ATTACHING MEANING TO CHINESE CHARACTERS

Jacob Frank Tønnesen
A special thanks to everybody who helped me gather data, either by participating or by helping me recruit participants.
Abstract

An increased demand for speakers of Mandarin Chinese (henceforth Chinese) in the West has lead to an increased need for high quality teaching of Chinese. In the present study, 24 native Swedish speakers with no prior Chinese language experience studied 60 Swedish words paired with their Chinese equivalents. Half of the pairs were learned by repeated studying (with the word pairs presented simultaneously) and half of the pairs by active retrieval (with the word pairs presented separately). The durations of repeated study and active memory retrieval were equated. All the Chinese characters were presented along with a phonological representation. After learning the 60 word pairs, a multiple-choice test was conducted to assess learning outcome. To test for the importance of the phonological representation, half of the word pairs learnt in each condition had the representation removed during testing. Logistic mixed models found that performance was significantly better for repeated studying compared to active memory retrieval and that having access to a phonological representation during testing did not yield significantly better test performance.

Keywords: Mandarin Chinese, second language learning, the testing effect

Abstrakt

En ökad efterfrågan på talare av mandarin kinesiska (hädanefter kinesiska) i väst har lett till ett ökat behov av högkvalitetsundervisning i kinesiska. I den här undersökningen studerade 24 svenska modersmålstalare utan tidigare kinesisk språkerfarenhet 60 svenska ord i kombination med deras kinesiska motsvarighet. Hälften av paren lärdes genom upprepade studier (där ordparen presenteras samtidigt) och hälften av paren genom aktiv minneshämtning (där ordparen presenteras separat). Varaktigheten av upprepade studier och aktiva minneshämtningen likställdes. Alla de kinesiska tecknen presenterades tillsammans med en fonologisk representation. Efter att ha lärt de 60 ordparen, utfördes ett flervalstest för att bedöma läranderesultatet. För att testa för den fonologiska representationens betydelse vid inlärning så presenterades hälften av ordparen från varje betingelse utan deras fonologiska representation under testningen. Logistik blandade modeller fann att prestation var signifikant bättre för upprepade studier jämfört med aktiv minneshämtning och att tillgång till en fonologisk representation under testning inte gav signifikant bättre testprestanda.

Nyckelord: mandarin kinesiska, andraspråksinlärning, the testing effect
Learning to attach meaning to Chinese characters:  
A retrieval practice study of learners with no knowledge of Chinese

Adult second language learners rarely approach the learning of a new second language solely communicatively or orally. Adult learners frequently find themselves learning lists of words between their language classes. They often learn these as lists of paired words without communicative context. In languages with a different script, e.g. Chinese, Hebrew, Hindi, Japanese, Russian, and Thai, the learning of new words is combined with learning a new way of written word visualization. The focus of this paper is the learning of Chinese characters.

In recent years more people have started learning Mandarin Chinese (henceforth Chinese), which is now being taught from elementary school through to university level ("Mandarin", 2017). It is important that instruction is based on research to assure high quality teaching of Chinese as a foreign language (Han, 2014). A core aspect of language learning is the learning of vocabulary, and for Chinese this means the learning of Chinese characters. Therefore, throughout this paper, Chinese vocabulary will also be used to refer to Chinese characters. To help improve the learning outcome for learners of Chinese, the present study aims to investigate the importance of phonological representations when learning to attach meaning to Chinese characters as well looking at the method of learning.

In order to investigate the importance of phonological representations for learning written Chinese characters and to investigate whether retrieval practice or repeated study is the superior learning approach an experiment was designed, following Bertilsson, Wiklund-Hörnqvist, Stenlund and Jonsson (2017). Specifically, the present experiment compares test performance on items learnt by using repeated studying (rereading) and retrieval practice (active retrieval) as well as the impact of having access to a phonological representation.

Earlier research suggests that retrieval practice, that is active learning with a retrieval component leads to better performance on tests as well as better retention of the material being learnt compared to restudy without active learning, e.g., re-reading. In the context of learning words in a foreign language, retrieval practice could take the form of reading the target word and retrieving the first language translation from memory and then controlling if it's correct or not. Restudy would simply be reading the list of word pairs without any active retrieval. Several studies have revealed that retrieval practice is an effective learning method in general (see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013 for an overview) and for learning foreign vocabulary in particular (e.g., Karpicke & Roediger, 2008). Bertilsson et al. (2017) confirmed that retrieval practice was a more effective way to learn words in second language than a restudy approach. They investigated these ways of learning new words in Swahili, a language that was unfamiliar to participants. Using a within-subject design, upper secondary school students were tested on their learning after retrieval practice and reading approaches to learning word pairs. The testing occurred immediately after the learning, one-week later, and four-weeks later. Swahili uses a Roman script and thus the learners could attach a phonological representation to the word during the learning phase, even if this phonological representation was not the “correct” Swahili pronunciation. In order to investigate the importance of the phonological representation for learning new written representations of words and to investigate whether retrieval practice was superior to restudy for the learning of Chinese characters, an experiment that followed Bertilsson et al. (2017) was designed.

