



The Journal of Language Teaching and Learning™

2017

Volume 7/Issue 1

Article 2

The Effects of Retrieval with Different Cues on Second Language Vocabulary Learning

Donggil Song, Sam Houston State University, donggil.song@gmail.com

Eun Young Oh, Seoul National University, milk@snu.ac.kr

Recommended Citations:

APA

Song, D., & Oh, E. (2017). The effects of retrieval with different cues on second language vocabulary learning. *The Journal of Language Teaching and Learning*, 7(1), 30-42.

MLA

Donggil Song and Eun Young Oh. "The effects of retrieval with different cues on second language vocabulary learning." *The Journal of Language Teaching and Learning* 7.1 (2017): 30-42.

The JLTL is freely available online at www.jltl.org, with neither subscription nor membership required.

Contributors are invited to review the Submission page and manuscript templates at www.jltl.org/Submitonline

As an online journal, the JLTL adopts a green-policy journal. Please print out and copy responsibly.





The Journal of Language Teaching and Learning, 2017(1), pp. 30-42

The Effects of Retrieval with Different Cues on Second Language Vocabulary Learning

Donggil Song¹, Eun Young Oh²

ARTICLE INFO

Article History:

Received September 01, 2016

Revisions completed December 12, 2016

Published January 16, 2017

Key Words:

Retrieval

Retrieval-based learning

Retrieval cues

Test

Second language teaching

ABSTRACT

Previous studies showed that learners' retrieval process in tests is one of the significant aspects for enhancing the process of learning. The effectiveness of retrieval process can be utilized as a second language learning strategy. The knowledge reconstruction from retrieval process and learner's performance may depend on retrieval cues. Retrieval cues may play a significant role in supporting the learners to avoid superficial engagement with the instructional materials. However, different types of retrieval cues may not be effective in the same way. If retrieval cues played a significant role in supporting learning, instructors would need to identify appropriate retrieval cues in order to benefit from the retrieval as an instructional strategy. This study was designed to extend the findings that show the effectiveness of retrieval in psychological experimental settings to the real classroom environment by providing several tests as an instructional strategy to learners in the second language course. Through a quasi-experiment, the present research targeted learners' performance using different types of retrieval cues in a second language course at a university. This study investigates an effect of retrieval on students' vocabulary learning, and the relationship between different types of retrieval cues and student performance in vocabulary retention and transfer. Implications for the design of retrieval and suggestions for future research were discussed.

Association of Applied Linguistics. All rights reserved.

One of the most frequently asked questions in second language instruction is how vocabulary should be taught for longer retention and effective transfer (Finkbeiner & Nicol, 2003). Thus, the much debated issue is to find effective and efficient ways for vocabulary acquisition, for example, whether letting students infer the meaning of an unknown word or giving students the translated meaning of the unknown word (Hulstijn, 1992). In order to address this issue, the present study considers the case of retrieval practice as an example of empirical findings from cognitive science with specific implications for the language education field. Through a quasi-experiment, the present research targeted learners' performance using different types of retrieval cues in a second language course at a university. This study investigates an

¹ Sam Houston State University, donggil.song@gmail.com, +1-936-294-2696

² Seoul National University, milk@snu.ac.kr

effect of retrieval on students' vocabulary learning, and the relationship between different types of retrieval cues and student performance in vocabulary retention and transfer.

Facilitating learners' acquisition of content knowledge has been focused in a wide variety of educational research areas, such as learning strategies (Warr & Downing, 2000), problem-based learning (Şendağ & Odabaşı, 2009), cognitive scaffolding (MacGregor & Lou, 2005), collaborative learning (Schoor & Bannert, 2011), and technology-mediated learning (Song & Bonk, 2016). In order to measure the acquisition of content knowledge, formative and summative assessments are usually designed in language instruction: a formative test for finding any existence of a gap between the learners' level and the instructional goal, and a summative test for acquiring evidences of the instruction (Taras, 2005). Tests have been employed as a means of measuring learners' knowledge acquisition. In a line of cognitive science studies, researchers have shown that learners' retrieval during tests is one of the significant processes for enhancing learning (Butler, 2010; Chan, 2010; Jacoby, Wahlheim, & Coane, 2010; Karpicke & Roediger, 2008; Rohrer & Pashler, 2010).

