Designing intelligent language tutoring system for learning Chinese characters

by

Jingyu Liu

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Industrial and Manufacturing Systems Engineering

Program of Study Committee:
Stephen Gilbert, Major Professor
Michael Dorneich
Volker Hegelheimer

Iowa State University

Ames, Iowa

2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am very thankful to my advisor, Dr. Stephen Gilbert, for his great and consistent instruction, understanding and patience throughout my graduate study. Also, I would like to thank my committee members, Michael Dorneich and Volker Hegelheimer, for their guidance and support throughout the course of this research.

In addition, I would also like to thank my friends, colleagues, the department faculty and staff for making my time at Iowa State University a wonderful experience. I want to also offer my appreciation to those who were willing to participate in my surveys and tests, without them, this thesis would not have been possible.

Finally, I would like to thank my parents, boyfriend and all the people for their consistent encouragement, help, and love to my life.

ABSTRACT

The purposes of this research are to explore 1) the design and usability of the interface for an intelligent tutoring system for recognition of Chinese characters, 2) the pedagogical effectiveness of different forms of information presentation and feedback. A prototype system (an iPad Chinese character tutor) was developed and was evaluated for its effectiveness and usability. In the evaluation test, two groups were given 34 Chinese characters and phrases to learn using two different versions of the system. Version A contained a metaphor-based pedagogy, feedback, and extra instructions; Version B did not. Participants' learning performance and survey results were used to measure the effectiveness and usability of the system. Learning performance of the group who used Version A was statistically significantly better than that of the Version B group. Participants surveyed rated Version A significantly higher than Version B on several constructs, including usability, satisfaction, functionality, and usefulness. This study lays the foundation for the development of an Intelligent Tutoring System (ITS) for Chinese learning.

CHAPTER 1 INTRODUCTION

1.1 Statement of the Problem

Since we entered the 21st century, Chinese, a language used by 20% of the global population in daily communication, has recently become more and more popular in the world (Shih, Chen, & Li, 2013). An increasing number of learners begun to choose to learn Chinese as a second language around the world (Chen et al., 2013). However, regarded as one of the most difficult languages (Huang & Ma, 2007), Chinese poses challenges for beginners, especially for learners whose native language is alphabetic-based. In the alphabetic writing system, phrases have relationships with their pronunciations, which is known as grapheme-phoneme correspondence (Shen, 2005). Chinese orthography, on the other hand, does not provide this connection. One cannot know a Chinese character's pronunciation by observing its representation. Lack of correspondence between grapheme and phoneme is one of the major obstacles for learning Chinese as a second language. Teaching and learning Chinese is still not easy and effective, although a lot of effort has been made teaching Chinese in the classroom and the overwhelming variety of learning instructions were rare until recently (Xing, 2006).

Technology is transforming education and traditional classroom instruction (Buckingham, 2007). Taking the advantages of technological instruction and the imitation of human tutor, Intelligent Tutoring Systems (ITSs) (Murray, 1999) have been successful at instructing students in various domains such as mathematics (Beal, Cohen, & Woolf, 2010) and physics (Graesser, Chipman, Haynes, & Olney, 2005; Graesser, VanLehn, Rosé, Jordan, & Harter, 2001; Hagge et al., 2015; VanLehn, 2011). A variety of studies and research have been complimented because of its capabilities for personalized feedback, assessment of students, self-learning etc. (Ahuja & Sille, 2013). However, ITS research in the language domain have rarely been touched (Robert Sottilare, Graesser, Hu, 2014) especially Chinese language. The rare number of studies and rapid ITS development demand provides a fertile field for a study of bridging ITS and Chinese language learning. Based on this current situation, the following session provides the research questions that will be addressed in this study.

1.2 Research Questions

The purposes of this study are 1) to explore the initial feasibility of an Intelligent Language Tutoring interface for learning Chinese characters to teach beginning Chinese-as-a-second-language learners and 2) to assess the system by investigating learning effectiveness, usability issues and users' attitudes towards the system.

More specially, this research attempts to answer the following questions:

1. Is there a difference between the learning of students who use the system with metaphor pedagogy and the students using the system without it?
2. Will beginning Chinese-as-a-second-language learners' interest be increased after using the system?
3. Will the interface be user-friendly and are there any usability issues?
4. Will users using the system with the metaphor pedagogy achieve better performances (higher scores in the quiz), more interest increased, better usability assessment than users using the system without metaphor pedagogy?

1.3 Brief Introduction to Intelligent Tutoring Systems

To further address the problem, it's worth exploring a brief introduction of ITSs, the components that they use, and especially how they have been used for language. In general terms, intelligent tutoring systems (ITSs) are computer-based instructional systems that are evaluate learners' responses and provide the personalized feedback to learners by imitating the performance of human tutors (Murray, 1999). The problem domains widely range from mathematics (Beal, Cohen, & Woolf, 2010), physics (Hagge et al., 2015; Vanlehn, Lynch, & Schulze, 2005), health care (Muñoz, Ortiz, González, López, & Blobel, 2010), or even game play (Baker et al., 2006).

According to Nkambou, Bourdeau, & Mizoguchi (2010), an ITS needs to have four basic components for teaching purposes.

1. The interface for communicating with learners (interface module)
2. The tutoring strategies (tutor module)
3. A representation of domain knowledge (expert module)
4. A way to represent student's knowledge (student module)

The graphical interface is very important for a language tutoring system because all the instructions communicated is through the interface. In order to guide students to learn properly within specific language settings, the interface or learning environment in a language ITS should use different media (graphics, animation, text, sound, video) to display language in meaningful, communicative situations (Swartz & Yazdani, 1992). This indicates the use of some multiple windows and multimedia design.

The tutor module represents tutor strategies to deliver instruction in the system. It teaches students by guiding them to solve problems within the system. The types of tutoring approaches selected should be based on the unique nature of foreign languages. The approaches can also differ depending on the skill to be learned and instructional purpose of the ITSs. For example, different strategies could be selected in order to emphasize different skill acquisition whether is vocabulary, pronunciation, grammar, writing, listening, all of them or some of these. Different skill level should be considered when selecting tutoring strategies, whether the users would be beginning, intermediate or advanced as well as instructional purposes if learning language for communicative purposes or other more professional uses.

The expert module contains the domain knowledge for the system. In foreign domains, certain type of approaches should be selected for representing the domain knowledge, given the understanding of the unique nature of linguistic knowledge acquisition (Swartz & Yazdani, 1992).

The student module evaluates the knowledge of students, and allows tutor module adaption feedback or instruction. In order to properly model the student, the most basic requirement of the system is to know about learner errors (Swartz & Yazdani, 1992). Nevertheless, whether to construct a “deep” student model is situation dependent. Sometimes a “deep” model may not be a priority in certain cases such as for beginning learning where the idea expressed is more important than actual grammatical structure.

The current research addresses Component 1 (the user interface) most directly, and touches on Component 2 (the tutor module) as well. More details about these components will be discussed in Chapter 3.

1.4 Study Overview

This study was designed to take a step in the direction of bridging the gap between the development of ITS for Chinese learning and the use of user interface design to improve learning. This current work focuses on the first steps that are necessary to construct an ITS: designing the interface and the feedback (Components 1 and 2 above). Specifically, this research explores 1) the design and usability of the interface for an intelligent tutoring system for recognition of Chinese characters, 2) the pedagogical effectiveness of different forms of information presentation and feedback. While creating a complete ITS for teaching Chinese characters was beyond the scope of the current project, this initial research lays the foundation upon which a complete ITS can be built.

In terms of the design of this ITS, the scope of the study and the unique settings of the language should first be considered. Although Chinese is regarded as difficult to learn, unlike the difficulty facing second language learners, native Chinese speakers learn speaking and writing in the different period of time, which distributes the difficulty of Chinese learning. Native Chinese speakers have been exposed to the language since birth. They do not learn Chinese characters until school begins. On the other hand, for those who speak Chinese-as-second-language, it would be very difficult when processing acoustic and visual information simultaneously. Additionally, the writing and speaking systems of Chinese are comparable separately (Chen et al., 2013). The proposed study focuses on teaching Chinese characters to beginners. The inference of teaching Chinese characters to beginners will be discussed in Chapter 2. Secondly, regarding visual complexity and a large number of characters of Chinese characters, previous research suggests that pedagogy that features the ideographic characteristic of Chinese characters integrated with metaphor and instruction will be more effective (Hsu, 2012). This pedagogy will be referred to as the "metaphor pedagogy" and will also be discussed in more detail below.

The formative evaluations will be conducted throughout the design process, and the usability testing will be conducted in the end to test the effectiveness of the current study. For the experimental design, the comparison will be made between two versions of the system, one with the metaphor pedagogy and one without. The quiz scores will be the measurement to compare whether the pedagogy would be useful and feedback would enhance learning. Also, usability surveys will be used to evaluate whether having the aforementioned features will lead to better user experience, and identify usability issues for future work.

1.5 Thesis organization

This current chapter discusses the current problems learning Chinese and the integration of Chinese learning within ITS. It also introduces the Intelligent Tutoring Systems and their components in the foreign language domain, and proposes the study and the questions that will be addressed. The rest of the chapter outlines the contribution to the current study on designing and evaluating Intelligent Tutoring System for Chinese characters. Chapter 2 discusses the previous work in multi-disciplinary work on the development of ITS, especially in language domain, and identifies the research gaps. Chapter 3 illustrates the methodologies used to design the system and explores the research questions. Chapter 4 presents data collected from the study. Finally, the discussion of the results in Chapter 4 and the areas of future work are presented in Chapter 5.

CHAPTER 2 LITERATURE REVIEW

In order to guide the design and address the problem, a comprehensive review of previous related work is required. This study is to take a step in the direction of bridging the gap between the developments of Intelligent Tutoring Systems for Chinese learning. In addition to a brief introduction to ITSs described previously, this section will briefly focus on language ITSs and their development. The nature of Chinese will then be introduced as well as Chinese instruction and pedagogical methods, which illustrate the challenges of learning and guide the design described in the next Chapter. Finally, the gaps will be identified based on the previous studies that the current study may address.

2.1 The Intelligent Language Tutoring System and Its Development

Intelligent Language Tutoring System (ILTS)

With the development of the computer and technology, the use of computers has been more and more adopted in language learning. The use of new media and information technologies for language learning, known as Computer-Assisted Language Learning (CALL) has become a research discipline on its own, which is a subfield of applied linguistics. In fact, CALL research has been developed for more than forty years (Hart, 1995). The integration of artificial intelligence techniques into CALL is called Intelligent Computer-Assisted Language Learning (ICALL) (Gamper & Knapp, 2002). ICALL, in general terms, is a computer program that is able to assess the learner's response and provide the feedback.

At the same time, the integration of artificial intelligence in education has led to the development of Intelligent Tutoring Systems (ITSs), which can provide learners feedback by imitating human tutor behaviors. An ITS in the language domain is called an Intelligent Language Tutoring System (ILTS). Hugh, Burns and Capps (1988) have differentiated ITSs from computer-assisted instruction (CAI) by describing an ITS's three intelligent modules as briefly introduced in the previous chapter. They are the capabilities to imitate the behavior of expert (expert module), evaluate the student's level of skills (student module), and implement instructions based on its pedagogical strategies (tutor module). Although some experts (Gamper & Knapp, 2002) differentiate an ICALL system from an ILTS by stating that the latter reflects its intelligence through its properties (the components of ITS) while the former one emphasizes error diagnosis (Levy & Stockwell, 2013), these two terms can be interchangeable to some extent (Amaral, 2007).

ILTS development

Most of the studies of developing ICALL systems focused on employing the AI techniques such as natural language processing (NLP) techniques and most recently automated speech recognition (ASR) and machine translation (MT) to test language skills, rather than taking into account about the foreign language acquisition. For example, Heift (2001) developed parsers to identify grammatical errors by incorporating NLP techniques in analyzing student input. His German Tutor provided error-specific feedback in answer processing. Meanwhile, the work of Wang and Carigliano (1992) demonstrated techniques of handling errors resulting from mother tongue transfer into a second language being learned. In other words, many ICALL systems often have very sophisticated language processing mechanisms, but they focus on very specific aspect of language, such as the syntax issue of clitic placement (Virvou & Tsiriga, 2001), instead of foreign language teaching and learning practice. For CALL tutors that focus on teaching foreign languages, English is the most frequently taught language, counted as 14 out of 40 systems in the review by Gamper & Knapp (2002). Some of the cases of language tutoring are Robo-Sensei (Nagata, 2002), and E-tutor (Heift, 2003; 2010). Nagata's system is designed for Japanese learning with a series of lessons while E-tutor was implemented for learners of German. Nevertheless, studies of systems developed for Chinese are rarely done. On the other hand, as a subfield of Intelligent Tutoring Systems research, the development of language ITSs has not been as popular as some sub-domains of science, such as math (Beal et al., 2010). The limited cases of ILTSs are often implemented and used in the field of military training for leadership training and foreign languages learning for military use. In general, they use game-based environment to make learners practice through dialog interactions and act as characters in order to learn foreign language and cultural skills. A famous example is tactical language training system (TLTS) (Johnson, Marsella,

Figure 1 Screenshot of Screen of TLTS (Johnson et al., 2004)

& Rey, 2004), which teaches learners basic communicative skills in foreign languages and cultures (see Figure 1).

This system has currently been used to teach Levantine and Iraqi Arabic and systems for other languages such as Farsi are being developed. Another similar project ELECT BiLAT (Lane, Core, Gomboc, Karnavat, & Rosenberg, 2007) is also a game-based simulation that offers soldiers a practice environment to negotiate in cultural contexts (Figure 3). The virtual environment cultural training for operational readiness (VECTOR) (Deaton et al., 2005) is designed to train military

Figure 2 Screenshots of Screen of ELECT BiLAT (Lane et al., 2007)

leaders and soldiers with skills in different cultural understanding (Figure 2). Trainees learn cultural skills in a virtual scenario where trainees act as characters using speech and gesture to communicate. The initial VECTOR was used in cultural training in Iraq. Another system called IN-TALE (the Interactive Narrative Tacit Adaptive Leader Experience) (Riedl & Stern, 2006) is another 3D game-based system teaching trainees in cultural skills and language. As a matter of fact, the main direction of ILTS development is primarily focusing on military use, which focuses on a narrow use, though they are highly effective and sophisticated. There are not many ILTSs developed for instruction of foreign languages, as noted above. Some of the ILTS teaching Chinese will be discussed in the latter of the chapter. Thus, the research and studies of commonly used foreign languages learning on a practical base will be a future direction.

In summary, there is relatively little research that bridges the overlapping fields of ILTS (arising from computer science and education) and ICALL (arising from linguistics, language

Figure 3 Screenshot of screen of IN-TALE (Riedl & Stern, 2006)

learning, and technology). Further still, the main target languages taught are English, Japanese, French, German and Spanish, while Chinese language is rarely taught. Therefore, an ILTS for Chinese language is needed. The increasing importance of Chinese and the nature of Chinese will be discussed in the following section.

2.2 Learning Chinese as a Second Language

The United States is a multicultural nation where many languages are spoken within the country. Other than official language English, the most worldwide popular spoken language (20% of world population) (Shih et al., 2013), Chinese, ranked the second popularly used right after Spanish in the United States, with 2.6 million (Census Bureau, 2013) people using it within the country. Nowadays, with the growing importance and influence of China's economy globally, Chinese, has gained more and more attention all over the world.

There is a growing interest in learning Chinese as a second language. Since the 90s, many universities, colleges and schools have established a Chinese degree program in order to satisfy the popular demand for Chinese. According to Chinese Ministry of Education ("Chinese Ministry of Education," 2014), there are 330 official institutions teaching Chinese in the worlds.

It's not unreasonable to predict that Chinese learning will be a rapidly growing subject in the 21th century. More instructions and technology related to Chinese learning will be needed.

The nature of Chinese characters

Written Chinese has been developed for more than 3,500 years in China, now it becomes the only one of the three known logographic systems in use (Karttunen & Crosby, 2006). Chinese language has long served as the initial writing system to other East Asian languages such as Japanese, Korean and Vietnamese (Park & Arbuckle, 1977). For example, "Kanji," regarded as Japanese version of Chinese characters (H Shu, Anderson, & Zhang, 1995), is derived from Chinese "Hanzi" (Chinese character as "汉字"). About 70% of Korean vocabulary came from Chinese (DeFrancis, 1984). Today, Chinese characters are still widely used in those countries, and most of people can read Chinese characters.

As a non-alphabetic language system, Chinese contrasts sharply to alphabetic systems such as English (Perfetti, 1999). Alphabetic languages use phonetic approach to record their sounds. You can pronounce the characters you see in the words. Characters/symbols themselves don't have meaning. On the contrary, there is no regular correspondence of Chinese spoken and written system. Thus many researchers classify written Chinese as logographic language system (Hung & Tzeng, 1981).

Each Chinese character forms a single syllable and can stand alone representing a distinct meaning. The meanings of Chinese characters can be directly delivered through its ideographic features instead of sounds (McEwen, 2006). In the ancient period, the Chinese pictographs were recorded to express meaning, which looked quite like the tangible objects. Over the years, more compounds are developed with several elements are combined (Hsu, 2012). The meaning of the compound characters can be derived from the meaning of its elements. The phonetic symbol system "pinyin," used to annotate pronunciations of Chinese characters, was generated until the modern time (Chung, 2003). The correlation between spoken form and written form of Chinese is comparatively independent.

The elements used to form the characters are known as "radicals." Most Chinese characters are compound characters, in which there are two or more radical components (Hua Shu, 2003) which can normally be categorized as semantic or phonetic radicals. The configuration of radicals in compound characters normally follows either a left–right or top–bottom structure (Wang, Perfetti, & Liu, 2005). For example, there are two parts in a compound character (e.g., 妈/ma 1/(mother)): one component is called a semantic radical (e.g., 女(female)), which demonstrates meaning, and another component is called a phonetic (马/ma3/), which carries information about pronunciation (Gitterman & Sies, 1994). All in all, there are around 200 semantic and 1100 phonetic radicals in Chinese writing system (Wang, Perfetti, & Liu, 2005). They can form the most of the Chinese characters.

The difficulty of learning Chinese characters

For learning Chinese as a second language, Chinese poses a challenge for alphabetic-based language speakers (Al-Mekhlafi, Hu, & Zheng, 2009).

One of the most representative characteristics of the Chinese writing system is its visual complexity. Huang and Ma (2007) have described Chinese difficulty as too many characters, too many pronunciations, and too many strokes. Leong (1989) stated the main difficulty is to learn a significant amount number of Chinese characters, and the only way is to rote memorization. Although the total amount of Chinese characters are truly large, the common used ones are only 3800 which can cover 99.9% of basic Chinese reading (Chen & Hsuan Chih, 1992). Even in these, a great proportion of them share common elements, which are used frequently. In fact, understanding hundreds of characters can help the comprehension of the basic conversations. Additionally, if the structural awareness of characters is obtained when learning, the logical inference of the possible meaning can be made even when meeting the new characters or new words.

In addition to the problems of character appearance and the large number of them, the lack of alphabet-like syllabary and the characters' common origins complicate the problem for alphabetic-based speakers (Huang & Ma, 2007). Some researchers (Chen et al., 2013) suggest these issues are grounded in the divorce between spoken and written Chinese, in which there lack the connection between systematic phonetic content and character structure. Taking English for example, its spoken form is consistent to writing system that records sounds (DeFrancis, 1984). So, it is comparatively convenient for people to learn. However, Chinese doesn't provide this convenience. Chinese establishes a direct link between form and meaning. Native Chinese speakers get the speaking skills since they were born, which set a solid foundation for learning characters after entering school. Generally, written and spoken systems are learned during different period of time for natives, which disperse the difficulties of learning. However, for beginning second Chinese language learners, it is very complex for them to pick up both written and spoken knowledge at the same time. The challenges posing the beginning Chinese-as-second-language learners are actually two aspects: spoken is one thing; written is another thing.

Pedagogy proposal for tutoring Chinese

Figuring out a proper pedagogical method for beginning Chinese-as-second-language learners is crucial. In fact, researchers have long ago proposed the strategies for Chinese-as-second-language learners. Rozin et al. (1971) conducted an experiment in which they taught 20 characters to eight second-grade students with severe reading difficulties. It turned out that those students gained characters quite quickly even though they could not even pronounce single consonant-vowel-consonant words. This study confirmed the hypothesis that Chinese characters bypass the phonetic approach.

To further understand the feature of Chinese written system, it is necessary to take a look at the difference between the information processes of reading different written systems. Liu (1978) diagrammed a figure to illustrate how different writing systems convey meaning.

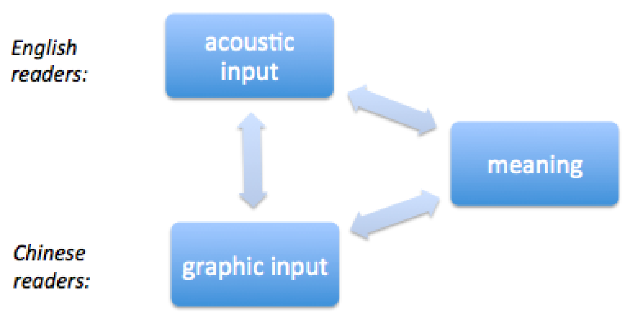


Figure 4 The information processing of English vs. Chinese (Liu, 1978)

As shown in Figure 4, Liu indicated that English readers perceive meanings phonologically while Chinese readers receive meanings directly from graphic input. Park and Arbuckle (1977) concluded that the visual appeal of Chinese characters is the most important factor of teaching strategy for American learners. Shi, a college professor who taught Chinese as a second language to non-native learners, also suggested that it would be easier to start with Chinese characters for non-native learners (Shen, 2005). (Chu-chang & Loritz, 1977) proposed the two strategies for reading process: 1) using a visual strategy when reading an unfamiliar language, and 2) using a more phonological strategy to read a native-like language. It infers that the visual approach is the first step when teaching Chinese characters to beginning Chinese learners.

In addition, as Chinese characters originated from pictographs, the mapping between the meanings and graphemes of characters can be utilized to make Chinese learning easy to start with. The interesting logic between characters' meanings and graphemes and the beauty of calligraphy can be attractive for beginning non-native learners. Giving the beginning learners the awareness of Chinese characters' structural properties is also an effective way to teach non-native speakers (Huang & Ma, 2007).

To conclude from above, this study will focus on tutoring characters to beginning Chinese-as-second-language learners by using visual appeal and structural awareness of Chinese characters, as justified effective by various research aforementioned.

2.3 Gaps in Previous Work

Current Chinese language instructions

A lot of effort of learning and teaching Chinese language to non-native speakers has been made (Hsiao, Chang, Chen, Wu, & Lin, 2013; Li, 2001). Traditional classroom and human tutoring is the most common method. Teachers plan out the process and content for students to learn. The great benefit of this method is that teachers can give immediate feedback whenever learners have problems throughout the learning process. Meanwhile, with the wide use of the Internet, more and more web-based online courses occurred, for instance, Yoyo's online courses (2012), which have detailed material as well as video or audio. Some other web-based system using animations to help rote memorization of the stoke orders. But those methods are mostly following the stereotypical educational routine in which the rote memorization and recitation strategies are adopted. These tutorials are not targeting beginning learners. Large amounts of learning materials embedded with every aspect of Chinese including spoken and written at the same time would overwhelm the beginning learners.

Yet, some researchers and instructors have noticed that instruction highlighting the elements (radicals) and the visual presentation of Chinese characters is more effective than focusing on rote memorization of stroke orders (Hue & Erickson, 1988). ShaoLan developed a visual-based method to teach Chinese characters, and published a book *Chineasy* (Hsueh, 2014). Her method made learning Chinese easy and fun possible by emphasizing the idea "building blocks" of characters. Besides, some studies such as Hsu's work (2012) and work by Ho et al. (2003) also adopted the way of using the structural feature of Chinese characters to improve Chinese character acquisition strategy. Still, as they are paper-based, lacking of embedded learning environment, learners could not be engaged or involved in the learning process.

Considering integration of etymology-based method and technology, namely, computer-assisted system, Hor's study (1991) was initial attempt to develop an etymology-based instruction in a hypermedia environment for beginning Chinese learners. Interactive videodisc was employed in the teaching of Chinese. It provides a foundation for pedagogical method; at the same time, takes advantage of the new instructional technology for teaching Chinese characters. Table 1 Chinese character acquisition instructions below shows the current instructions of teaching Chinese characters acquisition.

Table 1 Chinese character acquisition instructions

Representative instructions	Language skill taught	Delivery & Pedagogy	Method
Yoyo's online courses	Comprehensive skills	Video	Internet-based
***Chineasy* by Shaolan**	Characters acquisition	Visual based	Paper-based
Hsu, Ho	Characters acquisition	Radical (structural awareness)	Paper-based
Hor	Characters acquisition	Etymology-based	Hypermedia

Chinese ILTS

Today, ILTSs show great competency in the language teaching field by featuring the properties pf imitating human tutor and various forms of interaction through the user interface. Nevertheless, very limited systems have been developed for Chinese characters, with various methods and teaching targets employed. Among them, Massaro et al. (2006) applied the animated agent to produce realistic speech for Chinese speech learning, which didn't focus on Chinese characters. Kosek and Lison (2014) presented an ITS for Chinese words and grammar

Figure 5 Screenshot of interface of Kosek and Lison's work Kosek and Lison (2014)

learning while still not targeting beginners (Figure 5 Screenshot of interface of Kosek and Lison's

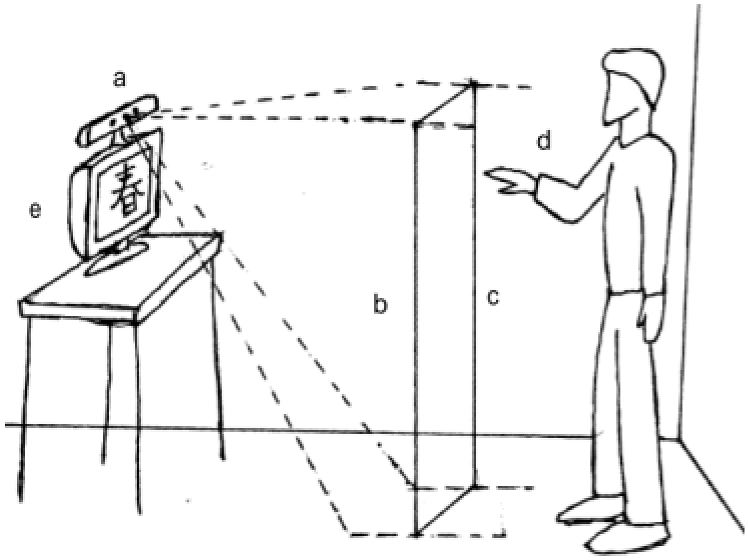


Figure 6 Figure from of Ji's work (Hsiao et al., 2013)

work Kosek and Lison (2014)). Another program was a mobile Chinese learning system using the application of TRIZ theory, implemented by researchers in China (Shih et al., 2013). But they didn't apply pedagogy for learning. Several other ITS work focus on the acquisition and diagnosis of Chinese characters by decomposing the parts from characters (Hsiao et al., 2013; Xu, Jiang, Lau, & Pan, 2007) of which Ji and his colleagues implemented a prototype system using gesture to teach stoke orders of Chinese writing (Ji, Yu, Li, & Shen, 2013) (Figure 6 Figure from of Ji's work (Hsiao et al., 2013)). Still, they didn't apply pedagogy of using etymology or Chinese structural features. Lam's work (Figure 7 Screenshot of Lam's work (Lam et al., 2001)) has fostered the structural awareness to teach Chinese characters within a CALL system (Lam et

Figure 7 Screenshot of Lam's work (Lam et al., 2001)

al., 2001) whereas further possible interactions between learners and system will still be acquired.

In summary, the several existing cases of Chinese ILTSs either don't focus on teaching characters or neglect the pedagogical method of applying ideographic characteristic. Therefore, there are gaps in previous work at different level of skills, instructional purposes and pedagogical methods. The study will be needed for an integration of pedagogical instruction within an ITS and targeting beginning learners. In order to address these gaps, this research will use the integration of metaphor and etymology as pedagogical method to teach Chinese character acquisition to Chinese-as-foreign-language beginners.

