If another set of material for a synonymous pair, for example, *win/beat*, which is comprised of B-1(14 citations), B-2 (24 citations), and B-3 (34 citations), what would they choose best; 3) why do they think so; 3) were there any difficulties during the activity?

When the teachers were asked which materials were likely to be used most effectively for the classroom activity, all of them answered those with higher numbers of citations were suitable for group work, while smaller ones were better for individual work. They answered that A-3 and B-3 would be appropriate for group work. However, they differed in responses regarding individual words. Teacher 1, who led Group 1, and Teacher 2, who led Group 2, answered that A-1 and B-1 were appropriate; Teacher 3, who led Group 3, answered that A-2 and B-1 were suitable. The teachers gave the reasons why they chose the best materials as noted below:

> Teacher 1: "….17 citations are too many for my students to handle. They are not good at English much. I think 12 are also too many. It will take too long for them to do the activity in class. Maybe, they are appropriate for group work since students can help each other…."
>
> Teacher 2: "…Some of my students had difficulty reading the texts. I think reading too many texts is difficult for them. They seem to be good for group work, not individual work. And I think a set of target is better than a single target, but texts are still too many and difficult…."
>
> Teacher 3: "….During the activity, I felt 30 minutes were not enough for my students to cover all the texts. I understand this activity is better for learning synonymous words, and the more texts, the better for DDL. But they wouldn't like the activity if there are too many texts in the materials."

According to the above, the teachers' answers were slightly different from the students'. While students preferred smaller numbers of citations regardless of the results of their performance, teachers thought they should be used differently based on types of activity. However, their choices were virtually the same as the students', since both teachers and students chose the numbers in view of learning fatigue.

Regarding difficulties during the class, the teachers mentioned text difficulty of citations, time limits for doing the activity in class, and the teachers' knowledge base in order to manage materials for their classes. Teachers 1 and 3 stated that texts in the materials were somewhat difficult for their students, and class time was insufficient for completing the activity. Teacher 2 emphasized that teacher's knowledge regarding DDL is essential

because a teacher knows the learning goal of the activity and can help students work successfully. The results from all these qualitative data explain why the students and teachers preferred the numbers of citations, and emphasize the learning fatigue when the materials are used.

Based on all the data from students' reports and all the participants' interviews, the results revealed that the numbers of students' findings from each of 12 and 17 citations were similar, whereas those from 7 citations differed from the other two. The results show that using the most numbers of citations did not make their performance better. However, in the survey, students responded that they preferred 7 and 12 citations. Based on both sets of results, it concluded that around 12 citations were appropriate for the students to work and could possibly reduce their fatigue from text reading. However, the teachers differed from their students in that they preferred different numbers of citations within the learning setting, higher numbers of citations for group work, and smaller ones for individual work. It is thereby assumed that teachers considered the effectiveness of learning as well as students' learning fatigue.

The results from students' findings partly show how language learners acquire vocabulary knowledge. According to Nation (1990), when learners guess words from context, nouns and verbs are usually easier to guess than adjectives and adverbs. Bauer and Nation (1993) insist that an understanding of affixes helps learners guess meaning and word classes of the acquired words. In this study, the students also recognized forms of verb and noun more, but were less successful in recognizing the phrasal verb and participles less. In addition, some failed to recognize the noun form as well as its meaning. Interestingly, the tables show that this failure occurred regardless of the types of materials. Alternatively, the tables also reveal that there were focus groups of students whose performance was dependent on types of the materials, i.e., those who recognized four items. They performed the activity better with the most numbers of citations. These summary data are essential to understand any patterns or tendency that might exist in the students' learning.

V. DISCUSSION AND CONCLUSION

The current study aimed to investigate needs from the material users as initial, basic research. Therefore, it focused on the individual responses with respect to the use of materials and examined them in depth, rather than on quantified information. The data from the interview and survey reveal that a particular citation and numbers of citations helped them to perform better in reasoning. They also provided detailed information on why many students failed to recognize all items and preferred the particular numbers of

citations. This information should be seriously considered when developing DDL materials. From the findings above, it is found that potential fatigue in learning affected their choice of material. It can also be deduced that not only the numbers of citations but also the contents of citations should be carefully selected when developing the materials.

