Aptitude Treatment Interaction in Intentional Vocabulary Learning: Different Effects of Context on EFL Learners at Three Stages of Lexical Development

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

Abstract

This study examined an Aptitude Treatment Interaction (ATI) between three methods of intentional vocabulary learning and learners’ lexical proficiency. Specifically, it was hypothesized that contextualized learning is more suitable for intermediate learners than for other learner groups. On the basis of past studies, including three prior studies of the present author’s (Hasegawa, 2012a, 2013a, 2013b), a new experiment was conducted to compare the following three conditions for translation-assisted vocabulary learning: (a) without context (Decontextualized), (b) with glossed sentences conveying a more concrete and comprehensive situation (Contextualized+IMG), and (c) with glossed sentences conveying very abstract information (Contextualized−IMG). A total of 84 undergraduates in Japan were divided into three proficiency groups and asked to learn a list of low-frequency words in one of the learning conditions (randomly allotted). Test results showed an interaction between learning condition and proficiency group, suggesting that high-intermediate learners in particular are best served by learning vocabulary in an appropriate context (Contextualized+IMG). It should be useful for teachers and researchers to know that learners at different proficiency levels have different extents of sensitivity to context.

1. Introduction

A common belief about learning vocabulary in a second or foreign language (L2) is that new words should be learned in context (Qian, 2008). Contextualized learning is necessary to develop the learner’s mental lexicon because new words learned apart from context will generally be “fossilized” (Jiang, 2000). It includes many types of learning such as incidental vocabulary gain through extensive reading and rote memorization using example sentences, and therefore has different definitions (see Section 2.1). However, it is now incumbent upon research on learning and teaching English as a foreign language (EFL) to reveal how learners’ vocabulary knowledge can be optimized through contextualized learning.
On the other hand, many EFL learners also take advantage of \textit{decontextualized learning} approaches such as rote learning of pairs of target words and their translations using word cards. In fact, past studies have revealed that the decontextualized learning is much more efficient than the contextualized approach (Folse, 2004). However, most comparisons of contextualized and decontextualized learning were based on a translation task using the new words (see Section 2.2). To accurately reveal context effects on vocabulary learning, further studies should include not only \textit{decontextualized} but also \textit{contextualized testing}, for instance gap-filling tasks embedded in a sentential context (e.g., Nation and Webb, 2011, p. 44).

Thus, to examine whether context presentation affects EFL vocabulary learning, the author of this article has conducted three experiments involving a contextualized test of newly learned words. The most important finding was that the effect of context types was different between proficiency groups. However, the results seem to be inconsistent among experiments: The context effect was observed in a higher proficiency group in one study but a lower group in another study. Therefore, through a new, larger-scale experiment including both contextualized and decontextualized learning and testing, the present study examined the hypothesis that the most effective way of learning differs by the learner’s proficiency level.

2. Literature Review

2.1 Contextualized Learning

There are a number of types or variations of vocabulary learning, and so researchers have established a number of classifications to describe the nature of different types of vocabulary learning. This section briefly reviews the following three sets of features: (a) \textit{intentional} vs. \textit{incidental}, (b) \textit{contextualized} vs. \textit{decontextualized}, and (c) \textit{translation-based} vs. \textit{guessing-based} learning (see Figure 1). This section also defines the types of vocabulary learning that this study will focus on.

![Figure 1. Classification of vocabulary learning. A typical task for each category is also presented. Black arrows shows comparisons made in three past studies (Laufer & Shmueli, 1997; Prince, 1996; Webb, 2007); the present study followed Webb’s (2007) comparison.](image-url)
Researchers have distinguished intentional from incidental learning of vocabulary since the 1980s at latest (Krashen, 1989; Laufer & Hulstijn, 2001). In spite of some arguments against the validity of this contrast (Bruton, Lopez, & Mesa, 2011), this classification has been approved by most researchers attempting to make a rough distinction between deliberate learning of lexical items (intentional learning) and skill-based or communicative activities, in which learners gain lexical knowledge as a byproduct of a main task (incidental learning). The present study focuses on the first mode of learning. According to Hulstijn (2005) and Schmitt (2010), the most typical situation of intentional learning can be created when the learner’s attention is placed on the memorization of a list of words and when the learner is aware that there will be a test on the material learned; the present study realized this situation.

