Available online at www.ejal.eu

Eurasian Journal of Applied Linguistics 1 (2015) 39–58

EJAL
Eurasian Journal of Applied Linguistics

The role of translation in vocabulary acquisition: A replication study

Michael John Alroe [a] *, Hayo Reinders [b]

[a] *Chulalongkorn University,254 Phayathai Road, Pathumwan, Bangkok Thailand*
[b] *Hayo Reinders, Unitec, 1 Carrington Road, Auckland 1010, New Zealand*

Abstract

This study aimed to test the findings of earlier research indicating that EFL students could better learn new vocabulary via translation from L1 rather than by encountering it in the context of L2 sentences. Over one thousand Thai freshmen students were allocated to one of three groups to learn 30 unfamiliar English words. One group studied translation pairs, a second studied the words in the context of English sentences with graphic illustrations and the third had English sentences, illustrations and Thai translation. After a brief delay participants were given a posttest involving gapped sentences and translations. Results showed the students who learned via translation did not do better than the two groups who learned contextually. The findings do not support the contention that switching to L1 in EFL classes to introduce new vocabulary is justified.

Keywords: Contextual learning; decontextualized learning; translation pairs; vocabulary knowledge

1. Introduction

The current state of knowledge of second language vocabulary acquisition has led language teachers to become more aware of the wide range of factors that influence the learning of new words. Vocabulary learning is deemed to be more effective if noticing and negotiation are involved, learners are required to retrieve the new words (repeatedly and productively) and there are clear links between form and meaning (Nation & Webb, 2011; Laufer & Hulstijn, 2001).

Specifically, learning via translation to L1 (first language) has received implicit endorsement in the recommendation that learners use word cards on which the learner has the target L2 (second or foreign language) word on one side and the L1 translation on the back (Nation & Webb, 2011, p. 29). Webb (2009, p. 361) directly addressed the question of whether students should learn new vocabulary through translation from L1 and concluded that the weight of opinion (here he cited

* Corresponding author name. Tel.: +61-421-934-215
E-mail address: malroe@hotmail.com
http://dx.doi.org/

authorities Hunt and Beglar, Nation, and Schmitt) was that this was acceptable. Nonetheless, whether teachers should resort to L1 to introduce new vocabulary in the classroom is contentious as it runs counter to the teaching approach that most second language teachers would espouse.

The emergence of the communicative approach to teaching L2 resulted in a general conviction that translation had 'little to contribute' to L2 acquisition (Marqué-Aguado & Solís-Becerra, 2013, p. 41). However, an increasing number of researchers now argue that this conviction is unfounded (Auerbach, 1993; Schweers, 1999; Cook, 2010). These researchers argue strongly for at least the limited use of L1 in English instruction. The Common European Framework of Reference for Languages has had some influence in this regard with its support of the practice of translating into and from students' L1 to, among other things, 'convey lexical…meaning' (Council of Europe, 2001, p. 44).

However, reliance on L1 in the EFL classroom where all students share the same first language with a non-native-speaking teacher can result in the L2 becoming the subordinate language (e.g., Murray, 2013. p.55-96; Chang, 2010, p. 67; Chinokul, 2012, pp.169-70, 176-77.) If L1 is the effective language of communication, the EFL class can only be a linguistics class where students study features of the L2 but do not develop communicative competence. The issue is relevant to the university in Bangkok where the present study was carried out. Informally, and on condition of anonymity, many of its NNS (Non-native-speaking) teachers argued for the need to rely on L1 in teaching especially with weaker students, despite admitting it was preferable to use more English in class. This indicated a need to understand if EFL learners could indeed be successfully introduced to new vocabulary through an English context rather than through L1. To the best of knowledge of the researchers Prince's (1996) study was most pertinent to the issue that we wished to investigate. This is largely because it 'demonstrated' the superiority of translation for studying new vocabulary on the one hand, and because it was widely recognized, being repeatedly cited by leading researchers in the area (Huckin & Coady, 1999; Groot, 2000; Nation, 2001; Hulstijn, 2008; Folse, 2006; Laufer & Girsai, 2008; Schmitt, 2008, 2010; and Webb, 2009). Careful examination of its methodology suggested weaknesses that may have predetermined the result that favored translation. Therefore the aim of this research has been to discover if EFL students could be as effectively introduced to unfamiliar vocabulary through studying the words in context as by studying via translation.

2. Literature review

2.1. Context or translation

A review of the related literature on whether the L1 could be justifiably used in the deliberate learning of new words reveals that Prince's (1996) study is the most recent and cited research (Webb, 2009). Prince examined whether French students of

English would do better to approach new vocabulary via translation from the L1 or through the context of an L2 sentence. Overall, his findings indicated that learning by translation was more effective, especially for weaker learners. Nevertheless, in Prince's interpretation of his results he did not propose that students confine themselves to that approach. Instead of endorsing vocabulary learning through translation, he argued students, particularly weaker ones, needed to be taught techniques to better utilize context in learning vocabulary and that they should exploit the advantages in both techniques to acquire new words. Despite Prince's even-handed approach, it was the evidence in the paper for the support of translation in L2 vocabulary acquisition that received most attention.

Despite the use of the word 'context' it is worth stressing here that both modes of vocabulary study in the Prince study involved intentional learning, for the context group was instructed to determine the meanings of the unknown words (in the context of a sentence) and learn them. There has been a tendency in the literature to take the article's title 'The role of context versus translation', as implying the study intended to contrast intentional learning via translation with incidental learning through context and providing proof that the latter was less effective (see Webb, 2009, p.361-2). If so, the study has been misinterpreted because both of Prince's learner groups were clearly engaged in intentional learning as were the participants of the present study. Undoubtedly, EFL students do acquire vocabulary incidentally despite their limited opportunity to do so. However, research also shows that such acquisition is modest in amount and is unlikely to provide students, such as those involved in both Prince's and the present study, with the range of vocabulary required for effective communication in English. As Groot argues: 'There is not enough time to copy the natural (largely incidental) L1 word acquisition process' (2000, p. 61). Read (2004) provides a good overview of research on this point. Further, to counter a view that intentionally learned language could never be truly acquired, a recent study by Elgort (2011, p. 1) exploited priming effects to convincingly demonstrate that intentional vocabulary learning could result in language use that showed a high degree of automaticity.