The phonological loop in vocabulary learning. Baddeley and Hitch (1974) proposed a simplistic model of working memory, which argued that limited amounts of information are held in short term memory with relatively little processing. The phonological loop is the part of working memory that deals with spoken and written material. It is specialized to retain verbal information over short periods of time and consists of a phonological store and a rehearsal process. The loop works by automatically retaining verbal information within a phonological
store, a sort of inner ear. Visually presented language can also be encoded into the phonological store by converting it into phonological code, which is then transferred to the phonological store. A second part of the loop consists of a rehearsal process, which helps maintain the phonological information and acts as an inner voice (Baddeley & Hitch, 1974). Parts of their working memory model have since been challenged, however one part of their model, the phonological loop, continues to form an element of many models of memory. According to the phonological loop, learning new and phonologically unfamiliar words, in either a first or a second language, depends on the availability of fitting representations of these sound patterns in the phonological loop.

Much research suggests that the phonological store plays a critical role in learning the phonological forms of new words (Baddeley, Gatercole & Papagno, 1998). It has been suggested to not only function for familiar words, but also to play a key role in learning new words, both during childhood but also during second language acquisition. Baddeley, Gatercole and Papagno (1998) argue that the phonological loop plays a crucial role when learning new vocabulary in a foreign language, stating that unfamiliar and phonologically similar items should be more confusable, thus leading to slower learning. They also state the possibility that the imitation of unfamiliar sounds may be a natural strategy which can be used to improve vocabulary acquisition by enhancing the representations of novel phonological structures within phonological loop (Baddeley, Gatercole & Papagno, 1998).

The process of reading Chinese characters differs from reading alphabetic scripts not only by having to distinguish more than 4000 logographs but reading Chinese appears to rely less on phonological processes and more on visual ones instead (Perfetti & Zhang, 1991, Tavassoli, 2002). Chinese characters evoke phonological representations in the reader as part of their identification, though, character identification is not mediated by phonemic processes (Perfetti & Zhang, 1991). Phonology alone can be an ambiguous cue in Chinese, at least for single character words without context, given the low number of syllables and large number of homophones in Chinese. Chinese may rely on a visuospatial buffer in the short term memory more so than reading alphabetic scripts do, because of the visual complexity of the characters (Tavassoli, 2002). These processing differences are supported by Huang and Hanley (1994), who found that even though learning to read Chinese is significantly correlated with phonological awareness, they did not find phonological awareness to be the primary cause of reading ability in children learning to read Chinese. Learning to read Chinese characters depends less on phonological awareness skills and more on visual skill than reading ability in English does. However, a lesser reliance on phonological representations does not mean that readers of Chinese view characters as purely visual representations as both logographic and alphabetic scripts involve phonological and non-phonological codes (Liu, Zhu & Wu, 1992).

**Learning Chinese.** Chinese uses a logographic writing system, with characters representing lexical morphemes. The characters are comprised of a number of strokes and stroke patterns, combined into square-shaped symbols. The principles behind the Chinese character can be described as the integration of sound, orthography and meaning, all of which are ideally integrated into the teaching and learning of Chinese (Zhang, Lin, Zhang & Choi, 2019). It is estimated that Chinese children learn more than 4000 different characters by the age of 12 (Huang & Hanley, 1994), with many of them looking visually similar (e.g., 木 / 林) making identification of the character a challenging task (Luo, Chen, Deacon, Zhang & Yin, 2013). Second language learners will be presented with a vast amount of Chinese characters that will have to be processed and learned throughout the study period. Unlike shallow orthographies, which have a close relationship between grapheme and phoneme (e.g. Spanish, Finnish, Turkish), the pronunciation of Chinese characters is almost completely opaque. Even though some of these stroke patterns, that Chinese characters consist of give a hint towards the pronunciation, these hints are far from reliable and it would be unlikely that someone would
be able to pronounce a word which they had not seen in print before (Huang & Hanley, 1994; Luo et al., 2013). This makes it a particularly challenging task for foreign learners of Chinese to create an extensive knowledge map that is able to connect graphic features and semantic values into their working memory. For this reason learners of Chinese are not only introduced to the characters themselves but also to the pinyin script connected with each character, adding another layer of information to process and retain (Wu, in Han, 2014). Typically introductory books for foreign learners of Chinese start out by familiarising the student with the pinyin system, subsequently, pinyin is provided, along with the English translation for new characters in word lists (Zhang et al., 2019).