2. Related Literature

2.1. Retrieval

In our real lives, we occasionally need to retrieve our prior knowledge and/or past experience. Specifically, the retrieval occurs when producing factual answers, explaining a concept, making an inference, and applying knowledge to a problem (McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011). A series of research conducted in cognitive science have shown that the retrieval process improves learning (Butler, 2010; Chan, 2010; Jacoby et al., 2010; Rohrer & Pashler, 2010). In their seminal paper, Karpicke and Roediger (2008) revealed the effectiveness of retrieval on learning. In their research, college students (i.e., native speakers of English) learned foreign language words (i.e., 40 Swahili-English word pairs) in the repeated study-test trials. Once the participants had correctly answered all the word questions (i.e., learners' initial encoding was completed), they were assigned one of three groups: (A) study-oriented group: repeatedly studied but dropped from further testing; (B) test-oriented group: repeatedly tested but dropped from further study; and (C) control group: students dropped from both study and test. In the results, repeated studying after learning (i.e., A group: study-oriented) had no positive effect on the recall test. However, repeated testing (i.e., B group: test-oriented) showed a significant positive effect on the recall test (Karpicke & Roediger, 2008). Similarly, previous studies showed that retrieval practices improve learning outcomes in terms of long-term retention more than does spending equivalent time repeatedly studying (Butler, 2010; Jacoby et al., 2010; Roediger, Agarwal, McDaniel, & McDermott, 2011; Rohrer & Pashler, 2010). Therefore, the retrieval process can be considered as an important aspect in enhancing learning.

If the retrieval process directly enhances learning, it could also mean that knowledge is reconstructed or the memory is changed when the learner retrieves prior knowledge. During the retrieval process, when taking a test, additional meanings and features can be added to test takers' prior knowledge (Karpicke & Smith, 2012). Thus, the retrieval might not only be an assessment tool to measure learners' current knowledge level, but also a tool for enhancing learning by supporting knowledge reconstruction.

2.2. Retrieval Cues

The knowledge reconstruction from retrieval process may depend on retrieval cues. A retrieval cue means a word, sentence, symbol, image, or sound used in a test item. Previous studies on learning and memory showed that learners' performance depends on retrieval cues (Anderson & Pichert, 1978; Tulving, 1983).

For example, as a type of retrieval cue, a question prompt is often used in classroom. An instructor uses a prompt when asking questions about the previous instruction in order to support students to retrieve their prior knowledge, usually at the beginning of a class (Huit, Monetti, & Hummel, 2009). Previous studies investigated the use of question prompts as a retrieval cue for student knowledge construction. Using a qualitative study method with multiple-case comparisons, Ge, Chen, and Davis (2005) investigated the effects of question prompts in ill-structured problem-solving tasks. The student task was to analyze an instructional design problem and make a suggestion. Eight graduate students were assigned one of the conditions: (A) question prompts were provided, and writing responses to the prompts were required; (B) the prompts were provided but writing responses were not required; and (C) no prompts were provided. During the problem-solving task, groups A and B received a list of question prompts as a retrieval cue. The qualitative findings indicate that the question prompts had positive effects on the student performance in ill-structured problem solving tasks (Ge et al., 2005). It was also reported that the prompts were not beneficial when a participant did not have relevant prior knowledge and sufficient schema. They claim that question prompts, as a retrieval cue, can be effective “in directing students’ attention to important aspects of the problem, activating their schema, eliciting their explanations” (Ge et al., 2005, p. 220). This might mean that retrieval cues play a significant role in supporting learners to avoid superficial engagement with the instructional materials (Papadopoulos, Demetriadis, Stamelos, & Tsoukalas, 2011). However, it should be noted that different types of retrieval cues are not effective in the same way (Chen & Bradshaw, 2007; Davis, 2003). If retrieval cues played a significant role in supporting learning, instructors would need to identify appropriate retrieval cues in order to benefit from the retrieval as an instructional strategy.

