

L2 Vocabulary Learning and Testing: Student Proficiency and the Use of L1 Translations versus L2 Definitions

Paul Joyce

Abstract

This study examines the relationship between student proficiency and the use of first language (L1) translations versus second language (L2) definitions in the learning and testing of L2 vocabulary. For this study, 48 Japanese L2 learners of English studied 200 lexical items from the Academic Word List over a ten-week period. The language in which the meaning of the target vocabulary was presented and tested was manipulated such that the learners were given half in their L1 and half in their L2. The results showed that the low proficiency group learnt significantly more vocabulary than the high group. However, while student proficiency did not interact with study language, it was significantly related to testing language. That is, both overall and over time, the higher ability learners did relatively significantly better on the L2 definition tests and worse on the L1 translation tests, and vice-versa.

Keywords: L2 vocabulary, L1 translation, testing, intentional learning, proficiency

Introduction

Vocabulary learning is essential to second language (L2) acquisition. And, among the many different elements of vocabulary knowledge, the form-meaning link is “the most essential lexical aspect which must be acquired” (Schmitt, 2008, p.333). In order to understand a range of written textual genres, it has been found that learners require comprehension of the 8000-9000 most frequent word families (Nation, 2006). One means of broadening L2 vocabulary knowledge is through intentional learning. Intentional vocabulary learning refers to working consciously to form associations between a word and its meaning. For learners from a wide range of proficiency

levels, it has been shown that the use of intentional study methods results in a greater vocabulary size expansion than a comparable time spent using an incidental approach (Laufer & Shmueli, 1997; Prince, 1996). Thus, as a complement to incidental vocabulary learning, intentional study is a valuable means of achieving greater vocabulary breadth.

The most prevalent forms of intentional learning involve associating target vocabulary with L1 translations or L2 definitions. The use of L1 translations has been shown to be of tremendous value (Latsanyphone & Bouangeune, 2009; Laufer and Shmueli, 1997; Prince, 1996; Ramachandran & Rahim, 2004). This is

likely due to translations being short, clear, and already existing in memory (Hall, 2002). Nevertheless, there remains a widespread belief that the L1 should not be used in the acquisition of L2 vocabulary (e.g. Gefen, 1987; Rivers & Temperley, 1978). This view frequently rests on the belief that L1 translations do not accurately represent the meaning of the word they are defining. While this sometimes can be the case, due to the tremendous conceptual commonality between languages, there is very often a close correspondence between the meaning of L1 and L2 words (Swan, 1997).

Nevertheless, there is an ongoing debate regarding the importance of learner proficiency in deciding whether to use an L1 translation or L2 definition in the intentional study of vocabulary. To address this issue, it is worth considering how L2 learners represent language. Psycholinguistic studies have revealed that when processing L2 lexis, language learners from a wide range of proficiencies access their L1 (e.g. Hall, 2002). A conceptual paradigm that accounts for this is the word association model (Potter, So, von Eckardt & Feldman, 1984). This theory holds that L2 words are tied to conceptual representations via the corresponding L1 word. Thus, when requiring access to a L2 vocabulary item, learners are forced to first activate the equivalent L1 item. While the word association model refers to lexical access, it suggests that L2 words are most efficiently learnt by way of their L1 equivalents. In contrast, it has also been hypothesized that rather than being sequentially related to one another, L1 and L2 words are each directly associated with non-linguistic concepts. This is known as the concept mediation model (Potter, So, von Eckardt & Feldman, 1984).

The above models of bilingual memory representation offer a rationale why L2 beginners should pursue L2 intentional

learning through L1 translations and more advanced students should use L2 definitions. However, since the psycholinguistic studies are based on single word access rather than sentence-based L2 definitions, and refer to word retrieval rather than acquisition, they do not necessarily provide a reliable guide to how students should conduct L2 intentional study.

