Contextualized Vocabulary Learning by Two Proficiency Groups:
Focusing on Imagery Instruction and Context Imageability

Yusuke HASEGAWA
Graduate School, University of Tsukuba/The Japan Society for the Promotion of Science

Abstract

Learning EFL vocabulary in context is an essential process for the enrichment of the mental lexicon; however, past studies have shown that the level of success of contextualized learning is often different between groups with different levels of lexical proficiency. The present study explores the conditions under which a lower proficiency group can gain as much vocabulary knowledge as a higher proficiency group, focusing especially on the mental process of generating an image from reading contexts. The experiment was carefully designed to take into account (a) learners’ lexical proficiency, (b) instruction on how to read the context, and (c) quality of provided context. A significant difference in scores on a context-cued word-recall task between the two proficiency groups was found only when the context was less imageable; when the groups were presented with more imageable contexts and instructed to try to imagine the situation described by the context (referred to as imagery instruction), the lower proficiency group’s scores were equal to the higher proficiency group’s. Based on the finding that imagery instruction did not reduce learning efficiency, no reason appears to exist to avoid vocabulary teaching that focuses on the meaning of a highly imageable context.

1. Introduction

Vocabulary knowledge plays an essential role in learning English as a foreign language (EFL). A common belief about EFL vocabulary learning is that new words should be learned in context (e.g., Qian, 2008; Read, 2000). Contextualized learning seems necessary to develop the learner’s mental lexicon, because new words learned apart from context will be “fossilized” (Jiang, 2000), a term referring to a situation where an EFL learner cannot access the meaning of an English word without undergoing a mental translation process (see Section 2.1).

In spite of the theoretical importance of contextualized vocabulary learning, there have been many studies suggesting that it is an ineffective or inefficient way of learning (see Folse, 2004). In particular, researchers such as Griffin (1992) and Prince (1996) have found that contextualized learning is likely to be unsuccessful when the learner does not have enough proficiency in the target language. Interestingly, the “lower proficiency” participants in
Prince’s and other researchers’ studies were not beginners; their typical proficiency level can be characterized as low- or high-intermediate, which suggests that most of them were able to comprehend the context but were not very fluent at processing that contextual information. On this point, recent studies have suggested that “lower proficiency” learners are incapable of associating contextual information with vocabulary knowledge (e.g., Nassaji, 2006). Therefore, the process of drawing information from context and associating it with vocabulary knowledge can be assumed to be quite difficult and to require considerable proficiency. However, there remains the possibility that teacher instruction can reduce this difficulty that learners experience in contextualized learning. One may argue that the effect of contextualized learning would be somewhat different if a teacher were to instruct his or her students to read context very carefully. The present study will reveal how instruction of this kind affects vocabulary gain among higher and lower proficiency groups.

2. Literature Review

2.1 Contextualized Learning of Vocabulary

Contextualized learning is often regarded as the best way of learning vocabulary, in spite of the potential difficulties mentioned above (see Prince, 1996). The view that EFL vocabulary should be learned in context emerges from two general observations. First is that contextualized vocabulary learning enables learners to encounter both frequent and infrequent words repeatedly (Nation, 2009). If the source of learning were limited to decontextualized approaches such as rote memorization using translations and a teacher’s explanation of word meanings, the number of new words encountered would be quite a bit fewer. This effect of contextualized learning was first proposed in studies on L1 vocabulary acquisition (e.g., Nagy, Herman, & Anderson, 1985), but has also been found important in the EFL situation and studied widely in that regard (e.g., Hunt & Beglar, 2005; Waring & Takaki, 2003).

The second reason for the importance of contextualized learning is rooted in the theory of mental lexicon development, which is more directly related to the present study’s focus than word frequency is. According to Jiang (2000), EFL words learned through rote memorization using translation cannot reach the “final level” of lexical development (see Figure 1), where the semantic, syntactic, morphological, and formal representations of vocabulary are integrated into the lexical entry. The potential differences between contextualized and decontextualized learning are exemplified in Figure 1. Although the figure is a little simplistic as a representation of this complicated theory, it is nevertheless useful as a starting point to help us discuss the influence of learning method on mental lexicon development. The model suggests that if, for example, a Japanese student learns English vocabulary without context, his or her lexical development will not completely close the gap between the L1 and L2, and as a result his or her lexical knowledge will be a compound made
up of English and Japanese lexical representations. In contrast, contextualized learning is assumed to fill in a lexical entry for a particular learner by associating a conceptual representation with a word-form representation. However, this facilitating effect assumes that the learner can generate an elaborate representation in his or her mind from reading context.

