What is the most effective second language vocabulary learning method?

Jason H. Moy, A13784035

Thesis submitted in partial fulfillment of the requirement for the degree of a B.S. in Clinical Psychology

University of California, San Diego

June 2020

Faculty Advisor: Timothy C. Rickard, Ph.D.

Abstract

Research suggested retention rate of an item increased when one was trained by retrieving the item. The present study will empirically investigate whether the benefit apply in learning second-language vocabulary. In this study, we asked native-Chinese speaker to study 48 English Chinese word pairs with 4 different training methods (productive vs. receptive vs. productive + picture, restudy). Their knowledge was assessed in two days using productive or receptive final test. Result indicated retention rate is higher using retrieval compared to restudy. Additionally, we found training in one-direction (e.g. productively) lead to substantial knowledge in the opposite-direction (e.g. taking a receptive posttest) except taking a productive posttest with receptive training. Our result also suggested that adding pictorial stimuli to productive training may be better than productive training alone. Theoretical and practical implication were discussed.

What is the most effective second language

vocabulary learning method?

Introduction

More and more people have started to learn a second language as a result of rapid globalization. For example, a report estimates that 300 million Chinese, almost the size of the U.S population, are learning English (Tan, 2015). When studying for a new language, it is necessary to learn its vocabulary. The present study will compare the efficiency of these four training methods, productive (i.e. first language \rightarrow second language), receptive (L2 \rightarrow L1), productive + picture (L1 + picture \rightarrow L2), and a restudy control using either a productive final test or a receptive final test.

Review of Productive and Receptive Vocabulary Learning Strategy

Researchers had investigated whether productive or receptive learning strategies are more effective. In productive learning, people will recall the L2 term when the L1 translation is shown to them (i.e. they produce the new term). In receptive learning, people will recall the L1 translation when the L2 term is shown to them. The previous literature has shown that productive training is better than receptive training when people were tested productively, and receptive training is better than productive training when people were tested receptively (Griffin & Harley, 1996; Mondria & Wiersma, 2004; Stoddard, 1929; Waring, 1997). For example, Griffin and Harley (1996) asked 110 American, native English speaking students to study 20 unfamiliar French words. First, participants were given a chance to see the French word and the English word together, one by one, for the entire 20 word-pairs. Then, they were trained productively for 10 words and receptively for the other 10 words. In the productive training session, participants

were shown an English word, then they were asked to write down the corresponding French word. After that, they were presented with correct the French word for feedback. In the receptive training condition, the procedures were the same, except they were shown a French word and wrote down an English word instead. Immediately after training, participants were assigned to take either a productive or a receptive final test immediately after the training session crossed with training type. The authors found that productive training is more effective than receptive training on a productive final test, and receptive training is more effective than productive training on a receptive final test.

A limitation of the Griffin and Harley (1996) study is that few people care about their vocabulary knowledge immediately after the training. Instead, they care more about their vocabulary knowledge several weeks and months after the training. Mondria and Wiersma (2004) addressed this limitation by adding a 2-week delayed final test. Their result was consistent with Griffin and Harley (1996) except that receptive training is not better than productive training on the 2-week delayed receptive final test.

The above pattern of results was shown to be true among participants that are proficient in the second language (Webb, 2009; Yanagisawa, 2016), that are familiar with the second language (Griffin & Harley, 1996; Mondria & Wiersma, 2004), and that have no prior knowledge of the second language (Stoddard, 1929). Virtually all of the above-discussed results can be explained by the transfer appropriate processing (TAP) theory which asserts that the degree of similarity between training and the final test determines performance on the final test. (Morris, Bransford, & Franks, 1977).

Review of Vocabulary Learning Strategies with Restudy Control

In the aforementioned studies that examined the effectiveness of vocabulary learning strategy, it was also argued that training receptively (or productively) led to a substantial amount of productive knowledge (or receptive knowledge). (Griffin & Harley, 1996; Mondria & Wiersma, 2004; Stoddard, 1929). Their argument was based on the finding that people were able to recall some target words when they trained receptively but tested productively, or vice versa. Since the participants did not know any of the target words prior to the study, any knowledge of the target word must come from the training itself. However, their evidence is not sufficient to support that argument since it is possible the mere exposure to the L1-L2 pair, instead of training, made participants learned the words. Therefore, it is necessary to compare the training condition with a restudy control to determine whether their argument is correct or not.