2.2. Prince's methodology

Prince's participants were 48 French university freshmen who were learning English as a foreign language. In his experiment, half learned 44 English words that were new to them by studying English/French translation pairs and the other half studied the same words by discerning and learning their meanings from the context of 44 English sentences. The post-test, given 50 minutes after the learning phase, involved the students inserting half of the target words into gaps in new sentences and translating the other 22 into English (half of the participants) or into French (the other half). This methodology is indicated in Figure 1. Prince's analysis considered the effect of mode of study (translation pair or context) and of student ability (advanced or weak) on the outcomes.

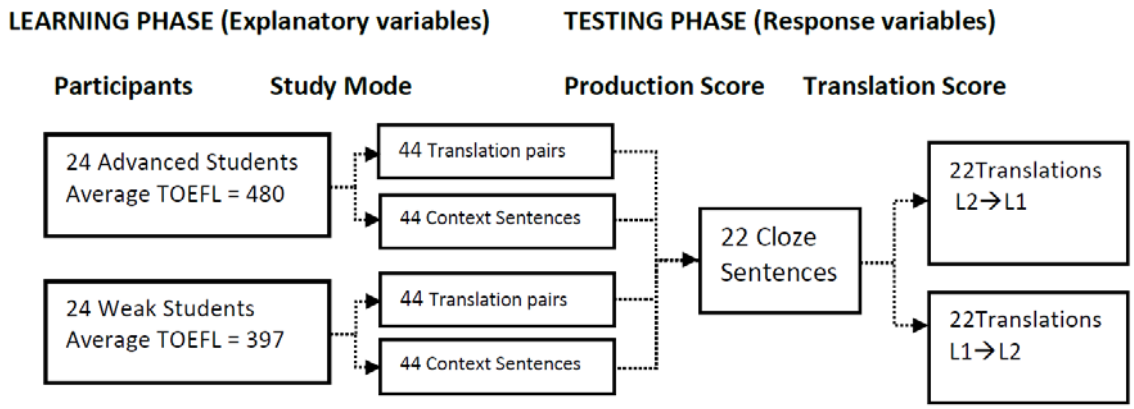


Figure 1: Prince's Methodology

Prince's analysis focused more on participants' ability than the mode of study. In particular, the weaker students' more limited ability in learning vocabulary was considered the explanatory variable affecting performance in the testing phase. But it is the mode of study that begs to be considered as the key explanatory variable. Indeed, it is the mode of study that has been most remarked upon in subsequent citations of Prince's work. Figure 2 plots mode of study against students' success in completing translation of the 22 words, whether French to English or vice versa.

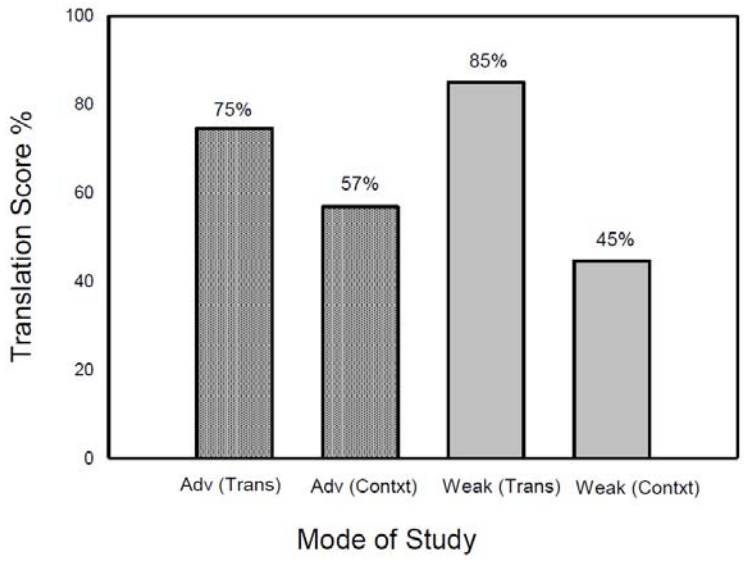


Figure 2: Success in translation by group and mode of study

Clearly, the two groups who studied by translation, irrespective of ability, outperformed the context groups in the translation task. Surprisingly, the weak translation group scored higher than either of the advanced groups. Moreover, it should not be forgotten that half of the students had to translate from L1 to L2 which requires greater productive capacity because the students had to recall the L2 word

from memory, not simply recognize it when it was presented to them. This result would seem to strongly support the effectiveness of learning by translation. In the other form of testing, students had to insert newly learned vocabulary items into gaps in new sentences. Again, the results supported the efficacy of learning by translation (Figure 3).