Despite these benefits of contextualized learning, recent studies have indicated that contextual presentation does not directly promote more effective learning in general (e.g., Laufer & Shmueli, 1997; Webb, 2007). Laufer and Shmueli (1997) compared intentional vocabulary learning in and out of context, using a sentential context and a passage from a course textbook. Their posttest scores were similar between no-context and sentential context conditions in a delayed test but significantly lower when new words were presented in a longer context. This suggests that L2 knowledge gain is inhibited when more contextual information is present.

In summary, in spite of its practical and theoretical necessity, the contextualized approach is not a very efficient way of learning EFL vocabulary when no special instruction is given on how to read context. However, past studies did not examine what will happen if a learning condition is fully arranged to enhance the learner’s elaborated comprehension. Thus, the next question is under what condition contextualized learning works the best.

![Figure 1. Plausible fruits of the two approaches. This model is adapted from Jiang (2000, pp. 53–55), with slight modifications by the present author. The example sentence was developed based on Webb (2008, p. 243), replacing a pseudoword with a real low-frequency word.](image-url)
2.2 The Proficiency Effect

Past studies have found that the effect of a contextualized task is quite different between higher and lower proficiency learners. For example, Griffin (1992) showed that contextualized learning was ineffective for lower proficiency students. Among participants aged 11 to 13, two proficiency groups were determined by the school they attended. A contextualized recall test, in which participants were given a learning context as a cue to recall learned word forms, showed that only students from the higher proficiency school performed better on the test when they had learned new words in context than out of context; those from the lower proficiency school showed the opposite tendency.

Griffin (1992) explained the cause of this difference between the two proficiency groups in terms of elaborative processing. Because contextualized learning requires more elaborated processing of information, the positive effect emerged only when the participants treated the task as a semantic rather than an episodic task. Thus, the lower proficiency learners in his study, who treated the task as episodic, were less able to learn what was written in context. It seems that the lower proficiency learners' focus was predominantly on the new words themselves, and thus that they could not semantically process the context information.

Similar results were reported in Prince (1996), who assessed university students' English proficiency according to their TOEFL scores. The average score of the higher proficiency group was 480 and that of the lower proficiency group, 397. Prince conducted several tests, including a contextualized recall task, using new contexts that were unfamiliar to the participants. In the contextualized recall task, neither the lower nor the higher proficiency group showed a significant difference between contextualized and decontextualized learning conditions. In fact, the higher proficiency group's scores after contextualized learning were better in the contextualized recall than the translation recall. This shows that reading a context can be effective for higher proficiency learners but a waste of effort for those with lower proficiency. Prince presented two possible explanations for this difference. First, the essential cause of the lower proficiency learners' inability to process context fluently might have been in their lack of lexical knowledge. Second, the lower proficiency group might have had trouble with syntactic analysis, which may have hampered their semantic processing. Again, it was difficult for learners with lower proficiency to process contextual content elaborately or fluently when their attention was mainly on the unfamiliar words themselves.

Research on contextualized vocabulary learning in Japan has not always kept the proficiency factor adequately in mind in studies focusing on the effect of a glosed context. For example, Webb's (2007) experiment was conducted in Japan, but he did not consider learner proficiency. One exception to this observation is Anezaki's (2003) study, which showed that a lower proficiency group learned the best in a decontextualized, translation-based condition and the least in a contextualized condition. It was interesting that
even though Anezaki's participants were beginner learners (junior high school students), the result was similar to those of Griffin (1992) and Prince (1996). However, the higher proficiency group's performance in Anezaki's study was maximized not in the in-sentence condition but in the in-collocation condition.

In summary, the generalization can be made from past studies that less proficient learners are not very good at utilizing context because they cannot cognitively afford to process contextual content fluently. However, it remains unclear when contextualized learning is the most effective among these slow learners. In particular, it has not yet been revealed what would happen if learners' attention were forced onto the context. The effect of contextualized learning among lower proficiency learners could be different under a condition of certain specific instructions about how to read the context. Based on Griffin's (1992) and Prince's (1996) implication that the substantial difference between proficiency groups in their studies was related to elaborated processing caused by lexical knowledge, there is a clear need for a new investigation of learners' elaborated comprehension of context. The next section discusses what kind of context comprehension can be considered an elaborated process.