Yanagisawa (2016) addressed this limitation by adding a restudy control condition. In their study, 18 Japanese undergraduate students learned 24 English words. They learned 8 words productively, 8 words receptively, and 8 words by restudying. In the receptive training condition, participants were shown the English words and asked to recall the corresponding Japanese meaning. After that, the correct Japanese translation was presented on the screen, and participants were asked to self-record whether they were correct in their recall. In the productive training condition, the procedure was the same as the receptive training condition, except that they were asked to recall the Japanese translation when English words were shown. In the restudy condition, participants were shown the Japanese and English words simultaneously. Both immediately following training and after a 2-week delay, each participant took both a productive and a receptive final test. The format of the final test was identical to the practice test used in training except feedback is not provided, and the order of the final test was counterbalanced.

Yanagisawa's (2016) findings on productive and receptive training were the same as prior studies (e.g. Mondria & Wiersma, 2004) except they found that training receptively is not different from training productively at an immediate receptive final test. In addition, they found that both the productive training and the receptive training is better than the restudy control in both the immediate and the delayed receptive final test. In the productive final test, they found that the productive training is better than the receptive training in the immediate final test, but the difference was no longer significant in the 2-weeks delayed productive final test.

There are several limitations to this study. First, on the final tests, participants were asked to self-report whether they successfully recall the word, which is not an objective measurement of participants' vocabulary knowledge. The second limitation is that all participants took both the receptive and productive final test using the same item. Although the order of the final test is counterbalanced, they did not investigate, or at least did not report, whether there is an order effect. For example, for participants who took the productive final test first, then their performance in the receptive final test may be inflated, due to learning that may have taken place on the productive test. The receptive final test is particularly vulnerable to that effect considering the prior literature which showed that receptive final test is easier than productive final test (Griffin & Harley, 1996; Stoddard, 1929; Waring, 1997). In the present study, we will improve on the prior work by making the final test condition a between-subject factor instead of a within-subject factor. The third limitation is that the words that participants learned in the restudy condition were tested in an immediate final test and a delayed final test. Since the final test can be a learning event, the delayed final test is unlikely to be a pure measure of relative learning in the training conditions. Therefore, in the present study, we administered only a 2-days delayed final test. Fourth, their sample of 18 participants which is relatively small

compared to other studies (Mondria & Wiersma (2004) had 198 participants; Stoddard (1929) had 328 participants; Griffin & Harley (1996) has 110 participants), raising the possibility of low statistical power.

Review of Efficiency

All of the aforementioned studies share the potential limitation that time per trial and per condition during training was equated. There are two advantages of to complete the training trials at their own pace. First, the training has higher ecological validity since it approximates how people learn vocabulary in reality. Second, it allows participants to be more engage with the training. Since some training conditions require less time than other, participants may be distracted from the task, result in less learning, or engage in other memory technique, such as creating mnemonics, result in more learning. Specifically, we hypothesized the benefit of receptive training may be inadvertently inflated due to the time constraint since prior studies found that people spend less time in receptive training session (Webb, 2005, 2009). This hypothesis is supported by the result of Webb (2005). He asked Japanese students to study novel English terms by a reading or a writing task. In the reading task, participants were shown the target English word, the corresponding Japanese translation, and three example sentences based on the target word. In the writing task, participants were asked to generate a new sentence with the target English word and the corresponding Japanese translation shown on screen. Participant knowledge was assessed using many measurements, and the most relevant experimental design was that they asked the participants to produce the L2 term given the L1 translation, and vice versa. In experiment 1, participants were required to spend 12 minutes on the training session. The result of experiment 1 showed that reading was better than writing across all measurements.

In the discussion of experiment 1, they made the same argument as we did regarding the benefit of reading task can be inflated due to the possibility of engaging in other memory technique. Therefore, they improved experiment 1 by removing the time constraint in experiment 2 and everything else were identical between the two experiments. In contrast to experiment 1, experiment 2 showed that writing was better than reading across all measurements. That change in pattern of result supports our hypothesis that the benefit on receptive training can be inflated when time per trial is controlled. Therefore, it is important to replicate prior studies with self-paced rather than time-controlled timing.

