

**How Knowledge of Context and Radicals Influences  
a Person's Ability to Learn to Read Chinese Characters**

by

Tanya M. Kortz

B.S., Electrical Engineering and Computer Science, Massachusetts Institute of Technology (2011)

Submitted to the Department of Electrical Engineering  
and Computer Science

in Partial Fulfillment of the Requirements for the Degree of  
Master of Engineering in Electrical Engineering and Computer Science

at the

Massachusetts Institute of Technology

June 2012

Copyright 2012 Tanya M. Kortz. All rights reserved.

The author hereby grants to M.I.T. permission to reproduce and  
to distribute publicly paper and electronic copies of this thesis document in whole and in part in  
any medium now known or hereafter created.

Author:

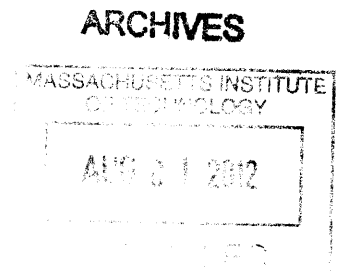
\_\_\_\_\_  
Department of Electrical Engineering and Computer Science  
May 21, 2012

Certified by:

\_\_\_\_\_  
Prof. Patrick H. Winston, Ford Professor of Engineering/MacVicar Fellow, Thesis Supervisor  
May 21, 2012

Accepted by:

\_\_\_\_\_  
Prof. Dennis M. Freeman, Chairman, Masters of Engineering Thesis Committee



**How Knowledge of Context and Radicals Influences  
a Person's Ability to Learn to Read Chinese Characters**

by

Tanya M. Kortz

Submitted to the  
Department of Electrical Engineering and Computer Science  
May 21, 2012

In Partial Fulfillment of the Requirements for the Degree of  
Master of Engineering in Electrical Engineering and Computer Science

**ABSTRACT**

Learning a new language is not always easy. This is particularly the case for native English speakers, who are used to letters combining to form words and tones expressing emotion, trying to learn Chinese. Mandarin Chinese is a logographic language, meaning that each character represents a word and voice tone is used to differentiate between what would otherwise be homophones. In this thesis, I present two hypothesized methods for improving a student's ability to learn to recognize Chinese characters. The first is the use of radicals, components that make up Chinese characters almost like letters make up English words. The second is the use of studying example sentences when learning new words. To test these hypotheses, I performed a pilot experiment using 10 MIT undergraduate students, none of whom were native speakers of Mandarin Chinese. There were not enough Chinese II students to test the second hypotheses, but the results suggest that the use of radicals does improve a student's ability to learn to recognize characters both more correctly and more quickly.

## **Acknowledgements**

My parents and my sisters, and my nephew Tony, for their endless support, care packages, and random phone calls.

Patrick Winston, for allowing me the flexibility to explore research questions far from his own, and for allowing me the opportunity to serve as a teaching assistant in his Artificial Intelligence class, a gift I will always cherish.

My boyfriend Chris, for being my rock and keeping me sane.

My Chinese teachers, Chen Laoshi, Chen Laoshi, Zhang Laoshi, and Yan Laoshi, for their patient instruction and for sharing this fascinating language with me.

My students in 6.034 and 6.00, for challenging me and feeding my addiction to checking my email.

## Contents

1. Introduction.....	7
1.1. Motivating Questions.....	8
1.1.1. Radicals.....	8
1.1.2. Sentence Context.....	9
1.2. Steps.....	9
1.3. News.....	10
2. Mandarin Chinese.....	11
2.1. Pronunciation.....	11
2.2. Pinyin.....	11
2.3. Radicals.....	12
2.4. Traditional and Simplified Characters.....	13
3. Experiment Design.....	15
3.1. Participants.....	17
3.2. Conditions.....	17
3.3. Measures.....	18
3.4. Apparatus.....	18
3.5. Procedure.....	18
4. Results.....	24
4.1. The Effectiveness of Studying Radicals.....	24
4.2. Alternative Explanations of Results for Radicals.....	27
5. Future Work.....	29
6. Related Work.....	30
7. Contributions.....	31
Appendix A: Character Sets.....	32
Radical Set.....	32
Data Set #1.....	32
Data Set #2.....	33
Appendix B: Results.....	35
Works Consulted.....	36

## **List of Figures**

1. Figure 1: The radical lesson, partially finished.
2. Figure 2: The first page of the character lesson, partially finished.
3. Figure 3: The first character quiz.
4. Figure 4: The second character quiz, with an answer selected for this character.
5. Figure 5: The third character quiz, with an answer selected for this character.
6. Figure 6: Average scores for the first two quizzes.
7. Figure 7: Average scores for all three quizzes.
8. Figure 8: The average time spent on each quiz.

## **List of Tables**

1. Table 1: Common transformations between traditional and simplified characters.
2. Table 2: Examples of characters with no consistent transformation between traditional and simplified characters.
3. Table 3: The radical set used in the study.
4. Table 4: The two character sets used in the study. The characters in blue (top 4 rows) are repeated in both sets, and the characters in pink (bottom left) are unrelated to the radicals.

## 1. Introduction

Some classes have monthly exams, others have nanoquizzes. In Chinese languages classes I have taken at MIT and Harvard, the regular monthly exams have been supplemented by weekly or twice-weekly character quizzes called “tīngxiě” (聽寫), which literally translates to “listen-write”, otherwise referred to as dictations.

Prior to these quizzes, we students would have been given a list of characters to memorize, and in class we had to recall and write the characters by pronunciation, usually with a strict time limit. In my early Chinese classes, the set of characters to learn for a tīngxiě was small, often only 15 or 20, and of these we would only have to write ten or twelve characters, mostly in two-word phrases, for the quiz. By third-year Chinese, the dictations were twice a week, and for each of these we needed to have mastered a character set often containing at least 75 characters. During the dictation, we would be asked to transcribe an entire 20- or 30-character sentence in only a matter of minutes, given only three repetitions of the sentence itself.

At first, I managed to temporarily memorize by rote the twelve or so characters that might be necessary for the dictation, and I could remember them for around an hour. This meant that I never really learned many of the characters in early Chinese classes, as I would forget them shortly after the dictation, re-learn them for the exam, and then forget them again. As the semesters passed, it became clear that I needed a different routine and a different set of heuristics than rote memory to maintain the character set needed for these tīngxiě, and for reading Chinese characters in general.

