**CHAPTER 3: Method**

This chapter describes the research methodology of the study. The information of the participants is detailed. Next, the instruments including text, target words and tasks are presented. Piloting and procedures of the study are also discussed. Finally, the details of data analysis are provided.

**3.1 Participants**

Ninety-eight 5th-grade students (48 boys and 50 girls) from four classes in a public elementary school in Seoul, South Korea participated in the experiment. All participants were Korean EFL learners. The mean age of the participants was 11.4. Initially, there were 107 learners, but later it was reduced to 98 participants due to the absence during treatment.

All participants have studied English as a compulsory subject at school from third grade. In fifth grade, they learn English for three hours a week, 40 minutes per class. A native English speaking teacher and Korean teacher teach English together for two hours, and a Korean teacher teaches English for an hour alone with the same English textbook published by Chunjae[[1]](#footnote-1). In 5th and 6th grade, they read short sentences below nine words except *and* or *but (*Ministry of Education and Human Resources, 2015).

As the treatment took place during normal English classes, participants were already assigned to intact classes. To examine the homogeneity of participants, one-way analysis of variance (ANOVA) was employed on the mean score of nation-wide 5th grade English proficiency test scores administered on March 2018. Table 2 shows the mean score and standard deviations of the scores of the English proficiency test. The results confirmed the homogeneity among four classes with no statistical difference ( = 1.48, *p* = 0.23). One of four tasks was assigned to each group of four homogeneous classes. The description of tasks is explained in Section 3.2.3.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 2: Descriptive Statistics of the English Proficiency Test** | | | | |
| Class | A (Task1) | B (Task2) | C (Task3) | D (Task4) |
| *N* | 27 | 25 | 24 | 22 |
| *M* | 20.56 | 22.32 | 21.92 | 22.50 |
| *SD* | 4.03 | 4.01 | 3.56 | 2.82 |

**3.2 Instruments**

This section describes the selection of reading text and target words. Four tasks designed for the study are subsequently illustrated.

**3.2.1 Text**

The reading text contained 293 words which was an excerpt from Nate the Great (1977) by Marjorie Weinman Sharmat that had been modified. It was chosen according to three criteria: familiarity, authenticity, and readability. A story written for children was used since narrative stories bring familiarity to young learners (Hughes, 2011). In terms of authenticity, a story that is not written only for EFL learners was selected. It would be problematic to choose an authentic text for language learners if the text is too difficult. For readability, the story was analyzed with Flesch reading ease test which scored 103.3 indicating a very easy text for participants. According to the 2015 revised national curriculum of English in Korea, elementary school students are required to read sentences with less than nine words except with the use of *but* and *and* (Ministry of Education and Human Resources, 2015). Therefore, the researcher checked the length of sentences to confirm the difficulty of the text. The English teacher of participants with 10 years of teaching experience checked the suitability of the reading text for the age group.

The original text includes 1572 words in total, but only 293 words were used for the treatment. In order to include target words, the text was slightly changed (Appendix 1). It was reviewed by a Korean English teacher and proofread by a native English teacher. Eight target words were provided a glossary in each task. As comprehending 95-98% of the words in a written text is considered to be beneficial for language learners to understand unsimplified written text (Laufer, 1989; Hu and Nation 2000), the reading text was analyzed with the VocabProfile tool in [www.lextutor.ca](http://www.lextutor.ca) (Cobb, 2016). The target words revealed to constitute 6.3% of the reading text which was relatively appropriate for 95% reading coverage.

**3.2.2 Target words**

Eight words were chosen for the treatments: *chick, ceiling, lawn, leash, porch, pebble, shovel,* and *whisker*. They were selected according to three criteria: number of syllables, relevance with the reading text, and unfamiliarity. The researcher selected target words from Neung Yule Junior English Dictionary (2008). For visual presentation, concrete nouns that can be easily visualized were chosen. The number of syllables was examined because of the readability. Words containing than three syllables were chosen. Additionally, words relevant to the content of the story were included. It was necessary that the words were unfamiliar to participants as the study has the purpose of examining the vocabulary acquisition through treatment. Therefore, words were checked through the level of frequency. According to Cobb (2016), *chick* is included in the range of BNC-COCA-1000 types, *ceiling* and *lawn* in BNC-COCA-2000 types, l*eash* and *porch* in BNC-COCA-5000 types, *pebble* and *shovel* in BNC-COCA-6000 types and *whisker* in BNC-COCA-8000 types. The target words were confirmed to be unknown to more than 85% of participants through the piloting and the pretest. All the target words were highlighted in bold print similar to previous studies (such as Hulstijn and Laufer, 2001; Keating, 2008; Kim, 2008; Silva & Otwinowska, 2017).