Review of Picture-word learning

In the present study, we added a picture + productive training condition in addition to word-based translation training and restudy training condition. The picture + productive training condition required participants to recall the L2 terms when a picture and the L1 term are presented. Adding this new condition is motivated by the dual coding theory. The dual coding theory (DCT; Clark & Paivio, 1991) proposes that there are two systems in memory, one specialized in coding non-verbal object information (e.g. picture), the other specialized in coding verbal object information (e.g. words). The two systems function independently, meaning that a person can forget one code without forgetting the other (Clark & Paivio, 1991; Paivio, 1990; Reed, 2012). For example, according to DCT, a person who has stored both the word "dog" and an image of a dog can recall the learned item successfully regardless whether they retrieve the image or the word. Therefore, if an item is coded by both systems, this item is more likely to be retrieved compared to the item that is only coded by the verbal system.

Running Head: EFFECTIVE LEARNING METHOD

DCT is empirically supported by some experiments that focused on incidental L2 vocabulary learning (Yeh & Wang, 2013; Yoshii, 2006; Yoshii & Flaitz, 2002). For example, Yoshii (2006) asked participants to read a book in which the target L2 words had 4 different type of annotations, definition in L1 only, definition in L2 only, definition in L1 plus picture, and definition in L2 plus picture. On the final test, they assessed participants' vocabluary knowledge by asking participants to provide the L1 definition of the target L2 word. Consistent with DCT, they found that people perform better when they were trained with a picture plus a definition.

In addition to incidental vocabulary learning, there is also some evidence to suggest that training using picture \Rightarrow L2 method is effective when people are learning vocabulary intentionally. Barcroft (2007) compared the effectiveness of picture \Rightarrow L2 training with restudy control. In this study, participants were assigned to a picture \Rightarrow L2 condition or a restudy condition. In the picture \Rightarrow L2 condition, each picture appeared alone without its corresponding word for 6 seconds and with its corresponding word as feedback for another 6 seconds. In the restudy control condition, the picture and its corresponding L2 term was presented simultaneously for 12 seconds. On the final test, participants were asked to write the L2 terms when the picture was presented, and they collected data on three final tests which were administered immediately, in 2 days, and in 3 weeks. They found that picture \Rightarrow L2 condition outperform the control condition in all three final tests. In light of the result from these studies, it is likely that picture + L1 \Rightarrow L2 training is better than L1 \Rightarrow L2 training alone in L1 \Rightarrow L2 final test.

The Present Study

In our study, we investigated which one of the four vocabulary learning methods is most efficient when native-Chinese speakers learning English. Those four vocabulary learning methods are recalling the Chinese definition given the English term (i.e. productive training), recalling the English term given the Chinese definition (i.e. receptive learning), recalling the English term given the Chinese definition and a picture (i.e. picture + productive training), and restudying the English term and its Chinese definition. Our procedure is similar to prior literature investigating vocabulary learning method (e.g. Mondria & Wiersma, 2004), except we added two novel conditions (i.e. restudy and picture + productive condition).

Method:

Participants:

We recruited 151 undergraduate students at University of California, San Diego. We excluded 15 subjects since they have prior knowledge of 10 or more of the target words. We also excluded 33 subjects due to technical difficulty. After exclusion, we analyzed the data of 103 subjects (n = 51 for receptive final test, n = 52 for productive final test).

<u>Material</u>

Based on the result of our pilot study, we selected 48 GRE terms which are translated into Chinese using The New Oxford English-Chinese Dictionary, Second Edition (see Appendix A for English and Chinese word pair). Additionally, we found 48 pictures that match with the meaning of the term.

Design & Procedure

All participants were given a consent form to read and sign before the experiment. In the training session, participants had the opportunity to go through the entire word pair list, one by one, and they were asked to type "y" if they knew the meaning of any word pair. Each word pair was presented for 10 seconds on screen which should provide sufficient time for them to determine whether they know the meaning of the English word.