By the time I was taking Chinese III, some of my friends were taking the “streamlined” version of Chinese I, which is designed for people with fluent speaking background but little writing knowledge. Despite their inexperience writing Chinese characters, when asked to

describe how the character for a specific word looked, they would not describe the characters in terms of pieces of other characters or the shape of the strokes, but rather in terms of *radicals*.

I had heard of radicals before, but had never formally been required to learn them. A radical is, at a high level, a piece that can combine with other radicals to make up a Chinese character, similar to how English letters combine to form words (although each Chinese character is, by itself, a full word). Not all characters are strictly combinations of radicals, and some radicals are characters themselves. They are also used to classify characters – a glance at a Chinese dictionary reveals the characters sorted by their main radical, much the way that English dictionaries are sorted by their first letter.

In watching my friends describe the characters to each other by structure, I realized this might be the learning tool that I was missing. This realization, along with the difficulties I had faced with *tīngxiě* consisting of sentences rather than short phrases, led to this thesis.

## **1.1. Motivating Questions**

There are two main questions that I sought to answer with my thesis work, both regarding how to ease the process of learning Mandarin Chinese. The first possibility is to appeal to a person's visual learning center by providing information about radicals, and the second is to engage a person's learning center by providing context as sentences containing the characters to be learned. If the answer to either of these questions is yes, it could help improve the way Chinese is taught in classrooms by changing the emphasis in early Chinese courses.

### **1.1.1. Radicals**

*Does knowledge of radicals improve a student's ability to learn Chinese characters?* If this is true, then both character recognition and character recollection should improve with knowledge

of radicals. In this thesis I only address the issue of recognizing characters, with the assumption that recognition is the first step to being able to write a character from memory.

### **1.1.2. Sentence Context**

*Does knowledge of sentence context improve a student's ability to learn Chinese characters?*

This question addresses the idea that perhaps, if characters are first introduced in a sentence rather than as character-definition pairs, then students might better recognize and recall the characters in the future.

## **1.2. Steps**

To answer these questions, I designed an experiment to measure the benefits of knowledge of radicals and/or sentence context. To this end, I implemented a computer program and designed two character datasets of twelve characters each, and used this software with ten MIT undergraduates to gather data.

In each experiment session, the software provided the participant with the twelve characters chosen from one of two datasets. For each character, its meaning and pronunciation and radicals providing sound and meaning were listed. Once the participant felt comfortable with the amount he or she had studied the characters, there were three quizzes (only two for the control group, who received no prior training on radicals or sentence context) to test the participant's ability to recognize Chinese characters and match them with their meanings and pronunciations.

This experiment resulted in data of the form of twelve <character given, meaning and pronunciation pair chosen to match, true/false> tuples for each quiz, as well as the time the participant took to complete the quiz.

The summary of the datasets and the results are in Appendices A and B, respectively.

### **1.3. News**

For the user study portion of this thesis, I met with ten MIT undergraduates, two of whom were in MIT's Chinese II course. The results from these experiments suggest that students who study radicals before trying to learn the characters composed of those radicals will be able to leverage that knowledge on character quizzes in the future. Specifically, these results suggest that students who study the radicals will, on average, be able to recognize those characters in the future both more correctly and more quickly.

Due to the lack of volunteers from MIT's Chinese II course, I was not able to study the learning improvements as a result of learning using sentences as context.

## 2. Mandarin Chinese

This section describes some different facets of Mandarin Chinese so that these terms can be used later. This includes descriptions of different character systems used in written Chinese, how Chinese characters are pronounced, and how these pronunciations are communicated in writing.

### 2.1. Pronunciation

Mandarin Chinese words are spoken with one of five tones:

1. First tone is pronounced with a flat, high pitch. In pinyin, first tone is indicated by a  $\bar{a}$  above the vowel: mā.
2. Second tone is pronounced with a rising pitch. In pinyin, it is depicted as a  $\acute{a}$  above the vowel: má.
3. Third tone is pronounced with a low pitch, which falls and then rises a little. In pinyin, this tone uses a  $\check{a}$  above the vowel: mǎ.
4. Fourth tone is pronounced with a falling pitch. In pinyin, a  $\grave{a}$  is placed over the vowel for fourth tone: mà.
5. Neutral tone is short, and the pitch depends on the previous syllable. In pinyin, the neutral tone is indicated by having no symbol above the vowel, like the second ma in the colloquial word for “mom”: māma.

Another way to represent the tones is to use numbers after each syllable instead of a symbol over the vowel. For example, the phrase “nǐhǎo” means hello. Using numbers instead of tone markings, this would be written as “ni3 hao3”.

### 2.2. Pinyin

There are multiple Romanization systems to represent Chinese words. Until the late twentieth century, a system called Wade-Giles was used in Mainland China. Despite the switch to Hanyu

Pinyin, some places and people still use the Wade-Giles system. For this reason, Beijing (Běijíng in pinyin) is sometimes spelled Peking, and Mao Zedong's name is sometimes still spelled Mao Tse-tung.

Pinyin uses four markings above the vowel (e.g. ā, á, ǎ, and à), or numbers after the word (e.g. a1, a2, a3, a4), to represent the non-neutral tones in Mandarin Chinese. The location of the tone marking follows a simple priority system: a, o, e, i, u/ü. Pinyin is currently used to teach Chinese, and also serves as an input method to enter Chinese characters using a regular QWERTY keyboard.

When providing the pronunciation to a user, my software application will provide those pronunciations via pinyin, using numbers to represent the tones.

### **2.3. Radicals**

Chinese is a logographic language, meaning that each character represents an entire word. These characters can also be combined to make other words. Therefore, unlike English, in which there are 26 letters to learn in order to be able to read the language, there are many thousands of characters that one has to be able to recognize in order to be able to read written Chinese.

Within most characters are components which give clues to the sound and meaning of the word. These components are called radicals. For example, consider the following traditional Chinese characters:

- 情 (qíng): feeling, emotion
- 請 (qǐng): to ask, to invite
- 清 (qīng): clear, quiet

Their similar pronunciation comes from the radical 青, which is pronounced qīng or jīng.