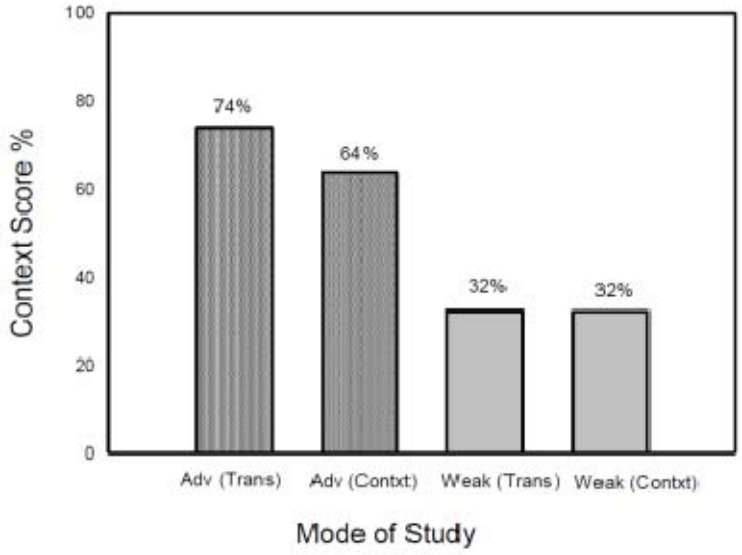


Figure 3: Success in cloze insertion by group and mode of study

In the discussion, Prince focused on accounting for why the weak (translation) group, who had performed so well on the translation post-test, performed poorly on the cloze test. However, more notable is the fact that the advanced (translation) group outperformed the advanced (context) group on this task (context gap) as well as the translation test and that the weak (translation) group performed no worse than the weak (context) group on the gapped sentences. All in all, the results testify to the effectiveness of learning new vocabulary by means of translation.

2.3. Limitations to Prince's study

Other researchers, however, do not consider the study to be conclusive evidence of the effectiveness of using translations in L2 vocabulary development. Takač (2008), for example, notes that evidence supporting the superiority of learning via translation 'is still in short supply' (p. 61). And our close examination of Prince's paper reveals methodological weaknesses that, we believe, justify a reexamination of his work to test the robustness of his results.

The first methodological issue to address is the selection of participants. Prince had 48 learners, suggesting that his translation and context groups contained only 24 in each which was further reduced to 12 when divided further into advanced and weak students. These are not numbers likely to produce a reliable statistical result (Cohen,

1990). Of potentially greater importance than the number of participants are problems with the materials provided for the groups learning from context. They seem to unduly handicap the groups who learned new items that way. For example, one of the target words was 'crush' which is contextualized in the following material provided in the study session:

Context groups: 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 crushed it.

Translation groups: to crush (écraser)

As pointed out in the study, the groups who learned from context had a more complex task in that they had to first determine the meaning of the unknown word (crushed). Prince speculated that they may have so focused on determining the meaning that they neglected the second step (learning it). But more to the point is why fairly weak students were given such a syntactically complex sentence. Having understood the clue that eggs inside the box were broken, students had to follow the anaphoric references to the 'box' through three conjoint clauses with an intervening ambiguous anaphor (it = eggbox/chair) to 'crushed it' and infer the meaning of the target word based on what happened to the eggs inside the box. Surely a syntactically simpler sentence would have sufficed to provide the context. For example: 'He put the grape in his mouth and crushed it between his teeth.'

Other sentences pose different problems. Some include a number of low frequency words other than the target word and it is hard to see that participants would be able to identify which word was meant to be learned. Consider the following where 'owl' is the target:

'The owl has the reputation of being a very intelligent bird because it often looks profoundly pensive.'

Low frequency words in this sentence include: "owl", 'reputation", "profoundly", and "pensive"

Of the ten sample sentences provided in the paper, we would argue that four are either syntactically overly complex or otherwise misleading. This, coupled with the difficulty students may have had identifying which word was to be learned (they were unmarked) and the relatively small number of participants in each group, suggests that the issue of whether to learn from context or from translation remains open. Thus the present study was designed to discover whether students provided with an enhanced context would learn unfamiliar vocabulary as well as if they studied translation pairs. In other words, with simpler syntax, with the target word highlighted to eliminate any confusion, with less confusing surrounding vocabulary and a more transparent context, might contextual learning match the results from translation pairs? For the purposes of the present study, it was determined that context enhancement could include an illustrative image. As Prince noted, Chun and Plass (1996) found that second year university students learned L2 vocabulary more

successfully when the definition of a glossed word was accompanied by a picture, a finding that echoed an earlier study by Kang (1995). Also, though a more complex procedure than learning via translation, context learning could have its own advantages by conveying an understanding of the appropriate environment of the new word that decontextualized translation learning may not provide. Context can offer information about a word – whether it is countable or not, transitive or intransitive, its collocations, and so on. With training students might be taught to be more sensitive to these aspects of usage when approaching an unfamiliar word in context.

3. Method

3.1. Setting and participants

In our replication study, first year university students in a major university in Bangkok studied vocabulary in one of three modes. One involved translation pairs in Thai/English. The others entailed studying the words in context. The procedure was modeled on Prince's but efforts were made to eliminate the weaknesses identified above. The aim was to retest the earlier result that favored translation pairs over learning in context. It was hypothesized that under the revised procedure, groups learning vocabulary in context would perform as well as the group learning with translation pairs.

One thousand and three students from a total candidature of approximately five thousand freshmen students enrolled in a general English course were recruited for this experiment by asking instructors teaching the course to volunteer to teach the selected vocabulary (which had already been set for study later in the semester) to their classes using the context or translation methods.

Thai students study English from the first grade in primary school so the participants would have completed 12 years of English instruction by matriculation but, because the university takes applicants from all over the country, freshmen students' English experience and proficiency are highly diverse. Shortly before the beginning of their first semester most of the participants sit an English proficiency exam called CuTep as part of the University's entrance procedures. The exam is modeled on the paper-based TOEFL test with a comparable level of difficulty and it reports results in a range of 310 to 677. The participants who took the CuTep exam scored an average mark of 488 with a standard deviation of 54.6. Thus, they were probably slightly more advanced than Prince's participants who averaged 440 on a comparable test. These CuTep scores help complete the English proficiency profile of the participants and were later used for analysis but, as they became available only after the experiment was completed, they were not used for allocation of students into treatment groups. To assist in the creation of comparable treatment groups for the experiment, a pre-test was given to the participants using one version of Nation's recognition test for the second thousand most common words in English (Nation, 1990).