2.3 Mental Imagery From Context

The mental process of reading has been a point of interest in psycholinguistics. The elaborateness of reading comprehension is explained in Walter Kintsch's theoretical work (e.g., van Dijk & Kintsch, 1983) and mentioned in other mental models such as the Construction-Integration Model, the Structure Building Framework, and the Landscape Model of Reading (see Grabe, 2009). What is widely acknowledged among researchers of reading is that there are different levels of comprehension. The most well known may be that among surface structure, propositional textbase, and situation model (Tapiero, 2007; van Dijk & Kintsch, 1983), in which the first level, surface structure, consists of the representation of the verbatim input; the second level, propositional textbase, relates to readers' propositional understanding of the text; and the third level, situation models, includes readers' interpretations and mental depictions of the text. A reader's process is most elaborated at the situation model level.

When an experimenter would like a reader to construct an elaborated representation in his or her mind, it is sometimes quite useful to give the reader instructions on how to read the text. A typical instruction may be to request the reader to formulate a mental image of the text. In previous studies on reading comprehension, this kind of imagery instruction was long used to help a reader construct highly elaborated representations (e.g., Alba, 1984; Horiba, 2002). In fact, less skilled readers can understand both explicit and implicit information in a text when undergoing imagery instruction (Gambrell & Jawitz, 1993). However, construction of an elaborate mental image like situation models might amount to effortful processing (Zwaan & Radvanski, 1998). Therefore, it is quite possible that such instruction may interrupt the
reading flow if the reader does not have enough cognitive sources available, or if the text is too difficult to raise an image.

The present study investigates the effect of imagery instruction on contextualized vocabulary learning, taking into account two other factors: learners’ lexical proficiency and contextual ease of evocation of a mental image. The proficiency effect has already been discussed in Section 2.2; the ease of evoking mental imagery on the basis of written or spoken materials is generally called imageability (e.g., de Groot, 2011; Steinel, Hulstijn, & Steinel, 2007). Studies have shown that words and texts with higher imageability are easier to retain mentally, even in an EFL situation (for examples of more and less imageable items, see Hasegawa, 2010, 2012). The present study compares more and less imageable contexts to examine whether the effect of imagery instruction differs according to the context quality.

2.4 The Focus of the Present Study

In summary of the above discussion, prior research has discovered that the effect of contextualized learning differs according to learner proficiency, but no one has definitively shown under what circumstances lower proficiency learners can effectively gain vocabulary knowledge. Although past studies have suggested that slow learners experience more difficulty than others in achieving contextualized learning, this approach to learning remains important for developing a mental lexicon. Using a contextualized word-recall task as the measure of rate of learning, this study examined the following three factors: learner proficiency, imagery instruction, and context imageability. Focusing on the elaborate processing of context comprehension, the following hypothesis (H) and research question (RQ) were addressed:

H: Japanese EFL learners with higher lexical proficiency are better at recalling unfamiliar words after contextualized vocabulary learning than those with lower proficiency.
RQ: How are higher and lower proficiency EFL learners’ recall of unfamiliar words affected by imagery instruction during learning and context imageability?

The hypothesis will be supported if the recall test scores were higher among the higher proficiency group than the lower proficiency group. This hypothesis will likely be upheld, because the current study is similar to those of Griffin (1992) and Prince (1996). However, it was important to replicate these studies carefully controlling the instruction and context types. To date, although many studies have examined contextualized vocabulary learning, almost no one discussed how the learners’ reading process was like. This study aimed to test whether the higher proficiency learners outperform the lower proficiency learners in any situation.

The research question was examined through an analysis of three factors (i.e., Instruction, Proficiency, and Imageability), as well as another factor that was necessary to be
taken because of a procedural reason (i.e., Test). If the instruction given the learners to read the context very carefully broadens the lower proficiency learners’ attention to context effectively, their recall scores will be higher than in the no-instruction condition. If, on the other hand, the cognitive capacity of the lower proficiency group cannot afford to process so much information, their scores will instead decrease, meaning that imagery instruction is an obstacle to learning. This will imply that the effect of imagery instruction is different according to learner proficiency. Another possible result would be an Instruction–Imageability or Proficiency–Imageability interaction. If these interactions were observed, the conclusion would be that the effect of contextualized vocabulary learning is predominantly controlled by context quality.