2.3. Language Learning, Retrieval, and This Study

As revealed in the previous studies on retrieval, the effectiveness of retrieval processes can be utilized as a second language teaching strategy. Snellings, Van Gelderen, and De Glopper (2002) investigated the effect of lexical retrieval training in a classroom environment on the gain in retrieval speed. One hundred 9th-grade students received a second language online training, which included the word translation task that requires lexical retrieval. The results indicate that the lexical retrieval resulted in positive effects which include higher hit scores and faster responses on lexical tasks (Snellings et al., 2002). Barcroft (2007) conducted an experiment for examining the effects of retrieval on second language vocabulary learning. The researcher compared the retrieval-oriented group (N=12) with the control group (N=12). The results show that providing undergraduate students with retrieval time positively affected their two types of second language vocabulary posttests (i.e., 2 days later and 1 week later) resulting in approximately 10% higher than the control group (Barcroft, 2007). Goossens, Camp, Verkoeijen, and Tabbers (2014) also investigated the effectiveness of retrieval practice on vocabulary learning in primary school students. The results show that there was a positive impact of the repeated retrieval practice on the recall test when compared with the repeated study condition (Goossens et al., 2014). Therefore, the previous studies on the effect of retrieval provide supportive results for the vocabulary instruction field. However, few studies have investigated the effect of different types of retrieval cues in tests on vocabulary retention and transfer. In addition, little has been done to apply the effectiveness of retrieval as an instructional strategy to a real classroom environment. Thus, the focus of the current study is on different types of retrieval cues which may have to do with the vocabulary learning, specifically, retention and transfer in a real classroom situation. This study investigates the effects of different types of retrieval cues on students’ vocabulary learning in a second language course. Specifically, two research questions were raised in this study:

1. Is the retrieval process positively associated with students' performance of vocabulary retention and transfer?
2. Are different types of retrieval cues differently associated with students' performance of vocabulary retention and transfer?

3. Method

3.1. Participants and Procedure

Twenty-three undergraduate students (14 females and 9 males, age between 20 and 23) in an intermediate Korean language course at a university in the Mid-west area of the United States participated in this study. A quasi-experimental study was conducted to investigate the impact of each type of retrieval cue on student performances in vocabulary retention and transfer tests. To determine the effectiveness of retrieval cues, the researchers looked at the extent to which students could retrieve words (i.e., which had been already tested in quizzes) in the retention and transfer tests.

A week before the word quiz, students were given a list of words, and were asked to learn those words. The instructor did not provide any specific strategies for learning the words, and the students learned the words on their own. According to the informal interview with the participants, they typically study a Korean word with its English translated word (i.e., "English - Korean" pair). The words were selected from the pool of words in the course curriculum, and each pool had the same level of difficulty. As shown in Figure 1, participants took the word quiz, and two weeks later, they took a retention test. Two months after the retention test, the participants took the transfer test. The cue condition of the word quiz is the independent variable and the scores of students in the retention test and in the transfer test are dependent variables. In the word quiz, retrieval cues were projected on-screen. The instructor asked students to write Korean words in paper. The retention and transfer tests were paper-and-pencil tests without using a projected screen. The instructor was present for the word quiz and tests for the administration.