3.2. Sampling procedures

Ideally, the 1003 students would have been randomly distributed into three treatment groups but it was not feasible to treat the participants as individual units in this experiment so, with approval of the relevant university authorities, 25 teachers, some of whom had more than one first year English class, agreed to use one of their scheduled English periods to run the experiment. Again, it was judged impractical to ask teachers to form three random groups within each class so the 33 classes were treated as individual experimental units, each assigned to one of three treatment conditions. To do this, the Nation vocabulary recognition test was given to all classes which were subsequently ranked from 1 to 31 on the basis of their mean scores (2 classes missed the pre-test due to teacher delay). Then the units were 'blocked' in 10 ability levels with each level containing three units. Classes ranked 1,2 and 3 were randomly assigned to the three treatment groups, then 4,5, and 6 and so on with the remaining classes in order until 30 had been allocated. (The three remaining classes were randomly distributed into treatment groups.) Thus, by a combination of blocking and random assignment of the classes we designed a procedure as close to an experimental design as the circumstances allowed. When the results of the CuTep test later became available and were analyzed, they did not contradict the assumption that the groups formed for the learning modes had been of similar ability. A one-way ANOVA comparing the CuTep means produced the result $F(2,852) = 1.174$, $p = .310$.

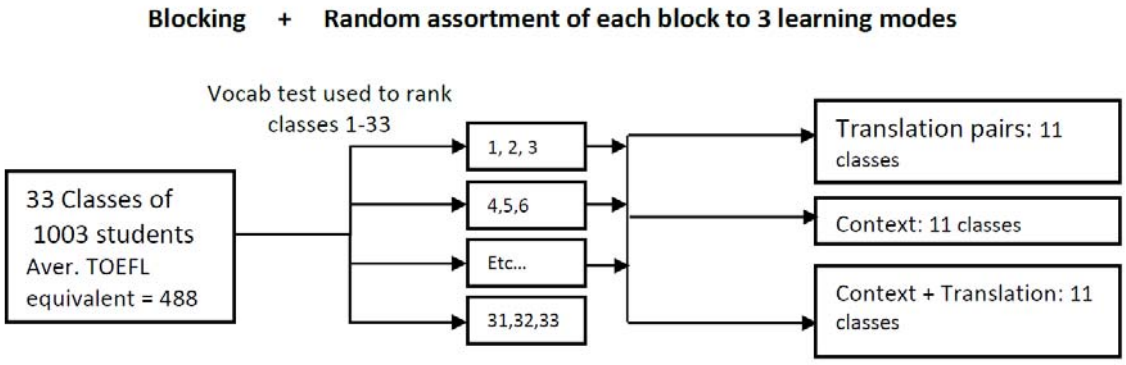


Figure 4: Assignment of groups to 3 learning modes

3.3. Materials

To accommodate teachers who were reluctant to devote teaching time to extra-curricular experimental procedures, 30 vocabulary items from future units in World Pass, the course text book (Stempleski et al., 2005) were selected. The 30 items were chose because previous years of experience suggested most students would not be familiar with them even though they might have encountered some of them before.[1] The target items are listed below with the frequency ranking given for those used

scored in the post-test, from the Corpus of Contemporary American English (Davies, 2008):

Words scored in the post-test	Frequency rank	Distractors not scored in the post-test
Ignored	1382	Solemnly
Fate	2941	Charms
Thumping	>5000	Misplaced
Consequences	1635	Protection
Tale	2260	Go after
Expectantly	>5000*	Interpret
Investigate	1816	Made up
Courage	3355	Verify
Anecdote	>5000	Go over
Specific	982	Cover (e.g. a story)
Haunted	>5000	Alter
Scary	4018	Paw
Spell (magic)	>5000	Journalist
Chronological	>500	Piece together
Embarrassed	4654	
Winked	>5000	

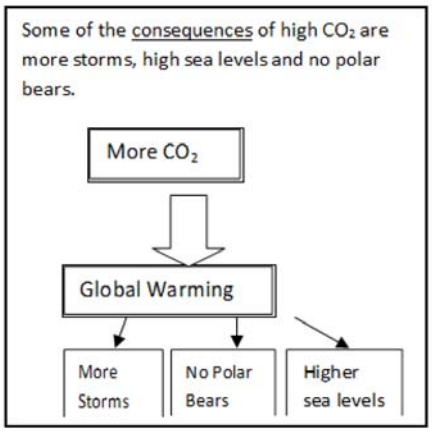
*the frequency rating for 'expect' was 406.

The study materials for each of the 3 treatments were as follows:

1. 30 English items listed above paired with a Thai translation
2. Each of the 30 words in an English sentence with an illustration and the target word(s) underlined
3. As for 2, with the addition of a Thai translation of the target

The three examples below illustrate the differing treatments:

1. Translation Pairs



2. Context (including graphic)

3. Both context and translation

Figure 5: Material illustrating the 3 learning modes

3.4. Procedure

Once the classes were allocated to the three treatment groups, teachers were given the study sheets and the test papers and instructed on the manner in which to conduct the experiment. The 11 classes in the Translation Pairs group were given a sheet with 30 English words with their Thai translations. The other two groups were each given a sheet that matched their treatment specification – either the same 30 words in the context of a sentence with an illustration, or the 30 words in a sentence with illustration and Thai translation. Teachers were asked to distribute the study sheets at the start of a lesson and to instruct the students to study the highlighted words. The group that studied via context with no Thai translation was told that they would need to work out the meanings of the words as well as study them. Teachers were asked not to help the students with the vocabulary items. However, they were informed that they could answer questions if students asked for clarification. No teachers reported having to do this. Twenty minutes were allowed for this activity. With five minutes allocated for distribution and collection, this meant an average of 30 seconds of study per word. Next, the teacher reverted to another classroom activity for at least 25 minutes, leaving 15 minutes at the end of the class for the posttest.