This study has evaluated both teachers' and students' needs that arise from learning fatigue when DDL materials are used in the classroom, as essential factors which be considered during DDL in the classroom. This fatigue might negatively affect students' performance. As a part of fundamental research for developing DDL materials and addressing their needs, the study provides practical information regarding how to place citations into the materials concerning their needs. By focusing on numbers of concordance samples, this research suggests a frame for displaying them. Whereas previous research (S. Chun, 2014a, 2014b) studied the content of DDL materials—types of texts related to genre and frequency in usage—it did so based on external conditions. Ample citations can work for valid inferences regarding general language rules. However, as the literature review has argued, these are not likely to be practical and effective for young learners at low English proficiency levels. Learning circumstances such as time limits and poor language skills might not allow lower level students to use the materials with many texts effectively in the classroom or even out of class.

To address existing and potential learning gap issues, this study has attempted to pinpoint appropriate sizes of concordance samples for successful learning. It has helped interpret probable needs of both teachers and students in terms of citation numbers for DDL materials. However, since the study was conducted with only small numbers of participants from three focused groups, with three different materials with one target word, other needs for maximum numbers could not be considered in this study. Findings therein should not be over-generalized and need to be supported by further research employing groups that are more diverse using more materials. Regardless of the limitations stated above, the needs analysis conducted in this study is an important first step when developing future appropriate materials.

REFERENCES

Adams, H. B., & Wallace, B. (1991). TASC: A model for curriculum development. *Gifted Education International, 7*(3), 104-113.

Adey, P. S., & Shayer, M. (1994). *Really raising standards: Cognitive intervention and academic achievement*. London: Routledge.

Ashman, A. F., & Conway, R. N. F. (1997). *An Introduction to cognitive education: Theory and application*. London: Routledge.

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253-279.

Bernardini, S. (2001). Spoilt for choice: A learner explores general language corpora. In G. Aston (Ed.), *Learning with corpora* (pp. 220-249). Bologna, Italy: Athelstan.

Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 165-182). Amsterdam: Rodopi.

Bernardini, S. (2004). Corpora in the classroom. In J. Sinclaire (Ed.), *How to use corpora in language teaching* (pp. 15-36). Amsterdam: John Benjamins.

Beyer, B. (1988). *Developing a thinking skills program*. Boston: Allynand Bacon.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4), 243-257.

Bloom, B. S. (1956). *Taxonomy of educational objective: The classification of educational goals, Handbook 1: Cognitive domain*. New York: Longmans Green.

Boulton, A. (2008). Looking for empirical evidence of data-driven learning at lower levels. In B. Lewandowska-Tomaszczyk (Ed.), *Corpus linguistics, computer tools, and applications: State of the art* (pp. 581-598). Frankfurt: Peter Lang.

Boulton, A. (2009a). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL, 21*(1), 37-51.

Boulton, A. (2009b). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, *35*(1), 1-28.

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, *60*(3), 534-572.

Boulton, A. (2012). Language awareness and medium-term benefits of corpus consultation. In A. Gimeno Sanz (Ed.), *New-trends in computer-assisted language learning: Working together* (pp. 39-46). Madrid: Macmillan ELT.

Breyer, Y. (2006). My concordance: Tailor-made software for language learners and teachers. In L. Flowerdew (Ed.), *Corpora and language education* (pp. 157-176). Basingstoke, UK: Palgrave MacMillan.

Chamber, A. (2007). Popularising corpus consultation by language learners and teachers. In E. Hidalgo, L. Quereda & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 3-16). Amsterdam: Rodopi.

Cheng, W., Warren, M., & Xun-feng, X. (2003). The language learner as language researcher: Putting corpus linguistics on the time table. *System*, *31*, 173-186.

Chun, Sooin. (2011). Learners as corpus researchers: The process of developing cognitive skills through hands-on concordancing. *Foreign Languages Education, 18*(3), 73-103.

Chun, Sooin. (2012). Effective concordance workshop for corpus learners. *Modern English*

Education, *13(*4), 67-88.

Chun, Sooin. (2014a). How to develop effective paper-based concordance materials for DDL. *Studies in English Education, 19*(1), 127-156.

Chun, Sooin. (2014b). Exploring the usefulness of Frequency List for developing concordance materials. *The Journal of Foreign Studies, 28*, 111-138.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System, 25*(3), 301-315.

Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning, 12*(4), 345-360.