Pinyin is a widely used romanized system, used phonetic alphabet for Chinese giving a representation of the pronunciation of the character including the tone. There are about 400 different syllables each with 5 different tones of which around 1300 are actually used. Given the nature of Chinese, where almost every single character (syllable) represents a word, there are many homophones (i.e. 时 (time), 十 (ten), 实 (real), 石 (stone), all with the same pronunciation, shí).

Visual processing plays a critical role in the acquisition of Chinese. A 2013 study by Luo et al. (2013) looked at the differences in visual processing of geometric shapes and Chinese character configurations and their contributions to Chinese reading. They found that geometric shape processing only explained variance in kindergarten children, whereas character configurations processing were better predictors of unique variance in older school children. Earlier studies found that learning Chinese depends much less on phonological awareness than for example English and learning to read Chinese was better predicted by the ability to learn unfamiliar figures measured by performance on tests of visual paired associates tests (Huang & Hanley, 1994). It is uncertain to what extent phonological knowledge plays a role in acquisition of Chinese.

This relative inconsistency in results from these studies has pointed to several questions on the topic that remain open for investigation (Luo et al., 2013). Some of the processing abilities of young Chinese children could in part be attributed to increased contact and exposure to Chinese characters in their environment, however it can also be argued that the same would be true outside of a Chinese context (i.e. Sweden), where, in the present study, participants are told that the characters they see are Chinese and are thus likely to be recognized as characters in a foreign language instead of just as random geometric shapes, thereby processing the visual stimuli as "words" opposed to "non-linguistic characters" (Huang & Hanley, 1994).

In a recent study Zhang et al. (2019) looked at Chinese foreign language learners' vocabulary knowledge and whether or not the presence of pinyin increased test scores. The aim of the study was to see whether presence of pinyin mattered in the testing of the learner's vocabulary knowledge and the experiment used to test this is highly relevant for the present study. During their experiment, Zhang et al. first presented the participants with a Chinese character without pinyin, then participants were asked to circle one of four pictures that best represented the meaning of the character. After answering participants were then presented with the Chinese character along with its pinyin and were allowed to revise their answer based on the additional information that the pinyin gave. Using this multiple choice test, the authors found that pinyin availability significantly better vocabulary test scores, as almost all items had successful corrections after pinyin was made available. Another study, conducted by Pierce and Hawthorn (2016), tested whether presentation modality had an impact on the testing effect. Here, 90 words were arranged in five lists, which the participants then studied either visually or auditorily before being tested on all five lists. They found the typical testing effect, but no significant testing by modality interaction. However, no condition by modality interaction was found, suggesting that testing had a similar effect regardless of the presentation modality.
The characters for the present study were chosen based on a more ecological approach, where the stimuli presented were characters that would normally be among the first characters you would encounter when beginning to learn Chinese. The characters were taken from the first chapters of Integrated Chinese (Liu, Yao, Bi, Ge & Shi, 2010), a commonly used learning book for Chinese as a foreign language. The chosen characters had a stroke count between three and fifteen.

**Retrieval practice.** Research on has long shown that testing memory not only is useful for measuring memory but the testing in itself also changes memory, often greatly improving retention of the learned material. Thus, testing by itself can be used as a strong learning tool enhancing later memory retention (Abbott, 1909; see Roediger & Karpicke, 2006 for a review). This learning occurs even when performance is far from perfect and no feedback is given. This form of retrieval practice is known as "practice testing" and "self-testing" (Dunlosky et al., 2013) most commonly called the testing effect (Roediger & Karpicke, 2006). The testing effect has been found across many different testing materials, including facts, paired associates, prose, statistics, and history (Carpenter, Pashler, & Cepeda, 2009; Roediger & Karpicke, 2006; see Bertilsson et al., 2017). The testing effect has also been observed for a wide range of language learning abilities, such as word lists, word definitions, foreign language vocabulary (Kang, 2010). Testing effect is also not limited to written/visual stimuli and has been found in experiments where stimuli are presented using multiple modalities (auditory vs visual) (Pierce & Hawthorne, 2016). It has also been shown to improve retention, even when the retrieval is done covertly, making it possible to implement these learning methods in situations where responding overtly is inconvenient or impossible (Kang, 2010).