There have been a number of intentional L2 vocabulary learning studies that have explored the benefits of L1 translation as an aid to incidental L2 vocabulary learning. Ramachandran and Rahim (2004) taught the meaning of 20 contextualised target vocabulary items to 60 Malaysian learners of English. While half of the learners were provided with L1 translations, the other half were given L2 explanations. The post-test results revealed the effectiveness of the translation method over the L2 approach. Similarly, in a study involving 169 Laotian learners of English, Latsanyphone and Bouangeune (2009) also found that the L1 translation group significantly outperformed those learners who had been provided with L2 definitions. However, while these studies suggest the benefits of L1 translation over L2 vocabulary acquisition, they were both conducted with elementary level participants. Therefore, beyond low level learners, they offer little guidance to the optimal relationship between proficiency and study method. In addition, in both of the above studies, participants were evaluated using L1 based test items at the post-test stage. However, L1 translation tests have been shown to be easier than L2 definition vocabulary tests and students have been revealed to score higher when the study language and testing language are matched (see Author, 2015). Therefore, due to contamination from testing language and matching effects, there is some doubt over the value of the results from these studies.

In a study involving a broader range of proficiency levels, Author (2015) addressed the interconnected relationship between study language, test language, and study time on the intentional learning of L2 vocabulary. It was found that while there were substantial benefits to intentional study, the language in which the target vocabulary was studied did not affect learning. However, since the L2 proficiency of the participants varied widely, and the role of proficiency in L2 acquisition was not encompassed by the study, the research did little to shed light on the impact of proficiency on intentional vocabulary learning style.

In a research study that did compare participants of different proficiency levels, Prince (1996) divided 48 French learners of English into low and high proficiency groups. The students all studied 44 target vocabulary items that were considered unknown to them. Half of each proficiency group was told to learn the lexis using a translation and half through context. The results indicated that the students scored higher in the translation learning condition than the context learning one. And, when evaluated through the context learning test, the advanced learners fared far better than the less proficient group. However, surprisingly, the low students that studied the target language using translations performed better on the translation post-test than the high group in the same condition. Despite the study not including an L2 definition element, the results suggest that there is an interaction between study method, testing method, and proficiency.

Elgort (2013) researched the relationship between testing method and proficiency further. In this study, the 121 participants' results on the regular Vocabulary Size Test (VST) (Nation & Beglar, 2007) were compared to their scores on a bilingual version (English-Russian). The findings showed that the

participants performed significantly more accurately on the bilingual test than the monolingual version. It was also found that as L2 lexical proficiency increased, the benefits of bilingual presentation declined. While this study shed light on the issue of testing language, the findings would benefit from being corroborated in different contexts, with different tests, and using alternative research methodologies.

From reviewing the literature, it is clear that the interaction between proficiency and intentional study methods remains insufficiently understood. As Folse (2004, p.68) notes, "Research should now move to...whether the value of L1 translation is as effective for higher-proficiency students as it is for lower-proficiency students". In addition, there is also a need to better understand the relationship between proficiency and testing language. Therefore, it was determined that the following research questions would be addressed:

1. At different levels of L2 learner proficiency, is there a significant difference in receptive L2 vocabulary learning through the use of L1 translations versus L2 definitions?
2. At different levels of L2 learner proficiency, is there a significant difference in receptive L2 vocabulary test scores when knowledge is evaluated through L1 translations versus L2 definitions?

Methodology

Participants

The study took place at a university specializing in foreign language studies in Japan. All of the 48 participants were native Japanese L1 speakers. They were enrolled as

freshman English language majors. Since one participant did not complete the study, results were only collected from 47 of the learners. The students who participated in the research were in two class groups that both used the same syllabus and course materials. However, in terms of their proficiency, one class could broadly be described as being of an intermediate standard while the other was other was at a beginner level. For the purposes of the study, the classes will be referred to as the low and high proficiency groups. The participants' English proficiency was measured through the university's in-house proficiency test; the Kanda English Proficiency Test (KEPT). Through standard multiple regression analysis, there has been found to be a high degree of predictability between scores on the KEPT and TOEFL tests (see Bonk, 2001). When KEPT scores are used to predict TOEFL scores, the TOEFL Paper-Based Test scores of the low group was predicted to average 444.69 (*SD* = 11.86) and the high group 492.87 (*SD* = 19.47) (see Bonk, 2001).