3.5. Posttest

Similar to Prince's study the present study had a posttest that consisted of two parts. The first part contained 15 gapped sentences in English. The first letter of the correct choice was given to guide students to use the target words.

Some of the c_____ of chemo treatment are loss of hair and loss of weight.

The second part consisted of 10 words for translation. Half the students received words in Thai while the others had to translate from English. If the student was required to translate a Thai word to English, the first letter of the English word was given, a concession to the fact that it is a more difficult way to translate than from L2 to L1. This was not done with the words to be translated into Thai as one of the study groups had not been given a Thai translation in the learning phase so it would have been unreasonable to expect them to select the specific Thai translation the other groups had studied. Thus, latitude was given for the Thai equivalent. Secondly, because in the Thai writing system a vowel may precede the initial consonant of the word in writing (so called 'preposed vowels'), in effect, it would be impossible to give the initial (spoken) letter for the Thai translation equivalent. English words had to be correct for part of speech, spelling and inflection. Such restrictions might cause underestimation of the amount of learning achieved, but it was considered important to provide a clear standard by which groups with large numbers could be compared.

Of the total 25 test items, 10 were distracters and 15 (10 cloze and 5 translations[2]) were marked and recorded. In order to prevent students copying from each other, the distracters varied for individual test-takers as did the order of items. Students were allowed fifteen minutes to finish the test, and the teachers reported that this time allocation was adequate.

4. Results

4.1. Mode of study

The gapped sentence scores for students according to their mode of study were compared. Inspection of the histograms indicated the distributions could be considered normal and so a one-way ANOVA was employed to determine if the differences between the groups were significant. The groups who studied L2 words in an illustrated English context either with or without a Thai translation performed better on the gapped sentences than the group who studied only translation pairs. There was a statistically significant difference at the $p < .05$ level: $F\,(2, 1000) = 6.2$, $p = .002$. The effect size, calculated using eta squared, was .01 which is small. Post-hoc comparisons using the Tukey HSD test indicated that the mean scores for the two context groups did not differ significantly from each other but both differed significantly from the group who studied translation pairs alone. The boxplot in Figure 6 shows the distribution of scores for the gapped or cloze sentences. (The bolded horizontal line indicates the median score and the shaded box represents the middle 50% of scores. The 'whiskers' extending from the boxes represent the remaining 50% except for outliers whose scores are represented by small circles outside the 'whiskers'.)

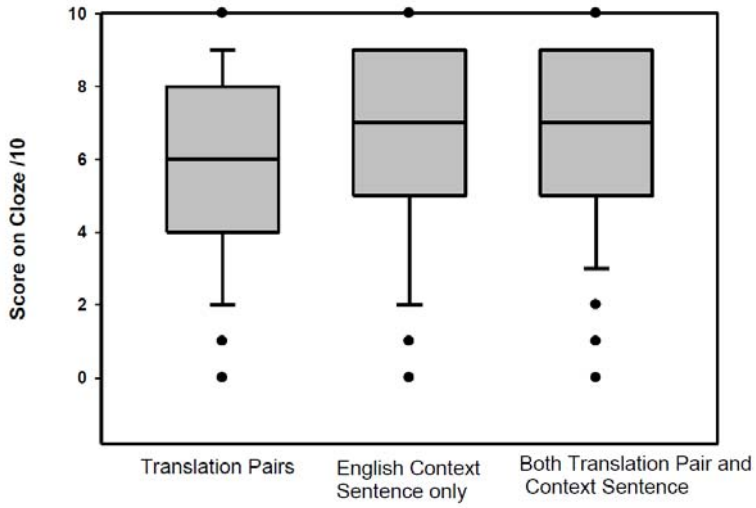


Figure 6: Effect of Mode of Study on Gapped Sentences Score

With the posttests in translation, the results were a little different. In this case the distributions did not appear normal so a Kruskal-Walls test was employed to compare the three groups. No significant interaction between mode of study and success in translating single words to English was found as the test of significance gave a p value of .86. However a significant interaction was found between mode of study and translation into Thai. Kruskal-Wallis tests showed a significant difference in scores for translation to Thai across the three modes of study, $\chi 2$ (2, n = 499) = 15.956, p = .001. Two Post-hoc Mann-Whitney U tests showed that the English only context group performed significantly worse than the English context group who had Thai translation as well. They also were outperformed by the group who studied translation pairs with the difference giving a p value of .042. But application of the Mann-Whitney U test here required a Bonferroni adjustment of the critical p level from .05 to .025. Thus we could not conclude that the difference was significant. It is likely that the English only group were outperformed simply because both the other groups had access to Thai translations in the study phase.

4.2. Skill level

Prince (1966) had compared advanced and weaker students by dividing his 48 students into two equal groups on the basis of a pre-test. He found that the advanced students had a significantly higher rate of total correct answers whatever the learning condition. However, that difference was wholly accounted for by the advanced groups' better performance on the cloze questions where they scored a mean of 68.97% compared with the weaker students' 32.30%. In the translation task there was no overall difference (65.69% to 65.17%) between the students of different skill levels. In the present study, skill levels were defined by temporarily discarding the scores of the 148 students who had not taken the CuTep and dividing the remaining 855 students into weak, average and advanced groupings according to their CuTep scores. Those who scored more than 0.5 of a standard deviation below the CuTep mean for all

participants constituted the weak group (n = 324), those who scored more than 0.5 of a standard deviation above the mean formed the advanced group (n = 265), and the remaining 266 were classed as average. With respect to success in completing gapped sentences, our findings accord with Prince's. A one-way analysis of variance between groups was conducted to explore the impact of skill level on success in completing the 10 gapped sentences. There was a statistically significant difference in the scores of the three skill groups: F (2,852) = 87.994, p = .001. The effect size, calculated using eta squared, was .17, which is large. Post-hoc comparisons using the Tukey HSD test indicated that the mean score for the advanced group (M = 7.6, SD = 1.858) was significantly better than that of the average group (M = 6.76, SD = 2.312) and the average group was significantly better than the weak group (M = 5.19, SD = 2.489