**3.2.3 Tasks**

Four tasks were designed to examine the effectiveness of vocabulary learning tasks based on Technique Feature Analysis (Nation & Webb, 2011). Four groups of participants were assigned one of four tasks. Table 3 shows the overall features of the tasks analyzed by TFA. Tasks are described in detail below.

**Task 1: Reading comprehension with text and picture glosses**

As the researcher read aloud the reading text, students were required to follow the text with glosses. Eight target words and four distracters were provided as glosses at the end of the text. The glosses consisted of an English word, Korean definition from Neung Yule Junior English Dictionary (2008) and a picture that depicts the meaning of the word. After reading the text, students were required to answer ten comprehension questions, each with three options in Korean. It was necessary to use the knowledge of target words to answer the questions. According to Technique Feature Analysis (Nation & Webb, 2011, p.319), the task-induced the total score of six by the sum of *noticing* (2), *generation* (1), and *retention* (3).

**Task 2: Reading comprehension with text and self-image glosses**

The overall design of the second task was equivalent to task 1 except the presentation of target words. Students read the text in the same speed as the researcher read aloud the reading text with glosses. The glossary of eight target words and four distracters were provided with an English word and Korean definition at the end of the text. After reading the text, students were given time to create their own mental images of the words and draw them beside the textual glosses. When they completed drawing images of the words, they were required to solve the same comprehension questions as Task 1. Since the TFA considers presenting words with picture or mental image as one point in the same criteria of *retention,* the task induced the combination of *noticing* (2), *generation* (1), and *retention* (3). Hence, the total score of Task 2 was six, the same as the Task 1.

**Task 3: Reading comprehension plus gap-fill with text-only glosses**

Learners were allowed to read the text as the researcher read aloud the text. The story and the comprehension questions were the same as in Task 1 and 2. However, the text included eight blanks to fill in with the target words. On a separate sheet of paper, a word list with 12 words was provided as gloss to fill in the blanks. The word list represented the text-only glosses including an English word and a Korean definition from Neung Yule Junior English Dictionary (2008). *Technique Feature Analysis* (Nation & Webb, 2011, p. 319) generated the score of seven involving the components of *motivation (2),* *noticing* (2), *generation* (1), and *retention* (2).

**Task 4: Reading comprehension plus gap-fill with text and picture glosses**

Task 4 was similar to Task 3 except the presentation of target words. After reading the text, students were required to fill in the blanks with the words given as a word list. The word list included eight target words and four distracters. It was considered as multimodal glossing type since the English word was presented by a Korean definition and a picture depicting the meaning of the word. The pictures were exactly the same as the Task 1. Therefore, the score of *Technique Feature Analysis* (Nation & Webb, 2011, p. 319) is similar in all other criteria with the Task 3. Adding a point in imaging of *retention*, the TFA score for Task 4 was eight.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 3: Analysing four tasks by Technique Feature Analysis (adopted from Nation and Webb (2011), p.318-319)** | | | | | | |
|  | Task1: Reading comprehension  with text and picture glosses | | Task 2: Reading comprehension with text and self-image glosses | | Task 3: Reading comprehension plus fill-in with text-only glosses | Task 4: Reading comprehension plus fill-in with text and picture glosses |
| **Motivation** | |  | |  |  |  |
| Is there a clear vocabulary learning goal?  Does the activity motivate learning?  Do the learners select the words? | | 0  0  0 | | 0  0  0 | 1  1  0 | 1  1  0 |
| **Noticing** | |  | |  |  |  |
| Does the activity focus attention on the target words?  Does the activity raise awareness of new vocabulary learning?  Does the activity involve negotiation? | | 1  1  0 | | 1  1  0 | 1  1  0 | 1  1  0 |
| **Retrieval** | |  | |  |  |  |
| Does the activity involve retrieval of the word?  Is it productive retrieval?  Is it recall?  Are there multiple retrievals of each word?  Is there spacing between retrievals? | | 0  0  0  0  0 | | 0  0  0  0  0 | 0  0  0  0  0 | 0  0  0  0  0 |
| **Generation** | |  | |  |  |  |
| Does the activity involve generative use?  Is it productive?  Is there a marked change that involves the use of other words? | | 1  0  0 | | 1  0  0 | 1  0  0 | 1  0  0 |
| **Retention** | |  | |  |  |  |
| Does the activity ensure successful linking of form and meaning?  Does the activity involve instantiation?  Does the activity involve imaging?  Does the activity avoid interference? | | 1  0  1  1 | | 1  0  1  1 | 1  0  0  1 | 1  0  1  1 |
| **Total score** | | 6 | | 6 | 7 | 8 |