Another viewpoint from which to differentiate the types of vocabulary learning is whether target words are presented in or out of context. Representative articles comparing contextualized and decontextualized learning within the intentional learning paradigm include Laufer and Shmueli (1997), Prince (1996), and Webb (2007). Laufer and Shmueli, for example, compared intentional vocabulary learning in a sentential context and a passage from a course textbook with learning under a no-context condition. Participants’ scores on a delayed test dropped significantly in a condition where new words were presented in a full paragraph context. It was suggested that contextualized learning is more effective in a sentential context, while decontextualized approach is also an efficient means of learning.

Another difference in classification can be observed in terms of how learners determine the meaning of target words. In particular, researchers have been interested in how knowledge of word meanings acquired from contextual information is different from that acquired from translations. In an intentional learning study, Prince (1996) compared two conditions: a contextualized condition, where learners were presented with a list of sentential contexts that included unfamiliar words, and a decontextualized condition, where they were given a list of unfamiliar words paired with translations. Generally, the decontextualized method of learning seemed equally effective regardless of learner group. In contrast, the results for contextualized learning were affected by both learner proficiency and test type as follows: Guessing-based contextualized learning was ineffective for learners with lower L2 proficiency; but for the upper learners, contextualized learning raised their scores on a contextualized test to a level that was closer to those for decontextualized learning.

These results are best understood in terms of Aptitude Treatment Interaction (ATI; Cronbach & Snow, 1977), which refers to a situation where a particular instructional condition is more or less effective depending upon the learner’s abilities or characteristics (see also Vatz, Tare, Jackson, & Doughty, 2013). However, as Figure 1 shows, Prince’s (1996) two learning conditions contained two different comparisons: not only decontextualized vs. contextualized but also guessing-based vs. translation-based learning. In
other words, it was impossible to determine whether the ATI (i.e., the learning condition \times\text{proficiency group interaction}) was caused by context presentation or guessing probability.

To solve this problem, Webb (2007) compared contextualized and decontextualized learning groups, presenting both groups with translations: Participants in the contextualized group used a list of glossed contexts consisting of example sentences and translations of the target words. Over 10 tests, he examined the difference between the translation-only and translation-and-context conditions. The results of his statistical analysis showed no significant overall difference, and he concluded that a single glossed sentence context may have little effect on vocabulary knowledge. However, one might wonder whether the lack of proficiency factors might have affected the results of his experiment; an effect of context presentation might have appeared if Webb’s participants had been divided into proficiency groups.

2.2 Contextualized Testing

Another momentous difference between the experiments by Prince (1996) and Webb (2007) was the types of tests implemented. As mentioned earlier, Prince conducted both contextualized and decontextualized tests, both specialized to assess learner knowledge of form–meaning mapping. More specifically, the contextualized test required participants to recall the target word, such as crush, to fill in a blank in a sentence such as Wine is made by \textit{____ grapes and letting the juice ferment} (p. 492). This was a new context for both the decontextualized (translation) and contextualized groups, as the latter group learned the target word 

\textit{crush} in a different sentence: “Why are those eggs all broken?” “Well, someone put the eggbox on the chair, and I didn’t see it, and I sat on it and\textit{crushed} it.” In contrast, using 10 tests, Webb measured five aspects of vocabulary knowledge (orthography, meaning and form, paradigmatic association, syntagmatic association, and grammar); however, none of them was contextualized except for the two grammar tests, where students were asked to write or recognize a grammatically correct sentence including a target word.