CHAPTER 3 METHODS

3.1 Overview

The purposes of this study are 1) to explore the initial feasibility of an Intelligent Language Tutoring interface for learning Chinese characters to teach beginning Chinese-as-a-second-language learners and 2) to assess the system by investigating usability issues, users' attitudes towards the system, and learning effectiveness. Briefly, this chapter introduces the design and evaluation methods used in this study. This study applies the user-centered design process described in by Dix et al. (2007), through which an iterative design process is initialized based on the requirements gathered at the very beginning to ensure the design aligns with the demands of target users. Then the design and development is followed as well as modification of design based on formative evaluation, and experimental evaluation of the system feasibility is discussed at the end. The brief process is shown in Figure 8.

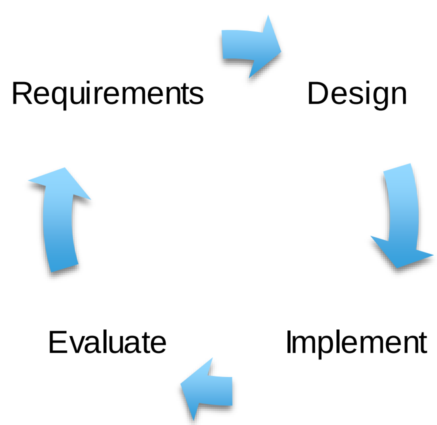


Figure 8 User-Centered Design Process per Dix et al. (2007)

Specifically, procedure of this study is shown in Figure 9. The process of designing the system follows a conceptual framework proposed by (Garrett, 2003), presented in Figure 10.

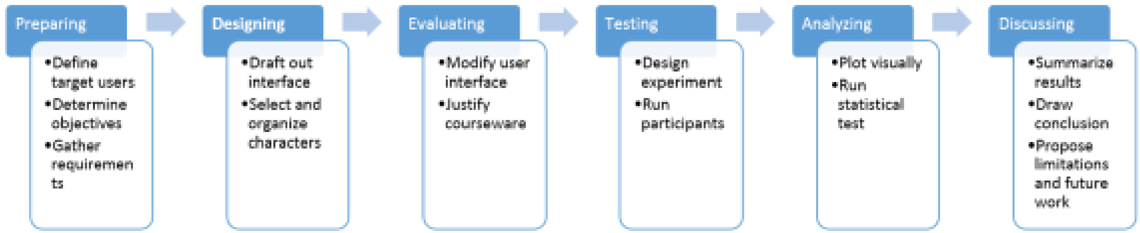


Figure 9 Study flow chart

The five planes - strategy, scope, structure, skeleton and surface - will be built from bottom to top to provide a better user experience to users. The process of following this conceptual framework will be discussed in the following sessions.

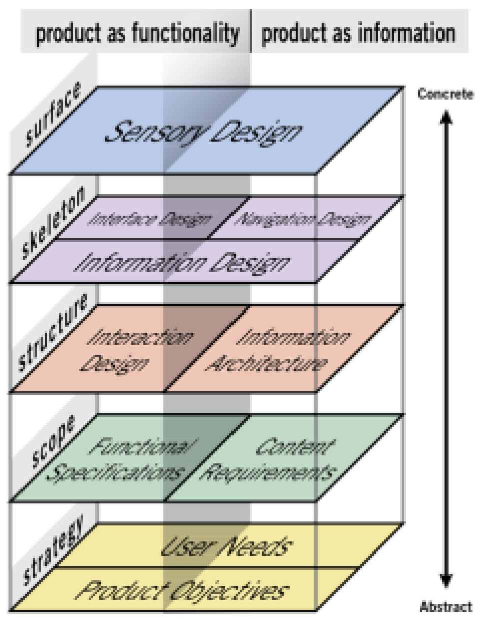


Figure 10 Conceptual framework - five planes (Garrett, 2003)

3.2 STAGE 1: Preparation for Content

The foundation of the design is a clearly articulated strategy (Garrett, 2003). In order to determine how the final system will provide the expected instructional purposes, knowing what to accomplish to the target users helps inform the decisions made about every aspect of the system. In this section, the content was prepared by 1) defining the target population of study, 2) determining the objectives of the courseware, 3) gathering requirements for guiding the design of the tutoring materials, and verifying the study design, 4) selecting characters and organizing characters, and 5) evaluating the tutoring materials.

Defining the target population to study

The target population is adults beginning Chinese-as-second-language learners (have no knowledge and background related to Chinese characters). Child learning and adult learning are very different (Xing, 2006). Younger learners can memorize faster but they lack endurance and analytical skills. Native speakers learning a language typically start with listening and speaking, while the focus of this study will be tutoring recognition of Chinese characters. Thus, adults will be the target population.

Determining the objectives of the courseware

The purpose of learners using this system is to increase their abilities to recognize Chinese characters and their interest in learning more characters or Chinese in the future. To figure out what content and functionality the system will offer to learners, the objectives of the courseware needs to be clarified, that is, the scope of the study is defined (Garrett, 2003).

1. Learners will be able to recognize the structural features of characters taught in the course.
2. Learners will be able to memorize the meanings of characters taught in the course.
3. Learners will be able to predict the meanings of unknown characters containing one of the elements taught in the course.
4. Learning won't require great difficulty (for example: higher mental workload).
5. Learners will have positive attitude through the learning process (for example: frustration is low).
6. Learners will have more interest in learning more Chinese characters.

Gathering requirements

The requirements were gathered at the very beginning of the design process in order to guide design. The requirements specification elicits the information from users' point of view, which would be used to analyze the learning experience for some Chinese-as-second-language learners, and to identify the challenges and potential needs for learners. This session describes the requirements gathering that was approved by IRB protocol 15-219 (see APPENDIX A).

1. Description of requirements gathering

An online survey containing 22 questions (See Appendix A) was conducted, in which both qualitative and quantitative data was collected from participants who studied Chinese as a foreign language. A convenient sample of students and staff was used. The reason why the participants are learners who have Chinese learning experience instead of beginning learners (target population) is that it would be more efficient gathering the requirements since the beginning learners wouldn't know how to learn a new language without any knowledge or background about the subject. On the contrary, those who have learning experience and are also second language learners would know better about challenges and difficulties during their learning process. In the survey, five demographic questions (academic role, gender, native language, non-Chinese foreign language learning experience and level of fluency of them), 13 slider questions on a scale of 1-100 about participants' related learning experience, and three open-ended questions regarding their challenges of previous learning experience and suggestions for learning materials were asked.

The questionnaire began by asking the participants how long they have studied Chinese and in what different ways they learn it. The following are examples of questions that were asked (the last two are open-ended questions):

- How long have you studied Chinese?
- What's the level of your Chinese fluency?
- When you started learning Chinese, which of the following did you begin with (choose all that apply)?
- How difficulty do you think learning Chinese characters is, and why?
- Could you briefly describe the way you learned Chinese characters, and comment on whether you think it was efficient?
- If your friends wanted to start learning Chinese characters/Chinese, what suggestions would you give them to help them start?

2. Results of Requirements Survey

A total of 53 available observations were collected with almost equal number of male (51%) and female (49%) and the majority of undergraduate students (84%) versus graduate students (16%). In the participants, most of them are native English speakers (82%), and 91% of the participants have learned more than two languages.

With more than 64% participants have studied Chinese for more than 1 year, and 83% of the total (44) have the experience learning Chinese characters. All responses about their Chinese learning experiences were helpful in general, and the responses from the 44 (83% of the total) were particularly useful for guidance on teaching Chinese characters.

Table 2 Level of fluency and difficulty of Chinese skills (n=53 on a scale of 1-100)

Skills	**Level of fluency**		**Level of difficulty**	
	Average value	Standard deviation	Average value	Standard deviation
Listening	41.06	30.75	65.89	35.30
Speaking	44.36	30.13	68.28	22.53
Reading	40.64	30.05	68.28	23.01
Writing	31.07	26.43	76.36	21.27

The most difficult part of Chinese is writing (characters are the means to present it), since its level of fluency is the lowest (31.07 out of 100) and its level of difficulty is the highest (76.36 out of 100) (see Table 2) among all the skills. Moreover, we can see both standard deviations are the lowest compared to other three skills, which means there is less variance between different individuals. In other words, writing characters is regarded the hardest part of Chinese so that their current learning of it is not effective. Acquisition of characters is the first step. Meanwhile, reading these characters is the second most difficult (68.28) and second lowest fluency (40.64). The standard deviations of writing are lower than the standard deviations of other skills, which indicates that people have a consensus of low level of fluency and high level of difficulty of Chinese. Also, in the question asking about their motivations of learning Chinese, the largest proportion (27%) among all the options is for learning Chinese culture. Chinese characters cover many aspects of the Chinese culture. According these results, finding an effective way to teach Chinese characters is crucial.

Interestingly, even though participants think reading and writing characters are the most difficult skills, and they don't learn them as well, reading and recognizing characters is one of the most adopted approaches when they started to learn Chinese (Table 3), meaning recognition of characters is among the basics of learning Chinese.

Table 3 Methods first used when learning Chinese (n=53)

# Number	Method	Response	Percent
1	Listen and understand what people say	33	66%
2	Speak some simple words learning from others	34	68%
3	Read and recognize some characters	34	68%
4	Write some characters	30	60%
5	Others	9	18%

When coming to the question of how Chinese characters are difficult, a relatively high score of 71.28 was given, and the responses on why they are so difficulty can be categorized into the following:

- Too many strokes and characters, especially rote memorization of them and their stroke orders
- So different with alphabetic languages
- Little connection between writing and speaking parts

The results align with the difficulties of Chinese characters described earlier, which also verifies the analysis of Chinese language nature and its pedagogical method in the previous chapter, and supports the proposed pedagogy aiming to address these difficulties. Some of the participants also pointed out that learning became much easier after they figured out the use of radicals, and even suggested on starting learning characters from radicals to find the meanings in the shapes that each character represents.

According to the participants' learning experience about the challenges and how they solve them, many of them commented that characters would be so difficult to learn if they didn't know the regular patterns of radicals and characters composed of them. But based on the responses of their current pedagogy adopted, the majority (78%) of them learn characters by memorizing and imitating the strokes without connecting meaningful structure of them. Thus, this is another finding to verify the proposed pedagogy would be effective and useful for beginning learners.

Table 4 describes participants' methods adopted for learning Chinese as well as their ratings for how useful these methods are. Participants may have chosen more than one method. There were 58 responses to traditional ways of learning (classroom and human tutor) versus 49 self-learners, which indicates the traditional way of learning is a little bit more prevalent while there is also a popular trend of self-learning. However, these traditional ways of learning are ranked the most useful, while self-learning from books or software also has the potential in teaching, but is ranked lower. Therefore, an intelligent tutoring system, if it could imitate the higher ranked human tutor, would be a solution to support learning. There is also a high interest score expressed (75.02) to try new learning software if a new system is designed.

Table 4 Methods of learning adopted (n=53)

Method	**Adoption**	**How useful are they?**		
		Not very useful	Useful somehow	Very useful
Learning in classroom	35	0	10	23
Learn from a tutor/friend	23	4	8	11
Self-learning from books	26	2	17	9
Self-learning from learning software	11	1	9	2
Self-learning from multi-media	12	3	8	2
Others	3	0	0	3

These results of the initial requirements survey were then used to guide the design of the interface for the future ITS.

Selecting and organizing characters

To establish effective and efficient instructional content, selecting and organizing tutoring materials is very important. As described in the previous chapters, there are a large number of Chinese characters; around 40,000 characters are recorded (Hsu, 2012) while 78% of them are obsolete or archaic. There are approximately 10,000 remaining current characters; thus, there is a need to select the representative characters to be taught.

Generally speaking, the criteria applied in this study is to choose characters which have:

- High frequency of use.
- Potential for generating new knowledge in the form of other characters. For example, many radicals and characters are commonly used and are components to form other characters or phrases. Those are the characters that can be considered to have potential to generate new knowledge.

The commonly used characters for understand basic literacy number about 2,000, in which 200 most popular ones are enough for daily reading. For the courseware in this study, to fit the study session in a regular 50-minute class period, eight most basic characters will be taught, from which about 30 characters and phrases will be derived.

Chinese characters are famous for their ideogram, in which the meanings conveyed directly from their shapes. So the radical of a single character always provides the clue to present the general meaning of characters containing the specific radical. The characters taught in this courseware are the representatives that have high potential for generating new knowledge. By knowing a couple of basic characters, hundreds of characters can be generated based on them. Thus, outside the current time-constrained study, this same principle would be applied to choose characters in a larger scaled system for teaching Chinese characters.

After selecting the characters, characters derived from one single character will be organized in a mini-course. The courseware shows from the visual presentation of the basic character to more explanation of mapping between its shape and meaning, then expand to other unknown characters based on it.

Evaluating of tutoring materials

To verify the feasibility of the courseware, a think aloud technique (Dix et al., 2007) was employed, and two informal reviews of the content on paper were conducted with two English speaking participants, both of whom have no knowledge of Chinese. They walked through the courseware without any external instruction, and they found the instructional content of the courseware is easy to follow and understand. This process gave initial validation of feasibility of the courseware to be used within the system that was under development.

3.3 STAGE 2: Design and Development

After the requirements were collected, what would be included in the system was clear; combining every aspect together to be a cohesive system was the next step. The following session introduces the design of the system by conceptualizing the structure, shaping the information design, and outlining the visual interface. Following the brief introduction of the four components in an ITS, described in Chapter 1, the approach to the design of the current system will be illustrated below by referring to these components: interface module, expert module, tutor module, and learner module.

The user interface & metaphor pedagogy

A user interface was created in Axure RP to create an app that would run on an iPad with a 9.7-inch screen. In the case of Chinese tutoring, characterized by its ideograms, the tutoring approach should be different with other alphabetic languages such as English. The characters selected to be taught were frequently used, and have potential to generate new knowledge. In total

Figure 12 Navigation (lesson selector menu) of the system

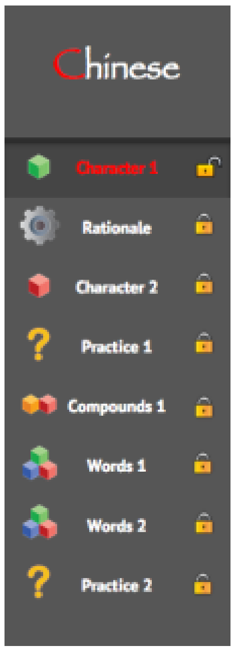


Figure 11 Progress indicator (left menu) of the system

34 characters and phrases derived from 8 characters (characters of “person”, “tree”, “sun”, “moon”, “mountain”, “mouth” and “door”) were taught within the interface. The instructional content was composed of 8 mini-courses, each based on one of these eight Chinese characters. Figure 11 shows the menu of navigating to each course. The number of characters can be expanded through arrows aligning each side of menu. Then, all the content presented within a course is organized in a

sequential flow (side menu presented in Figure 12), in which the order of content presentation is essential (Garrett, 2003). In this case, users will need to unlock each step to learn unknown characters based on the basic characters they've learned.

Furthermore, the tutoring approach in this system is dependent on the skill to teach and the instructional purposes. Within this system, the skill to teach is the recognition of characters and the target population is beginning learners. First, a brief introduction ( ) of Chinese characters will be displayed at the login page of the system to provide the basic knowledge about how to form Chinese characters.

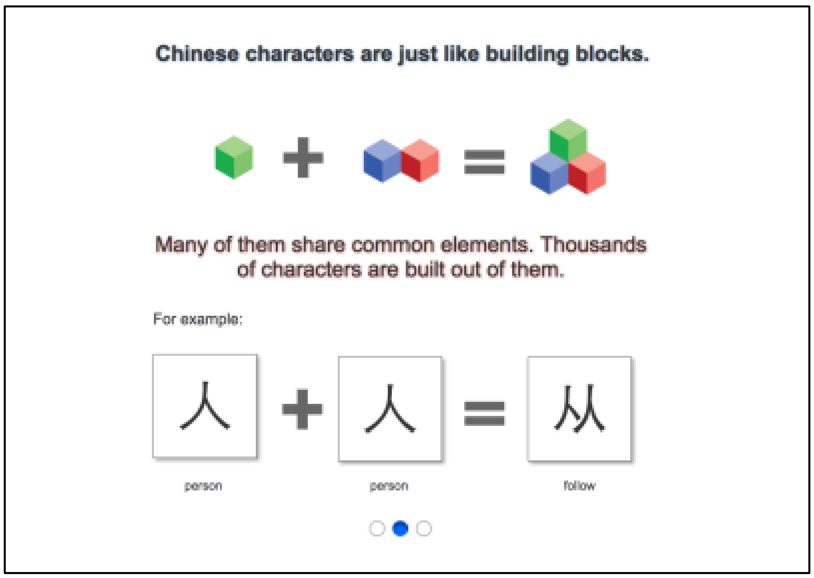


Figure 14 Introduction page

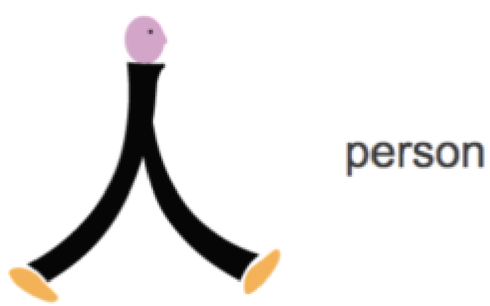


Figure 13 Example of metaphor

Second, given the nature of ideography of Chinese character, one of the approaches adopted is to use illustration to add more visual appeal, as discussed in the previous chapter that visual appeal is a proper method to teach Chinese characters to the beginners. For example, the Figure 14 represents character person, in which the illustration is added to make this character more intuitive. Then another approach used after the illustration is to providing more explanation highlighting the ideographic feature of the character. For example, an explanation box (Figure 15) describes the character of "person" just looks like the person it is.

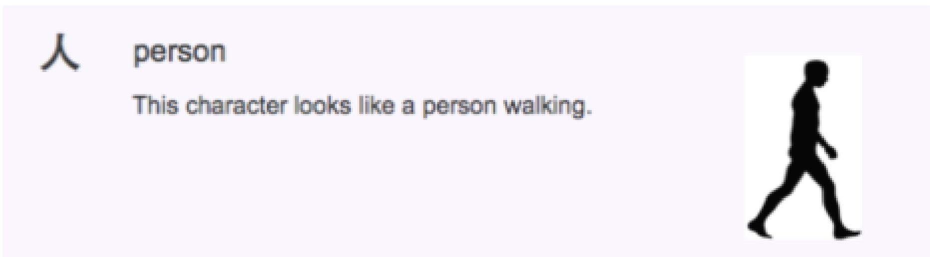


Figure 15 Annotation panel

Figure 16 Etymology example

Plus, etymology is provided to give users more ideas how characters evolved (Figure 16).

Next, use the concept of radicals is part of the tutoring approach to generate new characters or phrases. This gives learners the concept that simple characters can build new characters or phrases (Figure 17).

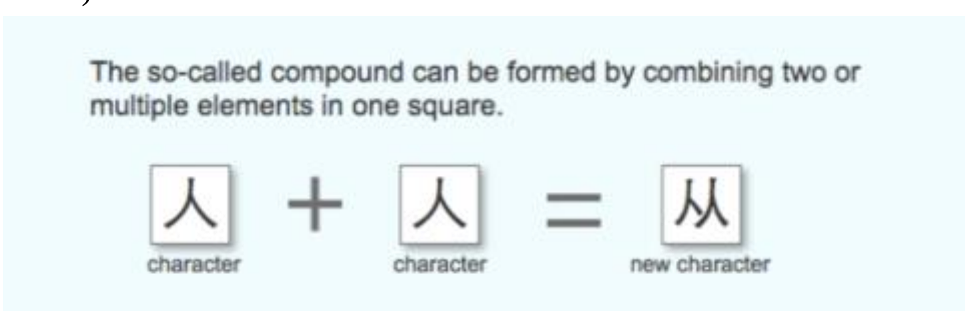


Figure 17 Concept of building blocks

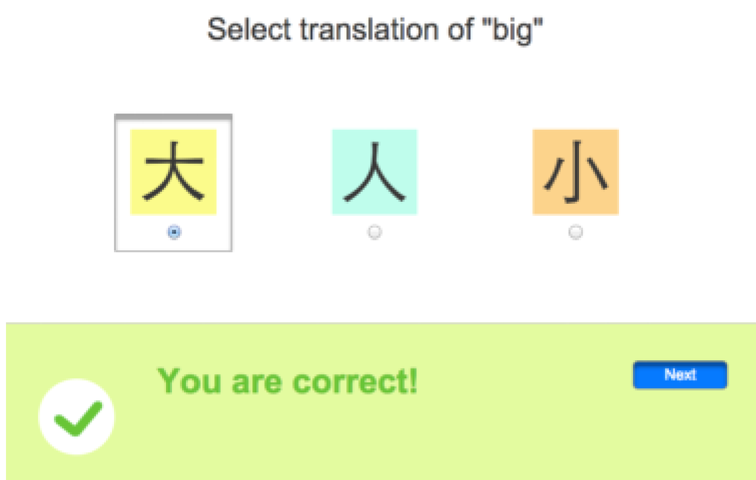


Figure 19 Learning practice

Figure 18 Panda cartoon

Additionally, the internal practices will be given in the middle of learning sessions in order to practice and reinforce students' knowledge (Figure 19). Lastly, game elements are employed to the design of this system. According to the idea proposed by (Aldrich, 2005), game elements increase enjoyment from educational experience. This system aims to increase the learning interest of learners by making learners have joys and less metal workload during learning process. Thus, a simulated cartoon instructor is employed in the design (Figure 19). A gamified path is created by releasing each lock on the side to unlock the current level and then go to the next learning session (Figure 20).

Figure 20 Gamified path

The user interface & design

At the top of the five-plan model, the visual design is the closest to users, through which the content, aesthetics and functionality come together to meet the goals of the other four goals. The design of buttons are based on the principles of using contrast and uniformity (Garrett, 2003). As mentioned, the important tool to draw attention is contrast, and uniformity provides effectively communication without distraction and confusion. For example, when a button is selected, it has its selected effects, which contrast sharply with other normal buttons (Figure 21). At the same time, the sizes, colors, locations and other properties follow consistency throughout system.

Figure 21 Button samples

Next, the main color palettes are gray on the menus plus the selected effects of buttons on them. Thus, a couple of more colors are used for icons and text on the menus to make them contrast with ambient color but not distracting. To make the presentation of characters and communication boxes more attractive, more colors are used for characters as to present flashcards to users, and some meaningful colors are presented as the green indicates right and red infers wrong (Figure 22).

Figure 22 Color palette samples

In order to properly model the student, the most basic requirement of the system is to know about learners' error (Swartz & Yazdani, 1992). In student module, the feedback interaction will be given based on students' performance on quizzes (Figure 23).

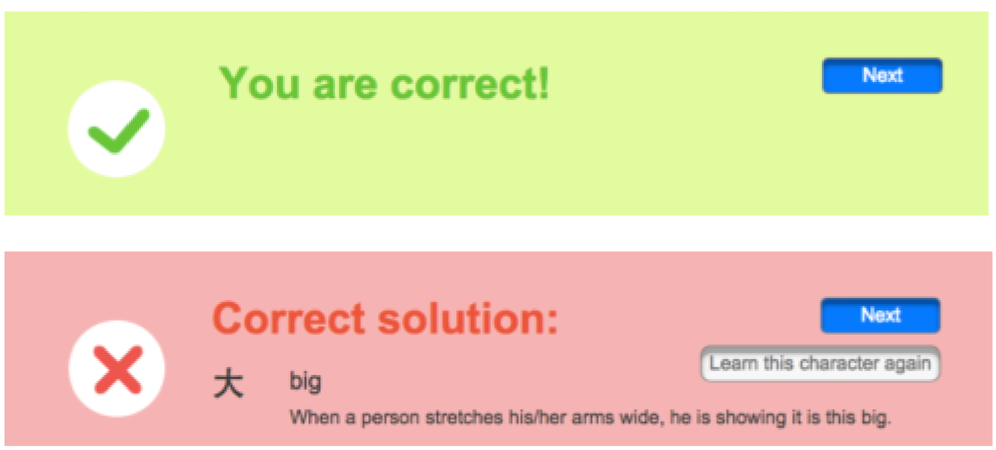


Figure 23 Feedback samples

Within this system, a complex computational diagnosis in the student module is not necessary since this system is focus on teaching recognition of characters instead of grammatical structure tutoring or speech diagnosis, and the feedback interactions are just dependent upon quizzes.

Formative evaluation of the interface

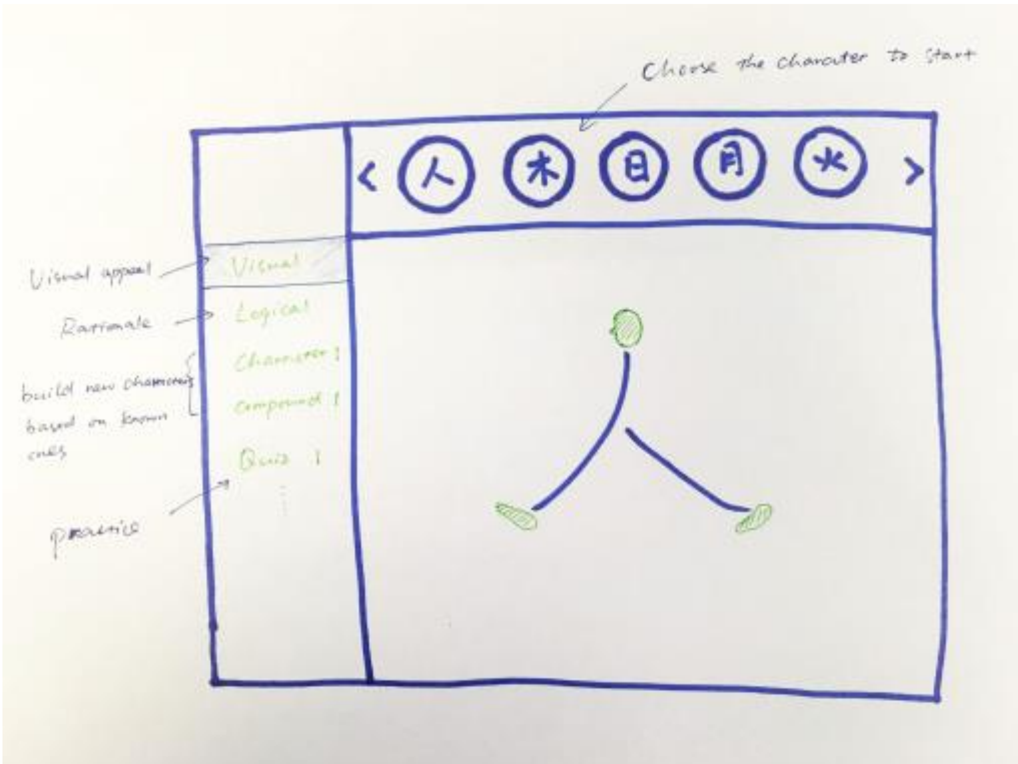


Figure 24 Paper prototype

Following the development of the first prototype (Figure 24), an informal heuristic evaluation (Dix et al., 2007) was conducted to identify any areas of the design that were likely to cause difficulties to users. Two human computer interaction faculty and three students were asked to do a cognitive walkthrough (Blackmon, Polson, Kitajima, & Lewis, 2002). They were asked to open up the iPad app, follow the instruction in the system, and perform the learning task using the system. Afterward, they provided feedback based on Nielson's ten heuristic (Dix et al., 2007), and suggestions on how to improve.

Based on this evaluation, the following modifications were made.

- Added a next button on each screen since users are more accustomed to click "next" to move to the next page rather than click the menu.
- Added more instructions on explaining better how the new characters and words are formed based on old ones so that the concept of ideograms for words vs. characters was not confusing.
- Changed the menu topics to more common names and added numbering to make it easy to following in a sequence.

System description and development

The introduction (Figure 25) comes up with when the application is opened. There are three pages to be swiped to get ready to move to the tour page. Then once users have moved to the tour page, a brief instruction on how to use the system is shown in Figure 26.