However, the characters have very different meanings. For these characters, the different

meanings are conveyed by the radicals on the left-hand sides of the characters. The radicals in these three characters are:

- 忄 : heart
- 言 : speech
- 氵 : water

Thus, knowing the radical for the pronunciation hints that the pronunciation of each of these three characters will be some form of “qing” (no information is given about the tone, however), and knowing these three radicals allows one to easily differentiate between the three characters based on meaning:

- heart – feeling/emotion
- speech – to ask / to invite
- water – clear/quiet.

## **2.4. Traditional and Simplified Characters**

Written Chinese has two forms. The first, called “traditional”, is the traditional way of writing characters, and dates back to the earliest known Chinese writing. Traditional characters are still commonly used in places like Taiwan and Singapore. In mainland China, the use of traditional characters was discontinued in the mid-twentieth century in favor of “simplified” characters.

The difference between the traditional and simplified forms of a word is often small.

Table 1 contains some common transformations and example traditional and simplified characters for those transformations.

Transformation	Traditional Character	Simplified Character	Pronunciation	Meaning
言→讠	課	课	kè	class
門→门	門	门	mén	door
車→车	東	东	dōng	east
馬→马	媽	妈	mā	mother
幾→几	機	机	jī	machine

Table 1: Common transformations between traditional and simplified characters.

Despite many common transformations, there are still some words for which the traditional and simplified characters are vastly different. A few of these are listed in Table 2.

Traditional Character	Simplified Character	Pronunciation	Meaning
這	这	zhè	this
還	还	hái	still
龍	龙	lóng	dragon

Table 2: Examples of characters with no consistent transformation between traditional and simplified characters.

Traditional characters also have a tendency to have evolved from pictorial representations of physical objects, and this imagery is lost in the simplified forms of some of these characters. For example, 龍, meaning dragon, could be seen to resemble a dragon with its feet and tail at the bottom right. However, the simplified form, 龙, has lost this resemblance. For this reason in particular, and in general because of the sometimes non-standard differences between traditional and simplified characters, I have used traditional characters in my work.

### 3. Experiment Design

The project began with designing the software used to teach and then quiz the participants during the experiment. This design process included designing the user interface, the data structures to represent characters and radicals and their pronunciation, definitions, and usages, and the data collection and reporting methods. The user interface design process was iterative, progressing from sketches, to a computer prototype, and then to the final implementation, with revisions made along the way to improve the testing experience.

The experiment was designed to replicate the dictation quizzes from Chinese courses. Therefore, students were presented with characters and given time to study them, and then quizzes on the characters they had just learned. This particularly resembled my experience with dictations because I so often just set aside time right before class to memorize all of the characters I thought were necessary, and took the dictation quiz only minutes after I finished memorizing.

While designing and implementing the software, I also determined the characters used. Because the nature of the experiment required that participants were not already familiar with the characters being tested, the characters used were selected in part from a later Chinese curriculum and in part from characters that had similar meanings and structures but were not used in first-year Chinese classes at MIT.

The experiment used two different sets of 12 characters each. The two sets drew from the same radicals and were thus similar in structure, but they had a few small yet relevant differences. In the first set, eight out of twelve of the characters were comprised mostly of the four radicals provided in the radical lesson (which was not given to the control group), and the other four characters were chosen at random from a third-year Chinese textbook used at Harvard.

These four characters did not have any visible relation to the others. In the second character set, all twelve characters were chosen for their similar set of radicals. Some of the characters in the first set are also in the second.

The four radicals used in the experiment are presented in Table 3, and are also presented in Appendix A with their meanings and pronunciations.

Radicals Used			
艮	手	心	各

Table 3: The radical set used in the study.

The two datasets are presented in their entirety in Appendix A, and are also in Table 4 to demonstrate the main differences between them. The first four characters listed for each set were the same. The last four characters from the first data set did not contain the radicals used for the experiment.

Character Set #1	Character Set #2
根	根
恪	恪
格	格
格	格
跟	抄
恨	炒
校	校
抬	砾
部	砾
兜	砂
胃	砖
光	抓

Table 4: The two character sets used in the study. The characters in blue (top 4 rows) are repeated in both sets, and the characters in pink (bottom left) are unrelated to the radicals.

### **3.1. Participants**

The participants in this study were all MIT undergraduate students. To increase the likelihood that participants had not learned any of the characters before, participants were non-native speakers who either had never studied Chinese or were in at most their second semester of study, MIT's Chinese II course (21F.102).

Ten MIT undergraduates participated in this study. Of that number, eight had never studied Chinese before, and the other two had just finished 21F.102.

### **3.2. Conditions**

The two independent variables this experiment was designed to test were the use of sentence context and the use of radicals. When radicals were used, the initial character lesson was preceded by a radical lesson, where each relevant radical was presented with its meaning and pronunciation. When sentence context was used, the initial presentation of the characters included an example sentence for each.

As detailed in the Procedure section, if either of these conditions was being used, the second character quiz was followed by a third in which the radical and/or example sentence for the given character was provided as an aid. The group of participants who did not receive a lesson with the radicals and did not see the example sentences was the control group.

Because only two of the volunteers in this pilot experiment had experience with Chinese, there would not have been enough data about sentence context, so the actual experiment as run did not use sentence context, and instead focused on radicals.

### **3.3. Measures**

The main dependent variable being measured in this study was the number of correct answers in the first, second, and (optionally) third character recognition quizzes. For each character, both the correct and given answers were recorded.

Beginning with the fourth user, the time taken for each quiz was also measured, and was used to determine whether the independent variables helped the student complete the quizzes more efficiently.

### **3.4. Apparatus**

The experiment was run using software written in Java, and used Java Swing to implement the interface. The experiment was conducted on a 1.50 GHz AMD laptop running 64-bit Windows 7.

### **3.5. Procedure**

The procedure for each user test was a character lesson, followed by a series of character recognition quizzes, in which the user was asked to match each character to its pronunciation and meaning. For most participants, the experiment took about twenty minutes, but that time was variable based on how much time the participant spent studying the characters before starting the quizzes. I chose not to limit the study time to more accurately represent real-life study situations in which one would have days to memorize characters, even though this might give those participants an advantage over those who chose not to spend as much time studying.