The contextualized recall format is useful for assessing a learner’s incremental knowledge gain. According to Sonbul and Schmitt (2010), who adopted the contextualized recall test together with some decontextualized tests, the combination of contextualized and decontextualized testing is suitable to assess the \textit{level of mastery} of the form–meaning link (see also Laufer, Elder, Hill, & Congdon, 2004; Schmitt, 2010, pp. 84–88). In their experiment, the contextualized test and decontextualized recall (i.e., L2–L1 translation) and recognition (i.e., multiple choice) tests of form–meaning association were conducted as measures of three different levels of mastery. These two types of decontextualized test, recall and recognition tests, have often been applied in research focusing on contextualized vocabulary learning (e.g., Pulido, 2007; Sonbul & Schmitt, 2010; Waring & Takaki, 2003; Webb, 2008), because researchers can assess partial knowledge of word meanings using these tests (Nation & Webb, 2011).
In spite of the advantage of conducting a contextualized test, a possible objection against Prince's (1996) approach to contextualized testing might be that the quality of the contexts was quite different between the learning and testing phases. For example, if learners had imagined that *crush* means something like "to squeeze the juice from fruit," they might feel that the meaning of *crush* in the testing context (*crush an eggbox*) was somewhat different from the one they had learned (*crush grapes*). Additionally, learners had to know two forms of *crush* to correctly answer to the contextualized test (*crushed* and *crushing*). Although we cannot of course equalize all characteristics of the context, there seems to be a preferable solution available to this problem: to conduct another contextualized test, using the same context as in the learning phase. This approach would make it possible to examine, for example, whether upper and lower proficiency learners can recall the target words when they encounter them again in their familiar context. To investigate this possibility, the present study conducted two contextualized tests, one using a new context and the other, the original context; interestingly, the ATI between learning condition and learner proficiency was found on this original-context test.

2.3 The Author's Prior Experiments

To test whether translation-based learning of vocabulary can be affected by differences in the context in which target lexical items are presented, the author of this article has conducted three prior experimental studies (summarized in Table 1 and Figure 2).

Table 1

<table>
<thead>
<tr>
<th>Prior Study</th>
<th>Proficiency</th>
<th>Major finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1 (2013a)</td>
<td>High-I</td>
<td>The lower proficiency group was sensitive to context quality</td>
</tr>
<tr>
<td>Study 2 (2012a)</td>
<td>High-I</td>
<td>A contextualized test of lexical proficiency should be used.</td>
</tr>
<tr>
<td>Study 3 (2013b)</td>
<td>Beginner</td>
<td>The upper proficiency group was sensitive to context quality</td>
</tr>
</tbody>
</table>

*Note.* The years of publication are shown in parentheses. High-I = High-intermediate.

![Figure 2](image-url) Correct answer rates (%) of contextualized tests in Prior Studies 1, 2, and 3. To make the figure as concise as possible, the most comparable results were adopted.
In Prior Study 1 (Hasegawa, 2013a), Japanese university students learned a list of pairs of unfamiliar English words and their Japanese translations, with example sentences. To examine the sensitivity of these EFL learners to context information in this learning condition, two types of context were prepared: more and less imageable contexts. Imageability is defined here as ease of evocation of mental imagery by written or spoken materials; it has been found to have a strong effect on ease of learning (de Groot, 2011; Hasegawa, 2013c). In addition, the study distinguished higher and lower proficiency learners were distinguished according to their scores on a contextualized test of lexical proficiency in English, the first subsection of the STEP Eiken test (STEP, 2010). Although this study was not designed to compare contextualized and decontextualized learning, the results showed a kind of ATI between context type and learner proficiency: The test scores of the lower proficiency group were reduced when they were given lower imageability contexts. Also, Prior Study 2 (Hasegawa, 2012a), which initially focuses on proficiency measures, found an ATI between context type and learner proficiency consistent with the results of Prior Study 1.

The results of Prior Study 3 (partially reported in Hasegawa 2013b) seem to be different from those of the earlier two studies, because the effect of context imageability was found in the upper proficiency group, not the lower group. This difference was considered to be caused by the difference in the overall proficiency level of participants between the studies; the participants in Prior Studies 1 and 2 were mostly high-intermediate learners, whereas those in Study 3 were beginners with a few exceptions. In fact, the “lower” proficiency groups in Studies 1 and 2 were actually more skilled learners than the “upper” group in Study 3. However, it was difficult to compare the learner groups directly across the studies because Study 3 did not conduct the same proficiency test, which seemed quite challenging for the beginners. In addition, the number of participants was small in Study 3, which makes it difficult to accurately define their proficiency level. Nevertheless, the results seem to indicate that there is a certain range of proficiency at which imageability affects vocabulary learning.