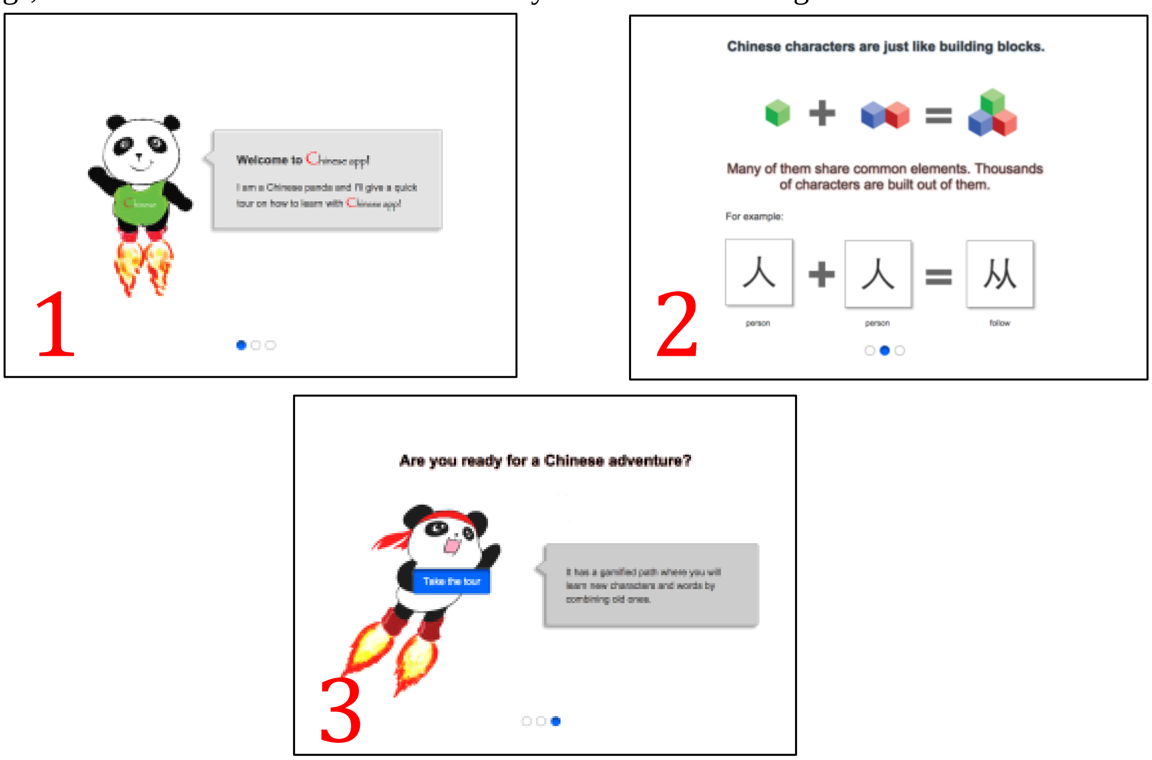


Figure 25 Introduction pages (from page 1 to page 3)

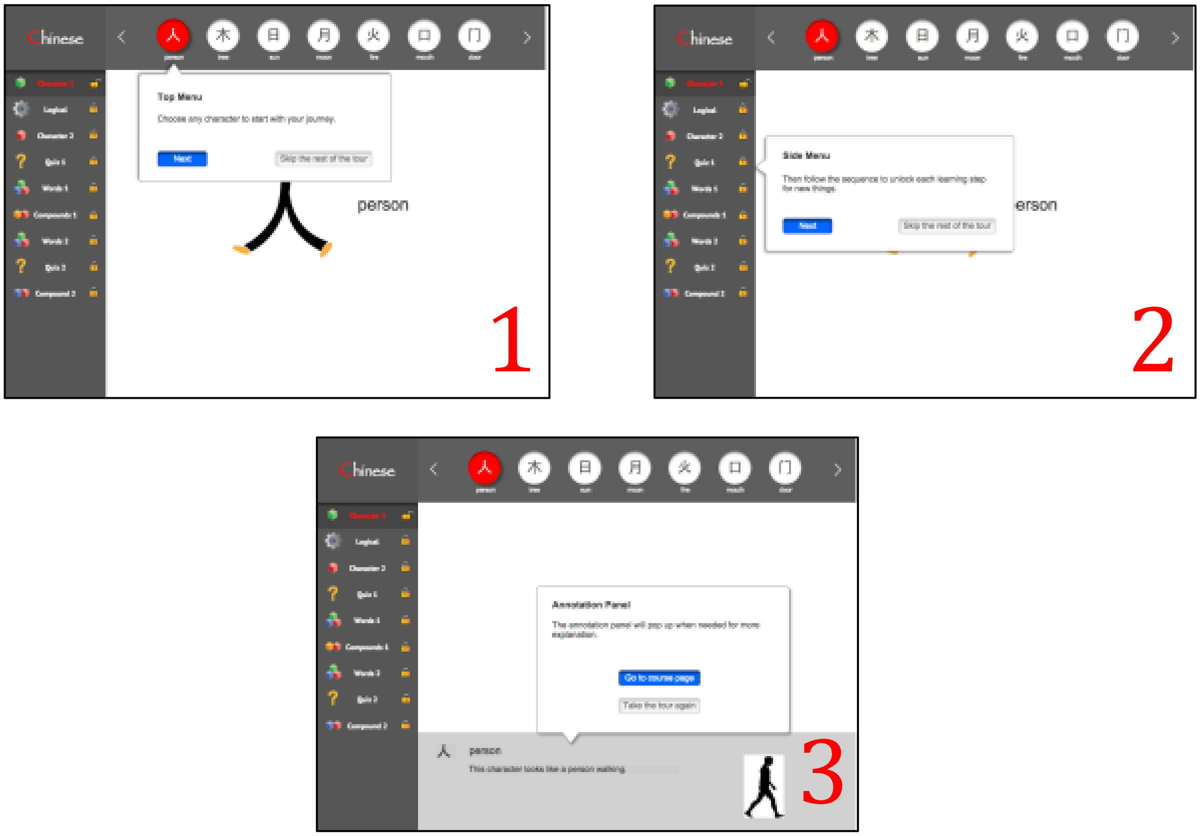


Figure 26 Tour pages (from page 1 to page 3)

The first page is customized by the first character with an illustration (Figure 27). A top menu containing every course and a side menu containing course content for each course is shown

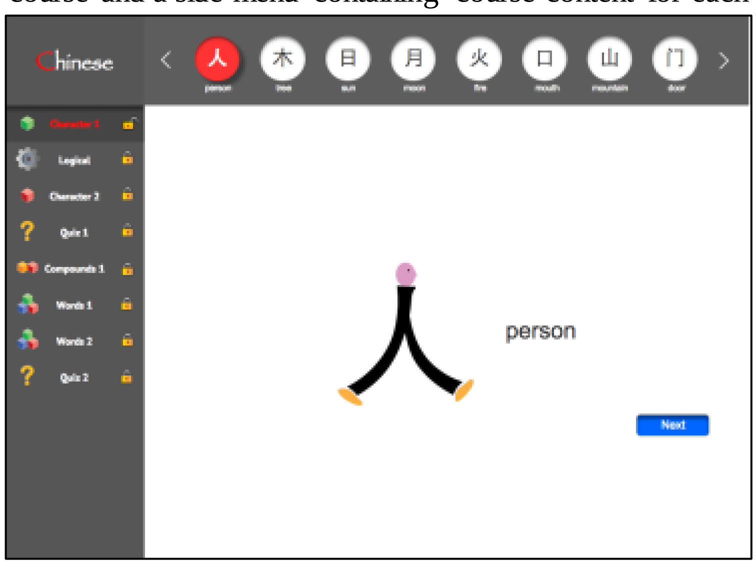


Figure 27 First page of "person" presentation

throughout the program. Navigating to the next step is executed by either clicking on the "next" button or directly choosing on the side menu.

The pages presented will show unlock on the sequence menu. The following pages after “person” page are showed in Figure 28. The design at the beginning cycle was made using Adobe Creative Suite. The final design was created using the prototyping tool Axure RP. This version was designed for the Apple iPad, but the layout could be used in other tablet formats or in a web based application as well. The prototype is functional in which the interactions between system and people were made through buttons, as well as having editable features.

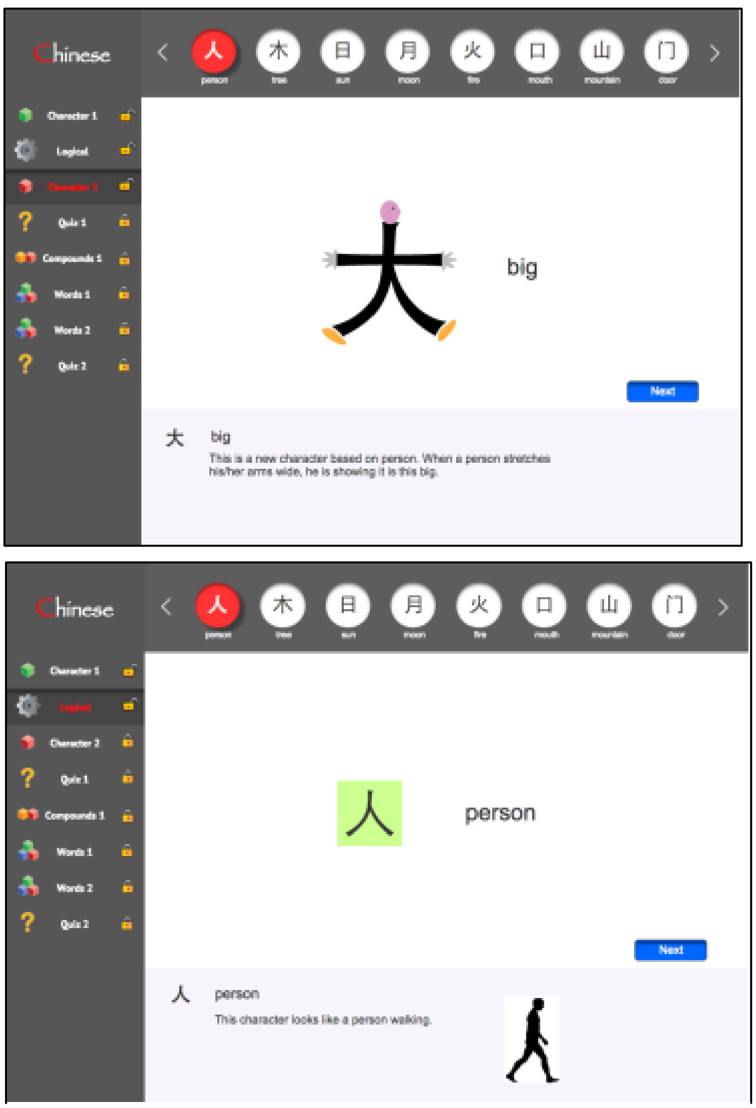


Figure 28 Following pages of “person” presentation

Other ITS modules

In a complete ITS, the expert module provides the domain knowledge that an expert tutor knows, while the tutor module represents the tutoring strategies used to deliver all the expert knowledge and instruction in the system. While developing these modules was beyond the scope of this project, the requirement gathering described above and the data gathered about the interface provide good input to these modules. In the full ITS, the expert module would contain feedback that would be customized based on learner behavior, and the tutor module would choose the best next character for the learner to learn based on his or her progress so far. The learner module would keep track, for each learner, which characters and skills of Chinese character learning have been mastered so far.

3.4 STAGE 3: Evaluation of the System

After the modification of the system, a formal controlled experiment was conducted. The experimental objective was to explore the feasibility of the system designed above and to evaluate the usability of it. In order to test the system that was built, the following hypothesis was explored. This section describes the experimental method that was approved by IRB protocol 15-597 (see APPENDIX B). For the purposes of this study, two versions of the iPad app were created: Version A was a "richer" version featured the full metaphor and etymology-based pedagogy described above, and Version B was "plain", with all metaphors and feedback removed.

Hypotheses

1. Users who use the whole system (Version A) will have better performance than the users using the system without feedback and pedagogical method (Version B).
2. Users who use Version A will have less difficulty when learning than the users using Version B.

3. Users who use Version A will have less mental workload while learning than the users using Version B.
4. Users' increase in interest in Chinese after using the system will be greater with Version A than with Version B when compared with the pre-survey baseline.
5. The interface of Version A will be user-friendlier than Version B.

Participants

Since this research is targeting adults beginning Chinese-as-second-language learners (they have no knowledge and background related to Chinese characters), a convenient sampling method was used to draw participants from the students' population at Iowa State University. The recruiting method was through mass email to college students and personal contacts.

Procedure

Prior to arriving to the experiment the participants completed an online informed consent form via the invitation email. Once they consented, they completed a pre-survey and signed up a time for the face-to-face portion of the study. When the participants arrived at the experiment location they were told to work with an iPad app designed to teach Chinese characters. With the completion of learning with the iPad, they took a quiz and completed a NASA-TLX survey and usability survey. The entire face-to-face portion of the procedure took 20-30 minutes. Participants were invited two at a time, so each participant might have had another participant nearby behind a divider, but participants were not introduced or encouraged to converse.

Table 5 Outline of experimental procedure

	PROCEDURE
1.	Online Consent form and pre-survey
2.	Introduction to experiment
3.	Learning session
4.	Quiz
5.	NASA-TLX survey
6.	Usability survey
7.	Survey
8.	Debriefing

Surveys

The surveys that were used in this experiment were a pre-survey (see Appendix B) and post-survey (see Appendix C) which contained the quiz, NASA-TLX survey and a usability survey. Each survey was presented online using Qualtrics.

Pre-survey

All of the participants completed a pre-survey before arriving to this experiment. The purpose of this survey was to gather some demographic information (e.g., gender, degree pursuing, etc.) and their Chinese related experience (exclude the participants have Chinese knowledge) and their interest of learning Chinese. The purpose of this pre-survey was to better understand participants' impression of Chinese language to see if their previous learning experience would influence their performance, and if their interests would increase after using the system.

Quiz

The quiz was included in the post-survey. There were 16 questions in total. Some questions were about translating the characters presented, and some were about select the translations of a given character or English word. At the end, three questions asked participants to guess the meaning of characters they had not been taught ("guess questions"). This approach was used to test if they have ability to predict the meaning of unknown characters based on what they learned from the radicals from the other characters.

NASA-TLX

NASA-TLX is a measure of workload (Sandra G. Hart & Staveland, 1988). This index has been used in many different experiments in many different fields. This index has six subscales: Mental, Physical, Temporal Demands, Frustration, Effort, and Performance. Hart stated (S. G. Hart, 2006) that workload could be represented by some combination of these six dimensions in NASA-TLX. In the current study, the NASA-TLX was given after completion of the learning session to be compared across the two versions of the system.

Usability survey

A usability survey was given after the NASA-TLX, which was also included in the post-survey. This survey contained the questions of System Usability Scale (SUS) (Bangor, Kortum, & Miller, 2008), which was used to test the usability. There were also other questions regarding the use of the system, which were more specific for the learning materials, and were categorized into overall satisfaction, ease of use, usefulness, interface satisfaction and functionality. At the same time, several open-ended questions were used to gather more personal comments. The purpose of this survey was to test overall usability through participants' learning experience using the system, as well as compare between two versions of system.

Independent variable and dependent variables

This experiment was a between-subjects design to avoid the learning factor. The independent variable was the two versions of the system: 1) The whole system (Version A), and 2) the system without feedback interactions and pedagogical methods (Version B). The comparison of two versions can be seen in Table 6, and screenshots of the two versions can be seen in Appendix D.

Table 6 Comparison between two versions

Features	Version A	Version B
Structure/navigation/flow	Yes	Yes
Idea of how to build characters	Yes	Yes
Metaphors	Yes	No
Pictures	Yes	No
Etymology & explanation	Yes	No
Practice feedback	Yes	No
Instructions (panda tips)	Yes	No

The dependent variables were performance (quiz score), time taken to complete each course, interest and related items measured in the usability survey. Details for dependent variables are shown in Table 7.

Table 7 Details for dependent variables

Construct	Measure	Method of Data Collection
Learning performance	Quiz score	Qualtrics
Time spent	The amount of time spent for each session	Manually counted with timer
Workload	Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration	NASA-TLX Survey
Usability	Likert scale	Qualtrics survey; questions based on SUS
Overall satisfaction	Likert scale	Qualtrics survey
Usefulness	Likert scale	Qualtrics survey
Interface satisfaction	Likert scale	Qualtrics survey
Functionality	Likert scale	Qualtrics survey

3.5 Data Analysis Plan

The number of the correct answers in the quiz will assess the learning performance. The predicted results, based on the literature about teaching Chinese characters, were that participants using Version A would have better performance. They were also expected to be more frequently correct on the guess questions, inferring the meaning of unknown characters. Since Version B doesn't offer much pedagogical scaffolding, the predicted results would verify the effectiveness of the metaphor-based pedagogy proposed in this study.

The workload will be measured by NASA-TLX survey. The usability questions will be measured on a scale from 0 to 100. Participants using Version A are expected to have lower mental workload and judge the iPad app to have better usability. Predicted results could be used to justify the feedback and pedagogical method as an effective foundation to support learning in a Chinese ILTS.

3.6 Limitations/Assumptions

There are assumptions and limitations within this experiment. First, it was assumed that there were no individual differences when learning the materials. Since it was not possible to do a within-subject experiment because of the learning effect, a between-subject method was used. Thus, the comparison was made between two groups of participants (comparing the mean). Thus, to compare the two versions of the app directly, we must assume that the two samples of population are similar. Second, it was assumed the participants took the similar amount of effort when learning. As we all know, learning is greatly influenced by how much attention you pay, how serious you are. But it was not possible to control participants' motivation during the learning session. Therefore, the learning performance and workload could be compared based on the assumption that participants have the similar state before test, and took similar amount of effort learning.

CHAPTER 4 RESULTS

The results and data analysis will be discussed in this chapter. As noted in Chapter 3, there were two groups of participants. The group who used Version A of the app (with metaphor and feedback) was Group A, otherwise were in Group B.

4.1 Participants

There were 86 participants in the experiment where equal number of participants (43) using version A and version B respectively (see Table 8). There were two comparison groups.

Table 8 Participant groups

	Group A	Group B
Version used	A	B
Features	Had metaphor and feedback	None
Number of participants	43	43

Demographic

The 86 participants were made up of 43 males (50%) and 43 females (50%) (Table 9). The role of the participants was as follows: 77% (66) were undergraduate students, 19% (16) were graduate students, and 4% (4) were faculty or staff (Table 10).

Table 9 Gender distribution

Version	Gender	Count	Percentage
A	Male	23	53%
	Female	20	47%
B	Male	20	47%
	Female	23	53%

Table 10 Academic role of participants

Version	Role	Count	Percentage
A	Undergrad	34	79%
	Grad	8	19%
	Faculty/Staff	1	2%
B	Undergrad	32	74%
	Grad	8	19%
	Faculty/Staff	3	7%

Figure 29 shows the native languages of participants. 85% (73) participants' native language is English. Among other languages Spanish and Hindi occupy 5% (4) respectively. Five other participants' native languages are Tamil, French, Malay, Sinhalese and Vietnamese. Regarding all participants' native languages, all of them are alphabet-based languages. Thus, all of the participants were alphabet-based native speakers.

Figure 30 shows the number of languages learned for all participants. Fifty-six percent of participants (48) have learned more than one language (including their native language). Among them, 54% (26) have learned two languages, 30% (14) have learned three languages, 15% (7) have learned four languages, and 1% (1) has learned five languages. In addition, three participants reported they had learned a little bit Chinese long ago and they had already forgot it. There were also two participants who reported they had learned Japanese, and another two participants who had learned Korean, but all of their fluency for these languages reported was under 10. Although Japanese and Korean share some common features with Chinese, they are still very different. Thus, they were eligible of having no knowledge of Chinese for participation.

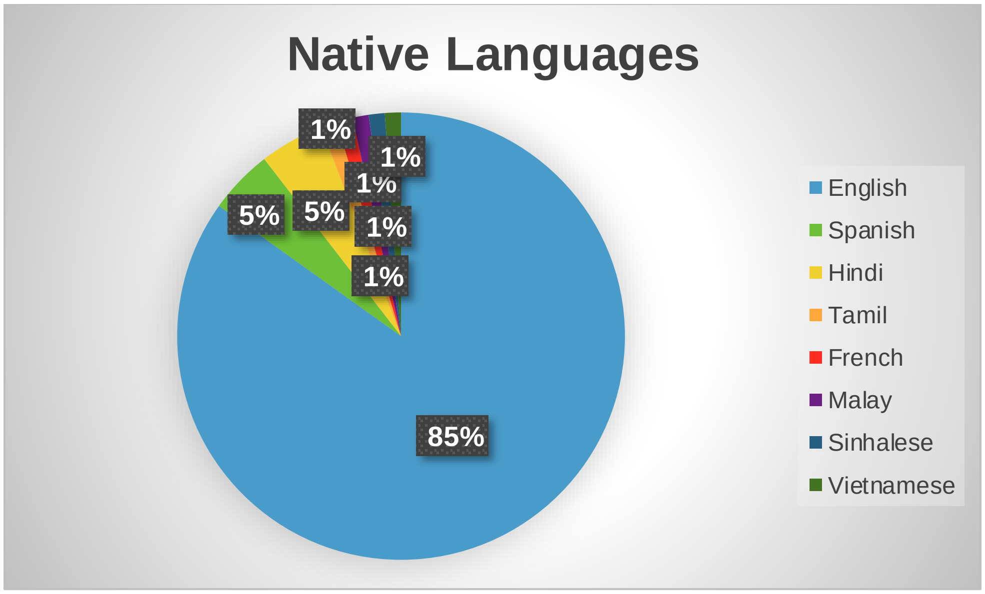


Figure 29 Participants' native languages

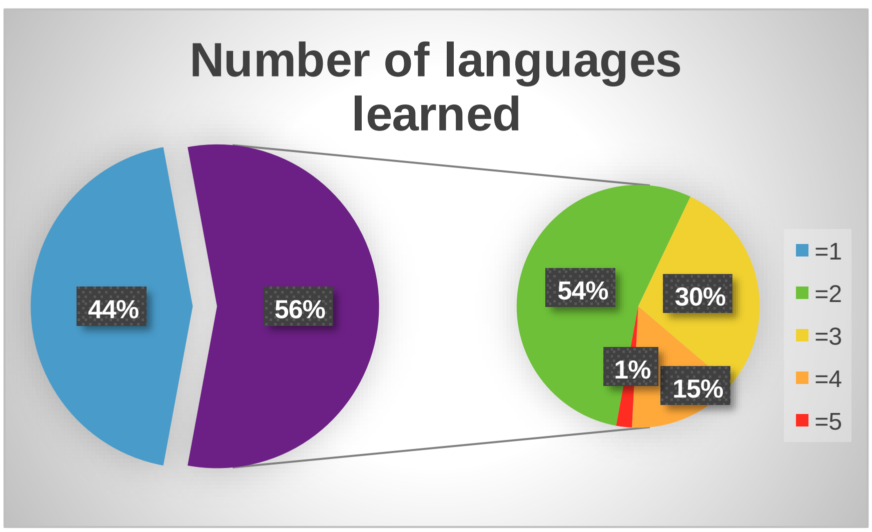


Figure 30 Number of languages learned

Attitudes toward learning Chinese

In asking how difficult they thought about Chinese characters, the mean difficulty rate out of the 100-point sliding scale was 79.0 (*SD*=14.42). The mean for their interest in Chinese was 63.3 (*SD*=19.03). Their likelihood of learning Chinese as a foreign language was 53.8 (*SD*=25.76).

The difficulty of Chinese characters posed to them, interest in Chinese language and the likelihood of them learning Chinese were reported in the bar chart below (Figure 31). Standard deviations are shown in the figure. Most of participants gave high difficulty rate for Chinese characters with range from 60 to 100. There were 51 out of 86 participants who rated difficulty above average. The interest and likelihood of Chinese were lower than the difficulty rate. Half of participants (43) designated their interest in Chinese as above the mean interest of 63.5. There were 33 participants who reported their likelihood of learning Chinese was below 50.

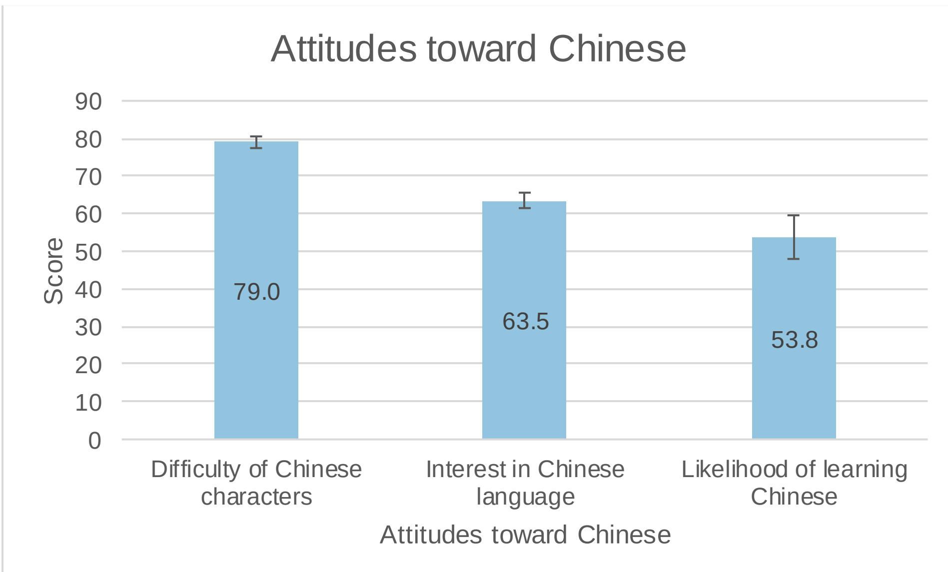


Figure 31 Attitudes toward Chinese

4.2 Learning Performance

In the study, all participants were tested using a quiz to measure their learning performance based on the iPad app. The quiz was formed of 22 questions with mixture of multiple-choice selection and translation questions. Each question is one point. The maximum quiz score was 22.

Learning performance comparison between two groups

In this study, participants who used Version B were the control group and participants who used Version A were the treatment group. The participants' quiz scores of two groups were plotted in the Figure 32 respectively. As seen in the boxplot, the majority of participants using Version A got higher scores than participants using Version B. Both groups contained participants achieving the maximum score, while the minimum score in Group A was 16 and the minimum in Group B was 8. The ranges from 25% quartile to 75% quartile scores for Group A and Group B were [19.5, 21] and [16.5, 21] respectively. More discussion of statistical difference of means and variances for both groups follows below.

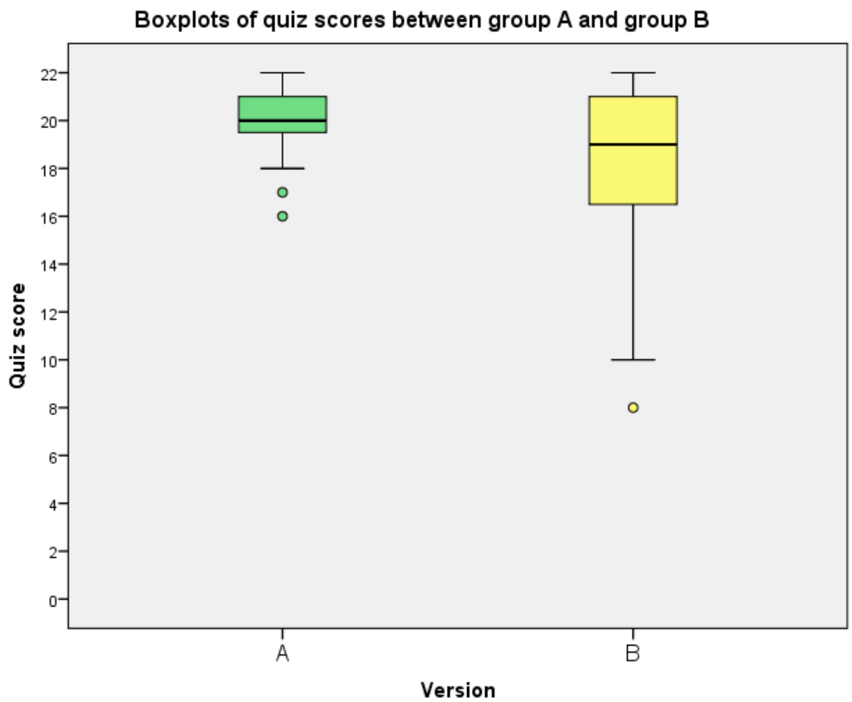


Figure 32 Boxplots of quiz scores between Group A and Group B

Independent-samples t-test

To further determine whether mean learning performance, measured in score points, differed between Group A and Group B, a two-sample independent t-test was run. The quiz score was the dependent variable, and the independent variable was group assignment.