As in Prince's (1966) study, we also found that weaker students equaled advanced students in their scores for word to word translation of target vocabulary into L1. A Kruskal-Wallis test was performed to compare the scores of the three groups in translating into Thai gave a p value of 0.76 indicating no significant difference. However, with respect to translating into L2, our results differed from Prince's study in that we found a significant divergence in the scores of the 3 groups: $\chi 2$ (2, n = 428) =13.03, p = .001. The distribution of scores of the three groups is shown in the boxplot below in Figure 7. As can be seen, at least 50% of the advanced and average students scored 5/5 on this task. The distribution of scores for the weakest group meanwhile covers a wider range with 50% getting 4/5 or lower.

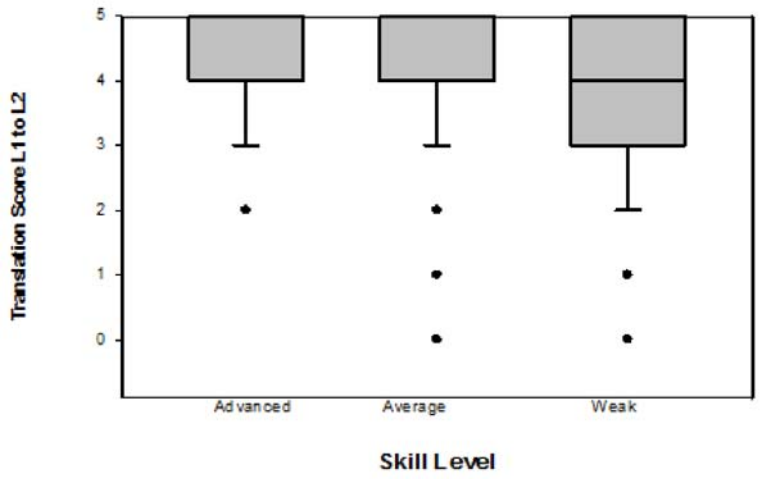


Figure 7: Direction of Translation

4.3. Direction of translation

It is commonly held that translating into L1 is easier than translating into L2 (Stoddard, 1929; Ellis & Beaton, 1993; Napier, Rohan & Slatyer, 2007.). Thus Prince was surprised that the direction of translation did not prove to be a significant factor in his study. Even though his weaker students averaged 94.6% in translating from English to French compared with 77.1% when translating to English the difference was not found to be significant. In the present study the direction of translation was

significant with scores for translation into L1 higher than the reverse. The mean scores for translation into L1 and L2 were 4.6/5 and 4.2/5 respectively. A Mann-Whitney U test gave a p value of .001 (Sample size = 501, U = 97925.000, z = -7.023). The effect size r = .22 was small to medium.

5. Discussion

There are three aspects of the results of the present study that merit discussion: the impact of learner ability level, direction of translation and, most importantly, mode of study. The first of these can be quickly explained in that advanced students performed better than weaker students in all respects except for translation of individual words into Thai. This suggests a connection with the second issue – direction of translation and so both will be considered together.

The findings of this study indicated that, as expected, advanced students performed better. In Prince's study the exception to this was the surprisingly good performance on the translation post-test of the weak students who had studied translation pairs. This result was not replicated in the present study. The translation results found in the present experiment are probably best accounted for by the claim that direction of translation (L1 to L2: L2 to L1) is of asymmetrical difficulty (Stoddard, 1929; Ellis & Beaton, 1993; Napier, Rohan & Slatyer, 2007.). Prince himself had expected that participants would be better at translating into their native French than into English but in fact, he found no evidence of this in his results. This is difficult to account for. It has long been accepted in the field that if a learner recognizes an L2 word it is easier to find a match for it in L1 (because of the depth and scope of one's knowledge of one's native language) than it is to match an L1 word with one in L2. The results from the present study are in accord with this. This was not seen in the scores of the advanced students, but this may be explained by a masking effect of their very high scores.

Prince found the students who had studied vocabulary in translation pairs outperformed those who studied vocabulary in context in every comparison between the two modes of study. Often that difference was statistically significant. That was not the case in the present study. We found that participants who studied words in context outperformed those who studied via translation pairs. This finding is in agreement with that of Laufer and Shmueli (1997, p.98) who had students learn vocabulary using various methods and found that learning words in the context of a sentence was significantly more effective in the short term than learning by translation pairs[3]. In the present experiment was designed the researchers expected that simplifying the contexts, identifying the target word and providing illustrations would enable groups studying from context to perform no worse on the cloze post-test than the group who studied translation pairs. That they would significantly outperform the translation group was unexpected. How can this result be explained and what implications might this have for pedagogy?