3. Method

3.1 Participants

A total of 32 undergraduate students in Japan participated in this study. They were majoring in various fields such as engineering, sociology, psychology, and international studies. Most of them were first-year students, but some were in higher years. According to their proficiency test scores, their English skills were estimated to be around the Grade 2 level or above in the Society for Testing English Proficiency’s (STEP) Eiken test. Although there are no formula available to convert STEP Eiken Grade into TOEIC scores, the 32 students seemed to resemble the participants in Prince (1996) in terms of English proficiency, with reference to the official data that students who have obtained STEP Eiken Grade 2 scored 517 points on the TOEIC test on average (Educational Testing Service, 2010). However, data from four participants were excluded from analysis because they knew most of the target words for the main experiment before undergoing the learning phase; the target words were no longer “new words” for them. Thus, data from 28 participants were analyzed.

The participants were divided into two groups according to score on the first subsection of the STEP Eiken test, Grade Pre-1 (2010). Table 1 shows the scores on this proficiency test.

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher</td>
<td>13</td>
<td>8.58</td>
<td>2.31</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>Lower</td>
<td>15</td>
<td>3.19</td>
<td>1.64</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>5.50</td>
<td>3.33</td>
<td>0</td>
<td>13</td>
</tr>
</tbody>
</table>

*Note. The maximum possible score was 15. The lower proficiency group’s minimum score became 0 after excluding the easiest items to improve the test reliability.*
This part of the Eiken test was chosen as the proficiency test because it measures lexical skills in the use of vocabulary in context; such skills were considered the most appropriate measure for this study’s purposes. The reason Grade Pre-1 was used instead of Grade 2 was that some students got nearly a maximum score when the experimenter presented this easier test in a pilot study. In the current study, the participants answered 25 items in total, but proficiency scores were calculated based on the 15 items with the highest reliability (Cronbach’s $\alpha = .81$). The 28 students were divided by score into the following two groups: the Higher group, who scored 6 or more ($n = 13$), and the Lower group, who scored 5 or less ($n = 15$).

3.2 Materials

Given the purpose of the current study, learning materials needed to be at an appropriate level of difficulty. Considering the participants’ proficiency level, target words and contexts were developed based on past materials used in STEP Eiken Grade Pre-1 (2010). These learning materials were all different items from the proficiency test items. There were 20 target words in total, including four nouns, eight verbs, and eight adjectives. All these words were presented to participants in context. In addition, translation words in Japanese that accorded with the context were selected, with reference to an English-Japanese dictionary (Konishi, Minamide, & Taishukan, 2001). The contexts had 25.30 words on average ($SD = 5.74$). However, some contexts described concrete situations, but others included relatively little information to imagine any situation; the imageability was different among them. To analyze the imageability difference in contexts, a norming study was conducted as follows.

The norming study was conducted with 10 graduate and undergraduate students who were majoring in applied linguistics or second language acquisition. They were asked to rate the imageability of each context on a 7-point Likert-type scale ($7 = $ the easiest to imagine a situation described in context; $1 = $ the most difficult to imagine anything about the context). The context imageability, or the IMG value, was determined according to the average rating. On this basis, the 20 contexts were divided into More and Less IMG groups, as in Table 2.

<table>
<thead>
<tr>
<th>Imageability</th>
<th>$k$</th>
<th>First rating ($N = 10$)</th>
<th>Second rating ($N = 10$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>More IMG</td>
<td>10</td>
<td>6.22</td>
<td>0.33</td>
</tr>
<tr>
<td>Less IMG</td>
<td>10</td>
<td>4.63</td>
<td>0.56</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>5.43</td>
<td>0.93</td>
</tr>
</tbody>
</table>

*Note.* The possible range of rating was between 1.00 and 7.00. The $k$ values represent the numbers of contexts. The same 10 raters participated in the first and second rating sessions.
To confirm the reliability of the imageability rating, another session was conducted three weeks after the first session. After the order of context presentation was shuffled, the same 10 students were asked to rate the imageability of the same 20 contexts. The ratings in the first and second session were quite similar ($r = .91$) and the groupings of More and Less IMG were exactly replicated. This showed the validity of the imageability ratings in terms of test-retest reliability.