1. Assumptions

There are six assumptions to be considered when using independent-samples t-test. The first three are: (1) continuous dependent variable, (2) independent variable is categorical with two groups, and (3) independence of observations. These three assumptions were guaranteed by the experimental design. The fourth assumption is (4) no significant outliers in terms of dependent variable. In the boxplot (Figure 32) shown above, there are several outliers. Since these outliers were the actual scores participants earned, removing them would not be appropriate. Thus, the test with and without outliers was conducted (sensitivity analysis), and the results were compared to decide whether two results differ based on the outliers. The fifth assumption is (5) the dependent variable should be approximately normally distributed for each group level of the independent variable. As shown in the following Tests of Normality table (Table 11) created in SPSS, quiz scores for each group weren't normally distributed, as assessed by Shapiro-Wilk's test ($p < .05$). Although the dependent variable was not normally distributed for each level of independent variable, non-normality does not affect Type I error rate substantially, and the independent-samples t-test can still be considered robust (Lund & Lund, 2013). The sixth assumption is (6) homogeneity of variances. The verification of equal variance in Groups A and B is discussed together with independent-samples t-test results down below.

Table 11 Tests of Normality

Tests of Normality

	Version	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
Total score	A	.179	43	.001	.900	43	.001
	B	.177	43	.002	.885	43	.000

a. Lilliefors Significance Correction

2. Results and interpretation

Table 12 Descriptive statistics for mean quiz scores

Group Statistics

	Version	N	Mean	Std. Deviation	Std. Error Mean
Total score	A	43	20.26	1.498	.228
	B	43	17.98	3.398	.518

As shown in the group statistics table above (Table 12), the mean quiz score was higher to Group A (*M* = 20.26, *SD* = 1.498) than Group B (*M* = 17.98, *SD* = 3.398).

Table 13 Independent samples test of mean quiz score

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
total score	Equal variances assumed	13.735	.000	4.025	84	.000	2.279	.566	1.153	3.405
	Equal variances not assumed			4.025	57.720	.000	2.279	.566	1.145	3.413

As seen in the independent samples test table above (Table 13), the assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances (*p* = .000). This large F value indicates, as can be seen from the boxplots in Figure 32, that the variance of the two groups is significantly different. Therefore, we use the t-test that has been adjusted for unequal variances. The results are as follows, with decimal precision adjusted appropriately based on the original score data. The mean quiz score of participants in Version A was 2.279 (95% CI [1.145, 3.413]) higher than mean quiz score of participants in Version B. There was a statistically significant difference in mean quiz score between participants in Group A and Group B, $t(57.72) = 4.025$, $p = .000$, $d = .87$. Therefore, we can reject the null hypothesis that the groups have the same score distribution and accept the alternative hypothesis that they differ.

In order to test if there are different results drawn from data without outliers (sensitivity analysis), another independent-samples t-test was run. The results showed the mean quiz scores for participants in Group A ($M = 20.44$, $SD = 1.266$) was higher than mean quiz scores participants in Group B ($M = 18.21$, $SD = 3.057$), a statistically significant difference, $M = 2.225$, 95% CI [1.200, 3.250], $t(54.934) = 4.350$, $p = .000$. This analysis yields a similarly significant result, and thus we can conclude that the outliers are not affecting the analysis.

Inference Questions

The design of the pedagogy of Version A aimed to help learners have the ability to make inferences about new Chinese characters they had never seen based on the knowledge they gained. To assess this goal, three of the quiz questions were "guess questions," which tested the participants’ ability to derive the meaning of a novel character that is a derivative of a character they learned. It is worth exploring whether the pedagogy aided this inference performance, and thus the two groups' scores on these questions are analyzed below. Table 14 shows the number of participants in each group who answer the guess questions incorrectly, along with the mean correct rate. The mean correct rate of all guess questions for participants in Group A was 90.7% (*SD*=9.3), while the mean correct rate in Group B was 85.3% (*SD*=22.9). To determine if the correct rate for participants in Group B was statistically significantly different than Group A, a binomial test was conducted. The binomial test table is shown below (Table 15).

Table 14 Correct rate for "guess questions"

Guess Question #	Group	Number participants incorrect	Correct rate
1	A	8	81.4%
	B	11	74.4%
2	A	4	90.7%
	B	3	93.0%
3	A	0	100.0%
	B	5	88.4%
Average across all guess questions	A	4.0	90.7%
	B	6.3	85.3%

Table 15 Binomial test for correct rate on "guess questions"

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)
Guess questions	Group 1	wrong	19	.147	.907	.000[a]
	Group 2	correct	110	.853		
	Total		129	1.000		

a. Alternative hypothesis states that the proportion of cases in the first group < .907.

The rate of getting guess questions correct for participants in Group B was .853 and the test proportion (correct rate of Group A) was .907 (Table 15). The correct rate of Group A was statistically significant higher than correct rate of Group B, evidenced by p = .000. Therefore, we can reject the null hypothesis that the groups have the same score distribution and accept the alternative hypothesis that they differ.

Factors influence learning performance

For participants in Group B, there was a large variance of their learning performance. We were interested in what factors led to highest (1st quartile) and lowest (4th quartile) scores in Group B.

1. Gender

Table 16 Lower and higher scorers in Group B (gender)

Level of learning performance	Gender	Count	Percentage
1st quartile (lower scores)	Male	3	27.3%
	Female	8	72.7%
4th quartile (higher scores)	Male	6	50.0%
	Female	6	50.0%

Shown in the Table 16, there were more males than females who got higher scores while more females got lower scores than males. Half of higher scorers were males and about ¾ of lower scorers were females. An independent-samples t-test was run to determine if there were differences between males and females. There was no statistically difference between males and females.

2. Academic role

Table 17 Lower and higher scorers in Group B (role)

Level of learning performance	Role	Count	Rate in the sample	Rate in Group B
1st quartile (lower scores)	Undergrad	6	54.5%	74.4%
	Grad	3	27.3%	18.6%
	Faculty/Staff	2	18.2%	7.0%
4th quartile (higher scores)	Undergrad	9	75.0%	74.4%
	Grad	3	25.0%	18.6%
	Faculty/Staff	0	0.0%	7.0%

As shown in the Table 17, two out of total four faculty/staff were in the lower score group. Their rate in 1st quartile (18.6%) was higher than their overall rate in Group B (74.4%).

Undergraduates had lower rate (54.5%) in 1st quartile compared with their rate in Group B (74.4%) while they had slightly higher rate (75%) in 4th quartile compared with their rate in Group B (74.4%). Graduates had both higher rates in 1st and 4th quartile than their rate in Group B.

In order to determine whether there are any statistically significant differences between the means of these three groups, a one-way analysis of variance (ANOVA) was run. There were no outliers in the data, as assessed by inspection of a boxplot generated by SPSS. As shown in the Figure 33 below, the quiz score decreased from the undergraduates (n = 5, M = 18.5, SD = 3.9), to graduates (n = 6, M = 16.33, SD = 6.1) to faculty/staff (n = 2, M = 13.5, SD = 2.1), in that order. There was homogeneity of variances, as assessed by Levene's test for equality of variances (p = .073). The group means were not statistically significant different ($p > .05$) and, therefore, we cannot reject the null hypothesis.

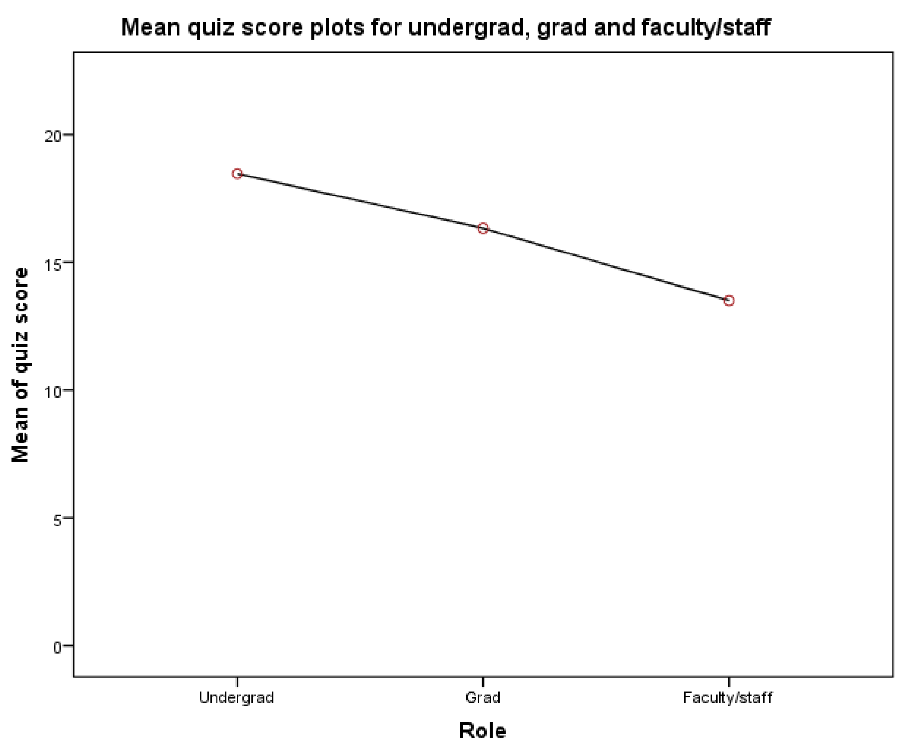


Figure 33 Means of higher and lower scores for undergrad, grad and faculty/staff (not significantly different)

3. Languages learned

Both in the higher score group and lower score group, except three participants were not native English speakers, the rest of them were all English speakers. Table 18 shows the number of languages learned for participants across 1st and 4th quartile of Group B.

Table 18 Lower and higher scorers in Group B (languages learned)

Level of learning performance	**Language learned (including native)**	**Count**	**Rate**
1st quartile (lower scores)	2	9	81.8%
	3	1	9.1%
	5	1	9.1%
4th quartile (higher scores)	2	8	66.7%
	3	4	33.3%

Another one-way analysis of variance (ANOVA) was run to determine if there were significant difference between the number of languages learned. As shown in the Figure 34, the quiz score increased from the participants who learned two languages (n = 17, M = 17.6, SD = 4.5), to who learned three languages (n = 5, M = 20.4, SD = 1.1), and decreased to who learned five languages (n = 1, M = 10.0), in that order. The group means were not statistically significant different and, therefore, we cannot reject the null hypothesis.

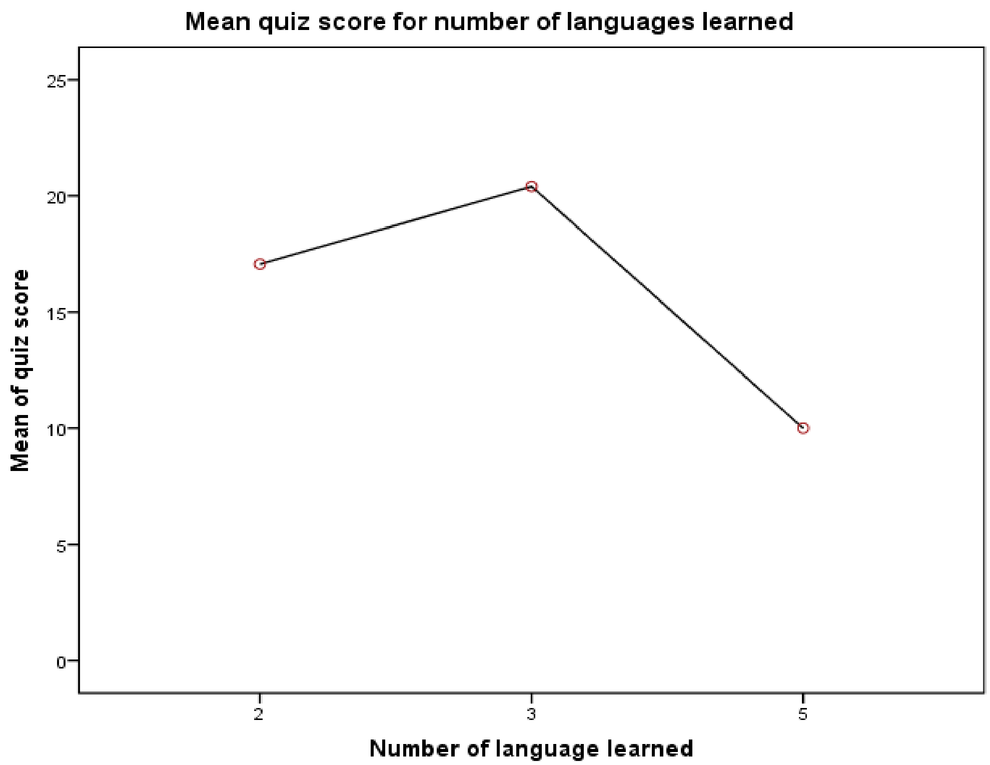


Figure 34 Means of higher and lower scores for number of languages learned (differences not significant)

4. Attitudes toward Chinese

We were also interested in whether participants' attitudes toward Chinese would affect their learning performance. In order to do that, a correlation test was run across all the participants. Although the variable of "how likely you would learn Chinese" was normally distributed, other variables were not, as assessed by Shapiro-Wilk's test ($p > .05$). Thus, Spearman's correlation was selected to run.

Table 19 Factors influences quiz score

Correlations

			how difficult Chinese character is	interest in Chinese language	how likely you would learn Chinese	total score
Spearman's rho	how difficult Chinese character is	Correlation Coefficient	1.000	.051	.100	-.043
		Sig. (2-tailed)	.	.644	.360	.695
		N	86	86	86	86
	interest in Chinese language	Correlation Coefficient	.051	1.000	.680**	-.028
		Sig. (2-tailed)	.644	.	.000	.797
		N	86	86	86	86
	how likely you would learn Chinese	Correlation Coefficient	.100	.680**	1.000	.041
		Sig. (2-tailed)	.360	.000	.	.709
		N	86	86	86	86
	quiz score	Correlation Coefficient	-.043	-.028	.041	1.000
		Sig. (2-tailed)	.695	.797	.709	.
		N	86	86	86	86

**. Correlation is significant at the 0.01 level (2-tailed).

Shown in the correlation table (Table 19) above, there weren't significant correlations of these factors to learning performance. There was a strong positive correlation between interest in Chinese and likelihood of learning Chinese (r = .680, $p < .001$).

4.3 Time Spent

In the study, participants were asked to go through the learning material in the app from introductory material and a course tour to each of the eight courses. Then they needed to finish a quiz based on the materials learned in the app and complete a survey at the end. The time spent for each learning session as well as time took for quiz and survey were recorded manually using stopwatches by the researchers viewing the iPad app via a mirrored screen. Times were plotted for each session in a side-by-side bar chart below.

Total time

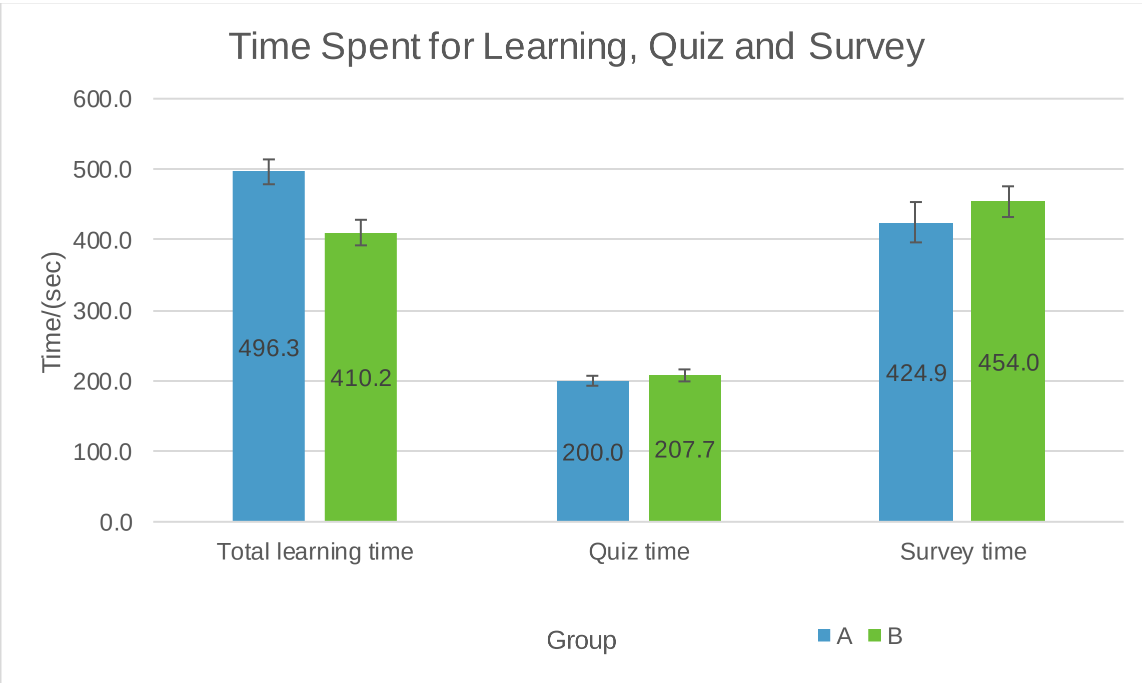


Figure 35 Time spent for learning, quiz and survey (error bars show standard error)

To compare the time spent between two groups, the mean of total learning time, quiz time and survey time was plotted in Figure 35. The total learning time starts from opening the app to end the learning courses. What shows in the figure was that the mean total learning time for Group A was longer than Group B while both of the quiz time and survey spent was shorter than that of Group B. Mean quiz time for Group A was 9.8 seconds shorter than Group B and mean survey time for Group A was 28.7 seconds shorter than that of Group B.

According to the independent-samples t-tests, differences of mean survey time and mean quiz time between two groups were not statistically significant. But mean total learning time of Group A (M=498.68, SD=115.49) was statistically greater than mean total learning time of Group

B (*M*=405.71, *SD*=118.69), a statistically significant difference, *M* = 86.10, 95% CI [43.00, 142.93], *t*(82.901) = 3.701, *p* = .000, *d* = .79. This mean difference of 86.10 seconds was an approximately 23% increase in time over the Group B average time.

Time spent per session

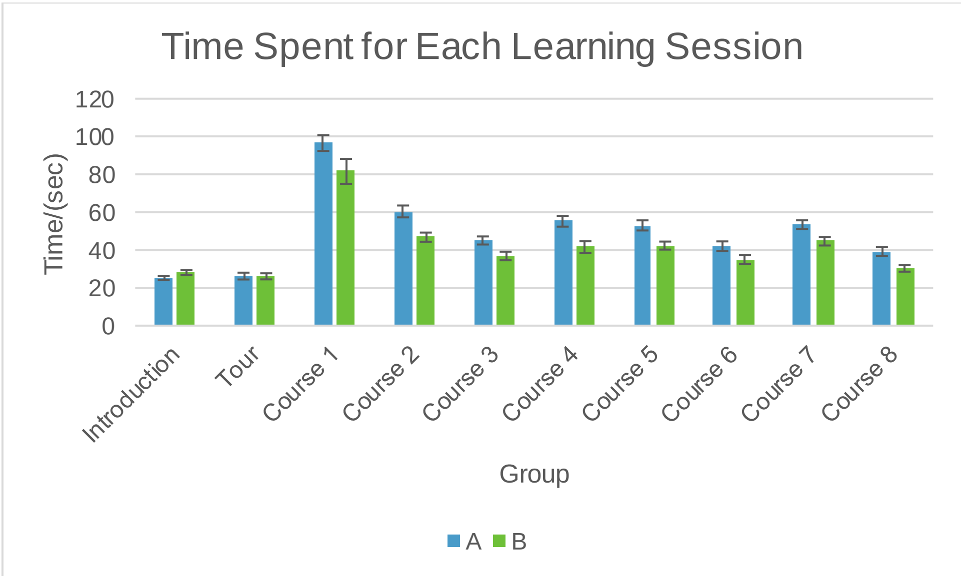


Figure 36 Time spent for each learning session (error bars show standard error)

To further to look at which part of learning session lead to longer time consumption for Group A than Group B, times per course were plotted. As we can see in the Figure 36, the mean time spent for the first two sessions, introduction and tour, were almost the same between two groups. The mean time spent for all the learning courses for Group A was higher than that of Group B, though statistical significance was not tested. It appears that the difference in the total time between Groups A and B was spread relative equally throughout the courses. Courses themselves varied in times because different courses varied slightly in the number of derivative characters that were taught.

4.4 NASA-TLX

Comparison of each measure by groups

In the survey, participants were asked to give ratings of NASA-TLX questions based on their learning experience. NASA-TLX was composed of six items: effort, frustration, mental, performance, physical, and temporal. They were all measured on a Likert scale from 0 to 10, as higher number represents higher workload. Performance was measured as the lower the more perfect. The following figure shows the comparison of boxplots of both groups across these six measures.

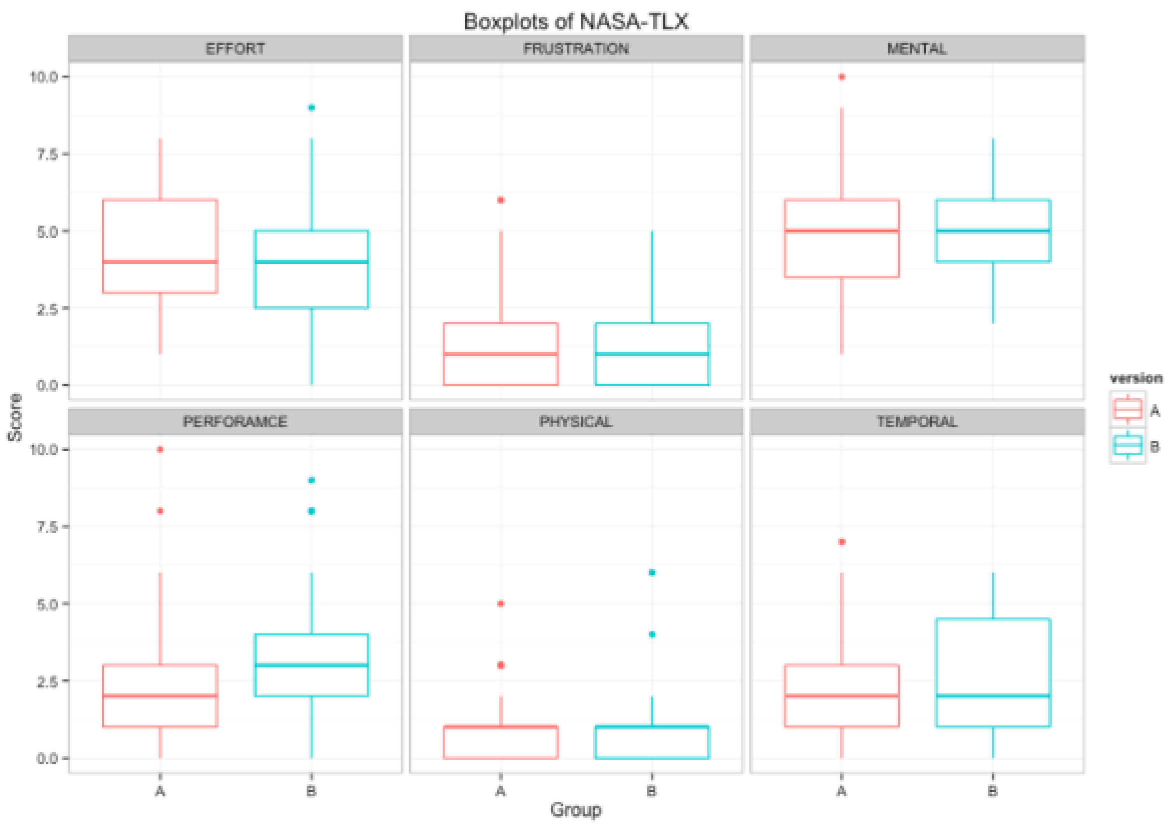


Figure 37 Boxplots of NASA-TLX

As shown in the boxplots (Figure 37), the distributions of frustration and physical demand for each group are similar. There were also not big median differences of effort, mental and temporal demand across two groups. But effort of Group A was slightly higher than Group B. Group A had bigger variances at effort and mental demand while less variance of temporal compared with Group B. the most obvious difference was that performance score in Group A was lower than Group B.

To further determine whether these measures were statistically significant by groups, independent-samples t-tests were conducted for each item. All of these measures have equal variance. The results shows participants in Group A (M=2.37, SD= 2.01) felt they had better performance than participants in Group B (M=3.49, SD=2.23), a statistically significant difference, M= -1.116, 95% CI[-2.027,-205], t(83.132)=-2.437, p= .017, d= .53. Other than performance score, there weren't statistically significance of mental (p=.478), physical (p=.990), temporal (p=.257), effort (p=.240) and frustration workload (p=.248) between groups.

4.5 Usability Questions

There were ten usability questions, which were based on the System Usability Scale (SUS) (Sauro, 2011). The scale in the current questions was a slider from 0-100. There were five positive questions and five negative questions in the initial survey. 1, 3, 5, 7 were positive, for which the higher the better while 2, 4, 6, 8, 10 were negative, the lower the better. In the side-by-side bar chart (Figure 38) below, all the negative questions were reversed into positive questions so that all the score for questions should be the higher the better. From the chart, in general we can see that scores of all the questions for Version A were higher than that of Version B.

Statistical test was done for further analysis.

An independent-samples t-test was run to determine if there were differences in these measurements between two groups. Over the pairs that are statistically significantly different, a bracket and a red asterisk * for $p < .05$ and ** for $p<.01$ were marked in Figure 38.

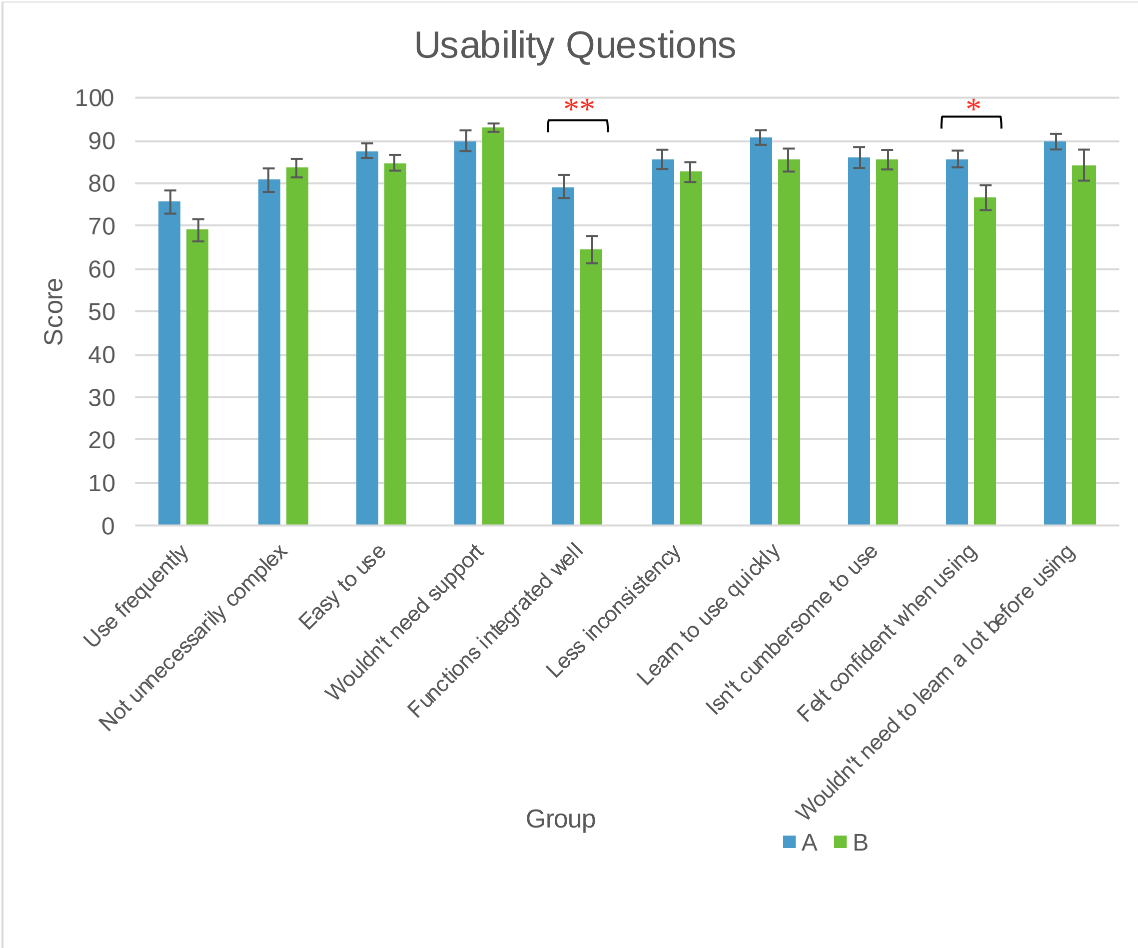


Figure 38 Bar chart of usability questions (error bars show standard error)

Participants in Group A (*M* = 85.7, *SD* = 12.8) were more confident than participants in Group B (*M* = 76.6, *SD* = 19.0), a statistically significant difference, *M* = 9.05, 95% CI [2.07, 16.02], *t*(73.607) = 2.584, *p* = .012, *d* = .56. In addition, the functions of Version A (*M* = 79.26, *SD* = 17.739) was integrated better than that of Version B (*M* = 64.47, *SD* = 20.982.0), a statistically significant difference, *M* = 14.791, 95% CI [6.455, 23.126], *t*(81.738) = 3.530, *p* = .001, *d* = .76. There wasn't statistical significance of “use frequently” (*p*=.082), “not unnecessarily complex” (*p*=.292), “easy to use” (*p*=.262), “wouldn’t need support” (*p*=.221), “less inconsistency” (*p*=.489), "learn to use quickly" (*p*=.102), "isn't cumbersome to use" (*p*=.876) and "wouldn't need to learn a lot before using" (*p*=.201) between groups.