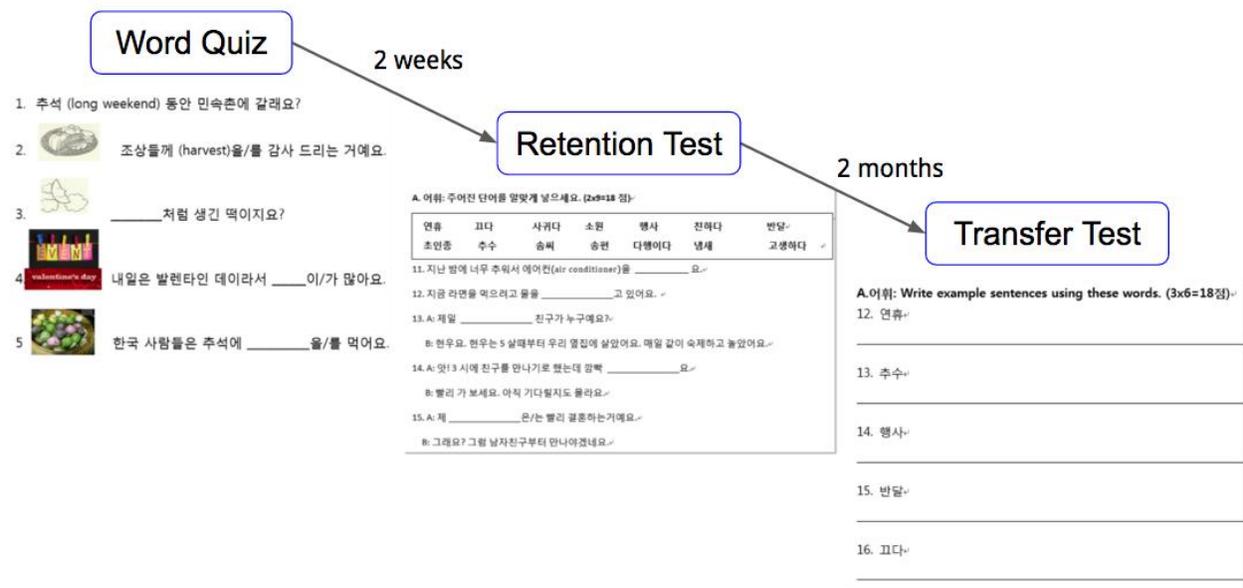


Figure 1. The word quiz, retention test, and transfer test.

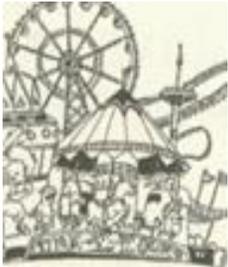
Since the research was conducted in a classroom environment and had to follow the scheduled curriculum, a single-factor within-subjects design was used. The researchers were not able to conduct the repeated measure with each condition at a time. All conditions of retrieval cue were presented in every measure to all participants. Thus, it should be noted that there might be unintended effects when the subjects were exposed all conditions at a time. This limitation is also addressed in the discussion section.

3.2. Instruments

3.2.1. Word quiz

In the word quiz, the instructor presented five different types of retrieval cues on a projected screen, and the students wrote their answers on paper. The participants took 9 word quizzes. Each quiz consisted of five questions. As shown in Table 1, five conditions of retrieval cue were: (C1) textbook sentence, translated word; (C2) textbook sentence and image, translated word; (C3) textbook sentence and image; (C4) new sentence, new image (which embeds translated word); and (C5) new sentence, new image (animated). New sentences and images mean that they were not used in the textbook, and had not been exposed to the students in the course before. The words were selected from the list of words that were given to the students prior to the word quiz. Each word was randomly assigned to one of retrieval cue conditions (i.e., C1 through C5).

Table 1.
Five Types of Retrieval Cues and Question Examples Used in the Word Quiz

Type	Question example	Correct answer [translation]
C1 Textbook sentence, Translated word	(Relatives) 분들을 위해서 선물을 준비하기도 한다.	친척 [relatives]
C2 Textbook sentence and image, Translated word	 (Amusement park) 에서는 연휴에 행사를 한다.	놀이 공원 [amusement park]
C3 Textbook sentence, Textbook image	 () 는 것을 좋아하는 사람도 있다.	사진(을) 찍 [taking a picture]

C4	New sentence, New image (Translated word)		누르 [push]
		길을 건너고 싶으면 버튼을 () 세요.	
C5	New sentence, New image (animated)		꺼 [turn off]
		고양이가 불을 ()요.	

*A translated word of the correct answer is embedded in the image.

**An animated image: a cat jumps to turn off the light.

3.2.2 Retention test

Two weeks after the word quiz, the participants took three retention tests. Each retention test consisted of 18 questions. Fifteen questions were randomly collected from the words that had been tested in the word quiz (i.e., 3 questions from each condition: C1 - C5). The rest 3 questions were the control condition (i.e., not-tested words in the word quiz). Thus, there were 54 questions in 3 retention tests.

3.2.3. Transfer test

Two month after the retention test, the participants took the transfer test, which consisted of 18 questions. Fifteen questions were randomly selected from the words that had been tested in the word quiz (i.e., 3 questions from each condition: C1 - C5). The rest 3 questions were the control condition (i.e., not-tested words in the word quiz and the retention tests).