3.3 Procedure

The 28 students each took part in a series of paper-based tasks. The overall procedures are summarized in Figure 2. The experiment included a proficiency test, two learning and testing phases, and pre- and posttests. The proficiency test and its results have been described in Section 3.1. This test was followed by a pretest, where the participants were presented with a list of 20 target words and asked to write their meanings. Almost all these words were unfamiliar to the participants, after the exclusion of four with whom many were familiar (see Section 3.2). If a participant knew any word in the pretest, the word was excluded from later analyses for that participant.

As shown in Figure 2, the main body of this experiment consisted of two sessions, which included contextualized vocabulary learning under Normal and Imagery instruction conditions respectively. In the first session, the participants were presented a list of 10 target words in context along with their meanings in Japanese (see Table 3). They were then asked to learn the word forms, meanings, and contexts as well as they could within five minutes. Because the tester’s intention was fully explained and the participants’ attention was on memorizing the list, this task can be regarded as an intentional learning task as described in Hulstijn (2005) and Schmitt (2010). No other special instruction was given in the first learning phase. After learning, the list was removed and a contextualized recall test was conducted in a gap-filling manner. In this contextualized recall task, the learners needed to remember the word forms matching the contexts, just as in Griffin’s (1992) test. In scoring,
Table 3

Examples of Learning and Testing Formats

<table>
<thead>
<tr>
<th>Phases</th>
<th>Format examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning 1 (Normal instruction)</td>
<td>bulky = 大きくて扱いにくい</td>
</tr>
<tr>
<td></td>
<td>The flight attendant tried to help the woman put her suitcase into the overhead compartment, but it was too bulky to fit.</td>
</tr>
<tr>
<td>Learning 2 (Imagery instruction)</td>
<td>bulky = 大きくて扱いにくい</td>
</tr>
<tr>
<td></td>
<td>The flight attendant tried to help the woman put her suitcase into the overhead compartment, but it was too bulky to fit. (least imageable) 1 2 3 4 5 6 7 (most imageable)</td>
</tr>
<tr>
<td>Testing 1 &amp; 2 (Contextualized recall)</td>
<td>The flight attendant tried to help the woman put her suitcase into the overhead compartment, but it was too ( ) to fit.</td>
</tr>
<tr>
<td>Pre-/Posttests</td>
<td>bulky = ( )</td>
</tr>
</tbody>
</table>

spelling was not a determining factor if the response could be clearly understood (see Webb, 2007). For example, close approximations of the target word bulky such as "bulty" and "yulky" were acceptable responses. When two independent raters marked all the responses, the interrater agreement was 97.67%. The disagreements were solved through discussion.

The second learning phase followed the first contextualized recall test. The participants received a new list containing the remaining 10 target words with contexts. To counterbalance any characteristic item effects on the test scores between the two learning conditions, the order of presentation for the two lists was reversed for half of the participants. The procedure was the same as in the first learning phase except for the instruction. As in the first session, the participants' primary task was to learn the word forms, meanings, and contexts within five minutes. However, in the second session, the participants were forced to consider context content and rate the imageability of each context, which confirmed that all participants properly followed the imagery instruction. The format and procedure of the second contextualized recall test was the same as the first. After these two learning/testing sessions, the same test as the pretest was conducted again to ensure that actual vocabulary gain had taken place. Although the scores on this posttest were analyzed together with the contextualized recall tests, the test-type factor was not as important as the other factors.

4. Results and Discussion

4.1 Results

The results of the contextualized recall test are presented in Table 4 and Figure 3; the posttest scores are in Table 5. To examine the effects of imagery instruction and context imageability on vocabulary learning, a four-way analysis of variance (ANOVA) was conducted using the recall scores of the higher and lower proficiency participants. The
independent variables were Proficiency (Higher and Lower), Instruction (Normal and Imagery), Imageability (More and Less IMG), and Test (Contextualized Recall and Posttest). Proficiency was a between-participants factor; the other three were within-participants factors. The alpha level was set at .05 for all analyses; $\eta^2$ was reported to allow future replication.

Table 4
Descriptive Statistics for the Contextualized Recall Scores (%)

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>$n$</th>
<th>Normal instruction</th>
<th>Imagery instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>More IMG</td>
<td>Less IMG</td>
</tr>
<tr>
<td>Higher</td>
<td>13</td>
<td>76.28 (28.97)</td>
<td>78.85 (23.64)</td>
</tr>
<tr>
<td>Lower</td>
<td>15</td>
<td>62.67 (35.96)</td>
<td>55.33 (33.14)</td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>68.99 (33.03)</td>
<td>66.25 (30.99)</td>
</tr>
</tbody>
</table>

Note. The standard deviations are in parentheses.