Before starting the experiment, each participant was given an overview of the experimental procedure (in the form of a waiver they were asked to sign). Each participant was also assigned to a character set (1 or 2), and using radicals or not. Because there were not enough participants with knowledge of Chinese, no participants were given example sentences.

The specific procedure of the experiment was as follows:

1. **Radical lesson:** If radicals were being used, the participant was presented with a short lesson on four radicals relevant to the character data set.

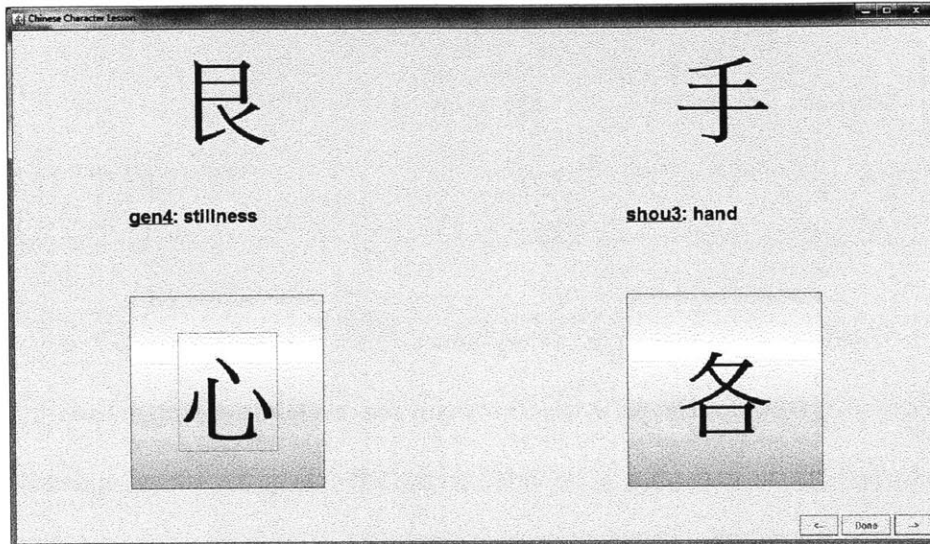


Figure 1: The radical lesson, partially finished.

2. **Character Lesson:** The participant was presented with the characters from the chosen dataset. Each character was displayed with its pronunciation in pinyin, using a number instead of a tone mark to indicate the tone, and its definition in English.
  - a. For most characters, the radicals providing meaning and/or sound were also given. This included their Chinese character, pronunciation in pinyin, and definition in English.
  - b. If sentence context was used, an example sentence was also shown with each character.

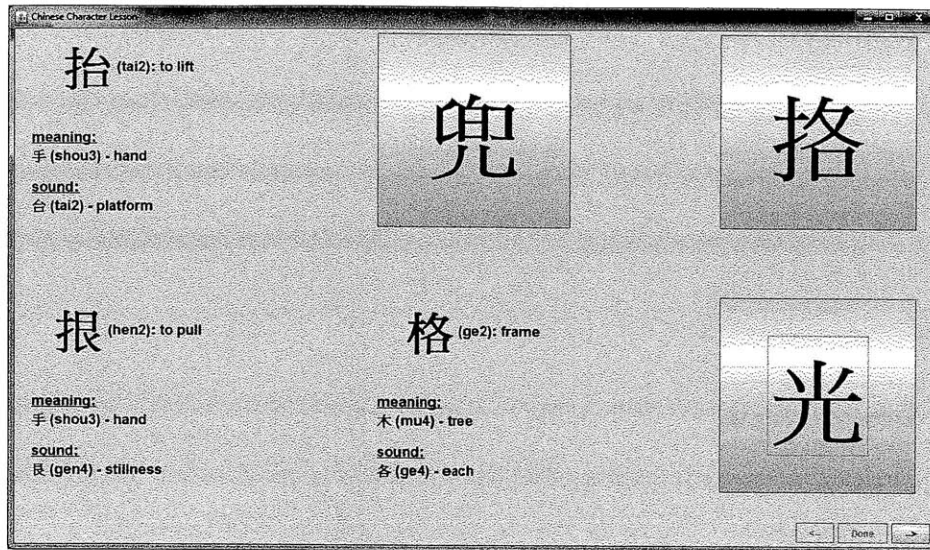


Figure 2: The first page of the character lesson, partially finished.

3. Character Quiz 1: The participant was quizzed on character recognition. Each character in the data set was provided on the left side of the screen, and each pronunciation and meaning pair was provided on the right side. Both lists were given in a random order, and the participant could see all characters and pronunciation/meaning pairs at once. In this quiz, the participant's task was to put the letter corresponding to each pronunciation and meaning into the text box next to the correct character. As a result, the user was not allowed to continue until each text box contained a unique, valid (i.e. corresponding to one of the available choices) letter for each character.

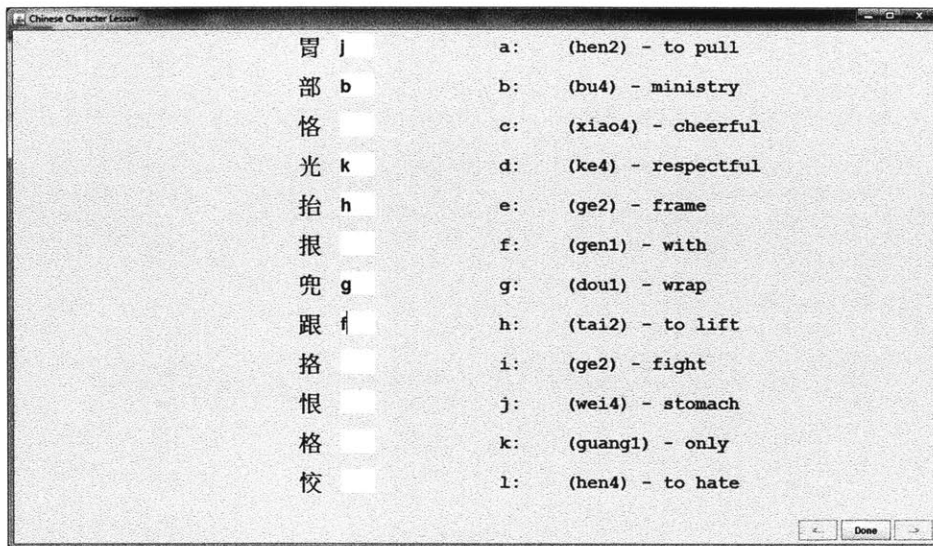


Figure 3: The first character quiz.