Following Tulving (1983) and Wiseman and Tulving (1976), Prince noted that 'performance should be better when subjects meet the same conditions in the recall phase as in the study phase' (p.5). This observation did not hold for Prince's own study as the advanced group that learnt via translation bettered the advanced context group in both translation and context testing. Moreover, the weaker translation group outperformed every other group on translation and did no worse than the weaker context group on the cloze passage. But that may have resulted from complexities in the sentences provided to the context group for the study period and lack of salience of the target words as pointed out above. In the present study those who studied using the context conditions did significantly better on context testing than those who studied in translation pairs. In translation to English (L1 to L2), there was no significant interaction between method of study and score. However, there was an interaction between method of study and translation into Thai with those who studied by context without translation equivalents performing worst while the two groups with access to Thai translations in the study phase performed better. Although this result was significant on a Kruskal-Wallis test ($\chi 2$ 8.68, $p<.03$) the difference was trivial with the 'context only' group getting 4.5/5, and the translation equivalent group getting 4.6/5, while the context plus translation group managed 4.8/5. Nevertheless, none of these results contradict the observation that participants did better when recall was tested with the same method as mode of study.

Taken together, these results suggest that if we want students to be able to use words in context we need to have them study vocabulary in context or in context with translation as well. As suggested earlier in this paper it may be that in focusing on words in context, students pick up additional meta-knowledge about the word e.g. part of speech, transitivity, countability for nouns, collocations, etc., that is not available when words are studied with decontextualized translation equivalents only.

6. Conclusions

6.1. Implications and significance

The aim of the present study was to interrogate the contention that EFL classes can better be introduced to new vocabulary through translation than by contextual study. In particular, this involved replication of Prince's 1996 study which has been used to authorize resort to translation to teach new words. Our findings indicated that students who studied new vocabulary in context could perform as well as those who used translation on immediate posttests involving gapped sentences and translation. Additional findings were that advanced students learned new words more successfully than weak students and it is easier to find a match for an L2 word in L1 than the reverse.

The implication of these findings is to cast doubt on a belief that teachers in an EFL context will be more successful in vocabulary instruction by 'code switching' into L1 to increase efficiency of instruction. It is true that the effect size in favor of contextual

learning was small, suggesting that the immediate advantage may not be great but still is significant. Furthermore, switching to L1 to 'explain' words in L2 is to reduce students' exposure to comprehensible input and the classroom may be the only opportunity they have to receive such input. And lack of such input clearly is the great disadvantage EFL students suffer in comparison to ESL students.

The wider significance of these finding is to support the orthodoxy that language classes are better conducted in L2. It is more difficult for the instructor who shares the students' L1 to find contexts to develop vocabulary and resorting to L1 can appear an attractive short cut. But, maintaining L2 as the language of communication in the classroom does not seem to prevent students learning new words even though the students might prefer to do so via L1and teachers worry that contextual learning might lead to misunderstandings. Furthermore initial contextual learning of vocabulary can be supported by measures such as programs using graded readers that can reinforce vocabulary learning and increase the amount of comprehensible input that students experience (Alroe, 2012).

6.2. Limitations and further research

This study was instigated primarily to determine whether it is feasible to introduce intermediate EFL learners to new vocabulary without code switching and so its findings may not be applicable to beginners or advanced learners. More importantly an objection may be raised that the present study (as was so with Prince's) does not have a delayed post-test and so does not indicate whether either approach to vocabulary acquisition has long term value. This is indeed a limitation, but as Hulstijn (2008, p. 371) argues there is value in trying to find the optimal way to learn words initially.

Moreover long term retention may involve more complicated mechanisms and cognitive processes than short term. The area provides opportunities for further research as for instance suggested in a 2008 paper by Karpicke and Roediger. They report an experiment involving study and testing in which participants whose L1 was English studied Swahili words under four different conditions. The results showed that repeated study of the 40 target words led to little long term retention whereas repeated testing was efficacious and the difference was huge - up to four standard deviations ($d = 4.03$). The researchers concluded that it must be retrieval of vocabulary items, for instance, through repeated testing that enables the transition from short term to long term memory. So a focus on methods to facilitate the transition from short term vocabulary learning to long term retention to embed the learning from initial encounters with new words would be a profitable direction for further vocabulary acquisition research.

Acknowledgements

Thanks are due to Jim Murphy for criticism and proofing, Punchalee Wasanasomsithi for logistical support, Sumpanee Chinnawongs for advice, Chatcharin Sitprasert for formatting, fellow teachers at Chulalongkorn University for implementing the experiment and the students for participating.

References

Alroe, M. J. (2012). Strategies to increase EFL students' mastery of the core vocabulary of English. *EJRT E Journal for Researching Teachers, 5*(1), 1–17.

Auerbach, E. R. (1993). Reexamining English only in the ESL classroom. *TESOL Quarterly, 27*(1), 9–32. **http://doi.org/bmdp57**

Chang, Y. Y. (2010). English-medium instruction for subject courses in tertiary education: Reactions from Taiwanese undergraduate students. *Taiwan International ESP Journal, 2*(1), 55–84.

Chinokul, S. (2012). *Development of EFL teaching skills: Knowledge from methodology course to real classroom practice*. Unpublished manuscript. Chulalongkorn University, Bangkok, Thailand.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80*(2), 183–198. **http://doi.org/c75t75**

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*(12), 1304–1312. **http://doi.org/bhd52q**

Cook, G. (2010). *Translation in language teaching: An argument for reassessment*. Oxford: Oxford University Press.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Davies, M. (2008). The corpus of contemporary American English (Coca): 450 million words, 1990-2012. Retrieved from http://corpus.byu.edu/coca/

Elgort, I. (2011). Deliberate Learning and Vocabulary Acquisition in a Second Language. *Language Learning, 61*(2), 367–413. **http://doi.org/br6fdn**

Ellis, N., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *The Quarterly Journal of Experimental Psychology Section A, 46*(3), 533–558. **http://doi.org/fcjznq**

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly, 40*(2), 273–293. **http://doi.org/dnhtfp**

Groot, P. J. M. (2000). Computer assisted second language vocabulary acquisition. *Language Learning & Technology, 4*(1), 60–81.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. [Review of the book *Incidental L2 vocabulary acquisition: Theory, current research, and instructional implications*, by M. Wesche & T. S. Paribakht (Eds.)]. *Studies in Second Language Acquisition, 21*(02), 181–193.