Design

The participants were divided into two vocabulary learning groups within their classes; Group A (23 members) and Group B (24 members). Group A contained 12 low and 11 high proficiency participants, and Group B included both 12 low and 12 high learners. The Academic Word List (AWL) (Coxhead, 2000) knowledge of each of the two proficiency sub-groups (low and high) was counterbalanced across Groups A and B. Further details on how this was done are given in the Placement Test section. Group A and B each studied the same 200 AWL vocabulary target items. Group A studied the meaning of the first 100 target words (List A) using an English definition (ED) and studied a second list of 100 words (List

B) using a Japanese translation (JT). On the other hand, Group B studied List A using a JT and List B through an ED. The learners' receptive knowledge of the target vocabulary was assessed through multiple choice pre-tests and post-tests. Half of these tests employed English definition test (EDT) items and half Japanese translation test (JTT) items. The language of testing was also manipulated within subjects. That is, the students conducted half of their study in the same condition as they were tested (i.e. ED-EDT and JT-JTT) and half in a different condition (i.e. ED-JTT and JT-EDT). The definitions and translations that were used in the tests were the same as those that were employed in the word lists. A summary of the research design is contained in Table 1 below:

Table 1: The research design

Group	Task	Vocabulary List A		Vocabulary List B	
A	Study	Use English definition (ED)		Use Japanese translation (JT)	
	Tests	Test 1a (EDT)	Test 2a (JTT)	Test 3a (EDT)	Test 4a (JTT)
B	Study	Use Japanese translation (JT)		Use English definition (ED)	
	Tests	Test 1b (JTT)	Test 2b (EDT)	Test 3b (JTT)	Test 4b (EDT)

Materials

Placement Test:The 36-item version of the Academic Vocabulary section of the Vocabulary Levels Test (VLT) (Schmitt, Schmitt, & Clapham, 2001) was used to evaluate the participants' prior knowledge of the AWL. The test was administered a week before the learning phase of the study (Cronbach's alpha = .90).To enable the results section to focus upon the answers to the research questions, the statistical basis for the placement of the students into Group A and B is provided in this section. Using a matched pair methodology, the participants were divided into Groups A and B such that the two sub-groups were counterbalanced for prior knowledge of the

AWL. The result of an independent samples t-test showed that there was not a significant difference between the low students in Group A (*Mean* = 15.54, *SD* = 4.64) and those in Group B (*Mean* = 15.30, *SD* = 4.56); $t(22) = -.13, p > .05$. Likewise, no significant difference was found between the high students in Group A (*Mean* = 26.00, *SD* = 4.38) and B (*Mean* = 26.00, *SD* = 4.73); $t(21) = .00, p > .05$. However, a statistically significant difference was found overall between the low (*Mean* = 15.42, *SD* = 4.50) and high (*Mean* = 26.00, *SD* = 4.46) proficiency learners; $t(45) = -8.06, p < .001$.

Target vocabulary and the consolidation task: As previously discussed, the study focused on the learning of 200 headwords from the AWL (Coxhead, 2000). The AWL is a compilation of 570 word families that occur with great frequency across a wide range of academic texts. Since the research centred on the initial stage of vocabulary learning, it was important that the target vocabulary was unknown to the participants. Therefore, the lexis considered least likely to be known was selected. The likelihood of being able to choose such vocabulary was considered high since experienced language teachers have been shown to be capable of predicting with which words students are unfamiliar (Brutten, 1981). For each of the two word lists (List A and List B), an equal number of verbs (60), nouns (20) and adjectives (20) was selected.

For each of the target words, the participants were provided with either an ED or a JT. The L2 definitions were based on learner dictionary entries. As the AWL assumes knowledge of the first 2000 word families, only words from this high frequency range of the BNC-COCA corpus

were included in the EDs. As words are frequently polysemous, care was taken to ensure that the L2 definition and L1 translation referred to the same aspect of meaning. This task was completed by a paid bilingual.

The 200 target vocabulary items were divided into 10 lists of 20 words. For 10 weeks, at the start of each week, the participants were each provided with a new word list that contained the same 20 new target vocabulary items. The only difference in the word lists given to the two groups was the language in which the meaning of the target vocabulary was provided. While one group received a list of 20 target words with an ED, the other studied the same lexis with a JT. Each week, for both groups, the language in which the meaning of the target vocabulary was given alternated. The participants were also provided with the target vocabulary item's part of speech, and an example sentence that contained the word. To encourage the participants to study the vocabulary, they were given a quiz at the end of each week. The 20-item quizzes required them to receptively match the target words with an ED or JT. The quizzes that the two groups received differed only in the language in which the meaning of the target words (the definition or translation) was supplied. When completing the quizzes, those students that had studied the target vocabulary that week using EDs were required to match the target vocabulary to EDs, and vice-versa.