Table 5
Descriptive Statistics for the Posttest Scores (%)

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>$n$</th>
<th>Normal instruction</th>
<th>Imagery instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>More IMG</td>
<td>Less IMG</td>
</tr>
<tr>
<td>Higher</td>
<td>13</td>
<td>82.05 (18.87)</td>
<td>82.05 (18.87)</td>
</tr>
<tr>
<td>Lower</td>
<td>15</td>
<td>79.78 (25.81)</td>
<td>85.78 (20.14)</td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>80.83 (22.47)</td>
<td>84.05 (19.29)</td>
</tr>
</tbody>
</table>

Note. The standard deviations are in parentheses.

Figure 3. Contextualized recall scores (%) as a function of Proficiency, Instruction, and Imageability.
ANOVA results show that the Proficiency × Imageability × Test interaction was significant, $F(1, 26) = 4.47, p = .044, \eta^2_p = .147$; it controlled the significant main effect of Test, $F(1, 26) = 17.68, p = .000, \eta^2_p = .405$. The interaction suggested that the combination of these factors was quite important and post hoc analyses would be necessary in order to interpret exactly what this interaction indicated.

No other interactions or main effects were significant: (a) a three-way interaction of Proficiency × Instruction × Imageability × Test ($p = .878, \eta^2_p = .001$); two-way interactions of (b) Proficiency × Instruction × Imageability ($p = .524, \eta^2_p = .016$), (c) Proficiency × Instruction × Test ($p = .161, \eta^2_p = .074$), and (d) Instruction × Imageability × Test ($p = .069, \eta^2_p = .122$); one-way interactions of (e) Proficiency × Instruction ($p = .577, \eta^2_p = .012$), (f) Proficiency × Imageability ($p = .259, \eta^2_p = .049$), (g) Proficiency × Test ($p = .051, \eta^2_p = .139$), (h) Instruction × Imageability ($p = .985, \eta^2_p = .000$), (i) Instruction × Test ($p = .527, \eta^2_p = .016$), and (j) Imageability × Test ($p = .427, \eta^2_p = .024$); and main effects of (k) Proficiency ($p = .315, \eta^2_p = .039$), (l) Instruction ($p = .706, \eta^2_p = .006$), and (m) Imageability ($p = .892, \eta^2_p = .001$). Although one of the one-way interactions (g) almost reached the alpha level (i.e., $p = .051$), it should be noted that it was controlled by the significant two-way interaction of Proficiency × Imageability × Test. The most important finding from these results is that no significant Instruction effects are indicated.

To interpret the significant interaction (i.e., Proficiency × Imageability × Test), post hoc tests were performed with Bonferroni adjustment. A number of statistical tests were actually done as post hoc analyses, but the meaningful results can be summarized as follows. First, for any factor, significance was found only in the contextualized recall test ($p < .05$); the posttest scores were equally high, around 80%, which indicates the insignificance of any effects on the test ($p > .10$). The posttest results confirmed that the participants in both proficiency groups learned the target items quite well within the limited time provided.

Second, with regard to the contextualized recall scores, a significant simple interaction was found between Proficiency and Imageability; specifically, the effect of Proficiency was significant in the Less IMG context ($p = .031$) but not in the More IMG context ($p = .480$), while the effect of Imageability was not significant in either the Higher ($p = .129$) or Lower ($p = .333$) learner group. No other effects reached the significance level ($p > .10$), except for the Proficiency effect within the Imagery Instruction condition, which showed what is called marginal significance ($p = .077$). This series of results indicates that the Lower participants' difficulty with contextualized learning was most remarkable in the Less IMG context.

### 4.2 Discussion

The present study took as points to be examined a hypothesis (H) and a research question (RQ). It was hypothesized that the Higher proficiency group would be better at recalling unfamiliar words than the Lower group, based mainly on the results of Griffin
(1992) and Prince (1996). The current result supported this assumption, but a further interesting finding was derived: The difference between the Higher and Lower learners was under the control of the Imageability factor, or the ease of raising a mental image of a context. The statistical results indicated that the difference between the two proficiency groups was maximized in contexts whose contents were difficult to imagine. Past studies have indicated that it is difficult for slow learners to learn vocabulary in context; however, the level of difficulty cannot be estimated unless the context quality is carefully considered. The results from the present study strongly suggested that it is essential to examine the ease of evoking mental imagery during learning. This issue is focused on by the RQ of this study.