Both retention and transfer tests used conventional cues (i.e., which are different from C1 - C5) that have been used in this course. As shown in Table 2, for the retention test, finding an appropriate word among the suggested words for a sentence with a blank was used. For the transfer test, writing a sentence using a suggested word was used. This conventional method has been used in this course for more than five years in tests or exams.

Table 2.
Examples of Retrieval Cues Used in Retention and Transfer Tests

	Question example	Correct answer [translation]
Retention test	Fill in the blank using one of the presented words*. Be sure to make it appropriate type. (... , 경기장, 즐리다, 마침, 다양한, ...)	졸려요 [feel sleepy]
A: 왜 이렇게 피곤해 보여요? 어제 잠 못 잤어요?		
B: 네, 새벽까지 파티했어요. 지금 너무 () .		

Transfer test	Write an example sentence using the suggested word. (끓이다) : _____	배고파서 라면을 끓이고 있어요**. [I am cooking a noodle soup because I am hungry.]
---------------	---	--

*25 - 30 words were presented in each retention test.

**An example of correct answer.

3.2.4. Scoring

A correct answer yielded 2 points in the word quiz and the retention test, and 6 points in the transfer test. The total score of each condition in the word quiz, retention and transfer tests is 18. The instructor and an associate instructor followed the conventional scoring process based on their existing criteria (i.e., which include partial grades) that have been used in this course for more than 5 years.

3.3. Data Analysis

All participants completed all the retention and transfer tests. Since two students missed a couple of word quiz, their retention and transfer test scores were excluded in the data analysis. Each student has 5 conditions for the word quiz (i.e., C1 - C5), and 6 conditions for the retention and transfer tests (i.e., C1 - C6). That is, 105 cases were gathered from the word quiz (i.e., 21 students and 5 conditions: C1 - C5), and 126 cases were collected from each retention and transfer test (i.e., 21 students and 6 conditions: C1 - C6). The instructor and the associate instructor who had served as reviewers of test instruments also served as graders. The graders followed the predefined criteria on how to assess each specific question. A one-way analysis of variance (ANOVA) and a one-way analysis of covariance (ANCOVA) were conducted to compare students' scores of each type of retrieval cues.

4. Results

Table 3 shows the average scores and standard deviations on the measures for all conditions.

Table 3.

Means and Standard Deviation: Retrieval Cue Types and Test Performance

	Quiz (N=105*)	Retention test (N=126*)	Transfer test (N=126*)
C1	16.70 (2.13)	15.67 (1.83)	13.79 (4.55)
C2	16.68 (2.33)	14.62 (2.87)	14.66 (5.08)
C3	15.52 (2.67)	16.71 (1.52)	16.17 (3.90)
C4	15.29 (3.40)	14.81 (1.94)	15.17 (2.60)
C5	14.31 (3.42)	16.14 (2.01)	16.91 (2.58)
C6	n/a	13.35 (3.64)**	15.67 (2.81)***

*N: The number of cases.

**The retention test included 5 types of retrieval cues (C1 - C5) and a control condition (C6: not-tested words in the word quiz).

*** The transfer test included 5 types of retrieval cues (C1 - C5) and a control condition (C6: not-tested words in the word quiz and the retention test).

4.1. First Research Question

The first research question was: Is the retrieval process positively associated with students' performance of vocabulary retention and transfer? C6 (i.e., not-tested words in the word quiz) was compared with the condition grouped with C1 through C5. On average, retention scores were higher in retrieval conditions (i.e., C1 - C5, $M=15.59$, $SD=2.19$) than the control condition (i.e., C6, $M=13.35$, $SD=3.64$). This difference was significant, $t(120)=3.52$, $p<.001$, and it represents a medium-sized effect ($r=.31$). However, transfer scores were not higher in retrieval conditions (i.e., C1 - C5, $M=15.34$, $SD = 3.96$) than the control condition (i.e., C6, $M=15.67$, $SD=2.81$). There was no significant difference between them, $t(120)=-.33$, $p=.74$.

A one-way ANOVA was conducted with all conditions (i.e., C1 - C6). As shown in Table 4, there was a significant effect of retrieval cues on the retention scores, $F(5, 116)=4.99$, $p<.001$, $\omega^2=.14$. A post hoc Tukey test showed that among retrieval conditions, the retention scores of C1, C3, and C5 were significantly higher than the scores of C6 ($p<.05$). However, there was not a significant effect of retrieval cues on the transfer scores, $F(5, 116)=1.84$, $p=.11$ (see Table 5).