Pre- and post- vocabulary tests: All of the participants took a pre and post-test. The tests assessed the students' minimalistic receptive word recognition. The pre-test that each of the two groups undertook was

identical to the post-test that they were given. There were 120 items on both the pre and post-test. As shown in Table 1, the pre and post tests were each sub-divided into four sub-tests. Each of the sub-tests contained 30 items and 50 answer choices. Each answer choice appeared only once in the pre-test and only once in the post-test. Therefore, all 200 target words appeared on the pre and post-test. Each of the four sub-tests was similar in design to the VLT in that the test items were divided into clusters. Each of the clusters contained six EDs or JTs which were identical to those provided in the word lists. The participants had to match the six EDs or JTs to the correct option from ten candidate target words. Within each cluster, the answer choices had distinctly different meanings, but were all from the same part of speech. An example of a cluster from an EDT is available in Appendix 1. Each of the sub-tests contained five clusters; three verbs, a noun, and an adjective test question cluster. As shown in Table 1, Group A was administered sub-tests 1a, 2a, 3a, and 4a, while Group B was given sub-tests 1b, 2b, 3b, and 4b. The only difference in the sub-tests that the two groups completed was the language in which the meanings (i.e. EDs or JTs) were provided. To control for any possible order effect, the sequence in which the students were administered the tests was carefully counterbalanced at both the pre-test and post-tests stages. The tests were delivered during class time and neither of the groups was informed in advance that they were to be administered. The purpose of the pre-test was to establish the students' baseline knowledge. And learning was defined as the difference between the pre-test and post-test scores.

Proficiency test: The participants' proficiency was measured through the Kanda English

Proficiency Test (KEPT). In essence, the KEPT is a university in-house norm-referenced general proficiency test of English as a foreign language. When the data was collected, it was comprised of five sections; listening (35-items), grammar (35-items), reading (35-items), writing, and speaking. It is administered routinely at the university where the research was undertaken. The test was administered within two weeks of the vocabulary post-test being given. To allow the Results section to concentrate upon the research questions, the results from the proficiency test are provided here. The Cronbach's alpha reliability of the test sections was found to be consistently high, and ranged between .75 and .8.

There was not found to be a statistically significant difference in the overall proficiency scores of the low students in Group A (*Mean* = 62.18, *SD* = 4.02) and those of Group B (*Mean* = 63.81, *SD* = 6.15), $t(22) = -.77, p > .05$. Likewise, on the reading section of the test, no statistical difference was found (Group A; *Mean* = 21.95, *SD* = 2.68, Group B; (*Mean* = 22.06, *SD* = 3.91), $t(22) = -.08, p > .05$). The same result was recorded for the high proficiency students across Groups A and B. There was neither found to be a statistically significant difference in their overall proficiency (Group A; *Mean* = 76.07, *SD* = 5.98, Group B; *Mean* = 77.48, *SD* = 5.22, $t(21) = -.60, p > .05$) nor their reading proficiency (Group A; *Mean* = 26.20, *SD* = 4.66, Group B; *Mean* = 26.35, *SD* = 4.80, $t(21) = -.08, p > .05$). On the other hand, as was expected, there was found to be a significant difference in the overall proficiency of the low (*Mean* = 63.00, *SD* = 5.15) and high (*Mean* = 76.81, *SD* = 5.51) proficiency groups; $t(45) = -8.88, p < .001$. And there was also discovered to be a difference in the reading ability of these two

sets of participants (Low; Mean = 22.01, SD = 3.28, High; Mean = 26.28, SD = 4.62, $t(45) = -3.67, p < .001$).

Results

As discussed in Methodology Section, in order to control for the participants' prior knowledge of the target vocabulary, the learners from the two proficiency groups were divided into groups of equal AWL vocabulary knowledge. To confirm that the group variable had been controlled within each of the proficiency sub-groups, a series of one-way repeated measures ANOVAs was conducted. The effect of the three within-subject's independent variables (study language, test language, and point in time (the start and end of the study)) and one between-subject's independent variable (proficiency group affiliation) on learning were compared. There was not found to be a significant effect of sub-group for the low students ($F(1, 22) = .09, p > .05$), the high students ($F(1, 21) = 1.61, p > .05$) or the total combined group ($F(1, 45) = .50, p > .05$). As a result, group affiliation (Group A and B) was removed from the analysis. However, when the two proficiency sub-groups (low and high) were compared to each other, there was a main effect for group ($F(1, 45) = 8.55, p < .01$). As expected, the more proficient students had outperformed the less able participants. The Cronbach's alpha reliability of the eight 30-item vocabulary tests ranged between .87 and .93.