On the instruction page, participants were informed that for the remainder of the experiment there was no trial time limit. In the productive training condition, participants saw the Chinese word and were asked to recall the corresponding English term; i.e. seeing a prompt of "Please type the corresponding English term of [Chinese word]." When they were finished, they received correct answer feedback. In the receptive training condition, participants saw the English term and were asked to recall the corresponding Chinese translation (i.e. seeing a prompt of "Please type the corresponding Chinese translation of [English word]." The correct Chinese translation appeared on the screen after they finished. In the picture + productive training condition, participants saw the Chinese translation and a picture with a prompt "Please type the corresponding English term." Feedback was the same as for the receptive training condition. In the RE condition, participants were instructed to click the next page button when they felt that they had mastered the word pair. The order of the training condition and materials were counterbalanced. Upon completing the training session, participants were instructed to come back in 2 days to study new word pairs, though in fact they were tested on the previously seen pairs).

In the testing session, participants were randomly assigned to take the receptive final test or the productive final test. The format of the final test is similar to the corresponding training

condition except feedback was not provided. Upon completing the final test, participants were thanked and excused.

Result:

Descriptive Result:

Participants spent on average 5.534 minutes in productive training condition (SE = 0.243), 5.498 minutes in receptive condition (SE = 0.273), 5.821 in picture + productive condition (SE = 0.242), and 2.519 minutes in restudy condition (SE = 0.158)

Strict Scoring Result

We analyzed the data using mixed factors Analysis of Variance (ANOVA). The between-subjects factor was the Test-Type (productive final test vs. the receptive final test), and the within-subject factor was Training Task (receptive vs. productive vs. picture + productive vs. restudy). The main effect for test-type was not significant, F(1,101) = 3.215, p = 0.076. The interaction effect between test-type and training task was significant, F(3,303) = 12.810, p < 0.001. We further analyzed the data using two within-subject ANOVA to examine the effect of training in productive and receptive final test.

Productive final test result is shown in Figure 1, Panel a, and the within-subject ANOVA model was significant, F(3,153) = 20.693, p < 0.001. Pairwise comparison using Tukey HSD adjustment revealed that proportion correct in the picture + productive condition (mean = 0.324; mean proportion correct is abbreviated as M hereafter) was significantly higher than that in both the receptive training condition (M = 0.145; p < 0.001) and the restudy condition (M = 0.158; p < 0.001). Additionally, proportion correct in the productive training condition (M = 0.264) was

significantly higher than in the receptive training condition (M = 0.145 p < 0.001) and restudy condition (M = 0.158; p < 0.001). No other significant effects were observed.

Receptive final test result is shown in Figure 1, Panel b, and the within-subject ANOVA model was also significant, F(3,150) = 23.5, p < 0.001. Pairwise comparison using Tukey HSD adjustment revealed that proportion correct in the receptive training condition (M = .349), the productive training condition (M = 0.312) and the picture + productive training condition (M = 0.336) was significantly higher than in the restudy condition (M = 0.159; all three p < 0.001). No other significant effects were observed.

In additional to the raw proportions correct, we were also interested in the efficiency of learning. Efficiency is defined as the number of words participants learned per minute (i.e. efficiency = number of learned words / amounts of learning time). Using that dependent measure, we conducted the same type of ANOVA that was described earlier. Productive final test result is shown in Figure 2, Panel a. The within-subject ANOVA model was significant (F(3,153) = 10.584, p < 0.001). Pairwise comparison using Tukey HSD adjustment revealed that restudy condition (mean efficiency = 0.999, mean efficiency is abbreviated as M hereafter) is better than receptive condition (M= 0.353 p < 0.001), productive condition (M = 0.569; p = 0.002) and picture + productive condition (M = 0.666; p = 0.025). Additionally, we found that picture + productive condition (M = 0.666) is better than receptive condition (M = 0.353; p = 0.041). No other significant comparison was found. Additionally, when using efficiency as dependent variable, the within-subject ANOVA model for receptive final test was not significant (F(3,150) = 0.468, p = 0.705) as shown in Figure 2, Panel b.

Lenient Scoring Result

We also analyzed the data using a more lenient criterion. Three research assistants who are proficient in both Chinese and English independently scored the data based on the following instructions: (1) For grading the Chinese answer, graders were instructed to give 0.5 points to answers that have similar meaning to the correct answer; (2) For grading the English answer, graders were instructed to give 0.5 points to answers that were spelt similarly to the correct answer. The graders agreed on 94% of items. We used the most common score among three graders when there was not complete agreement.