Table 4.

Analysis of Variance Summary (Retention Scores)

Source	Sum of Squares	df	Mean Square	F
Between Groups	138.93	5	27.79	4.99*
Within Groups	645.60	116	5.57	
Total	784.53	121		

* $p < .001$

Table 5.

Analysis of Variance Summary (Transfer Scores)

Source	Sum of Squares	df	Mean Square	F
Between Groups	128.96	5	25.79	1.84*
Within Groups	1627.98	116	14.03	
Total	1756.94	121		

* $p > .05$

4.2. Second Research Question

The second research question was: Are different types of retrieval cues differently associated with students' performance of vocabulary retention and transfer? In order to address this, the difference of quiz scores was investigated first. There was a significant difference between each condition in the word quiz scores (i.e., C1 - C5), $F(4, 100)=2.67$, $p=.037$, $\omega^2=.06$, which is medium (see Table 6).

Table 6.

Analysis of Variance Summary (Quiz Scores)

Source	Sum of Squares	df	Mean Square	F
Between Groups	86.07	4	21.52	2.67*
Within Groups	807.48	100	8.08	
Total	893.55	104		

* $p < .05$

Since there was a significant difference between each condition in the word quiz scores, a one-way ANCOVA was conducted for retention and transfer scores. As shown in Table 7, the covariate, the word quiz score, was significantly related to the participant's retention scores, $F(1, 99)=13.27$, $p<.01$, $r=.34$. There was also a significant effect of retrieval cue conditions on retention scores after controlling for the effect of word quiz scores, $F(4, 99)=5.44$, $p<.01$, *partial* $\eta^2=.18$. Planned contrasts revealed that having C3, $p=.30$, 95% CI [.13, 2.56], is positively related to the retention scores compared to having C1.

As shown in Table 8, the covariate, the word quiz score, was not significantly related to the participants' transfer scores, $F(1, 99)=2.89$, $p=.09$, $r=.17$. However, there was a significant effect of retrieval cue conditions on transfer scores after controlling for the effect of word quiz scores, $F(4, 99)=2.72$, $p=.03$, *partial* $\eta^2=.10$. Planned contrasts revealed that having C3, $p=.029$, 95% CI [.28, 5.02], and C5, $p=.004$, 95% CI [1.24, 6.11], is positively related to the transfer scores compared to having C1.

Table 7.

Analysis of Covariance Summary (Retention Scores)

Source	Sum of Squares	df	Mean Square	F	Partial η^2
Quiz	51.26	1	51.26	13.27*	.12
Condition	83.98	4	21.00	5.43*	.18
Error	382.45	99	3.86		

* $p<.001$

Table 8.

Analysis of Covariance Summary (Transfer Scores)

Source	Sum of Squares	df	Mean Square	F	Partial η^2
Quiz	42.60	1	42.60	2.89*	.03
Condition	160.66	4	40.16	2.72**	.10
Error	1459.46	99	14.74		

* $p>.05$, ** $p<.05$

5. Discussion

This quasi-experimental study presents the quantitative result which indicates: (1) the retrieval process in the word quiz had a positive effect on students' vocabulary retention, and (2) different types of retrieval cues had a different impact on vocabulary retention and transfer scores. The overall findings suggest that the retrieval process in tests have positive effects in facilitating vocabulary retention. The result is consistent with the previous studies on retrieval process (Butler, 2010; Chan, 2010; Jacoby et al., 2010; Karpicke & Roediger, 2008; McDaniel et al., 2011; Rohrer & Pashler, 2010). Different types of retrieval cues in the word quiz might be differently associated with students' performance in the retention and transfer tests. This finding is consistent with suggestions from previous investigations which show that the tests using different types of cues or prompts can guide learners through their thinking (Ge et al., 2005; Minstrell & van Zee, 2003).