4.6 Survey Questions

Other than usability questions that are used for global assessments of systems usability, there were also other survey questions more specific to this current system. These survey questions could be classified into several categories: overall satisfaction, usefulness, interface satisfaction and functionality. The scores for overall satisfaction, usefulness, interface, and functionality were all averages across two versions. The scale ranged from zero to one hundred, with zero being strongly disagree and one hundred being strongly agree.

Overall satisfaction

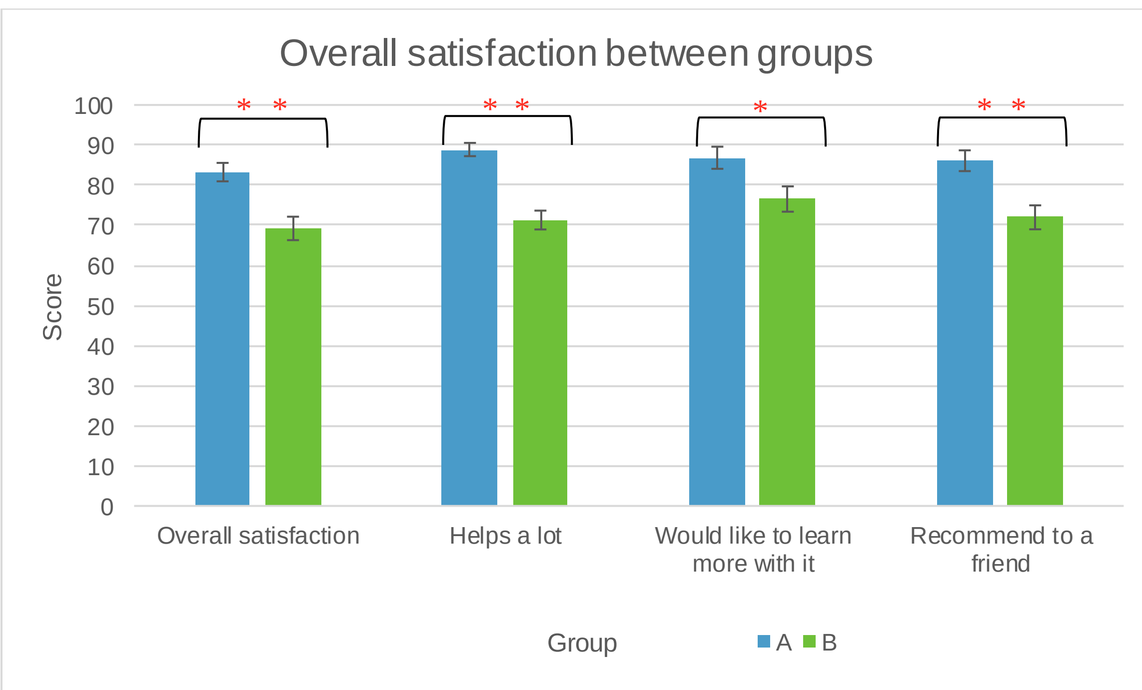


Figure 39 Bar chart of overall satisfaction (error bars show standard error)

As shown in Figure 39, first, participants in Group A (M = 83.26, SD = 15.079) had higher overall satisfaction than participants in Group B (M = 69.21, SD = 19.175), a statistically significant difference, *M* = 14.047, 95% CI [6.643, 21.450], t(79.576) = 3.776, *p* = .000, *d* = .81. In addition, participants indicated that Version A helped a lot more (*M* = 88.88, *SD* = 10.828) in tutoring Chinese than participants using Version B (*M* = 71.30, *SD* = 15.403), a statistically significant difference, *M* = 17.581, 95% CI [11.862, 23.301], t(75.363) = 3.530, *p* = .000, *d* = 1.32. Participants in Group A (*M* = 86.84, *SD* = 18.096) expressed a significantly higher willingness to learn more Chinese characters than participants in Group B (*M* = 76.53, *SD* = 20.713), a statistically significant difference, *M* = 10.302, 95% CI [1.959, 18.645], t(82.513) = 2.456, *p* = .016, *d* = .53. Finally, participants in Group A (*M* = 86.09, *SD* = 17.003) had a stronger desire to recommend this app to friends than participants in Group B (*M* = 71.98, *SD* = 15.595), *M* = 14.116, 95% CI [6.246, 21.986], t(82.363) = 3.586, *p* = .001, *d* = .86.

Moreover, Net Promoter Score (Garrity, 2010) was derived from the score of “would you like to recommend to a friend.” The net promoter score of Version A was 34.9% and the net promoter score of Version B was -11.6%. There were 29 “promoters” in Group A while 19 “promoters” in Group B. Fred calculated the average score of 400 companies across 28 industries and got the median Net Promoter Score was 16 (Reichheld, 2003). According to Net Promoter Network (Satmetrix Systems, 2015), the average score for software and app is 19.

Usefulness

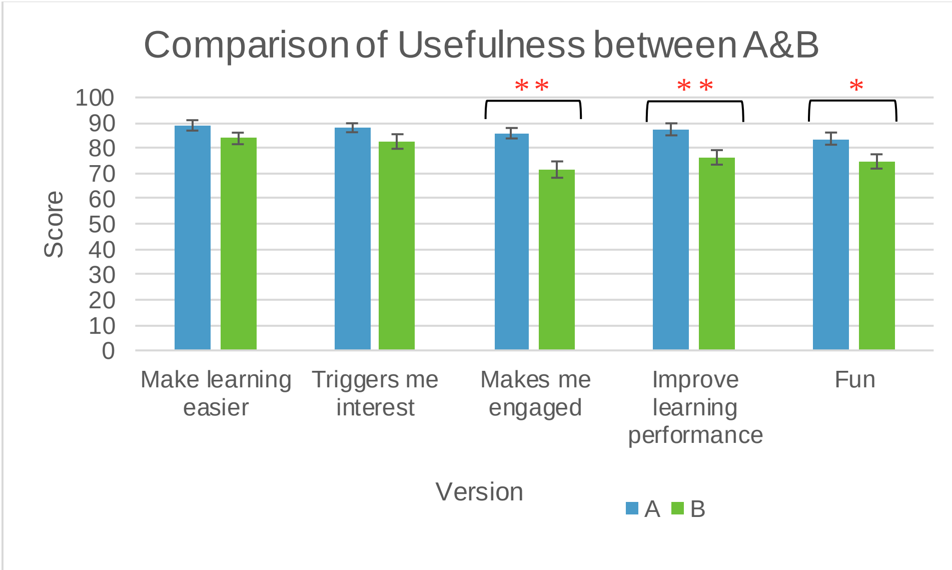


Figure 40 Comparison of usefulness by groups (error bars show standard error)

In the independent-samples t-test determining if there were differences in these measurements between two groups. The engagement, improvement and fun of Version A were statistically significant. A bracket and a red asterisk * for $p < .05$ and ** for $p < .01$ were marked over the pairs that are statistically significantly different in Figure 40 above.

Shown in Figure 40, Version A (M = 85.86, SD = 13.548) was more engaging than Version B (M = 71.49, SD = 21.199), a statistically significant difference, M = 14.372, 95% CI [6.723, 22.021], t(71.404) = 3.746, p = .000, d = .81. Participants in Group A (M = 87.40, SD = 15.625) thought they improved more learning performance than participants in Group B (M = 76.30, SD = 18.819), a statistically significant difference, M = 11.093, 95% CI [3.671, 18.515], t(81.253) = 2.974, p = .004, d = .63. Finally, Version A (M = 83.72, SD = 15.783) had more fun than Version B (M = 74.70, SD = 18.526), M = 9.023, 95% CI [1.640, 16.2407, t(81.932) = 2.431, p = .017, d = .52. There weren't statistically significance of "make learning easier" (p=.096) and "triggers interest" (p=.112) between groups.

Interface satisfaction

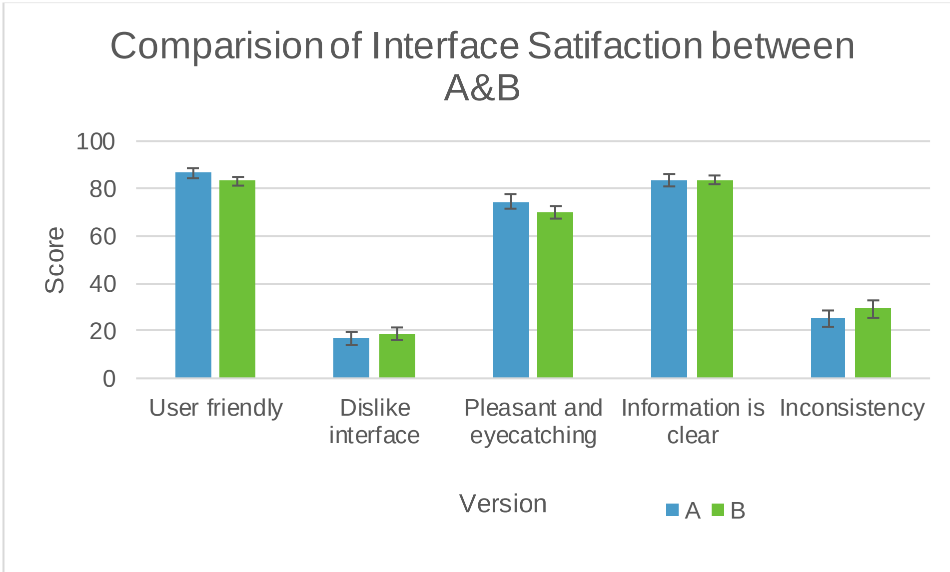


Figure 41 Comparison of interface satisfaction by groups (error bars show standard error)

As shown in

Figure 41, Version A was a slightly higher at user friendly score, pleasant and eye catching score compared with Version B. Version A was slightly lower at score of dislike of interface and inconsistency. But there wasn't statistically significance of constructs “user friendly” (*p*=.233), “dislike interface” (*p*=.607), “pleasant and eye catching” (*p*=.254), “information is clear” (*p*=.965) and “inconsistency” (*p*=.433) between groups.

Functionality

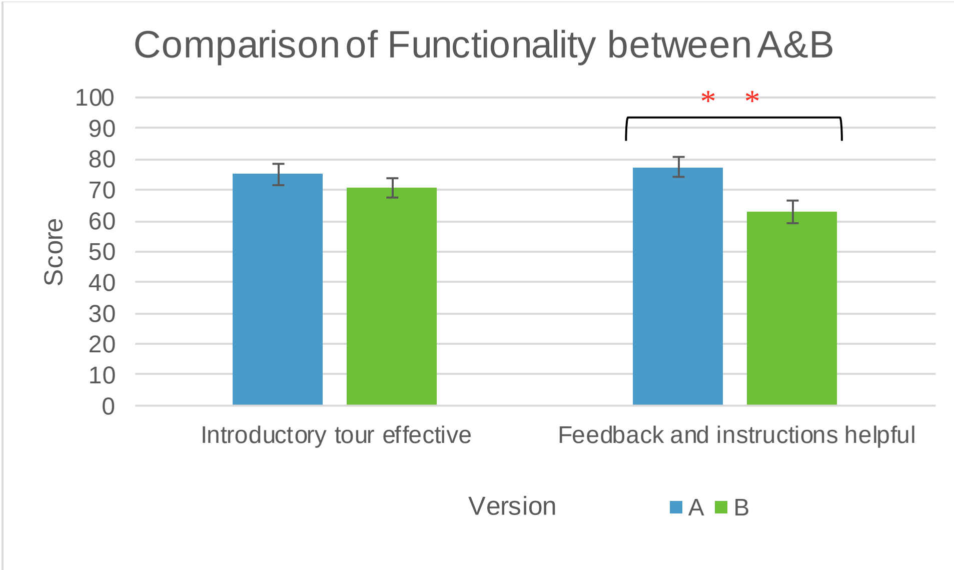


Figure 42 Comparison of functionality by groups (error bars show standard error)

As shown in the

Figure 42, Version A was both higher than Version B at effective introductory tour and helpful feedback and instructions. According to statistical t-test, the feedback and instructions of Version A (*M* = 77.47, *SD* = 21.207) were rated significantly more helpful than that of Version B (*M* = 62.86, *SD* = 24.123), evidenced by *M* = 14.605, 95% CI [4.862, 24.348], t(82.643) = 2.982, *p* = .004, *d* = .63. But there was no significant difference of the introductory tour (*p*=.354).

4.7 Before and After Comparison

Comparison of changes of attitudes between groups

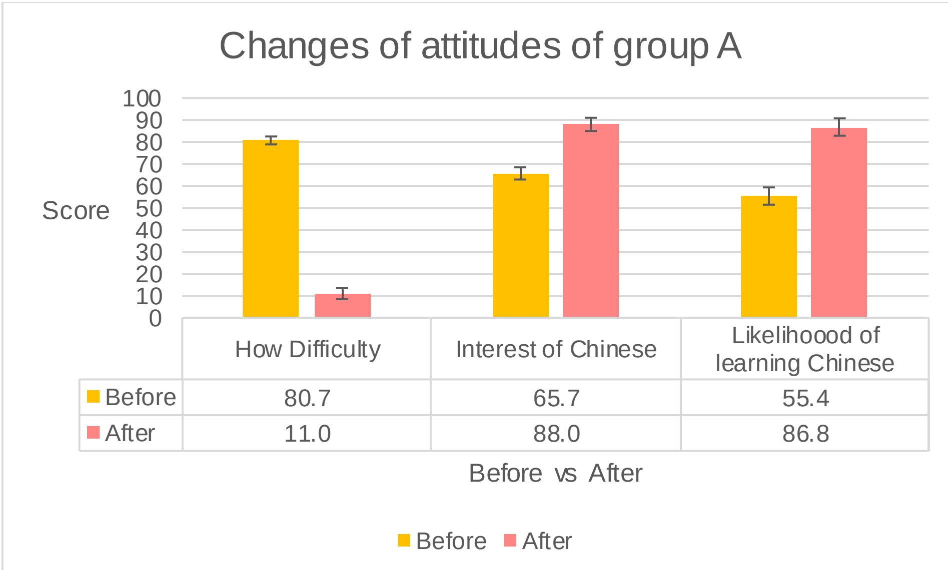


Figure 43 Changes of attitudes of Group A

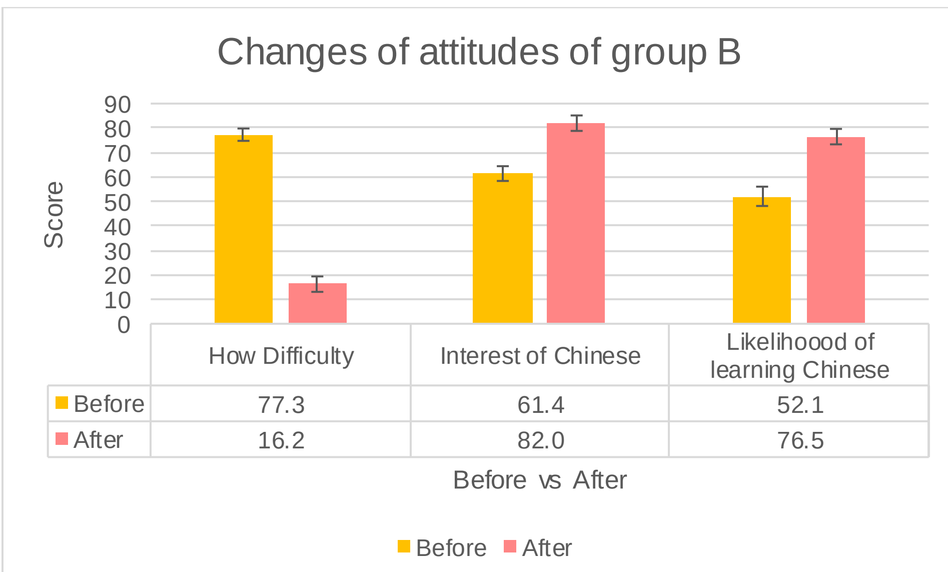


Figure 44 Changes of attitudes of Group B

In the pre-survey and the post-survey participants were asked several questions related to their attitudes about Chinese. Although the questions in pre- and post-surveys were not identical, we compared them as a measure of these attitude constructs. The questions are shown in (Table 20).

Table 20 Comparison of questions asked in pre- and post-survey

Construct	Pre-survey	Post-survey
Difficulty	How difficult do you think the Chinese characters are?	It makes learning Chinese much easier than I expected.
Interest	What's the level of interest you have in the Chinese language?	It triggers my interest in learning more Chinese.
Likelihood	If you were about to learn a second language, how likely is it that you would learn Chinese?	I would like to learn more Chinese using this system.

Thus, the changes of their attitudes were compared between two groups. Figure 43 and Figure 44 show the changes of attitudes from Group A and Group B respectively. The changes compared were how much difficulty decreased, how much interest increased and how much increase of likelihood of learning Chinese considered by participants in two groups. In order to compare the attitudes changes for both groups. A side-by-side bar chart showing attitude changes by groups was plotted and an independent-samples t-test was run to determine if there were significant difference between groups. Shown in the Figure 45, difficulty of Chinese challenging to Group A was sharply decreased by 69.6 on average while that of Group B was also largely decreased by 61.1. Both of the participants in two groups has an increase in their interest to learn more Chinese and likelihood of learning Chinese, but Group A had higher increase than Group B. The mean increase of interest was 22.4 for Group A and that for Group B was 21.2. The mean increase of likelihood of learning Chinese for Group A was 31.4 while that of Group B was 24.4.

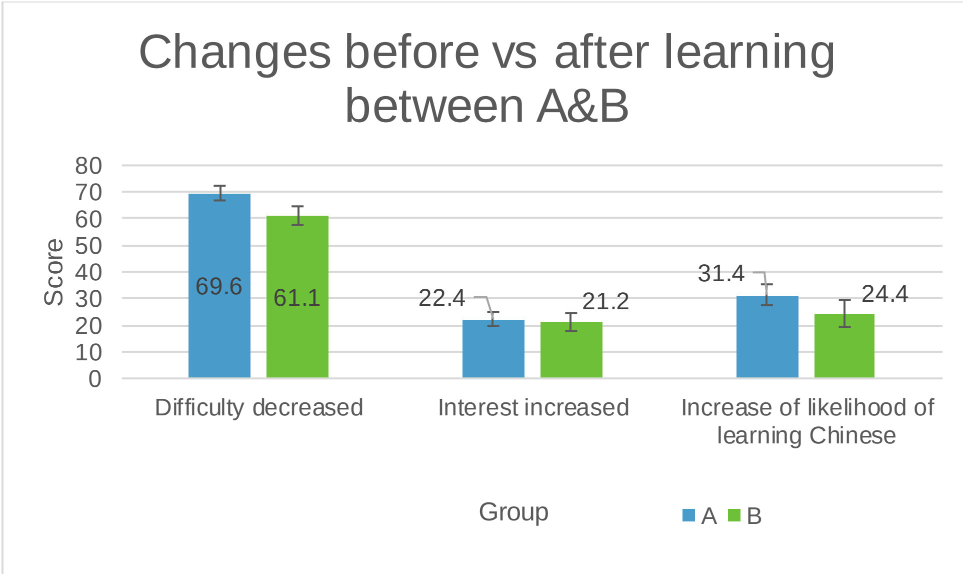


Figure 45 Changes before vs. after learning by groups (error bars show standard error)

The scores of difficulty decreased, interest increased and likelihood of learning Chinese increased for each level of group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). But there weren't statistically significance at decrease of difficulty (p=.061), interest increase (p=.777) and increase of likelihood of learning Chinese (p=.280) between groups according to the results of the independent-samples t-test.

Comparison of changes of attitudes within subjects

When comparing within subjects about their attitudes before and after using the system, although the differences were not significant per above, it is valuable to chart these differences to demonstrate that there is a noticeable of differences among individuals. Their agreement on how much easier it made learning Chinese was different. Figure 46 shows the difficulty decreases for each participant by group, ordered from smallest difference to largest difference. Almost all participants in Group A thought the degree of the app making learning Chinese easier was higher than that of participants in Group B.

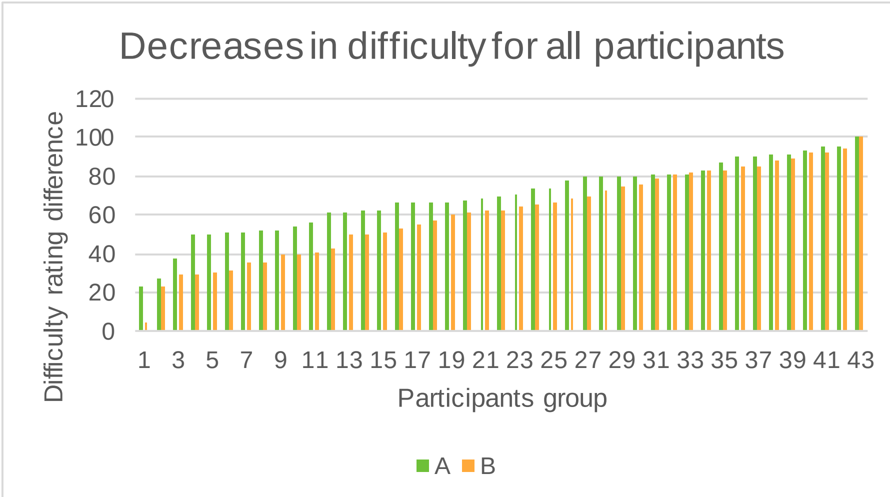


Figure 46 Decreases in difficulty for all participants, sorted from lowest to highest by group; this graph illustrates that differences vary notably by participant

There were some participants who had lower interest and likelihood of learning Chinese after they using the system. Figure 47 shows the interest increase for each participant. As noticed, there were ten participants had lower interest to Chinese compared with before using the app. Five were from Group A (represented as green bars) and five were from Group B (represented as orange bars). Among these negative increases of interest, Group B had three participants' interests dropped largely of 15, 17 and 20 compared with their pre-survey. Group A also had one participant's interest decreased largely by 10.

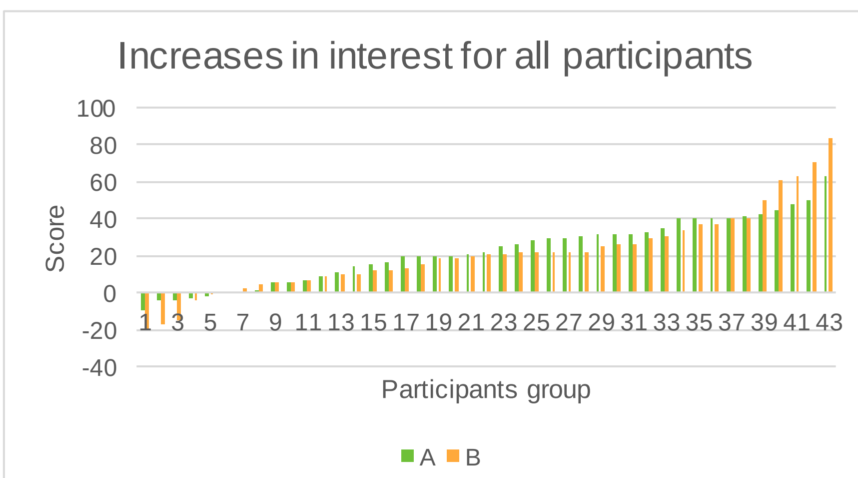


Figure 47 Increases in interest for all participants

Same as plotting interests increased for all the participants, a figure of increases likelihood was also plotted (see Figure 48). There were 7 participants in Group B had less likelihood of learning Chinese after using the app while there were four participants in Group A had less likelihood of learning Chinese after using the app. The biggest decrease was -67 who was in Group A. In general, the decrease of likelihood of learning Chinese after using the app of Group B was more obvious.

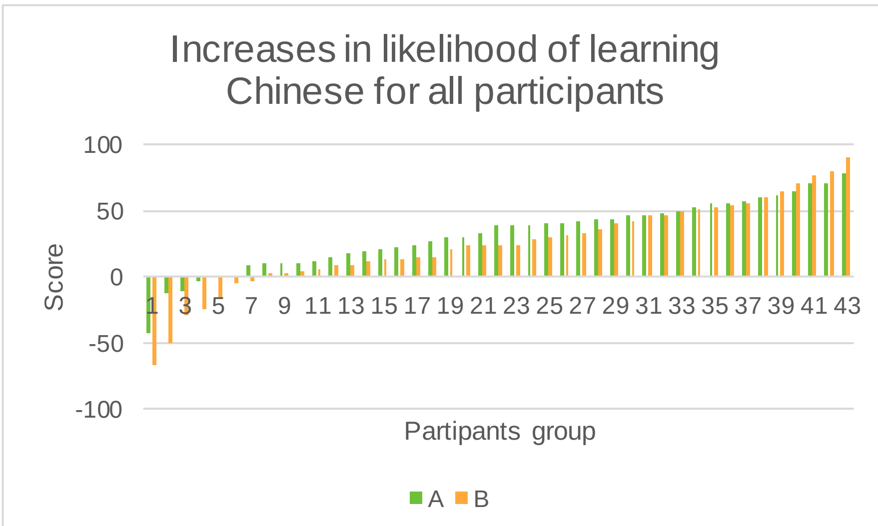


Figure 48 Increases in likelihood of learning Chinese for all participants

CHAPTER 5 DISCUSSION AND CONCLUSION

The purposes of this study were 1) to explore the initial feasibility of an Intelligent Language Tutoring interface for learning Chinese characters to teach beginning Chinese-as-a-second-language learners and 2) to assess the system by investigating learning effectiveness, usability issues, and users' attitudes towards the system. This chapter will discuss the conclusions drawn from the results presented in the previous chapter and how they relate to the research questions.

5.1 Predicted Outcomes

The comparison was made between two versions of the system. Version A was a “richer” version because it had metaphor pedagogy and feedback, and Version B was a “plain” version that did not have the features aforementioned. It was expected that the participants who used Version A would have better results than participants who used Version B. The results would be seen in the following ways:

1. Better learning performance;
2. More interest of learning Chinese;
3. Lower task load when learning;
4. Better usability assessment of interface.

More specially, this research attempted to answer the following questions:

1. Is there a difference between students who use the system with metaphor pedagogy and the students using the system without it?
2. Will beginning Chinese-as-a-second-language learners’ interest be increased after using the system?

3. Will the interface be user-friendly and are there any usability issues?
4. Will users using the system with the metaphor pedagogy achieve better performances (higher scores in the quiz), more interest increased, better usability assessment?