4. Chinese Background Questions: The user was asked if he or she had ever studied Mandarin Chinese before, and his or her motivations for doing so or reasons for not having studied it.
  - a. This data helped keep track of the students who were familiar with Chinese.
  - b. This pause to answer a question also served to distract the participant from any heuristics for memorizing characters or any other short-term memory tactics he or she was employing to retain the character pronunciations and definitions.
  
5. Character Quiz 2: The participant was again presented with each character, and asked to match it to the correct meaning/pronunciation pair. In this case, however, only one character was presented at a time, removing the ability of the participant to use the process of elimination to determine the correct meaning and pronunciation for the character. This also allowed for the participant to guess the same meaning and pronunciation for different characters, an option that had not been available in the first character quiz. As before, the characters were presented in a new, random order.

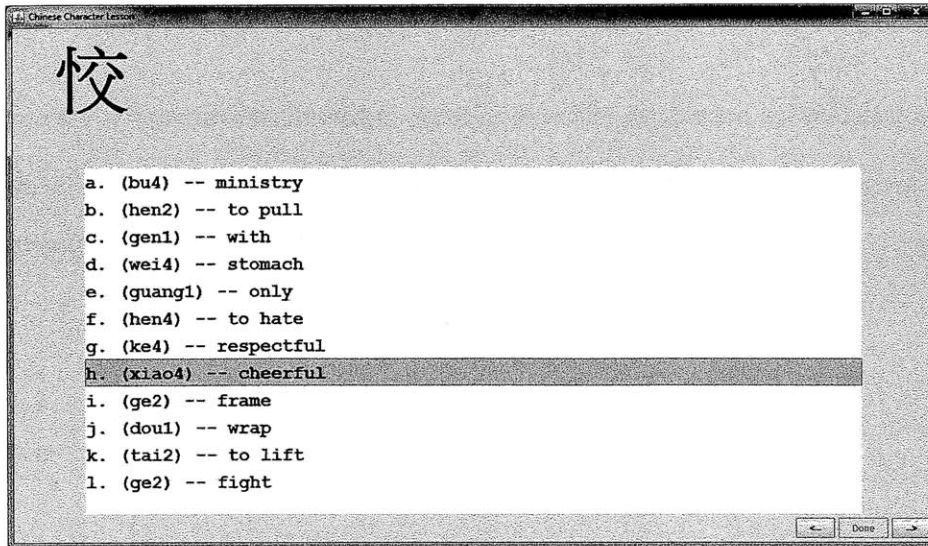


Figure 4: The second character quiz, with an answer selected for this character.

6. Character Quiz 3: If either or both of the independent variables of sentence context and radical lessons were selected for the given participant, the previous quiz (Step 5) was repeated, except that each character to be matched was presented with extra information as aids.

- a. If the experiment session started with a radical lesson, the radicals providing the meaning and/or sound were provided.
- b. If the experiment session was using sentence context, the example sentence from the character lesson (Step 2) was provided.

As in the previous two quizzes, the characters were presented in a new, random order.

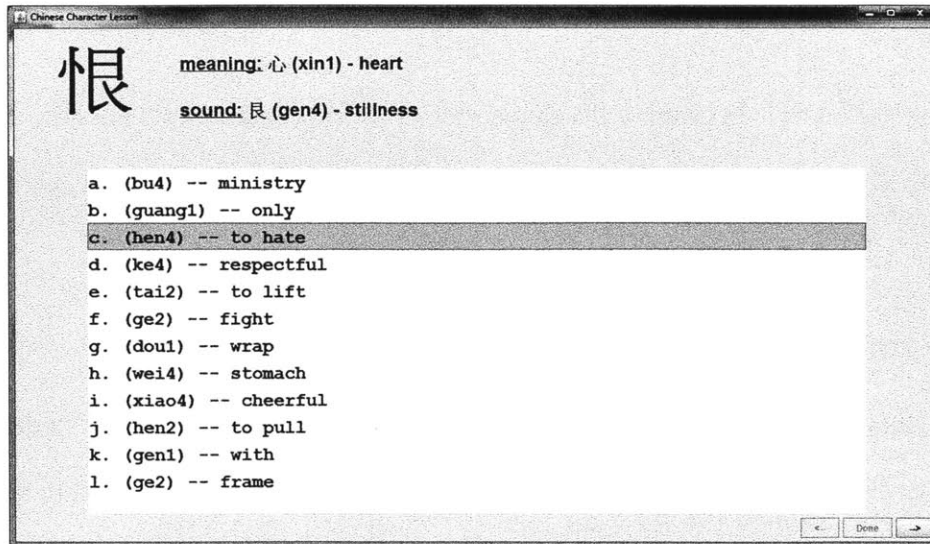


Figure 5: The third character quiz, with an answer selected for this character.

Once finished with the experiment session, each participant was debriefed. In the debriefing, participants were told in more detail the goals of the experiment, and the deception of the motivation question in Step 4 to disrupt their short-term memorization of the characters. Some participants also took this time to provide feedback on the experiment.

## 4. Results

Because only two students in Chinese II volunteered for the experiment, I did not have enough participants to have any statistical significance with sentence context, and focused instead on radicals.

### 4.1. The Effectiveness of Studying Radicals

Overall, providing a radical lesson at the beginning did not dramatically affect the students' abilities to learn to recognize the Chinese characters, as demonstrated in Figure 6, but it did have some impact.