If this is the case, one might argue that a learner will follow a process of lexical development featuring the following steps: (a) a first stage, where she/he is not able to deal with contextual information during intentional vocabulary learning; (b) a second stage, where she/he is at her/his most sensitive to context quality; and (c) a third stage, where she/he can always learn new words effectively, regardless of context type. This account was quite consistent with the ATI between context type and learner group found in the three prior studies; the “upper” group in Prior Study 3 and the “lower” group in Prior Studies 1 and 2 would all have been in the second stage. However, the three-stages hypothesis of lexical development described above lacks adequate evidence, since none of the experiments presented in Prior Studies 1, 2, or 3 treated low-intermediate learners as their main target. In addition, the discussion above does not compare contextualized and decontextualized vocabulary learning, only the learners’ sensitivity to the presented context types.
3. The Present Study

To conclude this investigation of the effect of context presence and context type on the intentional, translation-based learning of vocabulary, it was necessary to design a new experiment. The present study compared contextualized and decontextualized learning by utilizing both contextualized and decontextualized testing. For contextualized learning, two types of context—more and less imageable sentences (hereafter, \(+IMG\) and \(-IMG\) contexts, respectively; IMG stands for imageability)—were prepared, on the basis of the Prior Studies. The examples of \(+IMG\) and \(-IMG\) contexts are presented in Section 4.2. Figure 3 illustrates the overall research design, which is described in detail in Section 4.3.

![Figure 3. The overall procedures of this study.](image)

Given the results of decontextualized tests reported in Webb (2007), ATI between learning condition (i.e., presence or quality of contexts) and learner proficiency was not assumed on the decontextualized tests. Therefore, the following hypothesis was constructed:

**H1:** Scores on the decontextualized tests are affected not by learning condition but by learner proficiency.

With regard to contextualized test performance, the following two hypotheses were proposed:

**H2:** Compared to beginners and high-intermediate learners, low-intermediate learners are more sensitive to whether context is given. That is, intermediate learners' scores are higher after contextualized learning than after decontextualized learning.

**H3:** Low-intermediate learners are more sensitive to context quality than beginners or high-intermediate learners. That is, for low-intermediate learners, the more imageable a presented contexts, the higher the score they will get on a contextualized test.

In summary, the present study expected two types of ATI in the contextualized test results (that is, context-presence × proficiency and context-type × proficiency interactions) and no ATI in the decontextualized test results.
4. Method

4.1 Participants

A total of 87 Japanese undergraduates enrolled in EFL classes participated in this study. They were majoring in various physical and social sciences, including international studies, economics, and medical science. Data from three students were excluded from analysis because they happened to skip one or more of the experimental procedures. Thus, the number of participants whose data was retained for analysis was 84.

As shown in Table 2, the participants were divided into three groups according to scores on the first subsection of the STEP Eiken test (STEP, 2010; Cronbach’s α = .94 with 40 items in total), which measures lexical skills on the basis of the use of vocabulary in context. This test was selected as the proficiency measure because skills of this type were considered the most appropriate measure for this study’s purposes. Taking into consideration the participants’ estimated proficiency levels, Grades Pre-1, Pre-2, and 3 were employed.

Table 2

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>n</th>
<th>Grade 3 M</th>
<th>SD</th>
<th>Grade Pre-2 M</th>
<th>SD</th>
<th>Grade Pre-1 M</th>
<th>SD</th>
<th>Total M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-intermediate</td>
<td>26</td>
<td>99.74</td>
<td>1.31</td>
<td>99.23</td>
<td>2.72</td>
<td>70.26</td>
<td>12.82</td>
<td>89.74</td>
<td>4.19</td>
</tr>
<tr>
<td>Low-intermediate</td>
<td>28</td>
<td>97.62</td>
<td>5.51</td>
<td>93.21</td>
<td>8.19</td>
<td>36.67</td>
<td>12.10</td>
<td>75.83</td>
<td>4.98</td>
</tr>
<tr>
<td>Beginner</td>
<td>30</td>
<td>52.67</td>
<td>18.88</td>
<td>34.67</td>
<td>20.63</td>
<td>18.89</td>
<td>11.62</td>
<td>35.41</td>
<td>12.19</td>
</tr>
<tr>
<td>Total</td>
<td>84</td>
<td>82.22</td>
<td>25.04</td>
<td>74.17</td>
<td>32.49</td>
<td>40.71</td>
<td>24.40</td>
<td>65.70</td>
<td>24.75</td>
</tr>
</tbody>
</table>

On the test, Grade 3 is a benchmark for junior high school graduates, and is understood to correspond with level A1 (“beginner”) on the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001); participants’ scores on this grade suggested that the lowest proficiency group in this study could be assumed to be beginners. The highest and middle proficiency groups got almost perfect scores on Grade Pre-2, corresponding with CEFR level A2 (“elementary”), which suggests that these two groups were above the beginner level. However, their scores on Grade Pre-1, corresponding with CEFR level B2 (“upper intermediate”), were noticeably different; thus, the highest and middle proficiency groups were labeled as high- and low-intermediate learners, respectively.