With regard to the RQ, this study asked whether and how the Instruction and Imageability factors would affect Higher and Lower groups’ scores on the Contextualized Recall test. In addition to the finding described above that Imageability affected the contextualized learning performance, it is now worth discussing how imagery instruction works in contextualized vocabulary learning. Originally, there had been two different predictions about the RQ. The first was that imagery instruction would promote learning, because learners would be able to pay attention to contextual information under this kind of instruction. This idea was based on the prior suggestion that in a normal learning situation, EFL learners tend to utilize translation instead of context as a source of learning (Webb, 2007). Because the test format used in this study required participants to make a mental association between a target word and a context, the scores in the Contextualized Recall test would be expected to be higher in the Imagery condition than in the Normal learning situation. The second possibility was that imagery instruction would have a negative influence on word recall. Because vocabulary learners are not very efficient in processing a large amount of textual information (e.g., Laufer & Shmueli, 1997), it could have been troublesome for the participants to be forced to read the context carefully.

In fact, the study found no significant effects concerning the Instruction factor. However, two different interpretations of this finding can be offered. One possibility, of course, is that this kind of instruction has nothing to do with learning efficacy. However, although this was not supported statistically, Figure 3 illustrates that the Lower group’s scores increased as much as those of the Higher group only when contexts with good imageability were learned under the instruction to rate the context imageability. This kind of results cannot be explained by the first account. On this point, unfortunately, the present study only found that a marginally significant effect of Proficiency, which was found in the Normal Instruction condition, disappeared in the Imagery condition. This first interpretation might be rejected if future research finds significant difference between Imagery and Normal learning conditions.

The second possibility is that the imagery instruction actually had both positive and negative effects on learning, and that the positive effect offset the negative one in the experiment, which consequently counterbalanced the scores between Imagery and Normal
conditions. The combination effects of factors shown in Figure 3 appear to be consistent only with this second interpretation. If the second possibility is true, vocabulary learning under imagery instruction can be regarded as an effective method of learning, because this type of learning ensures that an advantage can be derived from contextualized learning (Jiang, 2000; see Figure 1) without any decrease in learning efficiency. Further research is needed to examine the exact reason why this kind of effort-requiring instruction did not affect overall test scores.

5. Conclusion

This study investigated the effects of learner proficiency, teacher instruction, and context quality focusing on contextualized vocabulary learning between higher and lower proficiency groups. The effects of these factors were carefully examined in terms of the mental imagery or semantic representation attached to the items learned, which was assumed to be important in the elaborated learning process. Vocabulary learning in context is supposed to be an important process in the development of a learner’s mental lexicon, although it is an effortful task that requires considerable proficiency from the learner and makes considerable cognitive demands.

In pedagogical terms, this study thus sought to discover how a language teacher can reduce students’ difficulty in vocabulary learning in context. It supported a rough prediction that Japanese EFL learners with higher lexical proficiency are better at recalling new English words learned in context than those with lower proficiency. However, focusing on context imageability, a surprising result was found: The difference in scores between the two proficiency groups was reduced in highly imageable contexts. Therefore, context imageability should be taken into consideration in future research concerning contextualized vocabulary learning. A pedagogical implication of this result is that teachers should be careful to maintain context quality. By presenting learners with image-evoking contexts, teachers may find that slower learners gain as much vocabulary knowledge as relatively advanced learners.

In the current study, this facilitation effect for the lower proficiency group seemed to be maximized when the imageable context was combined with the instruction to read context carefully and try to imagine the situation described. Further research should clarify the potential effect of learners’ effort to develop mental imagery for vocabulary learning contexts. Applying this research suggestion to an EFL classroom, the current results might be understood as evidence showing that a learner’s focus on the meaning of a context does not reduce the efficiency of learning, as far as university students are concerned. Based on the theory that semantic representation from a rich context promotes mental lexicon development, no reason appears to exist to avoid careful reading of a learning context if the learning material is well conceived and of high quality.
Acknowledgement

I would like to thank Professor Yuji Ushiro for his suggestions for this research. I am also indebted to the three anonymous reviewers for their helpful comments. This research was supported by the Japan Society for the Promotion of Science.

References