5.2 Discussion of Findings

Learning performance

In the data presented in the previous chapter, the overall learning performance (measured by quiz score) of participants in Group A was significantly better than learning performance of participants in Group B. The mean score difference between two groups was 2.279. Additionally, the large *F* value generated from independent samples t-test indicates that the variance of learning performance of Group B was significant bigger than learning performance of Group A, as can also be seen from the boxplots of both groups. Thus, we could safely conclude that participants who used Version A had better learning performance and more stable performance than participants who used Version B.

Other than the overall learning performance compared between two groups, there were also the questions focused on assessing participants' ability to make inferences about new Chinese characters they had never seen based on the knowledge they gained ("guess questions"). The results of the experiment showed that the correct rate on these questions of participants in Group A was significantly higher than the correct rate of participants in Group B. Therefore, we can conclude that the metaphor pedagogy, rationales and feedback provided in Version A were helpful for giving learners at least a basic underlying conceptual understanding of how Chinese characters are formed so that they could make inferences about new Chinese characters.

Based on the large variance of learning performance of Version B, further exploration of what might lead to higher scores and lower scores was made. The comparison between the 1st quartile (represented as lower scorers) and the 4th quartile (represented as higher scorers) was made to figure out what factors lead to these differences. But neither gender, academic role, nor number of languages learned were shown to be factors that make significant difference between higher scorers and lower scorers.

Lastly, we explored whether participants' attitudes toward Chinese would have influence on their learning performance. There were no statistically significant correlations of their attitudes toward their learning performance. The item that was significant correlated was the strong positive correlation between the interest of learning Chinese and the likelihood of learning Chinese. Thus, if this current work proved to increase learners' interest, they will be more likely to learn Chinese as a foreign language.

Time spent

The time spent learning, completing the quiz and survey was recorded and was compared across the two groups. As results showed, there was a significant difference between learning time used by Group A and learning time used by Group B. The mean learning time spent for Group A was 92.97 seconds longer than the mean learning time spent for Group B. On average participants in Group A spent 11.6 seconds more per learning course than participants in Group B. It was not surprising since there were more images and explanations with users of Version A than Version B. Thus, it took longer for users to learn the materials.

This result raises the issue of whether the high performance gains of Group A resulted simply from having more time on task with the learning material, rather than from the metaphor pedagogy itself. Group A did spend approximately 23% (93 sec) more time with the material than Group B. However, given the significantly higher ratings in Group A's usability ratings of the app (see below), we suggest that the cause for higher performance is not as simple as purely time on task; the content included in Version A made its users rank the experience more highly, which indicates that it had a specific effect on learners that is based on the content itself, not just the time it required.

Although learning time spent on Version A was longer than time spent on Version B, the time used for quiz and survey of Version B was longer (though not statistically significantly so). If this difference became significant in a future study with additional participants, a reasonable inference would be that participants in Group B needed more time to recall the learning materials they just learned. Since Version B lacked the metaphor pedagogy and rationales instilling the concept of building structures of Chinese characters, the most easily available method of learning the characters in Version B would be by rote memorizing, which didn't last as long, but required more time for recall. Also, a possible reason of more survey time for Version B would be that participants noticed some usability issues of Version B, so that they took more time to report them in the open-ended questions. The time taken for open-ended questions was not recorded; thus, future work can explore this issue to determine whether participants in Version B took a longer time in the open-ended questions than participants in Version A. But there were no statistical differences between quiz times and survey times in the two groups. At the same time, the time recorded may not be as accurate as desired since it was manually recorded, given observer differences in reaction times.

A final learning result worth mentioning is that within 10 minutes, more than half of the total participants (47) scored more than 90% correct on the quiz. There were 34 characters and phrases in total, which indicates more than half of them learned more than 30 characters and phrases within 10 minutes. Future work would explore the duration of this learning effect. Would participants still remember those characters after two weeks, for example?

Task load

Since the system was designed to help learners to learn easily with less effort, the task load was expected to be lower for participants in Group A than participants in Group B. The task load was measured by the NASA-TLX survey, in which there are six indices to assess workload of humans in terms of mental, physical, temporal, performance, effort, and frustration level. However, the task didn't require physical demand or time pressure. Thus, the result of physical or temporal demand was not the focus of interest. According to the analysis, performance (How successful were you at accomplishing the task) reported by participants in Group A was significantly higher than performance of participants in Group B, indicating that Version A gave users more confidence of how successful they were at the task than Version B did. There weren't significant differences for the other NASA-TLX indices.

Survey questions

The ten questions of the System Usability Scale (SUS) and sixteen other questions more specific to the system were used to assess the system effectiveness and usability. From the rating responses of these questions, an evaluation was done to determine which aspects of Version A were reported better or worse than Version B. There were two items for which Version B was rated better than Version A, but neither of them was significantly different. The following Table 21 shows the items for which Version A was significantly better than Version B.

Table 21 Significantly better features of Version A than Version B

Categories	Attributes	Mean difference, *p*-value
Usability	I found the various functions in this system were well integrated.	*M*=14.791, *p*=.001, *d*= .76
	I felt very confident using the system.	*M*=9.05, *p*=.012, *d*= .56
Overall satisfaction	Overall, I am satisfied with this learning system.	*M* = 14.047, *p* = .000, *d* = .81
	It helps me a lot to understand the learning content.	*M* = 17.581, *p* = .000, *d* = 1.32
	I would like to learn more Chinese using this system.	*M* = 10.302 , *p* = .016, *d* = .53
	I would recommend it to a friend.	*M* = 14.116, *p* = .001, *d* = .86
Usefulness	The system makes me engaged in the learning content.	*M* = 14.372, *p* = .000, *d* = .81
	Using the system would improve my learning performance.	*M* = 11.093, *p* = .004, *d* = .63
	It makes learning fun.	M = 9.023, p = .017, d = .52
Functionality	I found the feedback and instructions very helpful.	*M* = 14.605, *p* = .004, *d* = .63

According to the results in the table above, 2 out of 10 usability survey items, 4 out of 4 overall satisfactions, 3 out of 5 usefulness, 1 out of 2 functionalities, and 0 out of 5 interface satisfaction features of Version A were significantly better than those of Version B. The other survey items were not significantly different. A conclusion can be drawn that the overall satisfaction of Version A was better Version B. From a usability perspective, the functions were better integrated in Version A than Version B. Participants in Group A felt more confident using the system than participants in Group B. At the same time, Version A may be more useful than Version B since it's more engaging, more fun and improves learning performance more effectively. Furthermore, the Net Promoter Score was calculated based on the question "would you like to recommend to a friend." The Net Promoter Score for Version A was 34.9 and for Version B was -

11.6. Since the industry average of Net Promoter Scores for software is 19 ("Net Promoter System," 2013), Version A has a score notably higher than the average. Finally, since Version A has feedback and instructions but Version B hasn't, it showed that this form of feedback and instructions could be very useful during learning.

Discussion of responses to open-ended questions

Other than the aforementioned measurements used to assess the feasibility and usability of the system, there were two open-ended questions that asked participants to provide three things they liked and disliked about the system. One participant didn't give feedback in the open-ended questions. Two participants didn't give responses on things they disliked. All remaining comments were classified into categories based on the themes that emerged. The following Table 22 shows the positive comments sorted from most frequently mentioned to least frequently mentioned for both versions. Comments mentioned at least twice were shown in the table.

As seen in the table, the most commonly mentioned advantages of Version A were its pictures, explanations of how Chinese characters were involved/build together, ease of use, and the mini quizzes that can test their knowledge. It is not hard to notice that the good comments of Version A were emphasis of metaphors, how characters were explained, which were the added functionality that Version B didn't have, while the good comments of Version B focused on the good usability and clear structure of the system.

Table 22 Positive comments for each version

Version A			**Version B**		
Comments	Count	Rate	Comments	Count	Rate
Pictures help to visualize the characters	14	33.3%	Ease of use/simple	17	40.5%
Information integrated well/explain well	10	23.8%	Idea of how characters are build/relate together	11	26.2%
Explain how characters evolved	9	21.4%	Color/contrast	10	23.8%
How the characters build together	7	16.7%	The flow of learning/left menu to control progress	8	19.0%
Ease of use	7	16.7%	Mini quizzes	7	16.7%
Quizzes to test knowledge	7	16.7%	Navigation/layout	6	14.3%
Easy/quick to understand	7	16.7%	Learning flow/step by step	6	14.3%
Step by step instruction/left menu	7	16.7%	Engaged/fun	5	11.9%
Colors / Look / Art / Graphics / Animation	6	14.3%	Interface	4	9.5%
User-friendly and comfortable to use	6	14.3%	Comparison to other characters/ make them focus	3	7.1%
panda tips	5	11.9%	design is appealing/simple design to focus	3	7.1%
keeps you engaged / Interesting / Fun	4	9.5%	clues help remember	2	4.8%
interface good	4	9.5%	guess questions	2	4.8%
feedback of the quiz	3	7.1%	User Friendly	2	4.8%
rationale/logical	2	4.8%	structure is nice/organization	2	4.8%

Thus, a conclusion can be drawn that the extra explanations/metaphors of Version A helped users' learning and understanding, but they distracted the focus of the users from the flow and structure of the whole system. On the contrary, the clearer version (Version B) made participants feel it was easier to use and gave them control of their learning progress to learn on their own pace.

The following Table 23 shows the negative comments from the most commonly mentioned to least commonly mentioned across two versions. At least twice mentioned comments were shown in the table.

Table 23 Negative comments on each version

Version A			**Version B**		
Comments	Count	Rate	Comments	Count	Rate
Interface	9	22.0%	Quiz feedback is needed	21	48.8%
Too repetitive, more surprise, in depth questions are needed	7	17.1%	More explaining	9	20.9%
More useful characters	6	14.6%	Add rationales/logic	8	18.6%
Transition between courses confusing	4	9.8%	Too repetitive, more surprise, in depth questions are needed	8	18.6%
Write characters on iPad	3	7.3%	Add audio	6	14.0%
Add audio	3	7.3%	Appearance/color/add pictures	5	11.6%
Reevaluate if they get wrong	3	7.3%	Write characters on iPad	5	11.6%
More color, cooler animations	3	7.3%	Add more characters to learn	4	9.3%
Bugs/Errors	3	7.3%	Back button to review	4	9.3%
Mini quizzes	2	4.9%	More useful characters	3	7.0%

As can be seen in Table 23, the most frequently noted drawback was the interface of Version A, while the most frequently mentioned drawback of Version B was the lack of feedback of mini quizzes. This is another result that illustrates that the feedback is important for learning. In addition, 18.6% of participants who used Version B asked to add more explanation and rationales to explain characters. We can conclude that the metaphors and explanations in Version A were helpful and thus were needed by at least a certain percentage of participants. Participants in both groups thought the material was too simple and repetitive, so that they wanted more characters to be learned and those guess questions were complimented. Other common disadvantages of both versions were that participants want to add playable audio files to the characters when they went to higher levels, and they wanted to write on the iPad to learn writing.

5.3 Conclusions

In a conclusion, the results shown were almost consistent with the predicted outcomes. Metaphor pedagogy and quiz feedback of Version did help learning and give better user experience to users. Version A won Version B at the following advantages.

First, per the post-survey, information was integrated better in Version A, which was both evidenced by the statistical tests on the dependent variables and participants' personal comments in surveys. Then, the general assessment of Version A was better than B, reflected by not only participants providing higher satisfaction and thinking they understood better using the system, but also that participants in Group A had statistically higher willingness to learn more Chinese using the system and recommend to a friend. At the same time, Version A was more useful than Version B since users of Group A reported higher learning improvement and thought Version A was more engaging and fun. This was also supported in the user comments. The major benefits of Version A were its pictures and metaphors, which made participants engaged in the learning session. Lastly, the feedback of mini quizzes was crucial for learning. Both statistical testing and user comments all stated the importance of feedback.

But improvements were still needed for Version A. The user interface needed to be more user friendly. The design of the later ILTS would also need to create more in depth questions and learning material (but with questions not too difficult) to help reinforce knowledge gained. More design suggestions will be discussed in future work.

5.4 Limitations

There were some limitations of this study. First, as noted in the test, individual differences in the task involving learning were very obvious. At the same time, the attitudes of learners and how much effort they applied when doing the task were also factors that influenced learning behaviors. Meanwhile, all the survey measurements were collected using self-report survey, which can be doubted more than quantitative performance measures. Lastly, the time spent on each course within the app was manually recorded by a live observer nearby, which could have led to errors. Based on the limitations mentioned above, the improvements of both experiment and system will be discussed in the future work.

5.5 Future Work

Experimental design

The improvements regarding the experiment will be made as following in future iterations.

- A between-subject experimental design was selected to avoid the learning effect in the study. As noted previously, individual differences were a noteworthy issue. In future work, a counterbalanced within-subject design could be considered and could ask participants to learn two versions with different learning content. Then their learning performance based on two learning materials could be compared.
- Versions A and B could be more carefully calibrated to take the same amount of learning time, to avoid concerns about performance increases based on time on task.
- To test whether the learning gains of Version A are longer lasting, a quiz to test participants' knowledge could be added after a period of time (e.g., half a day, several days).

- As we can see in the results showing whether demographics influence learning performance, there was some imbalance of number of people across the higher score group and lower score group. But there wasn't a significant difference according to the statistical testing. A larger sample size would likely resolve this issue.
- The iPad app should have a time recording feature. In the current study, time spent was recorded manually that was not reliable and accurate. Thus, having a time recording will capture time spent accurately when participants click on a certain button.

System design

Based on the participants' suggestions and survey results reported by participants, the following improvements regarding the system design will be considered in future work.

- The interface of the system: Although results showed that the information of the system was integrated well, the interface was still not appealing enough. Some participants suggested adding more color and animations.
- Functionalities such as adding audio and drawing on the iPad need to be explored. Letting participants practice writing Chinese characters on the iPad will teach them how to write while learning to recognize them. Although based on the results in previous chapters, the visual appeal of the Chinese characters was an appropriate way to teach Chinese-as-foreign-language learners. But some participants wanted to know how to pronounce the characters when they see them. Thus, the sound of Chinese characters can be added when learners enter a higher level of learning.
- The system needs to be more intelligent. As in an intelligent tutoring system, customized feedback could be offered based on learners' current state as processed by its learner module. However, the feedback this system provided was based on whether

users had the correct answer or not. This type of feedback turned out to be helpful, but there were participants who thought the quizzes were too simple and redundant. Some participants also requested more characters to learn, while others were confused about the transitions between courses. In future work, a system based on the learner's current state could include more intelligent features such as content related to what they have learned already and adaptive quizzes. Also, the system should record learners' current learning progress so that they can always come back to continue what they were learning.

REFERENCES

Ahuja, N. J., & Sille, R. (2013). A Critical Review of Development of Intelligent Tutoring Systems: Retrospect, Present and Prospect. *International Journal of Computer Science Issues*, *10*(4), 39–48.

Al-Mekhlafi, K., Hu, X., & Zheng, Z. (2009). An Approach to Context-Aware Mobile Chinese Language Learning for Foreign Students. *2009 Eighth International Conference on Mobile Business*, (Cll), 340–346. http://doi.org/10.1109/ICMB.2009.65

Aldrich, C. (2005). *Learning by Doing: A Comprehensive Guide to Simulations, Computer Games and Pedagogy in E-learning and Other Educational Experiences*. San Francisco: John Wiley & Sons, Inc.

Amaral, L. A. M. do. (2007). *Designing Intelligent Language Tutoring Systems for Integration into Foreign Language Instruction*. Ohio State University.

Baker, R. D., Corbett, A., Koedinger, K. R., Evenson, S., Roll, I., Wagner, A. Z., … Beck, J. E. (2006). Adapting to when Students Game an Intelligent Tutoring System. *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, 392–401. http://doi.org/10.1007/11774303_39

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, *24*(June 2015), 574–594. http://doi.org/10.1080/10447310802205776

Beal, C. R., Cohen, P. R., & Woolf, B. P. (2010). Evaluation of AnimalWatch: An intelligent tutoring system for arithmetic and fractions. *Journal of Interactive Online Learning*, *9*(1), 64–77.

Blackmon, M. H., Polson, P. G., Kitajima, M., & Lewis, C. (2002). Cognitive Walkthrough for the Web. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves*, (1), 463–470. http://doi.org/10.1145/503376.503459

Buckingham, D. (2007). *Beyond Technology: Children's Learning in the Age of Digital Culture*. Cambridge: Polity Press. http://doi.org/10.1111/j.1099-0860.2007.00135.x

Burns, H. L., & Capps, C. G. (1988). Foundations of Intelligent Tutoring Systems: An Introduction. In *Foundations of Intelligent Tutoring Systems* (pp. 1–18).

Census Bureau. (2013). American Community Survey. Retrieved from http://www.census.gov/acs/www/

Chen, H., Hsu, C., Chang, L., Lin, Y., Chang, K., & Sung, Y. (2013). Using a Radical-Derived Character E-Learning Platform to Increase Learner Knowledge of Chinese Characters. *Language Learning & Technology*, *17*(1), 89–106.

Chen, Hsuan Chih, O. T. (1992). *Language Processing in Chinese*. Amsterdam: Elsevier Science B.V.

Chinese Ministry of Education. (2014). Retrieved from http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/moe_2812/index.html

Chu-chang, M., & Loritz, D. J. (1977). Phonological Encoding of Chinese Ideographs in Short-Term Memory. *Language Learning*, *27*(2), 341–348.

Chung, K. K. H. Effects of Pinyin and First Language Words in Learning of Chinese Characters as a Second Language, 12 Journal of Behavioral Education 207–223–223 (2003). http://doi.org/10.1023/A:1025560327860

Deaton, J. E., Barba, C., Santarelli, T., Rosenzweig, L., Souders, V., McCollum, C., … Singer, M. J. (2005). Virtual Environment Cultural Training for Operational Readiness (VECTOR). *Virtual Reality*, *8*(3), 156–167. http://doi.org/10.1007/s10055-004-0145-x

DeFrancis, J. (1984). *The Chinese Language: Fact and Fantasy*. Honolulu, HI: University of Hawaii Press.

Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2007). *Human-Computer Interaction* (3rd ed.). Harlow: Prentice Hall.

Gamper, J., & Knapp, J. (2002). A Review of Intelligent CALL Systems. *Computer Assisted Language Learning*, *15*(4), 329–342.

Garrett, J. J. (2003). *The Elements of User Experience: User-Centered Design for the Web* (2nd ed.). New York, NY: New Riders.

Garrity, J. (2010). Drive Customer Loyalty with Net Promoter Score. *ABA Bank Marketing*, *42*(9), 30–33. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=55381407&lang=de&site=ehost-live

Gitterman, M. R., & Sies, L. F. (1994). Psycholinguistic Implications for Linguistic Relativity: A Case Study of Chinese. *Journal of Neurolinguistics*, *8*(2), 157–161.

Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W., & Harter, D. (2001). Intelligent Tutoring Systems with Conversational Dialogue. *AL Magazine*, *22*(4), 39–51.

Hagge, M., Amin-Naseri, M., Guo, E., Gilbert, S., Jackman, J., Starns, G., & Faidly, L. (2015). Decision Based Learning for a Sophomore Level Thermodynamics Course. *122nd ASEE Annual Conference*.

Hart, R. S. (1995). The Illinois Plato Foreign Languages Project. *CALICO Journal*, *12*(4), 15–37.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. http://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, *52*(C), 139–183. http://doi.org/10.1016/S0166-4115(08)62386-9

Heift, T. (2001). Intelligent Language Tutoring Systems for Grammar Practice. *Zeitschrift Für Interkulturellen Fremdsprachenunterricht*, *6*(2), 15.

Heift, T. (2003). Multiple Learner Errors and Meaningful ICALL Systems. *CALICO Journal*, *20*(3), 533–548. http://doi.org/10.1558/cj.v20i3.533-548

Heift, T. (2010). Developing an Intelligent Language Tutor. *CALICO Journal*, *27*(3), 443–459.

Ho, C. S.-H., Ny, T.-T., & Ny, W.-K. (2003). A “Radical” Approach to Reading Development in Chinese: The Role of Semantic Radicals and Phonetic Radicals. *Journal of Literacy Research*, *35*(3), 849–878. http://doi.org/10.1207/s15548430jlr3503

Hor, S. (1991). *Etymology-Based Instructional System: Using Hypermedia to Teach the Recognition of Chinese Characters to Beginning Learners of Chinese as a Second Language*. Retrieved from ProQuest Digital Disserations. (Paper 9534).

Hsiao, H.-S., Chang, C.-S., Chen, C.-J., Wu, C.-H., & Lin, C.-Y. (2015). The Influence of Chinese Character Handwriting Diagnosis and Remedial Instruction System on Learners of Chinese as a Foreign Language. *Computer Assisted Language Learning*, *28*(4), 306–324. http://doi.org/10.1080/09588221.2013.818562

Hsu, J. F. (2012). *Learning Chinese Characters: A Comparative Study of the Learning Strategies of Western Students and Eastern Asian Students in Taiwan*. Retrived from ProQuest Dissertations and Theses.

Hsueh, S. (2014). *Chineasy: The New Way to Read Chinese*. New York, NY: Harper Design.

Huang, J. S., & Ma, M. Y. (2007). A Study on the Cognitive of Complexity and Difficulty of Chinese Characters when Reading and Recognizing. *Displays*, *28*(1), 8–25. http://doi.org/10.1016/j.displa.2006.11.002

Hue, C. W., & Erickson, J. R. (1988). Short-Term Memory for Chinese Characters and Radicals. *Memory & Cognition*, *16*(3), 196–205. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3393080

Hung, D. L., & Tzeng, O. J. (1981). Orthographic Variations and Visual Information Processing. *Psychological Bulletin*, *90*(3), 377–414. http://doi.org/10.1037/0033-2909.90.3.377

Ji, J., Yu, H., Li, B., & Shen, Z. (2013). Learning Chinese Characters with Gestures. *International Journal of Information Technology*, *19*(1), 1–10.

Johnson, W. L., Marsella, S., & Rey, M. (2004). Tactical Language Training System: Supporting the Rapid Acquisition of Foreign Language and Cultural Skills. In *Proceedings of InSTIL/ICALL2004-NLP and Speech Technologies in Advanced Language Learning Systems*. Venice.

Karttunen, F. E., & Crosby, A. W. (2006). Empires of the Word: A Language History of the World (Review). *Journal of World History*, *17*(2), 225–228. http://doi.org/10.1353/jwh.2006.0039

Kosek, M., & Lison, P. (2014). An Intelligent Tutoring System for Learning Chinese with a Cognitive Model of the Learner. In *EUROCALL 2014-CALL Design: Principles and Practice*.

Lam, H. C., Ki, W. W., Law, N., Chung, A. L. S., Ko, P. Y., Ho, A. H. S., & Pun, S. W. (2001). Designing CALL for Learning Chinese Characters. *Journal of Computer Assisted Learning*, *17*, 115–128.

Lane, H. C., Core, M. G., Gomboc, D., Karnavat, A., & Rosenberg, M. (2007). Intelligent Tutoring for Interpersonal and Intercultural Skills. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*.

Leong, C. K. (1989). Reading and Reading Difficulties in a Morphemic Script. In *Reading and Writing Disorders in Different Orthographic Systems* (Vol. 52, pp. 267–282). http://doi.org/10.1007/978-94-009-1041-6_16

Levy, M., & Stockwell, G. (2013). *CALL Dimensions: Options and Issues in Computer-Assisted Language Learning (ESL & Applied Linguistics Professional).* New York, NY: Routledge.

Li, J. (2001). Chinese Conceptualization of Learning. *Ethos*, *29*(2), 111–137. http://doi.org/10.1525/eth.2001.29.2.111

Liu, S. B. F. (1978). Decoding and Comprehension in Reading Chinese. In *Cross-cultural perspectives on reading and reading research* (pp. 144–156). Newark, DE: International Reading Association.

Lund, A., & Lund, M. (2013). Laerd Statistics. Retrieved from https://statistics.laerd.com

Massaro, D. W., Liu, Y., Chen, T. H., & Perfetti, C. (2006). A Multilingual Embodied Conversational Agent for Tutoring Speech and Language Learning. In *Proceedings of the Ninth International Conference on Spoken Language Processing (Interspeech 2006 - ICSLP, September, Pittsburgh, PA)* (pp. 825–828). Universität Bonn, Bonn, Germany.

McEwen, P. (2006). *Vocabulary Acquisition in CFL (Chinese as a Foreign Language) Contexts: a Correlation of Performance and Strategy Use*. Brigham Young University.

Muñoz, D. C., Ortiz, A., González, C., López, D. M., & Blobel, B. (2010). Effective E-learning for Health Professional and Medical Students: the Experience with SIAS-Intelligent Tutoring System. *Studies in Health Technology and Informatics*, *156*, 89–102. http://doi.org/10.3233/978-1-60750-565-5-89

Murray, T. (1999). Authoring Intelligent Tutoring Systems: An analysis of the State of the Art. *International Journal on Artificial Intelligence in Education*, *10*, 98–129.

Nagata, N. (2002). BANZAI : An Application of Natural Language Processing to Web-based language processing. *CALICO Journal*, *19*(3), 583–599. Retrieved from http://journals.sfu.ca/CALICO/index.php/calico/article/view/624

Net Promoter System. (2013). Retrieved from http://www.netpromotersystem.com/about/how-is-nps-related-to-growth.aspx

Nkambou, R., Bourdeau, J., & Mizoguchi, R. (2010). *Advances in Intelligent Tutoring Sytems*. Warsaw, Poland: Springer Science & Business Media.

Park, S., & Arbuckle, T. Y. (1977). Ideograms versus Alphabets: Effects of Script on Memory in "Biscriptual" Korean Subjects. *Journal of Experimental Psychology: Human Learning & Memory*, *3*(6), 631–642. http://doi.org/10.1037/0278-7393.3.6.631

Perfetti, C. (1999). Comprehending Written Language: A Blueprint of the Reader. *The Neurocognition of Language*, 167–208.

Reichheld, F. F. (2003). The One Number You Need To Grow. *Harvard Business Review*, *81*(12), 46–54, 124. http://doi.org/10.1111/j.1467-8616.2008.00516.x

Riedl, M. O., & Stern, A. (2006). Believable Agents and Intelligent Scenario Direction for Social and Cultural Leadership Training. In *Proceedings of the 15th Conference on Behavior Representation in Modeling and Simulation (BRIMS)* (p. Paper 27).

Rozin, P., Poritsky, S., & Sotsky, R. (1971). American Children with Reading Problems can Easily Learn to Read English Represented by Chinese Characters. *American Association for the Advancement of Science*, *171*(977), 1264–1267. http://doi.org/10.1126/science.171.3977.1264

Satmetrix Systems, I. (2015). Net Promoter Network. Retrieved from http://www.satmetrix.com/expertise/benchmarks/

Sauro, J. (2011). Measuring Usability with the System Usability Scale (SUS). Retrieved from http://www.measuringu.com/sus.php

Shen, H. H. (2005). An Investigation of Chinese-Character Learning Strategies among Non-native Speakers of Chinese. *System*, *33*(1), 49–68. http://doi.org/10.1016/j.system.2004.11.001

Shih, B.-Y., Chen, C.-Y., & Li, C. E. (2013). The Exploration of the Mobile Mandarin Learning System by the Application of TRIZ Theory. *Computer Applications in Engineering Education*, *21*(2), 343–348. http://doi.org/10.1002/cae.20478

Shu, H. (2003). Chinese Writing System and Learning to Read. *International Journal of Psychology*, *38*(5), 274–285.

Shu, H., Anderson, R. C., & Zhang, Z. (1995). Incidental Learning of Word Meanings while Reading: a Chinese and American Cross-Cultural Study. *Reading Research Quarterly*, *30*, 76–95.

Sottilare, R., Graesser, A., Hu, X., & Goldberg, B. (2014). *Design Recommendations for Intelligent Tutoring Systems*. Orlando, FL: the U.S. Army Research Laboratory.