We first analyzed the result using proportion correct as the dependent variable. The receptive final test result is shown in Figure 3, Panel b, we found that the pattern was identical to the strict scoring result. The productive final test result is shown in Figure 3, Panel a. Compared to strict scoring result, the proportion correct of the picture + productive training condition (mean proportion correct = 0.362) was slightly better than in the productive training condition alone (M = 0.292, p = 0.048) using proportion correct as the dependent variable. We also analyzed the result using efficiency as dependent variable. The pattern of results from receptive and productive final test were identical to the strict scoring result as shown in Figure 4, Panel a, b.

Discussion

Our findings are summarized as follows. In agreement with the prior literature, we found that productive training is better than receptive training on a productive final test (e.g. Griffin & Harley, 1996; Mondria & Wiersma, 2004; Yanagisawa, 2016). Additionally, our data did not detect any difference between receptive and productive training in a receptive final test. In contrast, Mondria & Wiersma (2004) found that receptive training is better than productive training on an immediate receptive final test, but the effect is no longer significant in the 2-weeks

delayed receptive final test. Both studies found that receptive training is better than productive training in a receptive final test without reaching statistical significance. One important difference between these two studies is that the final test of the present study was administered in 2 days. Therefore, it is likely that the memory benefit of training receptively for a receptive final test is not as large as the memory benefit of training productively for a productive final test, and the benefit of receptive training also diminish more quickly, possibly as soon as 2 days. In addition, we hypothesized the benefit of receptive training is inflated. We did not find any evidence to support that hypothesis.

In the current study, we have two novel conditions, the restudy control and the picture + productive training condition. Consistent with the literature on test effect (e.g. Carrier & Pashler, 1992), we found that the productive training and the picture + productive training condition is better than restudy control condition when using productive final test. However, in our data the receptive training proportion correct (M = 0.152) was indistinguishable from that of the restudy condition (M = 0.161) in the productive final test with strict scoring. Although that relationship needs clarification with future study, those results raise the interesting possibility that, in the current context, a training method which utilizes testing may not be better than restudying when the format of training and testing is different. In the receptive final test, we found that productive training, receptive training and picture + productive training is better than restudy condition, which is consistent with the testing-effect. Under strict scoring, the picture + productive training condition is better than productive training condition when using both the productive final test (0.06 difference) and the receptive final test (0.025 difference) without reaching statistical significance. Under lenient strict scoring, we found that picture + productive training is significantly better than productive training alone in productive final test. That results constitutes preliminary evidence that adding a pictorial cue to productive training condition may increase the memory benefit of that training.

Theoretical Implication

The results of the present study suggest that transfer appropriate processing (TAP) theory, which asserts that the degree of similarity between training and final test determine the achievement on the final test, applies to learning second language vocabulary. Additionally, we found evidence that even training in a reversed testing format compared to the final test (i.e. productive training on a receptive final test) may remain beneficial compared to restudy control. The present study also shows some preliminary evidence that the dual coding theory (DCT) applies to learning second-language vocabulary.

Practical Implication

Our study strongly suggest that productive training is best training method when students need do a productive task. For example, if the students want to write an essay, then the students should test themselves with recalling the L2 vocabulary when the L1 translation is presented.

One common option for that is flashcards. Students may also want to consider adding a pictorial cue to their flashcard to boost the memory benefit. If the students need to do a receptive task (e.g. reading), the current research cannot suggest which training method is most effective.

<u>Limitation & Future Direction</u>

Future work should examine whether adding pictorial cue to a training condition is better than the training condition alone in light of our preliminary evidence from the present study.

Clarification on that point can help guide recommendations on how to study vocabulary and would provide further tests of the applicability of coding theory to vocabulary learning tasks. In addition to pictorial cue, future studies could examine whether adding other cues, such as sound, will boost the memory benefit as expected according to the dual coding theory. The current (2 days) and prior work (2 weeks at max) only compared these training condition for relatively short retention interval, so future works may examine whether the memory benefit still exist in the long-term (e.g. 2 months).