4.2 Materials

The materials used in this study were (a) 10 target words, (b) 10 more and 10 less imageable contexts for learning, and (c) 10 additional contexts for a contextualized test. The target words were prepared with reference to Hasegawa (2012b), which originally used Webb’s (2008) pseudowords. However, for pedagogical reasons, the present study adopted
Table 3

<table>
<thead>
<tr>
<th>Target (nouns)</th>
<th>Translation</th>
<th>Synonym</th>
<th>Target (verbs)</th>
<th>Translation</th>
<th>Synonym</th>
</tr>
</thead>
<tbody>
<tr>
<td>esplanade</td>
<td>遊歩道</td>
<td>“street”</td>
<td>don</td>
<td>着用する</td>
<td>“wear”</td>
</tr>
<tr>
<td>gloaming</td>
<td>昼暮れ時</td>
<td>“evening”</td>
<td>indite</td>
<td>執筆する</td>
<td>“write”</td>
</tr>
<tr>
<td>repast</td>
<td>食事</td>
<td>“lunch”</td>
<td>mosey</td>
<td>ぶらりと訪</td>
<td>“visit”</td>
</tr>
<tr>
<td>sanatorium</td>
<td>療養所</td>
<td>“hospital”</td>
<td>reminisce</td>
<td>思い出にふ</td>
<td>“remember”</td>
</tr>
<tr>
<td>tram</td>
<td>路面電車</td>
<td>“train”</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>visage</td>
<td>顔つき</td>
<td>“face”</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. This study adopted low-frequency words instead of pseudowords used in Webb (2008).

Low-frequency words (see Table 3), which were not listed in JACET 8000 (Committee of Revising the JACET Basic Words, 2003), a standard EFL vocabulary list. Japanese translation that accorded with the context were selected from a dictionary (Konishi & Minamide, 2001).

Next, 30 contexts were prepared with reference to Hasegawa’s (2012b) more imageable, less imageable, and neutral contexts as well as example sentence in the English-Japanese dictionary (Konishi & Minamide, 2001). Each context was composed of around 10 running words (M = 10.33, SD = 0.71). The contexts used only words in the most frequent 3000 words in JACET 8000, with the exception of the target words. According to calculation by Microsoft Word 2010, the readability measures were as follows: Passive Sentences = 0%, Flesch Reading Ease = 81.4, Flesch–Kincaid Grade Level = 4.4.

Of the 30 contexts, 10 were selected for greater imageability (+IMG) and another 10 for less imageability (−IMG). For example, the +IMG context for the target word reminisce was Mike doesn’t reminisce about the Christmas day with his girlfriend; the −IMG context was Maybe it happened, but she did not reminisce about it. Through a pilot study with 28 university students with a low-intermediate proficiency level, the ease of evocation of mental imagery of these contexts was rated on a 7-point Likert scale (7 = easiest to imagine the situation described; 1 = most difficult to imagine anything about the context). As a result, their ratings were significantly higher for the +IMG contexts (M = 5.07, SD = 0.62) than for the −IMG contexts (M = 3.56, SD = 0.47), t(9) = 10.95, p < .001. The additional 10 contexts, which had a middle imageability, were used only in a contextualized test, by replacing a target word with a blank (e.g., Many people will _______ about him for a long time).

4.3 Procedure

The main experiment included a pretest, three types of intentional vocabulary learning task, and two contextualized and two decontextualized tests (see Figure 3). In the pretest, participants were presented with a list of 10 target words and asked to write their meanings. Almost all these words were unfamiliar to the participants; if a participant knew any word in the pretest, that word was excluded from later analyses for that participant.
As shown in Figure 3 (Section 2.4), the following three learning conditions were compared in this study: a no-context (i.e., Decontextualized) condition and two contextualized conditions with more and less imageable context, respectively (hereafter, the Contextualized+IMG and Contextualized–IMG conditions; IMG stands for imageability). Learners were randomly allotted to one or the other of the conditions. Participants in the Decontextualized condition were presented with a list of 10 target words paired with translations; in the two contextualized learning conditions, participants learned a list consisting of the target words and translations along with contexts. After the tester’s intention was fully explained, they learned as many of them as possible within five minutes.