Hulstijn, J. H. (2008). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381): Oxford: Blackwell.

Kang, S.-H. (1995). The effects of a context-embedded approach to second-language vocabulary learning. *System, 23*(1), 43–55. **http://doi.org/bx9fgv**

Karpicke, J. D., & Roediger, H. L. (2008). The Critical Importance of Retrieval for Learning. *Science, 319*(5865), 966–968. **http://doi.org/c2hd5f**

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics, 29*(4), 694–716. **http://doi.org/fbj8db**

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Applied Linguistics, 22*(1), 1–26. **http://doi.org/bc4rhp**

Laufer, B., & Shmueli, K. (1997). Memorizing New Words: Does Teaching Have Anything To Do With It? *RELC Journal, 28*(1), 89-108. **http://doi.org/bmcmhq**

Marqués Aguado, T., & Solís-Becerra, J. A. (2013). An overview of translation in language teaching methods: implications for EFL in secondary education in the region of Murcia. *Revista de Lingüística y Lenguas Aplicadas, 8,* 38-48. **http://doi.org/vsw**

Murray, M. (2013). *An exploration into what promotes or hinders beneficial English test washback on teaching and learning*. Unpublished manuscript. University of Southern Queensland, Toowoomba.

Napier, J., Rohan, M., & Slatyer, H. (2007). *Perceptions of bilingual competence and preferred language direction in Auslan/English interpreters*.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

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

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle Cengage Learning.

Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *The Modern Language Journal, 80*(4), 478–493. **http://doi.org/fn5zs4**

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24*, 146–161. **http://doi.org/dkgmsd**

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363. **http://doi.org/dh2bzh**

Schmitt, N. (2010). *Researching vocabulary*. Basingstoke: Palgrave Macmillan

Schweers, C. W. (1999). Using L1 in the L2 classroom. *English Teaching Forum, April-June, 37*(2), 6–9, 13.

Stempleski, S., Douglas, N., Morgan, J., Curtis, A. (2005). *World pass upper intermediate: Expanding English fluency (Bk.4).* Boston: Heinle.

Stoddard, G. D. (1929). An experiment in verbal learning. *Journal of Educational Psychology, 20*(6), 452–457. **http://doi.org/cqjjm2**

Takač, V. P. (2008). *Vocabulary learning strategies and foreign language acquisition*. Clevedon, UK: Multilingual Matters.

Tulving, E. (1983). *Elements of episodic memory*. Oxford: Oxford University Press.

Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal, 40*(3), 360–376. **http://doi.org/chrfsj**

Wiseman, S., & Tulving, E. (1976). Encoding specificity: Relation between recall superiority and recognition failure. *Journal of Experimental Psychology: Human Learning and Memory, 2*(4), 349–361. **http://doi.org/cddptc**

Footnotes

[1] It was expected that the 30 words chosen for study, and particularly the 16 used for testing, were words that the vast majority of first year students would have found problematic. However, when the treatment and posttest had been completed and it was found that most students had scored very highly on the posttest there was a concern that too many of the target words may have been already known to the participants before the treatment. Thus a 'post-control' group was chosen from a subsequent intake of first-year students and they were given the post-test on translation of the 5 words that had been used in the experiment at approximately the same stage of their Year 1 course as had experimental groups previously. They did not have a study period first as we wanted to see what prior knowledge they had of the test items. Half of the post-control translated into English and half into Thai. The post control consisted of three extant classes totaling 81 students with a CuTep average of 494/600 with a standard deviation of 39 compared to the experiment participants who had an average of 485 with a standard deviation of 55. They were roughly comparable and perhaps a little more skilled. Thus their results on the translation test should provide an estimate of what the participants would have scored without treatment. The post control group's medium score for translation into Thai was a little higher, as would be expected if translation into L1 is easier, but still well below the scores of the participants after they had completed their study phase and confirmed that the treatments had made a significant difference in experimental students' ability to translate the target words. The boxplot below shows the difference between the two groups for translation into English.

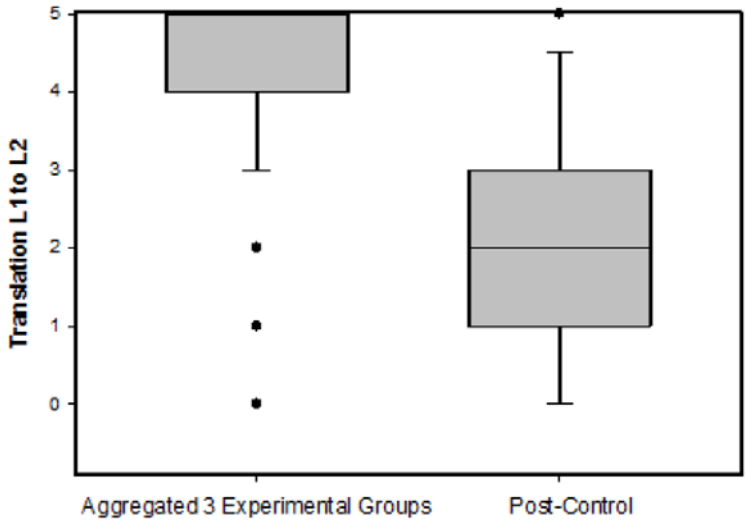


Figure 8:Experimental vs Post-Control Translation Score

[2] One word was dropped from the original 16 because in a post mortem discussion of the experiment in the faculty some Thai speakers raised doubts about the accuracy of the translation into Thai.

[3] On a delayed post-test (5 weeks later) to test long term retention Laufer and Shmueli found that the sentence context group was then equal to the translation pair group and both were significantly superior to the other methods of learning vocabulary.