Corno, L., & Randi, J. (1999). A design theory for classroom instruction in self-regulated learning? In C. M. Reigeluth (Ed.), *Instructional design theory and models: A new paradigm of instructional theory* (pp. 293-318). Hillsdale, NJ: Erlbaum.

Cresswell, A. (2007). Getting to 'know' connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo, L. Quereda & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 267-287). Amsterdam: Rodopi.

de Corte, E. (2000). Marrying theory building and the improvement of school practice: A permanent challenge for instructional psychology. *Learning and Instruction, 10*, 249-266.

de Corte, E., Verschaffel, L., & van de Ven, A. (2001). Improving text comprehension strategies in upper primary school children: A design experiment. *British Journal of Education Psychology, 71*, 531-559.

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston: DC Heath and Company.

Elliott, J. G. (2000). The psychological assessment of children with learning difficulties. *British Journal of Special Education*, *27*(2), 59-66.

Farr, F. (2008). Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness, 17*(1), 25-43.

Flowerdew, L. (2012). *Corpora and language education*. Basingstoke, UK: Palgrave MacMillan.

Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., & Schroeter, K. (2003). Enhancing third-grade students' mathematical problem solving with self-regulated learning strategies. *Journal of Education Psychology, 95*, 306-315.

Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (Ed.), *Learning with corpora* (pp. 108-137). Houston, TX: Athelstan.

Granger, S., & Meunier, F. (2008). Phraseology in language learning and teaching: Where to from here. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 247-251). Amsterdam: John Benjamins.

Hamers, J. H. M., & Overtoom, M. T. (1997). *Teaching thinking in Europe*. Utrecht: Sardes.

Johansson, S. (2009). Some thoughts on corpora and second-language acquisition. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 33-44). Amsterdam: John Benjamins.

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal, 4*(1), 1-16.

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 293-313). Cambridge: Cambridge University Press.

Johns, T., & King, P. (1991). *Classroom concordancing*: *English language research journal 4*. Birmingham, England: Centre for English Language Studies, University of Birmingham.

Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology, 5*(3), 77-90.

Krishnamurthy, R., & Kosem, I. (2007). Issues in creating a corpus for EAP pedagogy and research. *Journal of English for Academic Purposes, 6*(4), 356-373.

Leat, D. (1998). *Thinking through geography.* Cambridge: Chris Kingston Publishing.

Leat, D., & Higgins, S. (2002). The role of powerful pedagogical strategies in curriculum development. *The Curriculum Journal, 13*, 71-85.

Lipman, M. (2003). *Thinking in education*. Cambridge: Cambridge University Press.

Marzano, R. J., & Kendall, J. S. (2007). *The new taxonomy of educational objectives*. Thousand Oaks, CA: Corwin Press.

Meyer, B. J. F., Young, C. J., & Bartlett, B. J. (1989). *Memory improved: Reading and memory enhancement across the life span through strategic text structures*. Hillsdale, NJ: Erlbaum.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Massachusetts: Heinle & Heinel.

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

Newton, D. P. (2000). *Teaching for understanding.* London: Routledge-Falmer.

Osborne, (2004). Top-down and bottom-up approaches to corpora in language teaching. In U. Cornnor & T. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective* (pp. 251-265). Armsterdam: Rodopi.

O'Sullivan, I. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL, 19*(3), 269-286.

Prabhu, N.S. (1987). *Second language pedagogy*. Oxford: Oxford University Press.

Romer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future development. *Zeitschrift für Anglistik und*

Amerikanistik, 54(2), 121-134.

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

Shayer, M., & Adey, P. (2002). *Learning intelligence*. Buckingham: Open University Press.

Stiggins, R. J. (1987). Design and development of performance assessments. *Educational Measurement: Issues and Practice, 6*(3), 33-42.

Sun, Y. (2003). Learning process, strategies and web-based concordancers: A case study. *British Journal of Educational Technology, 34*(2), 601-613.

Taba, H. (1966). Teaching strategies and cognitive functioning in elementary school children (The U.S. Department of Health, Education, and Welfare, Cooperative Research Project No. 2404). San Francisco: San Francisco State College.

Wallace, B. (2001). *Teaching thinking skills across the primary curriculum: A practical approach for all abilities*. London: David Fulton Publishers.

Wood, D. J., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89-100.