Research question one: *At different levels of L2 learner proficiency, is there a significant difference in receptive L2 vocabulary learning*

through the use of L1 translations versus L2 definitions?

As can be seen in Tables 2 and 3, the participants from both proficiency sub-groups substantially increased their knowledge of the target L2 vocabulary over the study period.

Table 2: Low students - percentage of correct responses by study language, test language, and point in time

Study Language	Test Language	Pre-Test		Post-Test	
		mean%	SD	mean%	SD
English Definition (ED)	English Definition Test (EDT)	20.97	12.49	61.12	23.17
	Japanese Translation Test (JTT)	33.75	18.62	71.39	16.80
Japanese Translation (JT)	English Definition Test (EDT)	8.48	10.95	53.62	20.18
	Japanese Translation Test (JTT)	35.00	16.53	79.30	19.04

Table 3: High students - percentage of correct responses by study language, test language, and point in time

Study Language	Test Language	Pre-Test		Post-Test	
		mean%	SD	mean%	SD
English Definition (ED)	English Definition Test (EDT)	42.47	18.52	69.57	25.28
	Japanese Translation Test (JTT)	54.78	19.03	69.41	22.03
Japanese Translation (JT)	English Definition Test (EDT)	44.35	19.13	62.47	22.94
	Japanese Translation Test (JTT)	58.41	17.73	79.71	19.77

The results showed that the low proficiency learners gained far more than the high sub-group over the course of the study. While the low students achieved an average percentage increase of 39.31% points (pre-

test: *Mean* = 27.05%; post-test: *Mean* = 66.35%), the high group only improved their performance by 20.29% points (pre-test: *Mean* = 50.00%; post-test: *Mean* = 70.29%). When the relationship between time and proficiency was evaluated through a factorial repeated-measures ANOVA, a significant effect was recorded, $F(1, 45) = 19.73, p < .01$, partial $\zeta^2 = .31$. The result indicated that the low proficiency sub-group increased their scores by significantly more than the high sub-group.

Next, the relationship between study language, time, and proficiency was explored. On average, after studying the target vocabulary with EDs, the low proficiency participants scored 38.89% points higher on the post-test than the pre-test (pre-test: *Mean* = 27.36%; post-test: *Mean* = 66.25%). This compared with a 39.72% point gain for the same students on vocabulary studied using JTs (pre-test: *Mean* = 26.74%; post-test: *Mean* = 66.46%). The high proficiency group's progress was also very stable across the study languages. They gained 20.87% points on words studied using EDs (pre-test: *Mean* = 48.62%; post-test: *Mean* = 69.49%) and 19.71% points when using JTs (pre-test: *Mean* = 51.38%; post-test: *Mean* = 71.09%). When the relationship between study language, time and proficiency was calculated through a factorial repeated-measures ANOVA, there was not found to be a significant interaction effect, $F(1, 45) = .28$. In other words, the effect on the participants' scores of studying through EDs or JTs did not differ depending upon their proficiency.

Table 4: Within-subject's effects for study language, test language, time, and proficiency on test scores

Effect	F	Hypothesis df	Error df	p	partial ζ^2
Study * Proficiency	.86	1	45	.359	
Test * Proficiency	7.87	1	45	.007	.15
Time * Proficiency	19.73	1	45	.000	.31
Study * Test * Proficiency	.00	1	45	1.00	
Study * Time * Proficiency	.28	1	45	.598	
Test * Time * Proficiency	4.69	1	45	.040	.09
Study * Test * Time * Proficiency	.44	1	45	.512	

Research question two: *At different levels of L2 learner proficiency, is there a significant difference in receptive L2 vocabulary test scores when knowledge is evaluated through L1 translations versus L2 definitions?*