Swartz, M. L., & Yazdani, M. (1992). *Intelligent Tutoring Systems for Foreign Language Learning*. Springer-Verlag Berlin Heidelberg.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., … Wintersgill, M. (2005). The Andes Physics Tutoring System: Lessons Learned. *International Journal of Artificial Intelligence in Education*, *15*(3), 1–51. Retrieved from http://iospress.metapress.com/index/4QH80UBFDFT0G4YR.pdf

Virvou, M., & Tsiriga, V. (2001). Web Passive Voice Tutor: an Intelligent Computer Assisted Language Learning System over the WWW. *Proceedings IEEE International Conference on Advanced Learning Technologies*, 131–134. http://doi.org/10.1109/ICALT.2001.943878

Wang, M., Perfetti, C. A., & Liu, Y. (2005). Chinese-English Biliteracy Acquisition: Cross-Language and Writing System Transfer. *Cognition*, *97*(1), 67–88. http://doi.org/10.1016/j.cognition.2004.10.001

Wang, Y., & Carigliano, R. (1992). An Intelligent Language Tutoring System for Handling Errors Caused By Transfer. In *Intelligent Tutoring Systems* (pp. 395–404).

Xing, J. Z. (2006). *Teaching and Learning Chinese as a Foreign Language: A Pedagogical Grammar*. Hong Kong: Hong Kong University Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=1drHAQAAQBAJ&oi=fnd&pg=PP1&dq=difficulty+of+learning+chinese&ots=OiPyGUi8HY&sig=pjl4T01fkoWysBu6lEBUscpYpmA#v=onepage&q=difficulty of learning chinese&f=false

Xu, S., Jiang, H., Lau, F. C. M., & Pan, Y. (2007). An Intelligent System for Chinese Calligraphy. In *Proceedings of the 22nd AAAI Conference on Artificial Intelligence (AAAI '07), Association for the Advancement of Artificial Intelligence*. Vancouver.

Yoyo. (2012). Free Chinese Lessons. Retrieved from http://www.freechineselessons.com/

APPENDIX A - IRB 15-219 APPROVAL FORM

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
1138 Pearson Hall
Ames, Iowa 50011-2207
515 294-4566
FAX 515 294-4267

Date: 4/9/2015

To: Dr. Stephen B Gilbert, 1620 Howe Hall

CC: Jingyu Liu, 1620 Howe Hall

From: Office for Responsible Research

Title: Intelligent Language Tutoring For Chinese Characters

IRB ID: 15-219

Approval Date:	4/9/2015	**Date for Continuing Review:**	4/6/2016
Submission Type:	New	**Review Type:**	Full Committee

The project referenced above has received approval from the Institutional Review Board (IRB) at Iowa State University according to the dates shown above. Please refer to the IRB ID number shown above in all correspondence regarding this study.

To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use only the approved study materials** in your research, **including the recruitment materials and informed consent documents that have the IRB approval stamp.**
- **Retain signed informed consent documents for 3 years after the close of the study,** when documented consent is required.
- **Obtain IRB approval prior to implementing any changes** to the study by submitting a Modification Form for Non-Exempt Research or Amendment for Personnel Changes form, as necessary.
- **Immediately inform the IRB of** (1) **all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems involving risks** to subjects or others.
- **Stop all research activity if IRB approval lapses,** unless continuation is necessary to prevent harm to research participants. Research activity can resume once IRB approval is reestablished.
- **Complete a new continuing review form** at least three to four weeks prior to the **date for continuing review** as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a courtesy reminder as this date approaches.

Please be aware that IRB approval means that you have met the requirements of federal regulations and ISU policies governing human subjects research. **Approval from other entities may also be needed.** For example, access to data from private records (e.g. student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. **IRB approval in no way implies or guarantees that permission from these other entities will be granted.**

Upon completion of the project, please submit a Project Closure Form to the Office for Responsible Research, 1138 Pearson Hall, to officially close the project.

Please don't hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.

APPENDIX B - IRB 15-597 APPROVAL FORM

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
1138 Pearson Hall
Ames, Iowa 50011-2207
515 294-4566
FAX 515 294-4267

Date: 10/23/2015

To: Dr. Stephen B Gilbert
1620 Howe Hall

CC: Jingyu Liu
1620 Howe Hall

From: Office for Responsible Research

Title: Mobile App for Tutoring Chinese Characters

IRB ID: 15-597

Approval Date: 10/23/2015 **Date for Continuing Review:** 10/19/2016

Submission Type: New **Review Type:** Full Committee

The project referenced above has received approval from the Institutional Review Board (IRB) at Iowa State University according to the dates shown above. Please refer to the IRB ID number shown above in all correspondence regarding this study.

To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use only the approved study materials** in your research, **including the recruitment materials and informed consent documents that have the IRB approval stamp.**
- **Retain signed informed consent documents for 3 years after the close of the study,** when documented consent is required.
- **Obtain IRB approval prior to implementing <u>any</u> changes** to the study by submitting a Modification Form for Non-Exempt Research or Amendment for Personnel Changes form, as necessary.
- **Immediately inform the IRB of** (1) **all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems involving risks** to subjects or others.
- **Stop all research activity if IRB approval lapses,** unless continuation is necessary to prevent harm to research participants. Research activity can resume once IRB approval is reestablished.
- **Complete a new continuing review form** at least three to four weeks prior to the **date for continuing review** as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a courtesy reminder as this date approaches.

Please be aware that IRB approval means that you have met the requirements of federal regulations and ISU policies governing human subjects research. **Approval from other entities may also be needed.** For example, access to data from private records (e.g. student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. **IRB approval in no way implies or guarantees that permission from these other entities will be granted.**

Upon completion of the project, please submit a Project Closure Form to the Office for Responsible Research, 1138 Pearson Hall, to officially close the project.

Please don't hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.

APPENDIX C - REQUIREMENT SURVEY

Clicking "Yes" below indicates that you voluntarily agree to participate in this study, that the study has been explained to you, that you have been given the time to read the document and that your questions have been satisfactorily answered.

Clicking "Yes" also indicates that you are 18 years old or older. You may print a copy of this informed consent prior to your participation in the study if desired.

- ◯ Yes, I agree.
- ◯ No, thank you.

Chinese fluency/interest

How long have you studied Chinese?

- ◯ Less than 1 year
- ◯ 1-2 years

- 2-3 years
- 3-4 years
- 4-5 years
- More than 5 years

What's the level of your Chinese fluency?

	Poor		Fair		Good		Very Good		Excellent	
	0 10	20	30	40	50	60	70	80	90	100
Listening										
Speaking										
Reading										
Writing										

What's your motivation for learning Chinese?

- For communicating with Chinese friends
- For learning Chinese culture
- For travel purposes
- For job purposes
- Other

What's the level of interest you have in Chinese?

	None		Little		Some		A Lot				
	0	10	20	30	40	50	60	70	80	90	100
Interest											

Difficulties/methods characters

How difficult do you think the following skills are in Chinese?

	Very Easy	Easy	Somewhat Easy		Neutral	Somewhat Difficult		Difficult	Very Difficult	
	0 10	20	30	40	50	60	70	80	90	100
Listening										
Speaking										
Reading										
Writing										

When you started learning Chinese, which of the following did you begin with (choose all that apply)?

- ☐ Listen and understand what people say.
- ☐ Speak some simple words learning from others.
- ☐ Read and recognize some characters.
- ☐ Write some characters.
- ☐ Others, please specify

- 2-3 years
- 3-4 years
- 4-5 years
- More than 5 years

What's the level of your Chinese fluency?

Poor Fair Good Very Good Excellent

0 10 20 30 40 50 60 70 80 90 100

Listening										
Speaking										
Reading										
Writing										

What's your motivation for learning Chinese?

- For communicating with Chinese friends
- For learning Chinese culture
- For travel purposes
- For job purposes
- Other

Have you learned **Chinese characters**?

◯ Yes

◯ No

How difficult do you think learning **Chinese characters** is?

	Very Easy	Easy	Somewhat Easy		Neutral	Somewhat Difficult		Difficult	Very Difficult	
	0	10	20	30	40	50	60	70	80	90	100
Difficulty										

Why did you give this difficulty rating?

Learning methods

How do you learn to recognize **Chinese characters**? (choose all that apply)

☐ By trying to imitate each stroke many times

☐ By taking each part apart and constructing them according to their meanings

☐ By memorizing each character over and over again

☐ Other, please specify

What methods do you use to learn **Chinese characters**? (choose all that apply)

	Have you used the following methods?	How useful they are?		
	Yes	Not very useful	Useful somehow	Very useful
Learn in the classroom	◯	◯	◯	◯
Learn from a tutor / friend	◯	◯	◯	◯
Self-learning from books	◯	◯	◯	◯
Self-learning from learning software	◯	◯	◯	◯
Self-learning from multi-media	◯	◯	◯	◯
Other, please specify [ ]	◯	◯	◯	◯
Other, please specify [ ]	◯	◯	◯	◯

If there were a new learning software app that teaches **Chinese characters**, how likely is it that you would try it?

	Very Unlikely	Unlikely	Somewhat Unlikely	Undecided	Somewhat Likely	Likely	Very Likely
	0 10	20	30 40	50 60	70	80 90	100
How likely							

Open-ended questions

Could you briefly describe the way you learned **Chinese characters**, and comment on whether you think it was effective or efficient?

What challenges did you face when learning **Chinese characters**?

ISU IRB # 1 15-219
Approved Date: 9 April 2015
Expiration Date: 6 April 2016

If your friends wanted to start learning **Chinese characters / Chinese**, what suggestions would you give them to help them start?

Demographics

Demographics questions.

Click your current role:

- Undergraduate
- Graduate
- Faculty/Staff

Please select the gender with which you identify:

- Male
- Female

What's your native language?

- English
- Spanish

- German
- Korean
- Other, please specify

Do you have experience learning other foreign languages other than Chinese?

- Yes
- No

Please list all the languages you've learned and level of fluency.

Poor Fair Good Very Good Excellent

0 10 20 30 40 50 60 70 80 90 100

language 1										
language 2										
language 3										
language 4										
language 5										

APPENDIX D - PRE-SURVEY

Mobile App for Tutoring Chinese Characters: Pre-Survey

Click your current role:

- ◯ Undergraduate
- ◯ Graduate
- ◯ Faculty/Staff

How do you describe yourself?

- ◯ Male
- ◯ Female
- ◯ Transgender
- ◯ Do not identify with female, male, or transgender.

What's your native language?

- ◯ English
- ◯ Spanish
- ◯ German
- ◯ Korean
- ◯ Other, please specify

Is your native language alphabet-based or character-based language?

- alphabet-based (e.g., English, Spanish, Russian, Greek)
- character-based (e.g., Chinese, Japanese, Korean)

Please list all the languages you've learned and level of fluency.

	Poor		Fair		Good		Very Good		Excellent	
	0	10	20	30	40	50	60	70	80	90	100
language 1											
language 2											
language 3											
language 4											
language 5											

Do you know any of Chinese characters?

◯ Yes. (How did you learn them?)

◯ No

How difficult do you think learning Chinese characters is?

	Very easy	Easy	Somewhat Easy	Neutral	Somewhat Difficult	Difficult	Very difficult

0 10 20 30 40 50 60 70 80 90 100

Difficulty

ttps://iastate.qualtrics.com/SE/?SID=SV_1UqJJD7iN7AJ9Nr&Preview=Survey&Q_CHL=preview&BrandID=iastate

0/5/2015 Online Survey Software | Qualtrics Survey Solutions

What's the level of interest do you have in the Chinese language?

	None	Little	Some	A Lot

0 10 20 30 40 50 60 70 80 90 100

Interest

If you were about to learn a second language, how likely is it that you would learn Chinese?

	Very Unlikely	Unlikely	Somewhat Unlikely	Undecided	Somewhat Likely	Likely	Very Likely

0 10 20 30 40 50 60 70 80 90 100

How likely

>>

APPENDIX E - POST-SURVEY (INCLUDING QUIZ)

Mobile App for Tutoring Chinese Characters: Quiz & Post-Survey

Translate the following characters by writing their meaning under each character.

人

______

火

______

月

______

山

______

门

______

日

______

Please select the translation of **"big"**

- 人
- 大
- 小

Please translate the words below:

大众

______

大人

______

Select translation of 休

- person leaning the tree
- take a rest
- sleep

Select the translation of 森

- woods
- forest
- tree

Select the translating of **hot**

- 炎
- 林
- 火

Select translation of 旦

- sunrise
- moon
- sunset

Select the translation of **"tomorrow"**

- 旦
- 明
- 昌

Translate the word 大火

Select translation of 火山

- crater
- bonfire
- volcano

Select translation of **"sparkles"**

- 旦
- 昌
- 晶

Translate the word 休火山

Select translation of **"ask"**.

- ○ 闪
- ○ 问
- ○ 闭

Guess which of the following is the meaning of 品

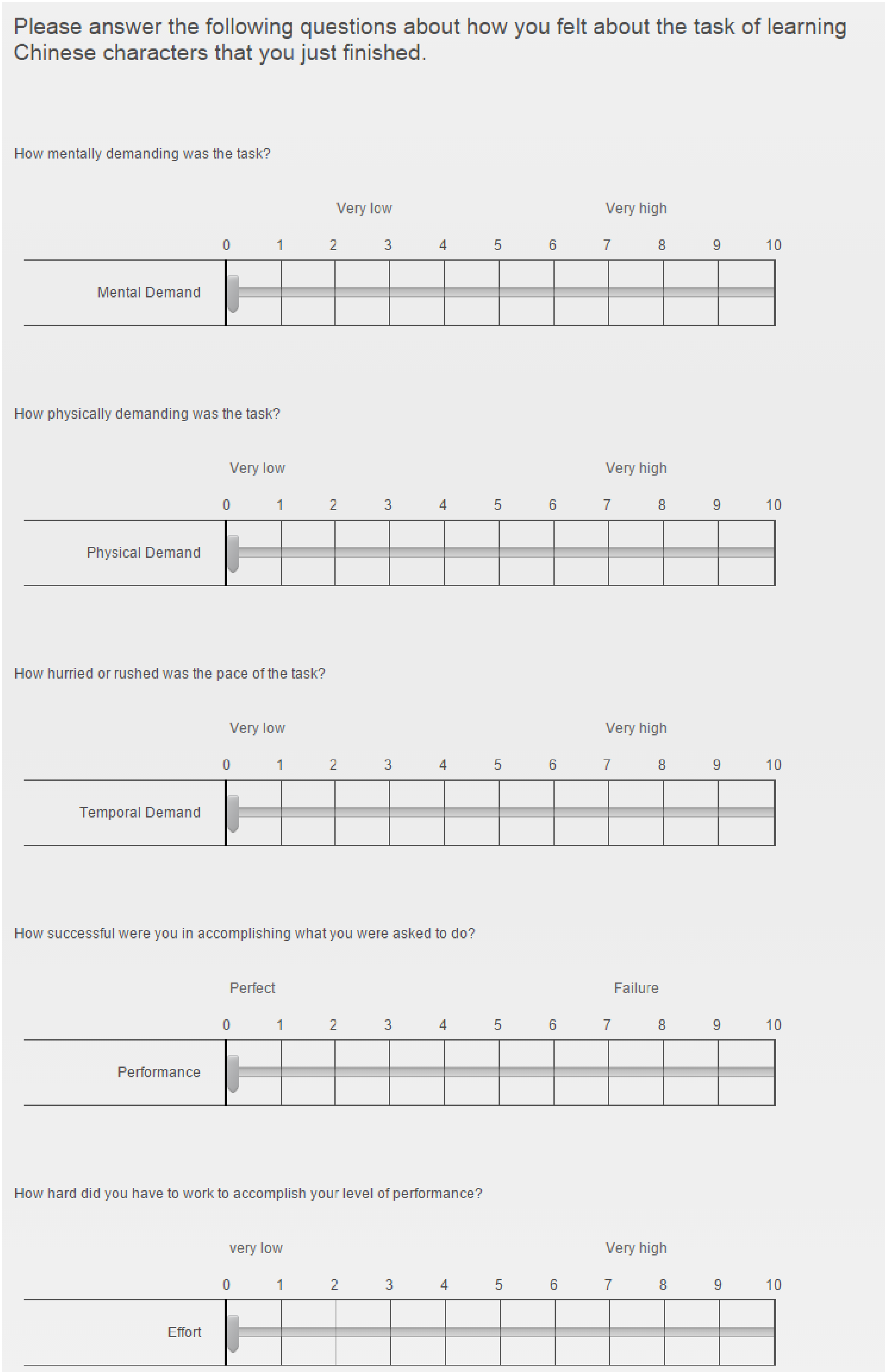
- gather
- taste
- crowd

Guess which of the following is related to **time** (hint: ancient people know time from the change of the sun).

- 灯
- 时
- 闩

Guess which of the following means **pine**?

- 松
- 汉
- 仲



Please answer the following questions about how you felt about the task of learning Chinese characters that you just finished.

How mentally demanding was the task?

Very low — Very high

0 1 2 3 4 5 6 7 8 9 10

Mental Demand

How physically demanding was the task?

Very low — Very high

0 1 2 3 4 5 6 7 8 9 10

Physical Demand

How hurried or rushed was the pace of the task?

Very low — Very high

0 1 2 3 4 5 6 7 8 9 10

Temporal Demand

How successful were you in accomplishing what you were asked to do?

Perfect — Failure

0 1 2 3 4 5 6 7 8 9 10

Performance

How hard did you have to work to accomplish your level of performance?

very low — Very high

0 1 2 3 4 5 6 7 8 9 10

Effort

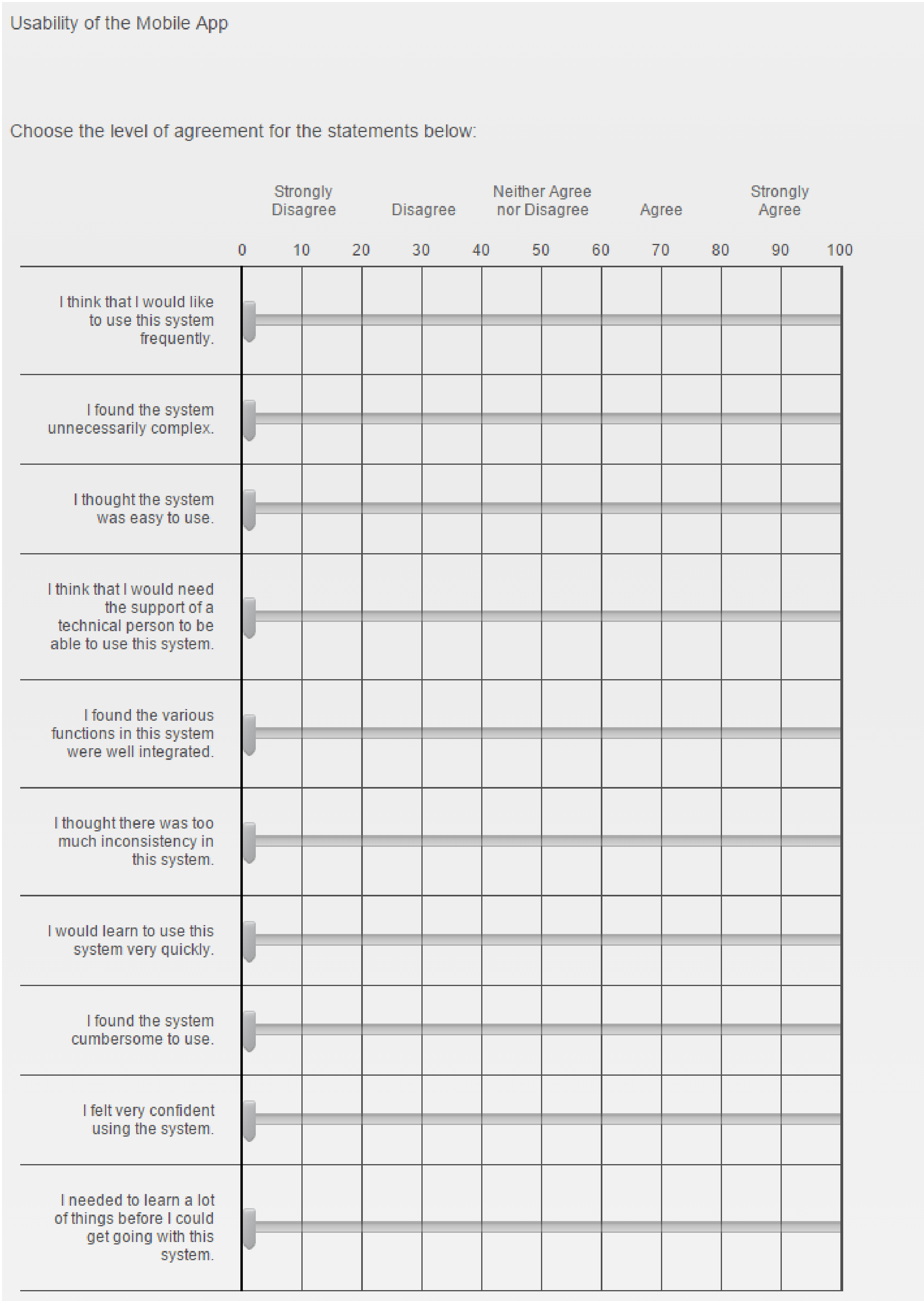
How insecure, discouraged, irritated, stressed, and annoyed were you?

	Very low							Very high			
	0	1	2	3	4	5	6	7	8	9	10
Frustration											

The next questions are related to your experience with the iPad app designed to teach you the Chinese characters. When the questions below say "system," think of that iPad app.

Choose the level of agreement for the statements below:

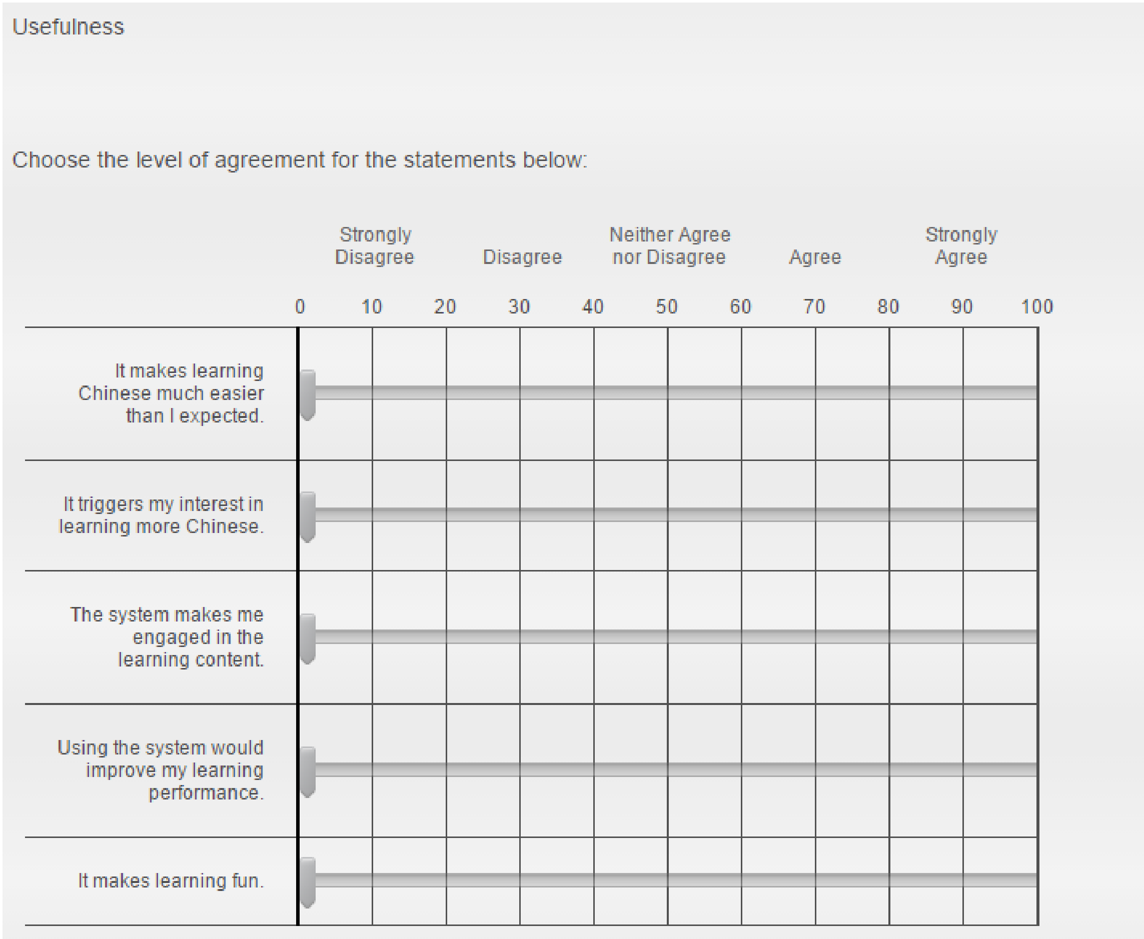
	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
	0 10 20	30 40	50 60	70 80	90 100
Overall, I am satisfied with this learning system.					
It helps me a lot to understand the learning content.					
I would like to learn more Chinese using this system.					
I would recommend it to a friend.					



Usability of the Mobile App

Choose the level of agreement for the statements below:

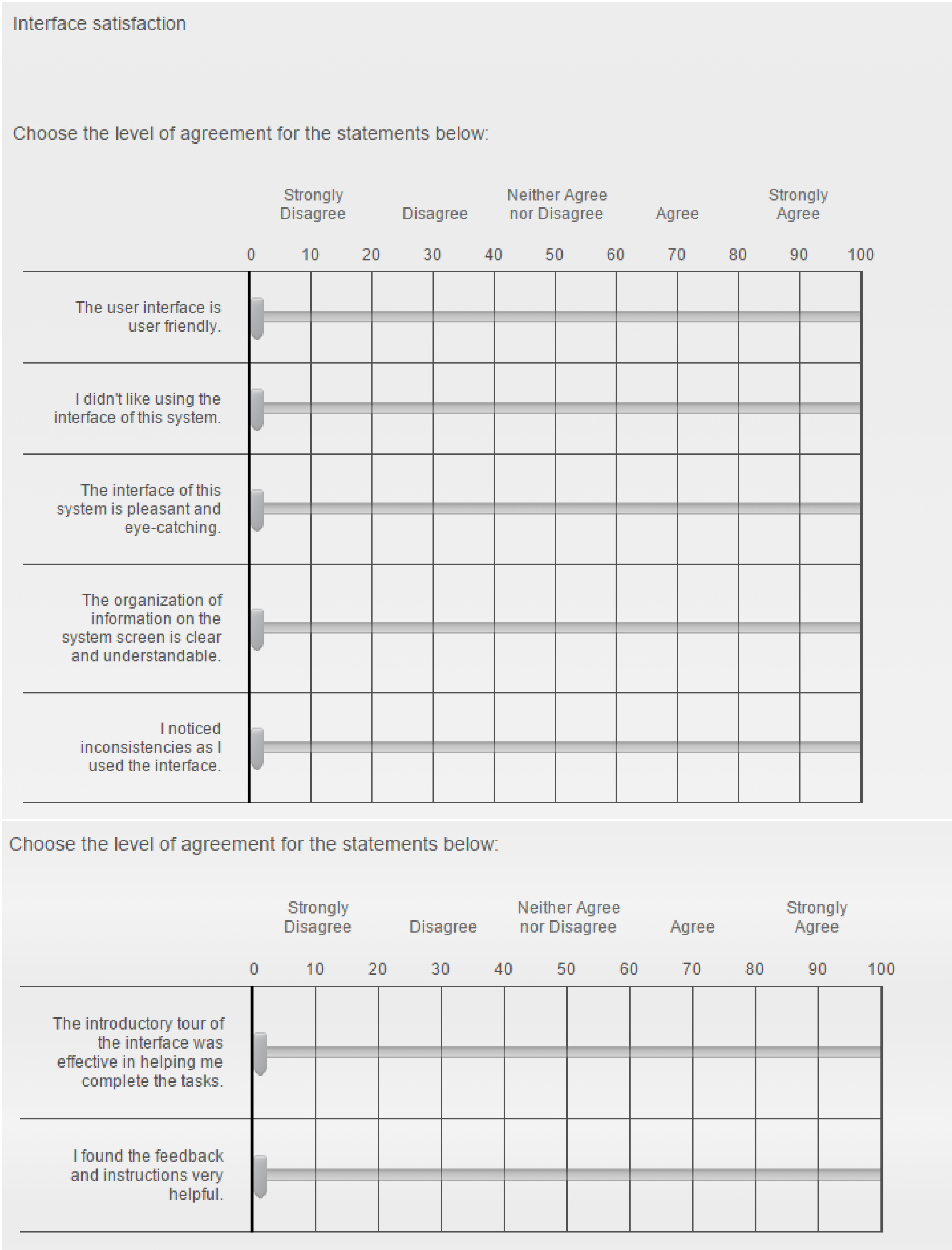
	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
	0 10 20	30 40	50 60	70 80	90 100
I think that I would like to use this system frequently.					
I found the system unnecessarily complex.					
I thought the system was easy to use.					
I think that I would need the support of a technical person to be able to use this system.					
I found the various functions in this system were well integrated.					
I thought there was too much inconsistency in this system.					
I would learn to use this system very quickly.					
I found the system cumbersome to use.					
I felt very confident using the system.					
I needed to learn a lot of things before I could get going with this system.					