**3.3 Piloting**

Two weeks before the actual treatment, a pilot test was administered to four fifth grade students with similar English proficiency and age in different classes who were not participating in the treatment. The pilot testing generally aims to improve internal validity of experiments (Nation & Webb, 2011). In this study, it had the purpose of checking whether the reading text, tasks, and vocabulary tests were appropriate for the treatment groups. Regarding the assessment, students seemed to check circles although they knew the meaning of the word in the Vocabulary Knowledge Scale test (Wesche & Paribakht, 1996). It was assumed that the measurement was unfamiliar to participants. Hence, a brief example of response was added on the top of the test sheet. The modification is briefly described in Figure 2.

|  |
| --- |
| **Figure 2: The format modification of Vocabulary Knowledge Scale test** |
|  |
|  |

**3.4 Procedures**

As the study was conducted during regular English classes at an elementary school in Korea, an official document was sent to the school to get permission for the experiment. The study was conducted over a period of four weeks (Table 4).

All participants took the vocabulary pre-test a week before the treatment. The pre-test aimed to check participants’ previous knowledge of the target words. However, in order to avoid any effect on learning from the pre-test, distracters were included along with the target words (Nation & Webb, 2011). Moreover, the researcher explained that the purpose of the vocabulary test was to select a storybook they would be reading in a week to maintain the incidental setting. From the pre-test, the researcher was able to check that the eight target words were unfamiliar to 85% or more of the students.

Like Hulstijn and Laufer (2001), one of the four tasks described above was randomly assigned to one of four intact classes. Class A (N = 27) performed Task 1; reading comprehension with text and picture glosses. Class B (N = 25) performed Task 2; reading comprehension task which presented glosses with text and self-image. Class C (N=24) was assigned to Task 3; reading comprehension plus gap-fill with text-only glosses. Class D (N = 22) was assigned to Task 4; reading comprehension plus gap-fill with text and picture glosses. All the treatments were conducted on the same day by the researcher. While one group was participating in the study, other groups were studying other school subjects with their homeroom teachers as usual.

At the beginning of the class, the researcher explained the purpose of the study as a story reading class focusing on general comprehension. Additionally, to guarantee the methodological aspect of incidental learning, students were not informed that they would take any vocabulary tests during the treatment. However, careful consideration of incidental vocabulary learning was needed as various types of glosses allow different levels of attention to target words.

The reading text was divided into two parts since 293 words seemed quite long for fifth-grade students. The researcher read aloud the reading text and told students to follow the written text in order to control the reading speed (e.g., Silva et al., 2017). After reading the first part, participants were asked to perform the assigned tasks. The second reading session continued with the same format as the first part. The reading time was controlled; 10 minutes in all groups. Regarding the time-on-task, the researcher put a digital clock on the TV screen in front of the classroom. Participants were told to report the time consumed on the task performance individually on their worksheets. The average time for each task was 14.03 minutes (Task 1), 18.32 minutes (Task 2), 12.90 minutes (Task 3), and 12.86 minutes (Task 4).

When learners completed the tasks, the researcher collected all of the materials to avoid any extra exposure to the target words. The English teacher taught learners with the textbook irrelevant to the treatment for 15 minutes. The learners took an immediate post-test at the end of the class. The immediate post-test took approximately five minutes. At the end of the class, the researcher required learners not to review the target words inside or outside the classroom. The delayed post-test was implemented two weeks after the treatment. It also took five minutes. In all three tests, participants were informed to take tests without any guesswork.

**Table 4: Procedure of the treatment**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **A** | **B** | **C** | **D** |
| **Week**  **1** | Pre-test | | | |
| **Week**  **2** | **Task 1** | **Task 2** | **Task 3** | **Task 4** |
| Reading comprehension with  text and picture glosses | Reading comprehension with  text and self-image glosses | Reading comprehension plus fill-in with text-only glosses | Reading comprehension plus fill-in with text and picture glosses |
| Immediate post-test | | | |
| **Week**  **4** | Delayed post-test | | | |