This study expands the previous literature because of its distinctive use of retrieval cues as a retrieval support tool that might allow higher levels of retention and transfer. In this study, different types of retrieval cues were used. C1, C2, and C4 included translated words of correct answers as a retrieval cue. As revealed in the informal interview, the participants usually study a Korean word with its English translated word (i.e., "English - Korean" pair). The students are familiar with these types of retrieval cues which are typical in their vocabulary learning. That might be the reason that the word quiz scores of C1 and C2 were higher than the other conditions. However, the effects of non-typical conditions (i.e., C3 and C5) on retention and transfer tests were found in this study. This might mean that the high performance

in retention and transfer cannot be guaranteed though learners acquired high scores in the word quiz. Previous studies suggest that retrieval process improve learners' short-term memorization for paired associates (e.g., Carrier & Pashler, 1992), which might lead to automated retrieval. The learner would be able to retrieve knowledge in an automated way when the learner was provided with typical cues (see Metcalfe & Finn, 2008). However, it may not be effective for long-term learning (e.g., retention and transfer) with typical cues because the automated retrieval may only be associated with short-term memorization. The conditions that participants do not use in their typical learning (e.g., C3 and C5) may encourage more deliberate retrieval process than the other types of cues. For example, C5 cues did not contain any translated words or textbook materials. During the test, students might have to deliberately think about the context of the presented cue and understand the meaning of the cue. This process may help the learner avoid the automated retrieval. As Roediger et al. (2011) argued, the effectiveness of retrieval could be originated from the learners' deliberate retrieval effort. Thus, learners may have to make more cognitive efforts to retrieve knowledge when they were given unfamiliar or inspiring types of cues rather than typical cues (e.g., translated words).

5.1. Implications

In classroom, instructions normally involve a series of questions planned before the class begins. Questions, usually with a retrieval cue, would encourage students to retrieve their prior knowledge. Retrieval process can be part of daily class activities such as quiz, test, and exam. Instructors need to make more efforts on creating questions and tests to utilize the effect of retrieval on student learning. In addition, identifying appropriate and effective retrieval cues needs to be considered.

Instructional materials need to be designed to avoid the automated retrieval because it might be associated with short-term memorization. Providing typical cues might cause the automated retrieval. As Minstrell and van Zee (2003) argued, the purpose of tests is to foster deep and long-term understanding rather than to check rote memorization. Presenting unfamiliar or inspiring retrieval cues would be important because such cues may encourage learners' deliberate elaboration opportunity. Through the deliberate retrieval process, the learner would make an effort to add diverse features and attributes to the existing knowledge. The learner selects relevant information and interprets the information from the prior knowledge (Mayer, 1992). Similarly, the learner would selectively focus on retrieval cues during retrieval processes.

Instructors need to understand some limitations of the effect of retrieval cues as the question prompts have (see Ge et al., 2005): (1) the retrieval process requires learner's prior knowledge in order to be retrieved, (2) if the learner had a misconception or bias, the retrieval process would not be meaningful even though the well-designed retrieval cues were presented, and (3) while a specific retrieval cue is effective for a learner, it might not be as effective for the other types of learners.

5.2. Limitations and Future Research

The conduct of a quasi-experimental study with samples in naturally occurring educational settings provides multiple threats to external validity, which are key to vary in replication studies in order to confirm the effects are not situation-specific. One of the limitations in this study is that the researchers were not able to conduct between-subjects design and the repeated measure with each condition at a time. Thus, it cannot be claimed that the learning performance in retention and transfer tests was caused solely by the retrieval cues. In addition, there could possibly be other external effects (e.g., the difficulty of words) on students' learning performances in the tests.

Another limitation is the small sample size, particularly in a second language course sample upon which the specific Korean language curriculum and its test data analyses were based. These might harm the power for all analyses, which naturally affects significance testing. In addition, there might be indirect effects. An indirect effect of retrieval (i.e., mediated by some other processes) might occur in test-related experiments (see McDaniel et al., 2011). For example, the participants might be motivated to study the instructional material in preparation for the quiz and/or tests. In addition, additional learning might be occurred when the participants reviewed their quiz/test results, and received feedback from the instructor.

The other limitation of this study is the lack of examining higher-level learning aspects (e.g., Bloom's Taxonomy). The researchers suspect that higher-level of learning may be viable ways to encourage retrieval practice. For example, rather than requesting one correct answer, instructors can use open-ended questions to elicit student thinking and reasoning. The role of retrieval process in this type of learning needs to be investigated rigorously.