After the learning phase, the list was removed and two contextualized tests (hereafter, Contextualized Tests A and B) and two decontextualized tests (Decontextualized Tests A and B) were administered (see Table 4). The Contextualized Tests were completed by answering gap-filling questions; the participants needed to remember the word forms matching the contexts (Prince, 1996). In Contextualized Test A, all participants were given a common list of 10 new contexts (middle imageability), while in Contextualized Test B, the 10 more imageable contexts were given after Contextualized+IMG learning and the 10 less imageable contexts after Contextualized–IMG learning, in the relevant groups, with each target word replaced with a blank. For the Decontextualized learning condition, the more imageable contexts were presented. In scoring, spelling was not a determining factor as long as the response could be clearly understood (Webb, 2007).

Decontextualized Test A required the participants to write the correct meaning of each word on a list of target words forms in Japanese. Decontextualized Test B asked them to choose the correct meaning from among four alternatives; the correct answer was the same as the translations used in the learning phase, and the three distractors were translations of other target words. The number of distractors was determined with reference to past studies (e.g., Sonbul & Schmitt, 2010; Webb, 2008). To avoid learning effects during testing, the four tests were administered in the following order: Contextualized Test B, Decontextualized Test A, Decontextualized Test B, Contextualized Test A.

Table 4

<table>
<thead>
<tr>
<th>Test</th>
<th>Cue</th>
<th>Task type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contextualized Test A</td>
<td>New context</td>
<td>Gap-filling</td>
<td>Many people will</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>about him for a long time.</td>
</tr>
<tr>
<td>Contextualized Test B</td>
<td>Same context</td>
<td>Gap-filling</td>
<td>Maybe it happened, but she did not</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>about it.</td>
</tr>
<tr>
<td>Decontextualized Test A</td>
<td>Target word</td>
<td>Translation</td>
<td>reminisce =</td>
</tr>
<tr>
<td>Decontextualized Test B</td>
<td>Target word</td>
<td>Multiple choice</td>
<td>reminisce = (a)... (b)... (c)... (d)...</td>
</tr>
</tbody>
</table>

*Note.* This is the example for a participant in the Contextualized–IMG condition; The cue context in Contextualized Test B was the same as in the learning phase.
5. Results and Discussion

5.1 Overall Results

This study had two contextualized and two decontextualized tests. To examine the effects of context and proficiency, two three-way analyses of variance (ANOVAs) were conducted using (a) the contextualized test scores and (b) the decontextualized test scores. The independent variables were Learning Condition (Decontextualized, Contextualized+IMG, and Contextualized–IMG), Proficiency (Beginner, Low-intermediate, and High-intermediate), and Test (A and B) for each analysis. Condition and Proficiency were between-participants factors; Test was a within-participants factor. The alpha level was initially set at .05.

The test results are summarized in Appendix. ANOVA results showed interesting interactions of factors in the Contextualized\(^2\) and Decontextualized Tests\(^3\); in particular, it was found that the results of Contextualized Test B should be discussed separately from the other tests. Section 5.2 explains what was found in Contextualized Test A and the Decontextualized Tests in terms of level of mastery, followed by Section 5.3 focusing on Contextualized Test B.

5.2 Level of Mastery (Contextualized Test A and Decontextualized Tests A and B)

For Contextualized Test A and Decontextualized Tests A and B, no significant (simple) main effects or interactions were found with respect to Learning Condition (ps > .10). However, the proficiency effects were different between the tests, suggesting that the level of mastery of the form–meaning link differed across the proficiency groups (see Figure 4).

For Contextualized Test A, the simple main effect of Proficiency was significant (p < .001). When the three proficiency groups were compared with Scheffe’s method, the High-intermediate group’s scores were significantly higher than those of the Low-intermediate (p = .011) and Beginner groups (p < .001), and the Low-intermediate group outperformed the Beginner group (p < .001). For the Decontextualized Tests, the differences on Test A were significant between Beginner and Low-intermediate (p = .029), Beginner and High-intermediate (p < .001), and Low- and High-intermediate (p = .028); Test B showed no significant differences between groups (ps > .10).