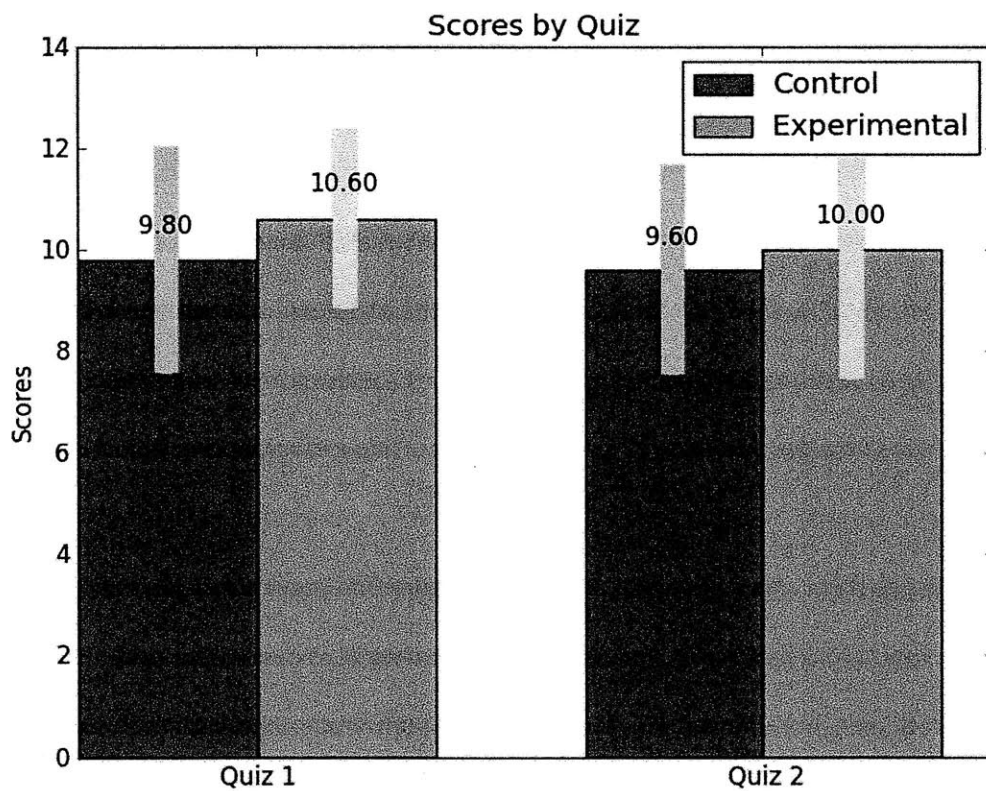


Figure 6: Average scores for the first two quizzes.

The large amount of overlap in the error bars (depicted as vertical bars overlapping the top of each bar in the graph) indicates that these results are not statistically significant. However, this

graph suggests that on average the students who had the radical lesson answered more questions correctly, with on average 0.8 more correct matches out of twelve in the first quiz, and 0.4 more correct answers in quiz 2, where they had to pair each character with its meaning and pronunciation on the same screen. For the second quiz, the difference was less pronounced, suggesting that perhaps the participants learned from the first quiz, and were able to apply that knowledge to the second quiz.

The drop in average scores between quiz 1 and quiz 2 potentially corresponds with the background questions asked, as mentioned in Step 4 of the Procedure section. This pause was designed to force the participants to think in English, possibly degrading some of the rote memorization and heuristics that they had developed during the character lesson and used during the first quiz. Given the drop in both groups, it seems likely that this degradation occurred.

The improvement of the experimental group between quizzes 2 and 3, shown in Figure 7, helps quantify those participants' abilities to extract the useful information from the radicals presented with each character.

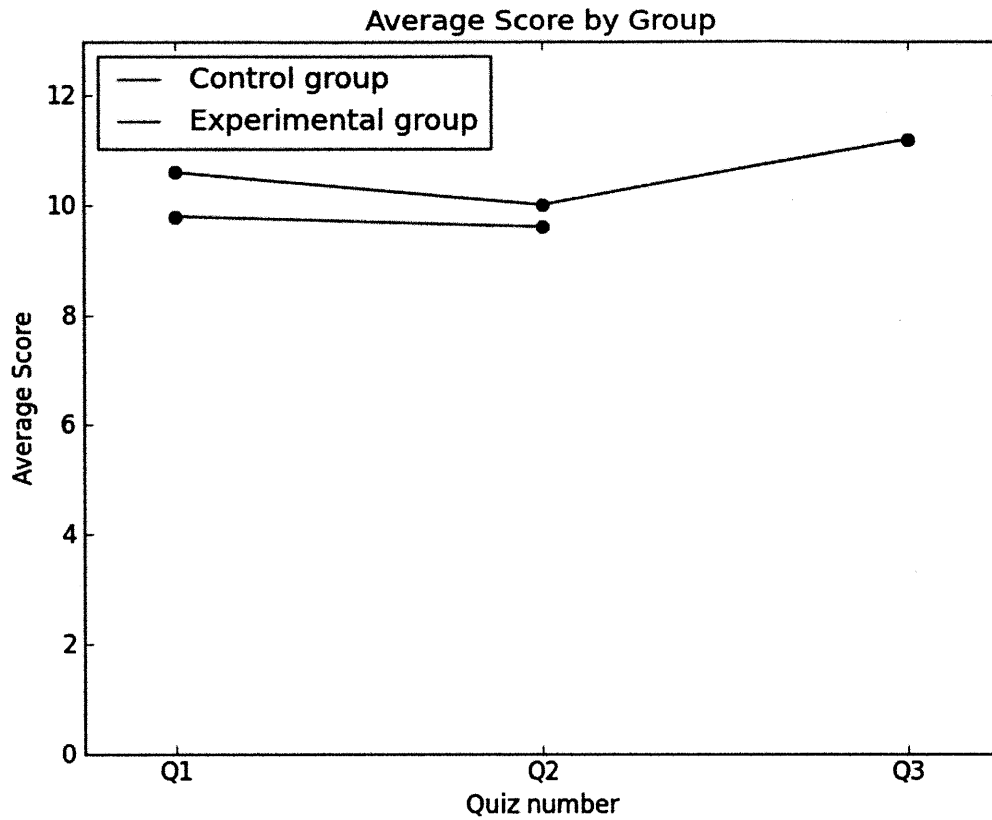


Figure 7: Average scores for all three quizzes.