Finally, given the benefits of retrieval in learning, it would be beneficial to provide learners with a variety of test opportunities. However, tests, mainly summative assessments, may have an issue of student anxiety or fear. Students with high levels of test anxiety tend to worry over potential failure too much, which leads them to utilize inefficient and ineffective cognitive processing skills (Cassady & Gridley, 2005). Thus, possible strategies to reduce students' test anxiety need to be investigated further. In addition, test process need to empower students to ultimately become self-regulated learners (Nicol & Macfarlane-Dick, 2006).

References

- Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, 17(1), 1-12.
- Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, 57(1), 35-56.
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1118-1133.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20(6), 633-642.
- Cassady, J. C., & Gridley, B. E. (2005). The effects of online formative and summative assessment on test anxiety and performance. *Journal of Technology, Learning, and Assessment*, 4(1), Available from <http://www.jtla.org>.
- Chan, J. C. (2010). Long-term effects of testing on the recall of nontested materials. *Memory*, 18(1), 49-57.
- Chen, C. H., & Bradshaw, A. C. (2007). The effect of web-based question prompts on scaffolding knowledge integration and ill-structured problem solving. *Journal of Research on Technology in Education*, 39(4), 359-375.
- Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *The Journal of the Learning Sciences*, 12(1), 91-142.
- Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics*, 24(3), 369-383.
- Ge, X., Chen, C. H., & Davis, K. A. (2005). Scaffolding novice instructional designers' problem-solving processes using question prompts in a web-based learning environment. *Journal of Educational Computing Research*, 33(2), 219-248.
- Goossens, N. A., Camp, G., Verkoeijen, P. P., & Tabbers, H. K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology*, 28(1), 135-142.
- Huitt, W. G., Monetti, D. M., & Hummel, J. H. (2009). Direct approach to instruction. In C. Reigeluth & A. Carr-Chellman (Eds.), *Instructional-design theories and models: volume III, building a common knowledge base* (pp. 73-98). Mahwah, NJ: Lawrence Erlbaum.
- Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 113-125). London: Macmillan.
- Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(6), 1441-1451.
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966-968.
- Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, 67(1), 17-29.
- Mayer, R. E. (1992). Cognition and instruction: Their historic meeting within educational psychology. *Journal of Educational Psychology*, 84(4), 405-412.
- McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, 103(2), 399-414.
- MacGregor, S. K., & Lou, Y. (2005). Web-based learning: How task scaffolding and Web site design support knowledge acquisition. *Journal of Research on Technology in Education*, 37(2), 161-175.
- Metcalfe, J., & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(5), 1084-1097.
- Minstrell, J., & van Zee, E. (2003). Using questioning to assess and foster student thinking. In J. M. Atkin & J. E. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 61-73). Arlington, VA: National Science Teachers Association Press.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.
- Papadopoulos, P. M., Demetriadis, S. N., Stamelos, I. G., & Tsoukalas, I. A. (2011). The value of writing-to-learn when using question prompts to support web-based learning in ill-structured domains. *Educational Technology Research and Development*, 59(1), 71-90.

- Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, 17(4), 382-395.
- Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher*, 39(5), 406-412.
- Schoor, C., & Bannert, M. (2011). Motivation in a computer-supported collaborative learning scenario and its impact on learning activities and knowledge acquisition. *Learning and Instruction*, 21(4), 560-573.
- Şendağ, S., & Odabaşı, H. F. (2009). Effects of an online problem based learning course on content knowledge acquisition and critical thinking skills. *Computers & Education*, 53(1), 132-141.
- Snellings, P., Van Gelderen, A., & De Glopper, K. (2002). Lexical retrieval: An aspect of fluent second-language production that can be enhanced. *Language Learning*, 52(4), 723-754.
- Song, D., & Bonk, C. (2016). Motivational factors in self-directed informal learning from online learning resources. *Cogent Education*, 3(1), 1-11.
- Taras, M. (2005). Assessment—summative and formative—some theoretical reflections. *British Journal of Educational Studies*, 53(4), 466-478.
- Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.
- Warr, P., & Downing, J. (2000). Learning strategies, learning anxiety and knowledge acquisition. *British Journal of Psychology*, 91(3), 311-333.