When the three proficiency groups were separately analyzed, the effect of Test was significant among the Beginner and Low-intermediate groups (both ps < .001) but not the High-intermediate group (p = .112). This indicates that all the learners...
gained knowledge of the form–meaning link at the initial stage of mastery through intentional vocabulary learning, but that the level of mastery was different across proficiency groups.

In summary, Contextualized Test A and Decontextualized Test A differentiated all three learner groups (Beginner, Low-intermediate, and High-intermediate), whereas Decontextualized Test B shows a ceiling effect. In terms of the level of mastery of the form–meaning link (Laufer et al., 2004; Schmitt, 2010), only High-intermediate learners were able to gain this kind of knowledge at the final level. It was too difficult for the beginners to apply their knowledge of newly learned words to a new context, but that as they improve their proficiency level, they will become gradually able to acquire vocabulary knowledge at the highest level of mastery. Furthermore, because learning condition did not affect the learners’ performance on the decontextualized tasks, the data fully support H1: Scores on the Decontextualized Tests are affected not by learning condition but by learner proficiency.

5.3 Aptitude Treatment Interaction (Contextualized Test B)

Scores on Contextualized Test B are displayed in Figure 5. ANOVA results showed that the Condition × Proficiency interaction closely approached the alpha level (p = .056), which would indicate that it controls the significant main effect of Proficiency (p < .001). When the results for the three proficiency groups were analyzed separately with Bonferroni adjustment, a significant simple main effect of Condition was found only in the High-intermediate group; Contextualized+IMG was significantly better than Contextualized–IMG (p = .007) but the difference between Contextualized+IMG and Decontextualized was only marginally significant (p = .060), and the difference between Contextualized–IMG and Decontextualized was not significant (p = 1.000). In contrast, the simple main effects of Condition were not significant in the Beginner and Low-intermediate groups (both ps = 1.000).

This result can be interpreted as showing that the positive effect of presenting good contexts during translation-based intentional learning of vocabulary was observed only when the learners were above a certain level of lexical proficiency. Thus, it supports Prince (1996) and Griffin’s (1992) contention that contextualized vocabulary learning is more suitable for upper-level learners, and further, suggests that there might be a kind of threshold between the Low- and High-intermediate groups in term of the learners’ sensitivity to context.

This observation was further verified when the differences between proficiency groups were examined across learning conditions. In the Decontextualized learning condition, the Low- and High-intermediate groups’ scores were significantly higher than
those of the Beginner group \( (ps = .019 \text{ and } 001, \text{ respectively}) \), but no significant difference was found between the Low- and High-intermediate learners \( (p = 1.000) \). Similarly, within the Contextualized-IMG condition, the Low- and High-intermediate groups were significantly better than the Beginner group \( (ps = .015 \text{ and } 003, \text{ respectively}) \) but there was no significant difference between the Low- and High-intermediate groups \( (p = 1.000) \). In contrast, within the Contextualized+IMG condition, the High-intermediate groups significantly outperformed the Low-intermediate and Beginner groups \( (p < .001 \text{ and } p = .002, \text{ respectively}) \); the difference between the Low-intermediate and Beginner groups became less clear \( (p = .062) \).

These results did not support H2 or H3; however, the ATI was observed in a different manner. That is, the statements should be modified as follows: Compared to the beginners and low-intermediate learners, high-intermediate learners are more sensitive to the presence and quality of context. The results from Contextualized Test B clarify the reason for the confusing results among the prior studies; an effect of context quality was consistently found in Prior Studies 1 and 2 because the participants in those studies fell into the range of high-intermediate learners. When learners at this proficiency level are further divided into upper and lower groups, the effect of learning condition (specifically, context imageability) is more noticeable in the lower group, because the upper-high-intermediate, or more succinctly, the advanced learners were able to learn the words on a given list regardless of context type. On the other hand, in Prior Study 3, a few students in the “upper” group got almost full scores on the beginner-level proficiency test. Those student might have performed similarly to the high-intermediate learners in the current study, affecting the overall results of Prior Study 3, which was a small-scale experiment \( (N = 17) \).