Wolff, D. (1997). Computers as cognitive tools in the language classroom. In A. K. Korsvold & B. Rüschoff (Eds.), *New technologies in language learning and teaching* (pp. 17-26). Strasbourg, France: Council of Europe.

Xiao, Z., & McEnery, T. (2005). Corpora and language education. In B. Lewandowska-Tomaszczyk (Ed.), *PLC 2001: Practical applications in language corpora* (pp. 479-492). Frankfurt: Peter Lang.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA: Academic Press.

APPENDIX 1

DDL Materials

Count

A-1

1	"How are we going to	**count**	these ballots ?"
2	The referee	**counted**	down from ten to one.
3	He said. " I 'm really	**counting**	on you, man. "
4	He	**counts**	on his binoculars to find the trees.
5	They would still be	**counted**	as children.
6	He was holding his breath for a	**count**	of ten.
7	The cell	**count**	dropped quickly.

A-2

1	“How are we going to	**count**	these ballots ?”
2	The referee	**counted**	down from ten to one.
3	They	**counted**	the number of beds in the rooms.
4	" I 'm going to	**count**	to ten .”
5	He said. " I 'm really	**counting**	on you, man. "
6	He	**counts**	on his binoculars to find the tree.
7	We	**count**	on continued donations .
8	"We 're not	**counting**	car rentals.”
9	They would still be	**counted**	as children.
10	He was holding his breath for a	**count**	of ten.
11	The cell	**count**	dropped quickly.
12	George is convicted on 12	**counts**	of murder.

A-3

1	" I 'm going to	**count**	to ten .”
2	“How are we going to	**count**	these ballots ?”
3	The referee	**counted**	down from ten to one.
4	They	**counted**	the number of beds in the rooms.
5	Scientists can’t always see and	**count**	the birds accurately in a rainforest .
6	“Can we really	**count**	on the U.N. to defend the world?”
7	He said , " I 'm really	**counting**	on you, man. "
8	He	**counts**	on his binoculars to find the trees.
9	Violinists can be	**counted**	on to have remarkable hands.
10	We	**count**	on continued donations .
11	"We 're not	**counting**	car rentals.”
12	They would still be	**counted**	as children.
13	The episode is too minor to	**count**	as an historical event.
14	He was holding his breath for a	**count**	of ten.
15	The cell	**count**	dropped quickly.
16	George is convicted on 12	**counts**	of murder.
17	The annual bird	**count**	has recently expanded into South America.

APPENDIX 2

Summary of Students’ Findings

TABLE A

Summary on the Numbers of Students’ Findings on *count*

Types of vocabulary knowledge	Group 1	Group 2	Group 3	Total
Form knowledge (part of speech/affix)	42	49	50	141
Meaning knowledge (definitional meaning/ register/association)	38	43	42	123
Application knowledge (frequency/ lexical chunks)	4	7	9	20
Total	85	99	101	284
Average	3.3	3.8	3.9	3.6

TABLE B

Summary on Types of Items Students Recognized

Types of items	Students from Group 1	Students from Group 2	Students from Group 3	Total
Verb	26	26	26	78
Noun	17	21	21	61
Participle	0	3	3	6
Meanings of verb	25	26	25	76
Meanings of noun	14	20	21	55
Phrasal verb	2	3	4	9
Frequency	0	1	0	1
Total	84	99	102	286

TABLE C

Summary on the Numbers of Students Who Recognized Each Item

Numbers of item	Students from Group 1	Students from Group 2	Students from Group 3	Total
5 items	4	6	7	17
4 items	6	11	12	29
3 items	9	6	4	19
2 items	7	3	3	13
Total	26	26	26	78

TABLE D

Summary of Students' Preferences for Appropriate Materials

Types of materials (numbers of citations)		A-1 (7)	A-2 (12)	A-3 (17)	Total
Numbers of students choosing the material	Group 1	12	10	4	26
	Group 2	9	12	5	26
	Group 3	8	12	6	26
	Total	29	34	15	78

Examples in: English
Applicable Languages: English
Applicable Levels: Secondary

Sooin Chun
Graduate School of Education
Dankook University
152, Jukjeon-ro, Suji-gu, Yongin-si, Gyeonggi-do, 16890, Korea
Tel: (041) 550-1378
Email: sooic@unitel.co.kr

Received 24 September 2015

Revised 1 November 2015
Accepted 20 November 2015