The testing effect has also been described by Bjork (Bjork, 1988; 1994). They argue that easier learning leading to easier forgetting, while learning that requires more attention leads to better memory retention. It is these desirable difficulties that are aimed for in the experiment in the present study. It is important to note, though, that to take full advantage of the testing effect/desirable difficulties, the learner needs to be motivated i.e. they have to want to learn (Roediger & Pyc, 2012).

**Retrieval practice contra learning without retrieval.** Earlier the standard assumption has been that learning "occurs while people study and encode material" (Karpicke & Roediger, 2008, p. 966). Hence, additional study trials should increase the learning outcome, whereas tests are just neutral assessments with no additional learning effects. However, several studies have found that repeated retrieval enhanced long-term retention, while once information can be recalled, repeated encoding through studying gave no additional benefit (Karpicke & Roediger, 2008). Karpicke and Roediger (2008) also found that the rate of forgetting information is not determined by the speed of learning, but rather, is greatly predicted by how it was learned and the type of practice used. In an experiment, they found that four different practice conditions produced similar learning curves, however, the rate of forgetting was slowed more when using repeated recall.

In a study by Dunlusky et al. (2013) practice testing and distributed practice received high utility assessments as these methods could be used by learners of different ages and skill levels on many different subjects and in different contexts with great benefits. Even though testing is likely to be viewed as an undesirable necessity by students it shouldn't conceal the fact that testing does improve learning. Here, practice testing means a low or no stakes practice or learning that occurs outside of class and exam situations and includes any kind of testing that a student would be able to do by themselves. There are two different kinds of testing effects, which can be accounted for theoretically. These are direct effects and mediating effects. The direct effects arise from taking a test itself, and mediated effects, arising from an influence the amount or kind of encoding that happens after testing (Roediger & Karpicke, 2006).
Carpenter (2009) proposed that the direct effects of practice testing stems from triggering elaborative retrieval processes, arguing that the retrieval of the target information activates related information which is then encoded along with it allowing for easier access to that information later. Pyc and Rawson (2010, 2012) similarly argued that practice testing facilitates encoding by connecting cues and targets during subsequent restudy trials, arguing for the mediating effect of practice testing.

Most research has shown benefits of practice testing when time allotted for practice testing and restudying is equated and even when the amount of time allotted is modest (Dunlosky et al., 2013). The effect of practice testing with feedback is very robust, outperforming restudying alone. Even testing alone without feedback has been found to yield better performance compared to restudying, though the accuracy is lower relative to when feedback is given. The testing effect has been proven across many different kinds of material, learners and testing formats giving it a broad applicability. Furthermore it is not very time intensive and can be implemented with minimal training.

**Purpose.** It is unclear to what extent access to a phonological representation plays a role in the learning and recognition of Chinese characters. It is also uncertain whether having access to a phonological representation, separate from the target learning item, impacts testing outcome. The aim of this paper is to evaluate the impact of learning method and access to pinyin on test performance. The present study investigates this by presenting the participants with stimuli consisting of Chinese characters, presented along with pinyin and either simultaneously with or separately from the Swedish translation. An experiment was designed that followed the experimental design of previous studies on retrieval practice and vocabulary learning (for example, Carpenter, Pashler & Vul, 2006; Kang, 2010; Wirebring et al., 2015; Bertilsson et al., 2017). In order to increase the ecological validity of the present study, character, that would normally be among the first characters encountered, from a common Chinese foreign language learning book are used as the stimuli. In this way, a greater ecological validity is achieved than that of Kang (2010) who used simplistic characters, consisting of between two and four strokes each, with some characters having more than 20 strokes (e.g., 龠).

Based on the above, two research questions were posed; (1) how does access to pinyin in testing affect memory retrieval of the meaning of Chinese characters in testing? (2) Does removing access to a phonological representation have an impact on the testing effect?

Two main hypotheses were proposed to answer these, namely, (1) memory retrieval will be better for characters where pinyin is provided and (2) characters learnt using retrieval practice will have better test scores and having access to a phonological representation removed will worsen test scores.

**Method**

**Participants**

The sample consisted of 25 adults recruited around a University in Northern Sweden. One participant was excluded due to a computer error during testing. The final sample, thus, consisted of 24 adults (12 female) between 19 and 46 years of age ($M = 26.42, SD = 6.90$). All participants were native or fluent speakers of Swedish and had no formal exposure to Chinese language or Chinese characters. The participants reported that on average they could have a conversation with a native speaker in two languages (range 1-4, $M = 2.33, SD = 0.70$). Written informed consent was obtained at the beginning of the learning session and participants were free to terminate their participation at any point without any consequences (see appendix A).
Participants received no monetary compensation for participation, but were offered refreshments.