Reference

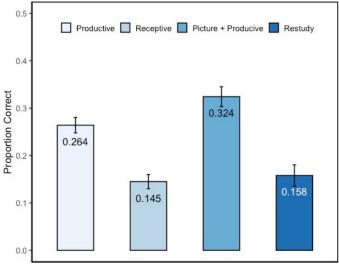
- Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, *57*(1), 35–56. https://doi.org/10.1111/j.1467-9922.2007.00398.x
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20(6), 633–642. https://doi.org/10.3758/BF03202713
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology**Review, 3(3), 149–210. https://doi.org/10.1007/BF01320076
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to

 Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *I*(1), 42–45.
- Griffin, G., & Harley, T. A. (1996). List learning of second language vocabulary. *Applied Psycholinguistics*, 17(4), 443–460. https://doi.org/10.1017/s0142716400008195
- Mondria, J. A., & Wiersma, B. (2004). Receptive, productive, and receptive + productive L2 vocabulary learning: what difference does it make? In P. Bogaards & B. Laufer-Dvorkin (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing,* (pp. 79–100). https://doi.org/10.1075/lllt.10.08mon
- Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

- Paivio, A. (1990). Mental representations: A dual coding approach. Oxford University Press.
- Reed, S. K. (2012). Cognition: Theories and Applications. CENGAGE learning.
- Stoddard, G. D. (1929). An experiment in verbal learning. *Journal of Educational Psychology*, 20(6), 452–457. https://doi.org/10.1037/h0073293
- Tan, H. (2015). China is losing interest in learning English. Retrieved December 11, 2019, from https://www.cnbc.com/2015/11/02/china-boosts-chinese-usage-slips-in-global-english-proficiency-ranking.html
- Waring, R. (1997). A study of receptive and productive learning from word cards. *Studies in Foreign Languages and Literature*, 21(1), 94–114.
- Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27(1), 33–52. https://doi.org/10.1017/S0272263105050023
- Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal*, 40(3), 360–376. https://doi.org/10.1177/0033688209343854
- Yanagisawa, A. (2016). The Effects of Receptive and Productive Word Retrieval Practice on Second Language Vocabulary Learning. *KATE Journal*, *30*, 139–152. https://doi.org/10.20806/katejournal.30.0 139
- Yeh, Y., & Wang, C.-W. (2013). Effects of Multimedia Vocabulary Annotations and Learning Styles on Vocabulary Learning. *CALICO Journal*, *21*(1), 131–144. https://doi.org/10.1558/cj.v21i1.131-144
- Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, 10(3), 85–101. Retrieved from http://llt.msu.edu/vol10num3/yoshii/

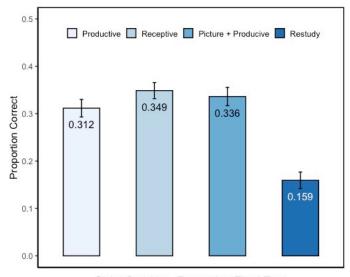
Yoshii, M., & Flaitz, J. (2002). Second Language Incidental Vocabulary Retention: The Effect of Text and Picture Annotation Types. *CALICO Journal*, 20(1), 33–58.

A.



Strict Scoring - Productive Final Test

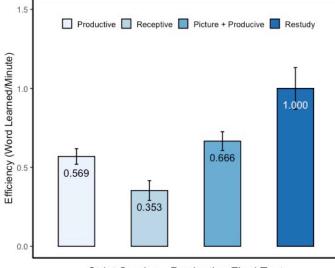
B.



Strict Scoring - Receptive Final Test

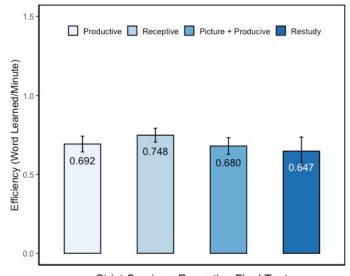
Figure 1. Performance on final test with proportion correct as the dependent variable using strict scoring. Panel a: productive final test. Panel b: receptive final test. Productive indicates training with recalling English term given Chinese definition. Receptive indicates training with recalling Chinese definition given English term. Picture + Productive indicates training with recalling English term given Chinese definition and a picture. Error bar are standard error calculated based on the Cousineau-Morey method (Cousineau, 2005; Morey, 2008).





Strict Scoring - Productive Final Test

B.



Strict Scoring - Receptive Final Test

Figure 2. Performance on final test with efficiency (i.e. word learned per minute) as the dependent variable using strict scoring. Panel a: productive final test. Panel b: receptive final test. Productive indicates training with recalling English term given Chinese definition.