**3.5 Assessment**

There are several kinds of vocabulary scale tests that measure learners’ knowledge of vocabulary (e.g., D’Anna, Zechmeister, & Hall, 1991; Wesche & Paribakht, 1996; Zimmerman, 1997). Wesche and Paribakht's (1996) Vocabulary Knowledge Scale (VKS) test was frequently used in previous studies to investigate the gained knowledge of vocabulary (Folse, 2006; Keating, 2008; Yang, Shintani, Li, & Zang, 2017; Zou, 2016). However, the original VKS test should be carefully applied to estimate receptive and productive knowledge of vocabulary simultaneously as some researchers have different opinions about measuring them together (Waring, 2000; Meara, 1990). Nevertheless, it is still valuable in that it divides the level of vocabulary knowledge by including not only the performance of knowledge but also learners’ self-perceived knowledge. Therefore, it is possible to analyze partial knowledge gained by empirical experiments which may not be revealed from L2-L1 translation test or sentence making test. Since the study allowed learners to be exposed to target words for no more than three times, it was reasonable to focus on investigating their receptive knowledge that can be gained partially. Therefore, the modified version of the VKS test by Waring (2000) was used (Table 5).

**Table 5: The modified version of Vocabulary Knowledge Scale**

|  |  |  |
| --- | --- | --- |
| **Score** |  | **Self-report categories** |
| **0** |  | I do not know this word. |
| **1** |  | I have seen this word before, but do not know its meaning. |
| **2** |  | I have seen this word before and know its meaning a little. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| **3** |  | I know this word. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

The same format of VKS test was used in the pre-test, immediate post-test, and two-week delayed post-test. Considering the proficiency level of participants, the vocabulary scales and the instructions were provided in Korean. Because of the unfamiliarity of the testing tool, the researcher put a brief example of answering on the test sheet. There were 8 target words and 2 distracters making 10 test items in total. The order of the items was aligned randomly to avoid any memorization in the pre-test, immediate post-test, and delayed post-test.

For scoring, if participants did not know the meaning of a word it was scored 0. When participants checked *I have seen this word before, but do not know its meaning,* they got the score of 1. For respondents providing synonyms, they were scored 2. Finally, participants who knew the correct definition, they were scored 3. As there are 8 target words, the maximum score is 24.

**3.6 Data Analysis**

Initially, there were one hundred and seven participants. The data from nine participants who missed one or two tests during the experiment were excluded from the final analysis. Final collected data from 98 participants were examined. As mentioned earlier, the data was classified by four groups assigned to one of four tasks.

To answer research questions 1 to 3, the IBM Statistical Package for the Social Sciences (SPSS) version 23 was used for the data analysis. Before parametric tests, assumptions were checked. Post-hoc analysis was reported and interpreted by the confidence interval (CI).

For the homogeneity of previous knowledge of target words, the pretest scores of four groups were investigated. One-way analysis of variance (ANOVA) was conducted with tasks (4 levels) as the independent variables, and the pre-test scores of Vocabulary Knowledge Scale (VKS) test as the dependent variable.

To answer the research questions, the whole data was first submitted to a repeated measure ANOVA with task (4 levels) and time (2 levels) as independent variables and the scores of VKS tests as the dependent variables.

The first research question was to reveal the effect of various types of glosses on the initial vocabulary learning and medium-term vocabulary retention. The comparison was carried out within the same task type: Task 1 vs. 2 (reading comprehension with glosses), and Task 3 vs. 4 (reading plus gap-fill with glosses). Two series of independent samples *t*-test were used with the task (2 levels) as the independent variable and the results of the VKS test as the dependent variable. As the assumptions for parametric tests were violated, alternative analysis by bootstrapping was implemented.

To answer the second research question, the four tasks were compared to verify the predictions of TFA on the initial vocabulary learning and medium-term vocabulary retention. The data was submitted to a one-way analysis of variance (ANOVA) with tasks (4 levels) as independent variables and the scores of the VKS tests as the dependent variables. With the violation of assumptions for parametric tests, the bootstrapped ANOVA was used.

The third research question was to find out whether the scores of TFA have differential impact on the initial vocabulary learning and medium-term vocabulary retention when time-on-task is considered. The immediate and delayed scores of VKS tests were divided by their individual time on task to obtain the gained scores per minute. A one-way analysis of variance (ANOVA) with the tasks (4 levels) as the independent variables and the scores of VKS tests per minute as the dependent variables. Bootstrapped ANOVA was conducted as the assumptions for parametric tests were violated.

For research question 4 to reveal students’ opinions, the data from the student questionnaire was analyzed based on their responses.

1. Elementary School English 5 by Chunjae [↑](#footnote-ref-1)