Design and Materials

Participants were presented with sixty Chinese-Swedish word pairs, with each pair being shown four times each. The word pairs were presented either as repeated reading pairs that required no retrieval on the part of the participant, or as retrieval practice pairs where the participant first saw the Chinese character after which the correct Swedish counterpart was presented. Each Participant learned 30 word pairs by retrieved practice and 30 word pairs by re-reading. To test for the importance of a phonological representation 15 of the word pairs learnt by retrieved practice and 15 of the words learnt by re-reading were presented without access to a phonological representation under the Chinese character during the testing phase. Further to reduce variance, no homophones were included and all the Chinese words presented consisted of a single character. The study used a within-subject design with learning condition, with presence of pinyin and test score as the main factors. Each participant learned 30 of the word pairs through repeated study, and 30 through retrieval practice. The learning of the 60 characters was assessed within subjects by means of an immediate post test.

The participants were instructed to focus on the Chinese character as they would be tested on retrieving the correct Swedish meaning matching the Chinese character. Participants were also told that the pinyin was only meant as a help and that they would not be tested on it. Unlike Kang (2010) and Bertilsson et al. (2017) there was no initial learning phases identical across subjects.

Stimuli. The to-be-learned Chinese characters for the 60 word pairs were chosen from the first four chapters of the textbook Integrated Chinese Level 1 part 1 (Liu et al., 2010) with fifteen single characters chosen from each chapter. The 60 word pairs were separated into two sets of 30, with the characters from one set being learned in one condition and the other half learned in the other (i.e. first set word pairs for retrieval and the second set for restudy). The characters within each set were matched for complexity by the number of strokes for each character in each set (set 1: range 3-15, M = 7.67, SD = 3.54; set 2: range 4-13, M = 7.70, SD = 2.60). The meaning of each character was translated into Swedish by the author and approved by a colleague with knowledge of both languages.

Testing material. The testing material consisted of the 60 stimuli characters.

Procedure

Learning phase. The 60 Chinese-Swedish word pairs were presented using PowerPoint on a 15 inch laptop.

For the restudy condition the word pairs were presented side by side with pinyin written below the Chinese character. Participants were instructed to not focus on the pinyin, that they would not be tested on it and that they did not have to remember it. Each word pair was presented one by one for 7 seconds (see figure 1).

![Figure 1: Learning phase without retrieval element.](image-url)
For the retrieval condition the word pairs were presented separately on two slides (see figure 2). First, the Chinese character with pinyin written below was presented for 4 seconds, then the Swedish translation was shown on a second slide for 3 seconds, to ensure that all word pairs were given an equal amount of exposure. For both conditions a 1 second "+" was presented between each word pair to indicate a new pair was about to appear.

First participants studied 30 characters from one condition followed by the remaining 30 characters in the second condition (i.e. restudy followed by retrieval or vice versa). Each of the 60 characters were studied four times for a total of 8 rounds (30 characters x 2 conditions x 4 repetitions) lasting 32 minutes. Allowing each subject to experience both learning conditions affords a stronger test of the hypothesis that participants have more correct answers on characters learned in the retrieval condition and that the presence of pinyin also increases test performance. The characters were pseudo-randomized between each round. Before learning the Chinese characters an initial practice run was conducted using four dummy-characters to ensure that the instructions and the method were clear to the participants. In the practice run two characters were shown for each condition followed by a multiple choice test on the four dummy-characters. Each character was shown along with five different options (see figure 3). No overt behavioral responses were required during the learning phase.

Figure 2: Learning phase with retrieval element.

Testing phase: After the learning phase, a multiple choice test was conducted on the same laptop as in the learning phase to test the participants ability to remember the correct meaning of each of the 60 characters. During the test all of the 60 learned characters were presented one by one along with 5 options (i.e. A through E) with all words being translations of characters learned within the same condition. Participants were then instructed to circle the correct answer on a piece of paper. Half of the characters in the test were presented without a phonological representation by having the pinyin removed. These 30 characters consisted of 15 characters from each condition spread across the 4 different chapters from which they were chosen. A second test set was created where the characters that had the pinyin removed were reversed. Each participant received one of the two tests at random. The test was not timed and participants were given as much time as they needed to complete the test. No feedback was given during testing and participants could not go back to revise their answers.

Figure 3: Test phase with and without pinyin respectively.