Receptive indicates training with recalling Chinese definition given English term. Picture + Productive indicates training with recalling English term given Chinese definition and a picture. Error bar are standard error calculated based on the Cousineau-Morey method (Cousineau, 2005; Morey, 2008).

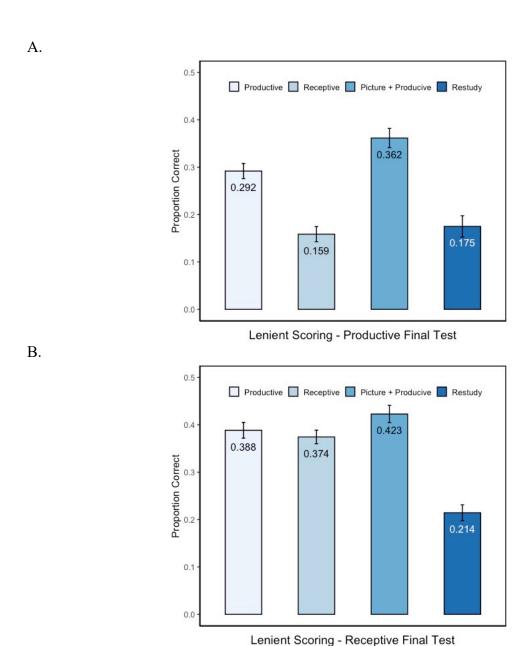
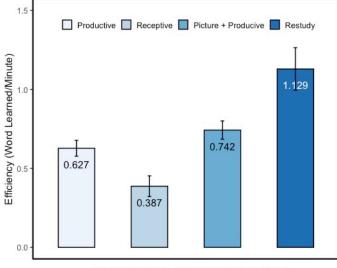


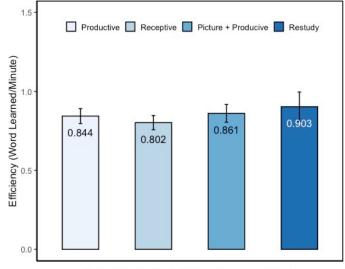
Figure 3. Performance on final test with proportion correct as the dependent variable using lenient scoring. Panel a: productive final test. Panel b: receptive final test. Productive indicates training with recalling English term given Chinese definition. Receptive indicates training with recalling Chinese definition given English term. Picture + Productive indicates training with recalling English term given Chinese definition and a picture. Error bar are standard error calculated based on the Cousineau-Morey method (Cousineau, 2005; Morey, 2008).





Lenient Scoring - Productive Final Test

B.



Lenient Scoring - Receptive Final Test

Figure 4. Performance on final test with efficiency (i.e. word learned per minute) correct as the dependent variable using lenient scoring. Panel a: productive final test. Panel b: receptive final test. Productive indicates training with recalling English term given Chinese definition.

Receptive indicates training with recalling Chinese definition given English term. Picture + Productive indicates training with recalling English term given Chinese definition and a picture. Error bar are standard error calculated based on the Cousineau-Morey method ((Cousineau, 2005; Morey, 2008)).

Appendix A:

English Chinese Word Pair

ENGLISH TERM	CHINSES TRANSLATION
Agog	渴望
Apprise	通知
Appurtenance	附属物
Augur	预言
Blazon	徽章
Bridle	束缚
Caricature	滑稽
Countenance	拥护
Dabble	玩水
Dawdle	偷懒
Don	教授
Dullard	笨蛋
Effervesce	兴奋
Ensconce	安置
Fulminate	责骂
Garble	篡改
Gauche	粗鲁
Gripe	牢骚
Guzzle	狂欢
Halcyon	宁静
Hale	强壮
Headlong	鲁莽

Heckle 质问
Inveigle 诱骗
Irksome 烦躁

Jabber 闲聊

Macerate 浸泡

Malaise 不适

Panache 炫耀

Paroxysm 急救

Piddling 少量

Potable 可饮用

Proselytize 改变宗教

Ragged 破旧

Remonstrance 抗议

Requite 补偿

Rile 生气

Sepulchral 阴森

Singe 烤焦

Squall 尖叫

Strand 遗弃

Umbrage 不快

Warble

Wayward 刚愎自用

Wheedle 哄骗