6. Conclusion

The present study found an ATI between the contextualized learning of vocabulary and learner proficiency, in an intentional, sentence-based, and translation-given learning situation. The results suggest that even in a translation-based learning condition, contextual information is important for high-intermediate EFL learners. In contrast, beginners and low-intermediate learners might depend on a given translation regardless of the given context; in terms of the three-stages hypothesis proposed in Section 2.3, learners at these levels correspond to the first stage. The results also indicated that these groups’ gain in the form—meaning knowledge was at the halfway level of mastery. The context effect might also disappear among the advanced learners, who might be in the third stage, if they are able to learn the target words almost perfectly. In Prince (1996), the upper and lower proficiency groups’ TOEFL scores were 480 (equivalent to CEFR level B2) and 397 (level B1), respectively; these results are quite consistent with the assumption mentioned above if we regard Prince’s upper and lower groups to be equivalent to high- and low-intermediate learners.
In pedagogical terms, it will be informative for teachers’ practice to be aware that learners at different proficiency levels have different degrees of sensitivity to context. When students have developed their lexical proficiency to a certain level, they might be able to compare a new word’s meaning with the situation described in context. In such a case, their vocabulary knowledge gain will be maximized when a good context is provided. The current results indicate that a short sentence made up of around 10 frequent words and conveying richly specific and meaningful information may facilitate the students’ future performance on a contextualized task. On the other hand, decontextualized learning might be an efficient approach for beginners who are not able to deal with contextual information during intentional vocabulary learning (Griffin, 1992; Prince, 1996).

However, of course, intentional, translation-based learning is not enough for adequate vocabulary development. Future research should examine whether the ATI found in the present study can be applied to other approaches to vocabulary learning.

Notes

1. It is a term referring to a situation where a learner cannot access the meaning of an L2 word without first undergoing a mental translation process from their L1.
2. The ANOVA results for the Contextualized Tests showed that the Condition × Proficiency × Test interaction was significant, F(4, 75) = 2.63, p = .041, ηp² = .123; it controlled the significant interaction of Proficiency × Test, F(2, 75) = 6.32, p = .003, ηp² = .144, and the significant main effects of Proficiency, F(2, 75) = 46.27, p < .001, ηp² = .552, and Test, F(1, 75) = 10.05, p = .002, ηp² = .118. No other interactions or effects were significant (p > .10).
3. For the Decontextualized Tests, the interaction for Condition × Proficiency × Test was not significant, F(4, 75) = 2.05, p = .096, ηp² = .099. Instead, the Proficiency × Test interaction was significant, F(2, 75) = 14.78, p < .001, ηp² = .283; it controlled the significant main effects of Proficiency, F(2, 75) = 10.48, p < .001, ηp² = .218, and Test, F(1, 75) = 94.43, p < .001, ηp² = .557. No other interactions or main effects were significant (p > .10).

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

impracticable term? TESOL Quarterly, 45, 759–768.

Committee of Revising the JACET Basic Words (Ed.). (2003). JACET list of 8000 basic words. Tokyo, Japan: the Japan Association of College English Teachers.


### Appendix: Correct Answer Rates (%) of the Four Tests

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>n</th>
<th>Contextualized Test A (new context)</th>
<th>Contextualized Test B (same context)</th>
<th>Decontextualized Test A (recall)</th>
<th>Decontextualized Test B (recognition)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-L</td>
<td>26</td>
<td>90.00 (12.25)</td>
<td>100.00 (0.00)</td>
<td>85.56 (18.10)</td>
<td>100.00 (0.00)</td>
</tr>
<tr>
<td>Low-I</td>
<td>28</td>
<td>76.67 (34.28)</td>
<td>67.00 (21.63)</td>
<td>75.56 (16.67)</td>
<td>90.67 (7.07)</td>
</tr>
<tr>
<td>Beginner</td>
<td>30</td>
<td>42.00 (30.11)</td>
<td>50.00 (39.72)</td>
<td>73.00 (29.83)</td>
<td>85.00 (22.24)</td>
</tr>
<tr>
<td>Total</td>
<td>84</td>
<td>68.57 (33.63)</td>
<td>70.36 (33.16)</td>
<td>77.86 (22.50)</td>
<td>92.14 (19.12)</td>
</tr>
</tbody>
</table>

*Note.* The standard deviations are in parentheses. De = Decontextualized; High/Low-I = High/Low-intermediate.