Statistical analysis. Analyses were performed in R using the lme4 and lmerTest packages. Three different logistic mixed models were run. A logistic mixed model (also known as mixed effects logistic regression) is used to model binary outcome variables (e.g.
correct/incorrect) using both fixed and random effects. Fixed effects are factors that are fixed across individuals, whereas random effects are factors that allow us to regard multiple responses from the same individual to be independent from each other by assuming different intercepts for that variable (Winter, 2013).

For all three models the outcome variable was a dichotomous measure of performance, i.e. whether a character was answered correct or incorrect. The first model is a simple model with learning condition as the only fixed effect and participant and character as random effects. The second model included the same random effects and furthermore included learning order (i.e. retrieval or restudy first), presence of pinyin in the test, number of strokes, number of languages spoken, sex and age as fixed effects. Here pinyin is of main interest, the other factors are mostly included as to control for any underlying differences not directly influenced by the experiment. The third model includes the same fixed effects as the second but also considers learning condition x presence of pinyin interactions. Number of strokes was included to determine whether stroke count was a good indicator of difficulty.

Table 1: **Overview of the three logistic mixed models.**

<table>
<thead>
<tr>
<th>Model 1, 2 and 3</th>
<th>Model 2 and 3</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>performance ~ learning condition</td>
<td>order + pinyin + strokes + languages + sex + age</td>
<td>condition x pinyin</td>
</tr>
<tr>
<td><strong>Random effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>performance ~ character(intercept) + participant(intercept)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Results**

Mean test performance for characters learned through restudy was 81.11% correct \((SD = 20.47)\), whereas test performance for characters learned through retrieval was 74.31% \((SD = 22.77)\). Looking at test performance with and without the phonological representation, having access to pinyin yielded slightly higher results with an average performance of 79.17% \((SD = 23.45)\) and 76.25% \((SD = 18.60)\) without access to pinyin (as shown in figure 4).

Average performance for each character between both conditions can be seen in the graph below. Performance was significantly better in the restudy condition, which is visualized in figures 4 and 5.

![Figure 4: Mean percentage correct for learning conditions (left) and testing conditions (right) with standard deviation.](image-url)
Figure 5: Mean percentage correct for each of the 30 individual characters split between learning conditions (retrieval left and restudy right).
Model 1, the simple model, revealed the effect of learning condition to be a significant ($p < 0.001$), with restudy being better than retrieval. Model 2, included more fixed effects and again restudy was significantly better than retrieval ($p < 0.001$), with age also found to be a significant ($p < 0.05$), with older participants performing best. Presence of pinyin was near significant levels ($p = 0.079$). For model 3 condition ($p < 0.001$) as well as age ($p < 0.05$) were significant, showing the same patterns as models 1 and 2, however, no significant interactions between learning condition and presence of pinyin was found ($p = 0.35$). There were no significant effects of the order of the conditions, number of strokes or number of languages spoken for either model 2 or 3.

Table 2: Results from all three logistic mixed models.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th></th>
<th></th>
<th>Model 2</th>
<th></th>
<th></th>
<th>Model 3</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std. error</td>
<td>p-value</td>
<td>Estimate</td>
<td>Std. error</td>
<td>p-value</td>
<td>Estimate</td>
<td>Std. error</td>
<td>p-value</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>2.705</td>
<td>0.4145</td>
<td>6.73E-11</td>
<td>2.66641</td>
<td>1.60683</td>
<td>0.097031</td>
<td>2.86642</td>
<td>1.62345</td>
<td>0.077458</td>
</tr>
<tr>
<td>Condition</td>
<td>-0.5488</td>
<td>0.1524</td>
<td>0.000316</td>
<td>-0.63011</td>
<td>0.1691</td>
<td>0.000194</td>
<td>-0.76686</td>
<td>0.22507</td>
<td>0.00656</td>
</tr>
<tr>
<td>Pinyin</td>
<td>0.27737</td>
<td>0.15806</td>
<td>0.07928</td>
<td>-0.16468</td>
<td>0.60141</td>
<td>0.742581</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>-0.20026</td>
<td>0.17261</td>
<td>0.245995</td>
<td>-0.19565</td>
<td>0.17278</td>
<td>0.257485</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strokes</td>
<td>-0.05394</td>
<td>0.04837</td>
<td>0.264764</td>
<td>-0.05375</td>
<td>0.04845</td>
<td>0.267269</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of languages</td>
<td>-0.56494</td>
<td>0.39544</td>
<td>0.153106</td>
<td>-0.56499</td>
<td>0.39583</td>
<td>0.153474</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>-1.13117</td>
<td>0.62691</td>
<td>0.071176</td>
<td>-1.13213</td>
<td>0.39583</td>
<td>0.071201</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.09976</td>
<td>0.0471</td>
<td>0.034168</td>
<td>0.09993</td>
<td>0.04714</td>
<td>0.034008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition x Pinyin</td>
<td>0.28554</td>
<td>0.30766</td>
<td>0.353354</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion

The aim of this paper was to evaluate the impact of learning method and access to pinyin on test performance. The otherwise well established testing effect was not replicated in this study, however additional delayed tests (one, two or three weeks after the initial test) might have revealed a testing effect, with better retention/slower rate of forgetting for characters learned in the retrieval condition, as found by previous studies (Dunlosky et al., 2013; Roediger & Karpicke, 2006; Karpicke & Roediger, 2008). On the contrary seeing both the Chinese character and the Swedish translation at the same time yielded significantly better results during testing compared to seeing them separately, arguing against the idea that more effortful learning leads to better memory as well as disproving the main hypothesis. Though the retrieval component was absent for the first round of retrieval during the learning phase, as it was the first time participants saw the given characters, there would likely be retrieval involved during the subsequent three repetitions of those characters requiring more effort to encode - theoretically leading to better retention. Given the simplistic design, having an initial learning phase using a word list, displaying all 60 word pairs, as Kang (2010) and Bertilsson et al. (2017) did, could be seen as an additional restudy trial. This would lead to an unequal exposure to each condition and therefore, no initial word list was presented.

Regarding pinyin, test performance was better when pinyin was displayed along with the characters, however the difference was only near significant levels using model 2, giving no strong support for the suggestion that having access to pinyin during testing allows for better test performance. It should be noted, though, that Zhang et al. (2019) suggested that having
characters presented together with pinyin during testing could be a more effective way of assessing Chinese learner's vocabulary knowledge and lead to better written vocabulary knowledge tests, as pinyin is often removed from tests and study material once the character has been learned. Though looking at the findings in the present results, access to pinyin does not yield significantly better performance in the test. As mentioned above, this does not necessarily mean that Chinese readers view the characters as pictorial representations of certain concepts (Tavassoli, 2002), however, it does argue that the phonological loop is not as important for the acquisition of Chinese, as for languages using alphabetic scripts. For alphabetic languages, access to a phonological representation is closely tied to the visual representation such as in orthographically shallow languages, which are often used in experiments with similar setups (i.e. Swahili). Therefore, there is the possibility that participants are really learning another written representation of, in this case, a Swedish word, without any knowledge of how it is pronounced, ignoring the phonological loop.

If the phonological loop plays a role in attaching meaning to the characters, it could also be that there is a cognitive overload, when introducing three different aspects for each character (i.e. visual, semantic and phonological representations), which could lead to pinyin not mattering and retrieval practice not being the optimal learning method. Further study is needed to fully draw conclusions regarding the importance of the phonological representation in attaching meaning to Chinese characters. Using auditory stimuli could help explore the role of the phonological loop in logographic scripts.

Lack of any significant effect from number of strokes indicates the that number of strokes in the Chinese character is not a good indicator of how difficult a character is to remember, at least for learners without any prior knowledge of Chinese. Therefore, other things would need to be changed if difficulty was to be increased or decreased. Had there been a significant effect of number of strokes, an additional interaction between stroke count and condition could have been included in model 3 to check for differences in difficulty between the two Chinese character sets.

There was some concern regarding whether the visual stimuli would be processed as words as opposed to arbitrary sign with no linguistic value, however previous findings have shown differences in the processing of arbitrary symbols and Chinese characters (using Chinese children as participants) (Luo et al., 2013), suggesting that even though the characters may be unknown to the participant, they are still processed as a word. To further ensure this, the instructions before the experiment also made it clear that these were actual Chinese characters they were about to learn.

The choice of recruiting participants from and around a university was made to increase the likelihood that those recruited were either currently studying or had studied at University, and performing the experiment in a university setting making it close to a normal learning environment, gives the study reasonable ecological validity.

**Further research.** The present experiment was done on a small scale with a relatively small sample of 24 participants. This is a limiting factor and could be a threat to the internal validity as well a limiting the power of the statistical analyses. Therefore, replication on a larger scale would be encouraged in future studies.