In this graph, one can see that the average score for the experimental group jumped from 10 out of 12 correct in quiz 2 to 11.2 correct answers in quiz 3. This indicates that given the radical information for a character, participants were able to effectively utilize that information to correctly match the character with its meaning and pronunciation in most cases.

The time measurements also provide some insight into the effectiveness of studying radicals. Figure 8 displays the time measurements for all three quizzes, separated by group.

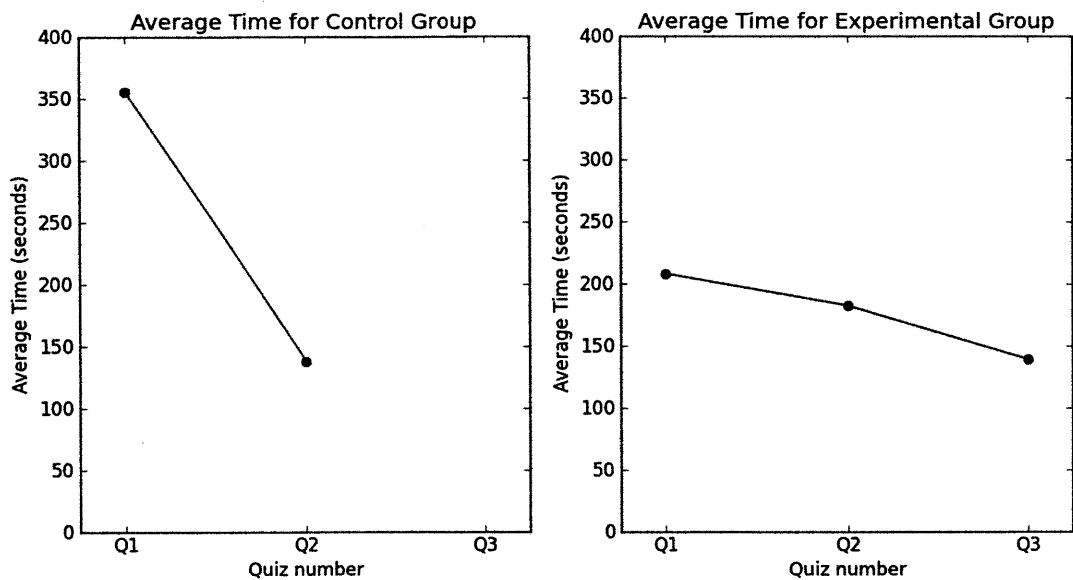


Figure 8: The average time spent on each quiz.

As this plot illustrates, students who had studied radicals before studying the characters completed quiz 1 about two and a half minutes faster than students in the control group, suggesting that the students in the experimental group had made for themselves more effective heuristics to remember each character. The time differences between each quiz are likely a result of learning effects; because each quiz was comprised of the same character set, participants would have been able to re-use any heuristics they had learned to match the characters in the previous quiz(es).

Overall, the data suggests that learning radicals before learning the characters comprised of those radicals improves one's ability to recognize characters more effectively and more quickly.

#### 4.2. Alternative Explanations of Results for Radicals

While the results mentioned in the previous section are promising, they must be taken with a grain of salt. The following are some possible explanations of the results other than the hypothesized effectiveness of studying radicals.

- The participants were all tested near the end of the semester, when some were likely sleeping less than normal. Therefore, some students might not have performed to their fullest potential, while others might have been done with assignments and well-rested.
- All participants were given the radical information, but only some were given the actual radical lesson. Because of this, any particularly perceptive participants in the control group could have studied the radicals just as much as those in the experimental group, and likely scored just as well.
- There was no time limit on the quizzes. This does not perfectly map to the classroom dictation experience, where students are given a time limit. It is possible that the control group did as well as they did because they took more time to complete the quiz than the experimental group, and that if both groups were given a time limit, the control group might have done worse.
- During the debriefing after the experiment session, some of the participants explained that their rationale for remembering characters had nothing to do with the characters' pronunciation, and everything to do with the meaning of the characters and their radicals.

## 5. Future Work

There are many possible extensions to this work, first and foremost being a broader user study. Given many more student participants, data for sentence context could be collected, as well as for radicals.

As mentioned in the Results section, time limits could be added to the quizzes. This would more accurately reflect the classroom experience during a dictation, and would perhaps implicitly normalize the scores for both the control and experimental groups.

The radical information could be removed completely from the character lesson. This would force the students to notice any patterns between characters entirely on their own. This models the traditional textbook style of giving a character with just its meaning and pronunciation. However, it does not necessarily model students' learning styles for these dictations, where any possible resource is valid during studying, as it is done outside of class.

A potentially more interesting change would be to train the participants on one character set, and then give everyone a third quiz using a different character set. The expected result is that participants in the control group would have memorized characters as units, and would thus be less able to extend their memorization heuristics to a new character set, while those in the experimental group would ideally have memorized the structure of characters, and thus be able to apply that knowledge to any new characters with the same radicals.

In addition to testing a user's ability to recognize characters, it would be interesting to test a user's ability to recall characters and draw them on the screen. Handwriting recognition software could be incorporated to make this scalable. This would provide a much more realistic parallel to the classroom dictation experience, but it would limit the possible participants to those with some Chinese background.

## 6. Related Work

There has already been a great deal of work done to improve the ease with which someone can learn a new language. In terms of software, both Rosetta Stone and Skritter have been particularly influential in how I learn Chinese. Rosetta Stone is a popular tool for learning any of a variety of different languages, and Skritter is a web-based flashcard program that helps people learn Chinese and Japanese.

In Rosetta Stone, a user is presented with a series of images, often four to a page. Each image is accompanied with a phrase or sentence, which is read to the user, and a button to repeat the phrase. In Chinese, there are options to display the text in pinyin, in simplified Chinese, or with both. There are also reading lessons which teach specific characters, usually two at a time out of a lesson of eight or ten, and speaking lessons which force the user to attempt to speak the language. There is no information given about radicals. Furthermore, nowhere in Rosetta stone does it use English; instead, the user must infer the meanings of the words (including vague concepts, like “this” or “the”).

Skritter uses an entirely different approach. Unlike Rosetta Stone, which teaches users to speak and read a foreign language, Skritter focuses on the writing aspect. It provides a user with the pinyin and the definition for the character which he or she is to write, as well as an example sentence with a blank for the character the user is to enter, and a list of character mnemonics others have written. A user can also click a button to hear the pronunciation of the specific phrase being taught. As time progresses, it keeps track of the characters users have trouble with.