As can be seen in Tables 2 and 3, the high proficiency students generally outperformed the low proficiency learners. However, the pattern of outperformance differed depending on proficiency. When the pre and post test scores are combined, the difference between the low proficiency sub-group's performance on the JTTs (*mean* = 54.86%) is 16.17% points higher than their EDTs scores (*mean* = 38.54%). On the other hand, for the high proficiency sub-group, the difference between their JTT (*mean* = 65.58%) and EDT (*mean* = 54.71%) scores was only 10.72% points. As shown in Table 4, there was consequently a significant interaction effect between test language and proficiency $F(1, 45) = 7.87, p < .01$, partial $\zeta^2 = .15$. Furthermore, this result could also be seen over time. That is, the low group increased their pre-test to post-test scores by more on the JTTs (+40.97% points) than the EDTs (+37.64% points). In contrast, the high sub-group followed the opposite pattern (JTT: +17.97% points, EDT: +22.61%). This pattern of relative performance on the two types of test differing over time depending upon proficiency was found to be statistically significant $F(1, 45) = 4.69, p < .05$, partial $\zeta^2 = .09$. Therefore, the results show that both overall and over time, the advantage of

taking the bilingual test declined as proficiency increased.

Discussion and conclusion

The results from this study showed the relationship between proficiency, study language, test language, and study time on the intentional learning of L2 vocabulary. The intentional approach to L2 vocabulary learning was confirmed to be an effective means of vocabulary learning for both beginner and intermediate level L2 students. This result is consistent with previous direct vocabulary learning studies (e.g. Latsanyphone & Bouangeune, 2009; Laufer & Shmueli, 1997; Prince, 1996).

However, although the vocabulary knowledge of both groups was aided by intentional vocabulary study, the low students learnt significantly more than the high sub-group. This result is both consistent with Prince (1996) and accords with the widely held belief that shallower vocabulary learning activities are better suited to beginners (e.g. Schmitt, 1997). The reason for this could relate to the simplicity of the activity. For less proficiency students, since the target lexis was decontextualized, their limited cognitive resources were sufficient to retain the language. Furthermore, the lower group may have felt more suited than the higher proficiency group to the narrow task of establishing a form to meaning relationship between a word and its meaning.

Regarding the first research question, there was not found to be a significant difference in the amount of vocabulary learning achieved at different proficiency levels depending on whether the vocabulary was presented through L1 translations or L2 definitions. That is, each proficiency group

increased its scores by a similar degree in both the L1 translation and L2 definition learning condition. These results diverged from previous studies (Latsanyphone & Bouangeune, 2009; Ramachandran & Rahim, 2004) in which participants were found to learn more vocabulary when studying through L1 translations than L2 definitions. However, as discussed in the Introduction, the value of the findings from these studies is undermined by their methodological shortcomings.

The answer to research question one can be better understood with reference to the models of bilingual memory representation that were discussed in the Introduction. Given that both proficiency groups in this study acquired a comparable amount of vocabulary regardless of the language in which the meaning of the target vocabulary was given, the results suggest that the learners had surpassed the ability threshold of the word association model (Potter, von Eckardt & Feldman, 1984). Therefore, consistent with the concept mediation model, both groups seem to have been sufficiently skilled to establish direct links between the L2 lexicon and conceptual understanding. However, the development of such connections may depend upon the mode of learning. As found in a previous study, although the degree of learning was not influenced by the language of study, there was a matching effect between study and test language (Author, 2015). In other words, when learners studied in the same language as they were tested, they were advantaged. Therefore, once learners exceed the word association model proficiency threshold, to help facilitate automaticity in L2 vocabulary retrieval, there is a basis to recommend vocabulary study through the L2.

As shown by this study, owing to the benefits accrued from direct vocabulary learning, this form of acquisition should form part of foreign language study, especially for elementary level students. Yet, despite the widely recognized benefits of intentional vocabulary study (see Folse, 2004), there is reason to believe that the percentage of words taught explicitly is extremely low (Tang & Nesi, 2003).

Concerning research question two, the low proficiency group was found to have been more advantaged by the bilingual tests than the high proficiency group. This finding matched that of Elgort (2013). Furthermore, during the study period, the low proficiency group improved by relatively more on the JTTs than the EDTs, and the high group progressed by relatively more on the EDTs than the JTTs. This result has implications for language testing. Owing to the ease with which large numbers of vocabulary items can be administered and scored, vocabulary testing is often used to indirectly measure general language proficiency. In such instances, such as in the case of many Japanese university entrance tests, L2 vocabulary receptive understanding is often evaluated through L1 translations. As previously discussed, both in this study and that of Elgort (2013), elementary students were advantaged by bilingual test items. Indeed, at the post-test stage, the lower proficiency group's average score on the JTTs (75.35%) exceeded that of the higher ability group (74.57%). Therefore, the results indicate that when vocabulary knowledge is being used as a proxy for overall proficiency, especially when students have the opportunity to study in advance, it is much better to administer an EDT.