Since the results gathered from the statistical models go against an otherwise well established finding there is a risk that the results could be invalid. A way to test this would be to introduce a post test to see if memory retention would be better for characters learned via the retrieval condition, this is only speculation though and would need to be examined in another study. Another way to increase the strength of the results would be to subdivided the participants into more groups with different conditions, such as participants not having access pinyin during learning conditions or having no pinyin at all both during learning and testing. Future research would also include more stimuli and introducing them in different ways, such
as having pinyin displayed along with each character for both learning and testing phases or having no pinyin at all. This could also be extended to having longer training periods in an ecological setting - e.g. using actual learners of Chinese applying the learning methods and tests in class.

Another consideration would be to include compound characters (i.e. words with more than one character), which could lead to a higher performance when giving more clues for the same word however, it would also be interesting to see interactions between different words with some of the same characters, such as 注意 vs 意思 vs 意义. This also leads into the point about homophones and minimal pairs - words or syllables that differ in only one phonological element, be it phoneme or tone - of which there are a lot in the Chinese language. Homophones and minimal pairs were also present in the first few chapters of the learning material used in the present study, however, these were intentionally left out from the experiment, since these were thought to be too difficult potentially leading to floor results, if using participants with no prior knowledge of Chinese. Though, seeing as performance for a portion of the participants was at or near ceiling and overall performance being relatively high at more than 75% as well it might be fine to increase the difficulty in future experiments either by having more characters that are more similar in either form or meaning or as mentioned above, by introducing compound character, homophones and minimal pairs.
Reference list


Information och informerat samtycke om deltagande i studie om att lära sig glosor/nyckelbegrepp som ingår i magisteruppsatsen ”learning to attach meaning to Chinese characters”

1. Bakgrund och syfte
Ett ofta aktuellt område är hur vi ska få en bättre undervisning. Denna studie syftar till att undersöka olika metoder för att lära sig kinesiska tecken.

2. Förfrågan om deltagande
Läs igenom hela informations materialet och bestäm sedan om du kan och vill delta i studien. För svar på frågor utöver denna och den muntliga informationen så vänder du dig till ansvarig forskare Jacob Frank Tønnesen (se kontaktinformation i slutet av detta brev). Om du vill delta så fyller du i och skriver under det informerade samtycket (i två exemplar) som är bilagda. Det ena ska du behålla själv och det andra samlar vi in.

3. Hur går studien till?
Studien innebär deltagande vid 1 tillfälle som tar ca 1 timme. Det kommer också att vara möjligt att delta i ett post-test som tar ca 10 minuter, kring en vecka efter det första tillfället. Vid första tillfället kommer du att lära kinesiska tecken och sedan testas du i hur bra du kan komma ihåg betydelsen av tecknen. Om du deltar i post-testet, kommer du testas igen. Vid första tillfället får du också fylla i enkäter där vi frågar efter till exempel språkkunskaper, kön och ålder.

4. Vilka är riskerna?
Det finns inga kända risker att delta, uppgifter kan ibland upplevas som lite ansträngande, men är i den bemärkelsen inte jobbigare än ett vanligt prov i skolan.

5. Finns det några fördelar?
Många personer säger att de tycker att det är intressant och lärorikt att delta i forskningsprojekt. Detta är den främsta fördelen vi ser med deltagande.

6. Hantering av data och sekretess
Dina resultat sparas på dator under ett anonymt kodnummer. Dina svar och dina resultat kommer att behandlas så att inte obehöriga kan ta del av dem, i enlighet med GDPR. Studiens resultat kommer att presenteras på gruppnivå (till exempel medelvärden). Du har rätt till ett utdrag ur detta register och begära rätten av eventuellt felaktiga uppgifter. Dina resultat från studien sparas i vanliga fall i 10 år och raderas sedan.
7. Hur får jag information om studiens resultat?
Studiens resultat kommer att presenteras i en magisteruppsats som går att få tillgång via DiVA.

8. Försäkring, ersättning
Sedvanlig försäkring gäller under studiens genomförande. Det kommer inte att vara någon monetär ersättning för ditt deltagande.

10. Frivillighet
Deltagande i forskningsprojektet är frivilligt och du har när som helst, utan särskild förklaring, rätt att avbryta deltagandet. Du har rätt att begära att dina resultat förstörs eller märks så att de inte längre är möjliga att spåra till dig. Ändrat samtycke av personuppgifter samt resultat anmäls till ansvarig forskare (se nedan).

11. Ansvariga
Forskningshuvudman, personuppgiftsansvarig och ansvarig forskare för studien är Kirk Sullivan (supervisor) och Jacob Frank Tønnesen (MA student).

Kontaktinformation:
Jacob Frank Tønnesen,
institutionen för psykologi, Umeå Universitet
901 87 Umeå
jacob-frank@homail.com
+45 31 13 18 19