## **7. Contributions**

The work described in this thesis helps provide a better understanding of what helps people learn to read Chinese. To this end, I have:

- Framed the questions of whether learning radicals and learning characters using example sentences helps students build more lasting memory pathways to recognize those characters in the future.
- Demonstrated the effectiveness of learning radicals before learning characters composed of those radicals through an experiment using ten MIT undergraduates.
- Developed software to model language learning in a college classroom setting. The software can read in a character set and display information about those characters to a user, and automatically generate matching quizzes to test the user's ability to recognize the characters afterwards.

It is my hope that the results described in this thesis can help improve the effectiveness of early Chinese curricula, both at MIT and around the world.

## Appendix A: Character Sets

During the experiment, the tones were indicated in pinyin by numbers following the words rather than markings above the vowels. Also, only the second character set was given sentences, as some of the characters in the first character set were part of the early Chinese curriculum.

### Radical Set

Radical	Pronunciation	Meaning	Purpose for character set
艮	gen4	stillness	sound
各	ge4	each	sound
手	shou3	hand	meaning
心	xin1	heart	meaning

### Data Set #1

Character	Pronunciation	Meaning	Meaning Radical	Sound Radical
扞	hen2	to pull	手 (shou3): hand	艮 (gen4): stillness
恪	ke4	respectful	心 (xin1): heart	各 (ge4): each
拞	ge2	to fight	心 (xin1): heart	各 (ge4): each
格	ge2	frame	木 (mu4): tree	各 (ge4): each
跟	gen1	with	足 (zu2): foot	艮 (gen4): stillness
恨	hen4	to hate	心 (xin1): heart	艮 (gen4): stillness
佼	xiao4	cheerful	心 (xin1): heart	交 (jiao1): to hand over
抬	tai2	to lift	手 (shou3): hand	台 (tai2): platform
部	bu4	ministry		
兜	dou1	wrap		
胃	wei4	stomach		
光	guang1	only		

Data Set #2

Character	Pronunciation	Meaning	Meaning Radical	Sound Radical
扌	hen2	to pull	手 (shou3): hand	艮 (gen4): stillness
我扌我的朋友。 / 我扌我的朋友。 (I pull my friend.)				
恪	ke4	respectful	心 (xin1): heart	各 (ge4): each
他很恪。 / 他很恪。 (He is very respectful.)				
格	ge2	to fight	心 (xin1): heart	各 (ge4): each
我跟他格。 / 我跟他格。 (I fought with him.)				
格	ge2	square	木 (mu4): tree	各 (ge4): each
這個格很大。 / 这个格很大。 (This square is big.)				
抄	chao1	to make a copy	手 (shou3): hand	少 (shao3): few
我抄我的朋友作業。 / 我抄我的朋友作业。 (I copied my friend's homework.)				
炒	chao3	to stir-fry	火 (huo3): fire	少 (shao3): few
今天我吃炒飯。 / 今天我吃炒饭。 (Today I ate fried rice.)				
校	xiao4	school	木 (mu4): tree	交 (jiao1): to hand over
她在學校。 / 她在学校。 (She is at school.)				
砾	le4	rocky	石 (shi2): rock	
這個地方很砾。 / 这个地方很砾。 (This place is very rocky.)				
砾	ge4	to press painfully	石 (shi2): rock	各 (ge4): each
他在砾我的手。 / 他在砾我的手。 (He is pressing my hand very painfully.)				
砂	sha1	sand	石 (shi2): rock	少 (shao3): few

這個砂很熱! / 这个砂很热!  
(This sand is hot!)

磚

zhuan1

brick

石 (shi2): rock

專 (zhuan1): to  
monopolize

我的宿舍有很多磚。 / 我的宿舍有很多砖。  
(My dorm has a lot of bricks.)

抓

zhua1

to grab

手 (shou3): hand

爪 (zhua3): claw

我抓他的手。 / 我抓他的手。  
(I grabbed his hand.)

## Appendix B: Results

Timing code was added after the third user test. Because only two students in Chinese II volunteered for the experiment, I did not have enough users to have any statistical significance with sentence context, and focused instead on radicals.

The data is sorted by which users were given the radical lesson, and given user numbers to reflect this sorting. Users 1-5 did not receive the radical lesson, and users 6-10 did. Users 5 and 10 were the Chinese II students.

The times are given in seconds.

User #	Quiz 1 Score	Quiz 1 Time	Quiz 2 Score	Quiz 2 Time	Quiz 3 Score	Quiz 3 Time
1	10	--	9	--	--	--
2	9	--	8	--	--	--
3	6	511.1	7	174.8	--	--
4	12	131.2	12	74.34	--	--
5	12	423.9	12	165.2	--	--
6	9	--	8	--	9	--
7	12	267.4	12	221.2	12	130.1
8	8	368.2	6	243.6	11	198.1
9	12	117.6	12	92.7	12	88.8
10	12	79.4	12	172.3	12	140.8

## Works Consulted

- Everson, Michael E. "Word Recognition among Learners of Chinese as a Foreign Language: Investigating the Relationship between Naming and Knowing." The Modern Language Journal 82.2 (1998): 194-204.
- Garrett, Nina. "Computer-Assisted Language Learning Trends and Issues Revisited: Integrating Innovation." The Modern Language Journal 93 (2009): 719-740.
- Hayes, Edmund B. "Encoding Strategies Used by Native and Non-Native Readers of Chinese Mandarin." The Modern Language Journal 72.2 (1988): 188-195.
- Ke, Chuanren. "An Empirical Study on the Relationship between Chinese Character Recognition and Production." The Modern Language Journal 80.3 (1996): 340-349.
- Shen, Helen H and Chuanren Ke. "Radical Awareness and Word Acquisition among Nonnative Learners of Chinese." The Modern Language Journal 91.1 (2007): 97-111.
- Shu, Hua and Richard C. Anderson. "Role of Radical Awareness in the Character and Word Acquisition of Chinese Children." Research/Technical. Center for the Study of Reading, Urbana, IL, 1995.