In terms of study limitations, it should be noted that the time students spent studying the target vocabulary was not controlled in this study. Also, since the results are based on a sample of Japanese university students, the characteristics of this research population may reduce the generalisability of the findings to other learner groups with different L1s, cultural backgrounds or educational environments.

Given the paucity of research in this area, it should be stressed that the findings of this study are tentative and that further research is required to expand and deepen our understanding. For example, to address the study limitations mentioned above, it would be instructive to explore whether the results hold for learners of other nationalities, L1s, ages, and in different educational settings. Furthermore, there would be value in investigating the relationship between study method and part of speech. For example, could a potentially closer correspondence between the meaning of L2 nouns and their L1 translations mean that they are most suited to being studied through a learner's mother language? In addition, by collecting qualitative data, such as from questionnaires and interviews with students, a richer account of the relative advantages of L1 and L2 vocabulary study methods could be derived.

References

- Bonk, W.J. (2001). Predicting paper-and-pencil TOEFL scores from KEPT data. *Research Institute of Language Studies and Language Education, Kanda University of International Studies*, 12, 65-85.

- Brutten, S.R. (1981). An analysis of student and teacher indications of vocabulary difficulty. *RELC Journal*, 12(1), 66-71.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Elgort, I. (2013). Effects of L1 definitions and cognate status of test items on the Vocabulary Size Test. *Language Testing*, 30(2), 253-272.
- Folse, K.S. (2004). *Vocabulary myths: Applying second language research to classroom teaching*. Ann Arbor, MI: University of Michigan Press.
- Gefen, R. (1987). Increasing vocabulary teaching in Israel schools. *English Teachers Journal*, 35, 38-43.
- Hall, C. (2002). The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *International Review of Applied Linguistics*, 40(1), 69-87.
- Latsanyphone, S. & Bouangeune, S. (2009). Using L1 in teaching vocabulary to low proficiency level students: A case study at the National University of Laos. *English Language Teaching*, 2(3), 186-193.
- Laufer, B. & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal*, 28, 89-108.
- Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63 (1), 59-82.
- Nation, I.S.P. & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31 (7), 9-13.
- Potter, M.C., So, K.F., Von Eckardt, B. & Feldman, L.B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of verbal learning and verbal behavior*, 23(1), 23-38.
- Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *Modern Language Journal*, 80, 478-493.
- Ramachandran, S.D. & Rahim, H.A. (2004). Meaning recall and retention: The impact of the translation method on elementary level learners' vocabulary learning. *RELC Journal*, 35(2), 161-178.
- Rivers, W.M. & Temperley, M.S. (1978). *A practical guide to the teaching of English as a second or foreign language*. Oxford: Oxford University Press.
- Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 199-228). Cambridge: Cambridge University Press.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12, 329-363.
- Schmitt, N., Schmitt, D. & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18, 55-88.
- Swan, M. (1997). The influence of the mother tongue on second language acquisition and use. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 156-180). Cambridge: Cambridge University Press.

Tang, E.& Nesi, H. (2003). Teaching vocabulary in two Chinese classrooms: Schoolchildren's exposure to English words in Hong Kong and Guangzhou. *Language Teaching Research*, 7(1), 65-97.

The author: Paul Joyce is an Associate Professor at Kindai University in Osaka, Japan. He is from Britain. He undertook his M.A. (TESL/TEFL) studies from

Birmingham University in England and conducted his Ph.D. in Applied Linguistics at Surrey University in England. His research interests include vocabulary acquisition, testing, and second language listening.

Appendix-1

Example Item Cluster

- | | |
|--|----------------|
| 1) A group of people or things that are similar to each other | a) aspect |
| 2) The central or most important part of something | b) category |
| 3) Numbers that have been collected in order to provide information about something | c) core |
| | d) criteria |
| 4) The general rule that you use when you make a decision or form an opinion about somebody or something | e) guideline |
| 5) One of the qualities or parts of a situation, idea, or problem | f) instance |
| 6) The printing a book, magazine, and making it available to the public | g) philosophy |
| | h) psychology |
| | i) publication |
| | j) statistic |