Usefulness

Choose the level of agreement for the statements below:

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
	0 10 20	30 40	50 60	70 80	90 100
It makes learning Chinese much easier than I expected.					
It triggers my interest in learning more Chinese.					
The system makes me engaged in the learning content.					
Using the system would improve my learning performance.					
It makes learning fun.					



Interface satisfaction

Choose the level of agreement for the statements below:

	Strongly Disagree		Disagree		Neither Agree nor Disagree		Agree		Strongly Agree		
	0	10	20	30	40	50	60	70	80	90	100
The user interface is user friendly.											
I didn't like using the interface of this system.											
The interface of this system is pleasant and eye-catching.											
The organization of information on the system screen is clear and understandable.											
I noticed inconsistencies as I used the interface.											

Choose the level of agreement for the statements below:

	Strongly Disagree		Disagree		Neither Agree nor Disagree		Agree		Strongly Agree		
	0	10	20	30	40	50	60	70	80	90	100
The introductory tour of the interface was effective in helping me complete the tasks.											
I found the feedback and instructions very helpful.											

What are three things did you like about the system?

What are three things that need to be improved?

Thank you for your time helping us create a great learning experience!

APPENDIX F - SCREENSHOTS OF VERSION A AND VERSION B

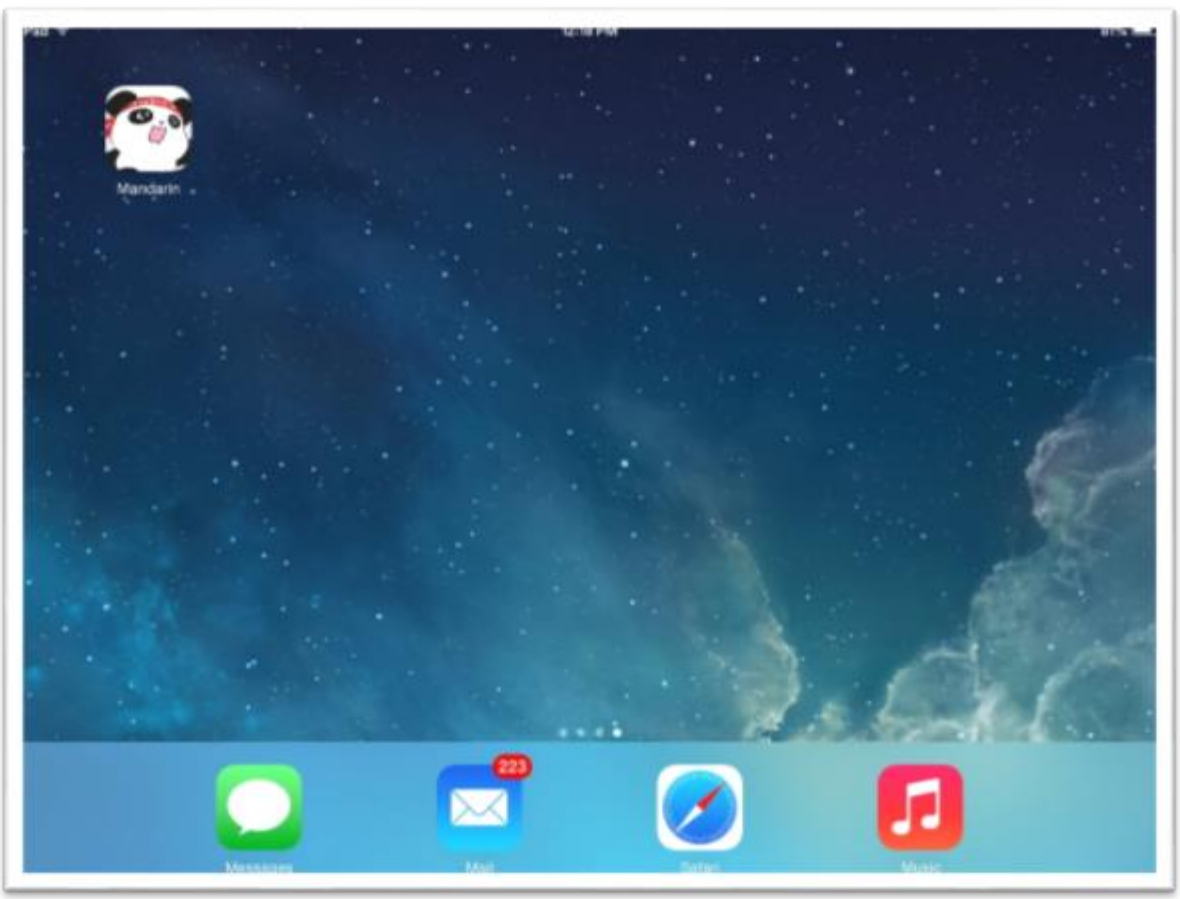
Mobile App for Tutoring Chinese Characters: Screenshots of Mobile App

This document contains screenshots of the mobile app participants will be using. There are two versions: one with more pedagogical interactions such as feedback for the quizzes and metaphor-based explanations.

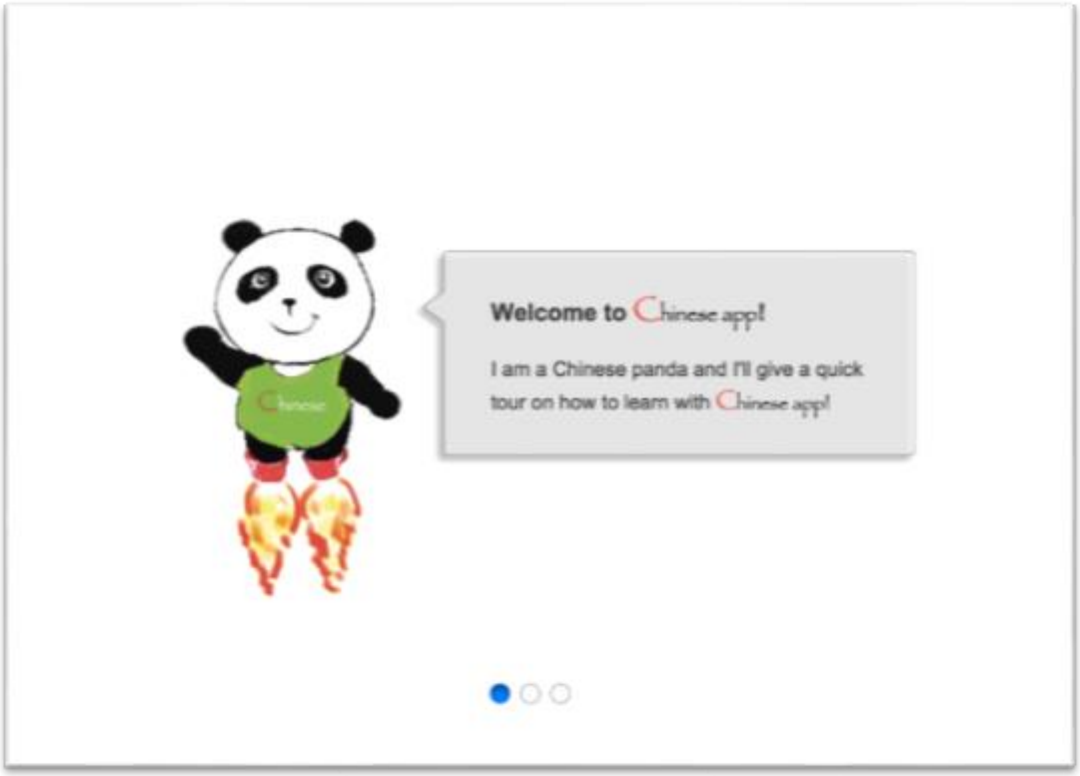
There are 8 mini-courses, each based on a single Chinese character. There are over 100 screens. This document does not show all screens, but instead shows the introduction screens for the app and then shows all screens for the first mini-course. The subsequent mini-courses vary slightly in structure, e.g. some how more compound words and others have fewer, but the first mini-course is the longest, and gives a representative impression of the app for the other characters.

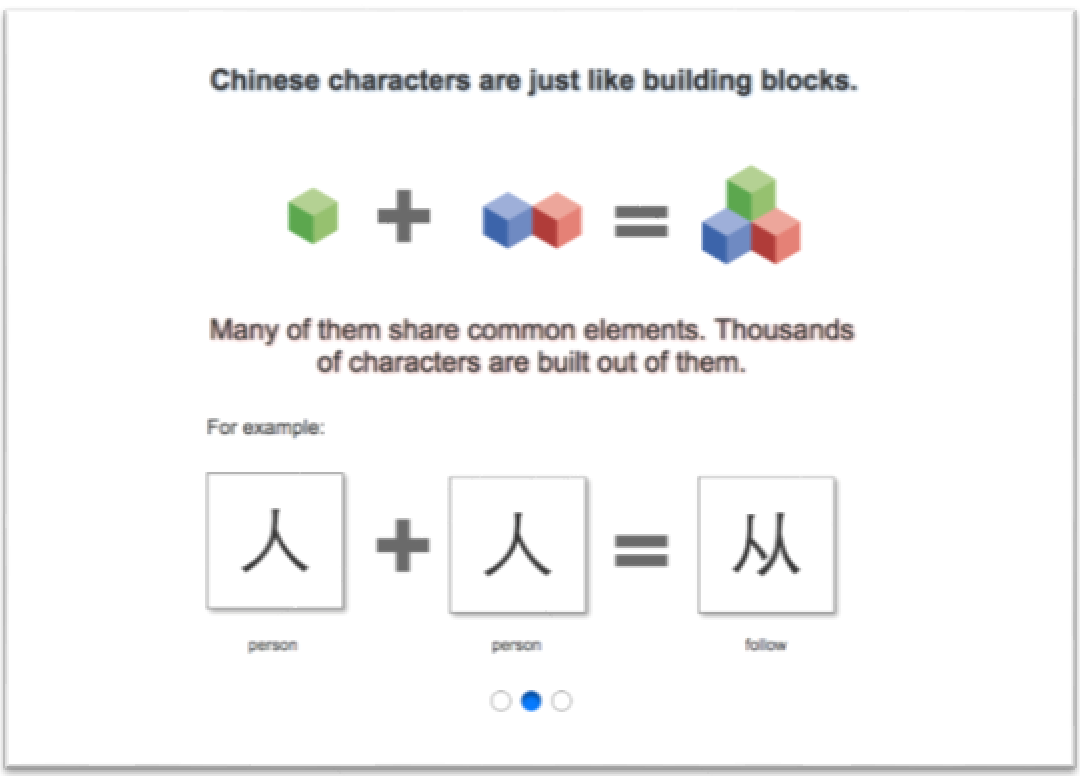
VERSION A: More pedagogical interaction

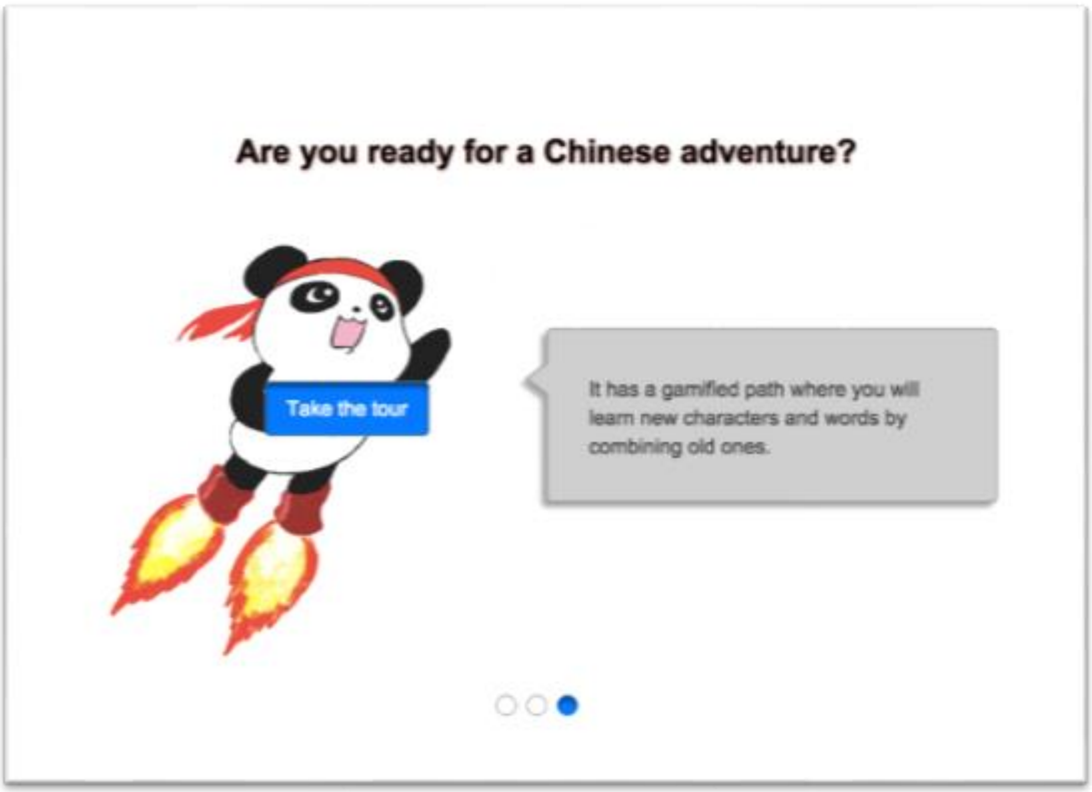
SCREENSHOT OF ICON



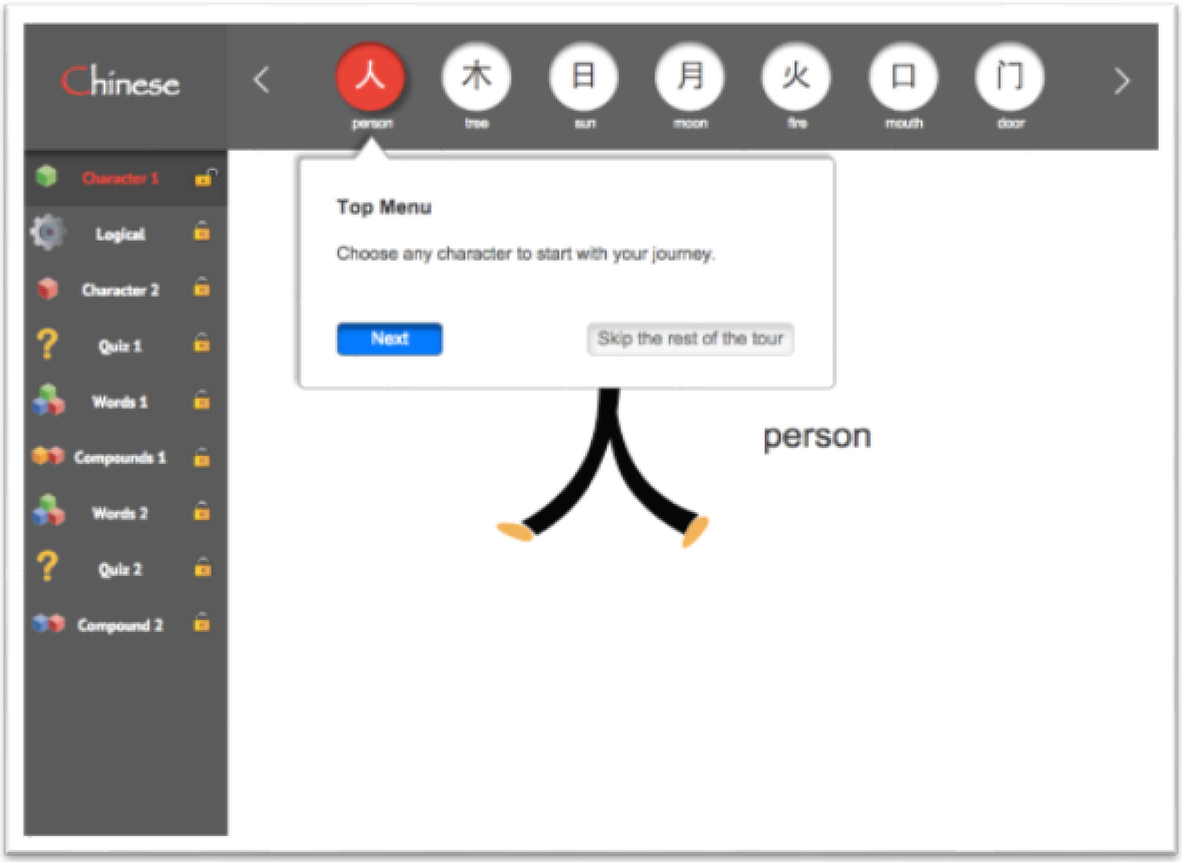
INTRODUCTION PAGES

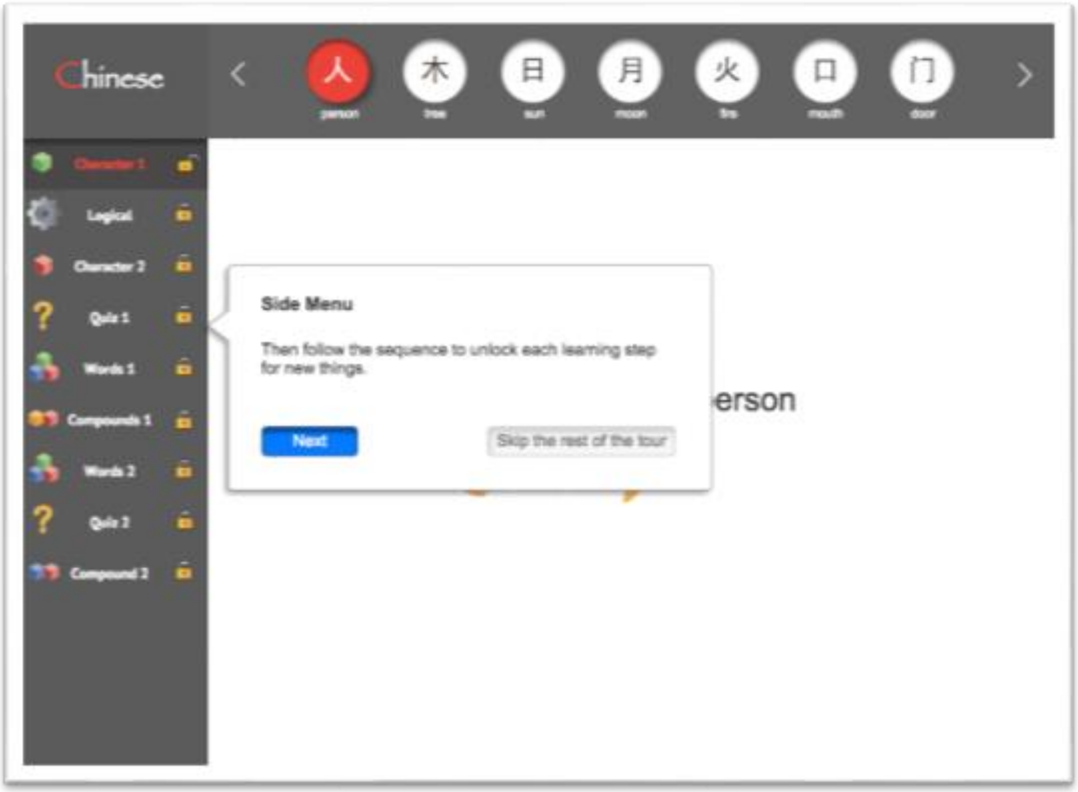

Welcome to Chinese app!
I am a Chinese panda and I'll give a quick tour on how to learn with Chinese app!

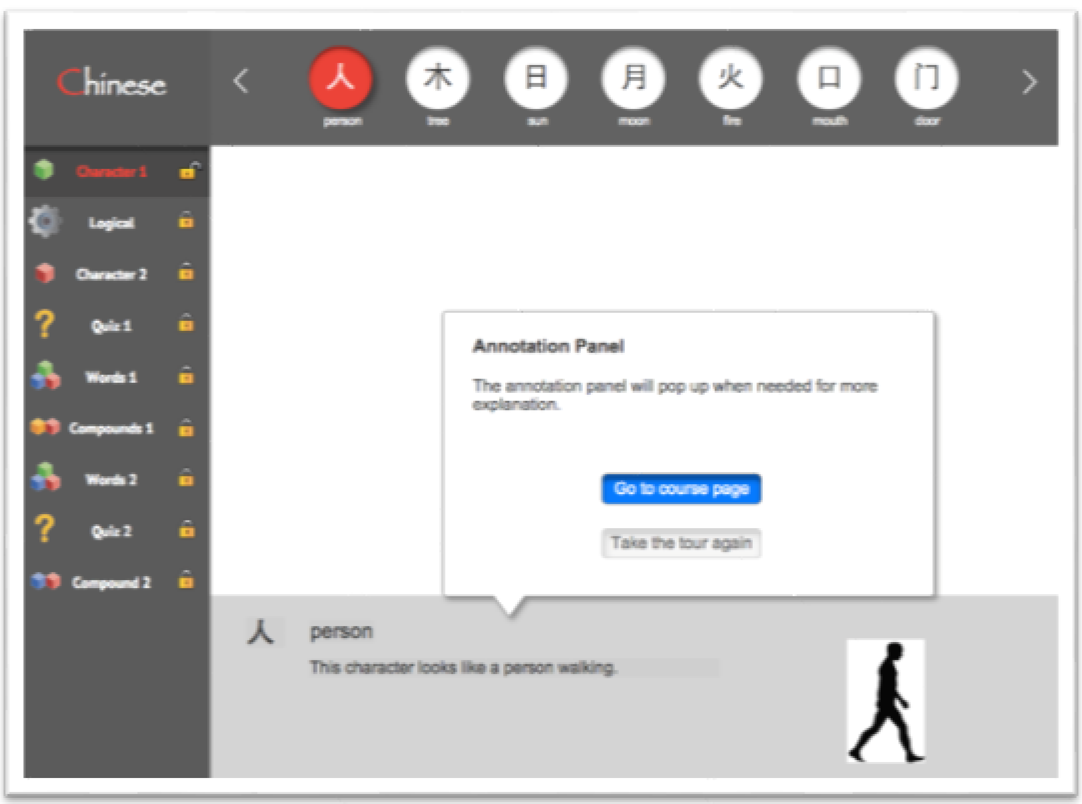

Chinese characters are just like building blocks.
Many of them share common elements. Thousands of characters are built out of them.
For example:
人
人
从
person
person
follow


Are you ready for a Chinese adventure?
Take the tour
It has a gamified path where you will learn new characters and words by combining old ones.

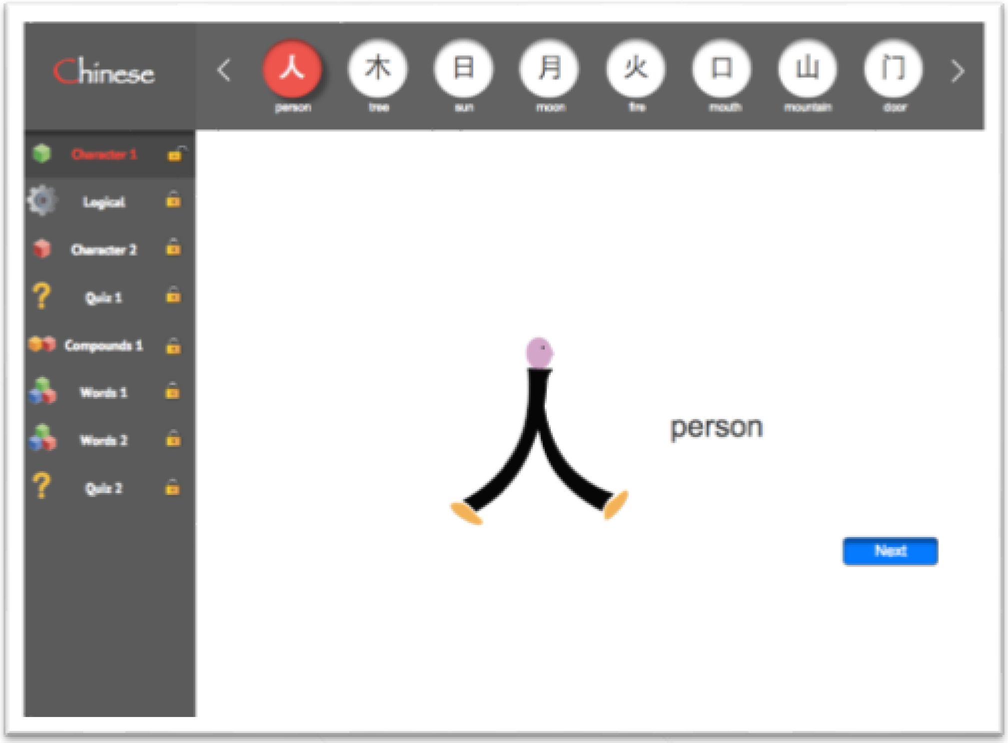
TOUR PAGES

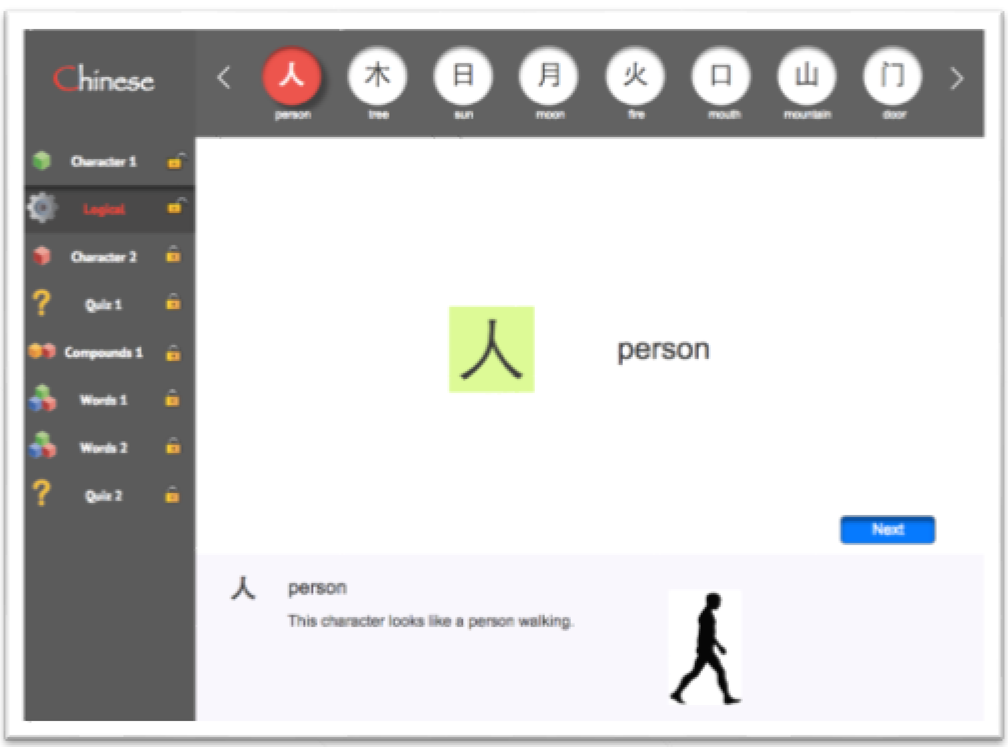

Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
门
door
Character 1
Logical
Character 2
Quiz 1
Words 1
Compounds 1
Words 2
Quiz 2
Compound 2
Top Menu
Choose any character to start with your journey.
Next
Skip the rest of the tour
person

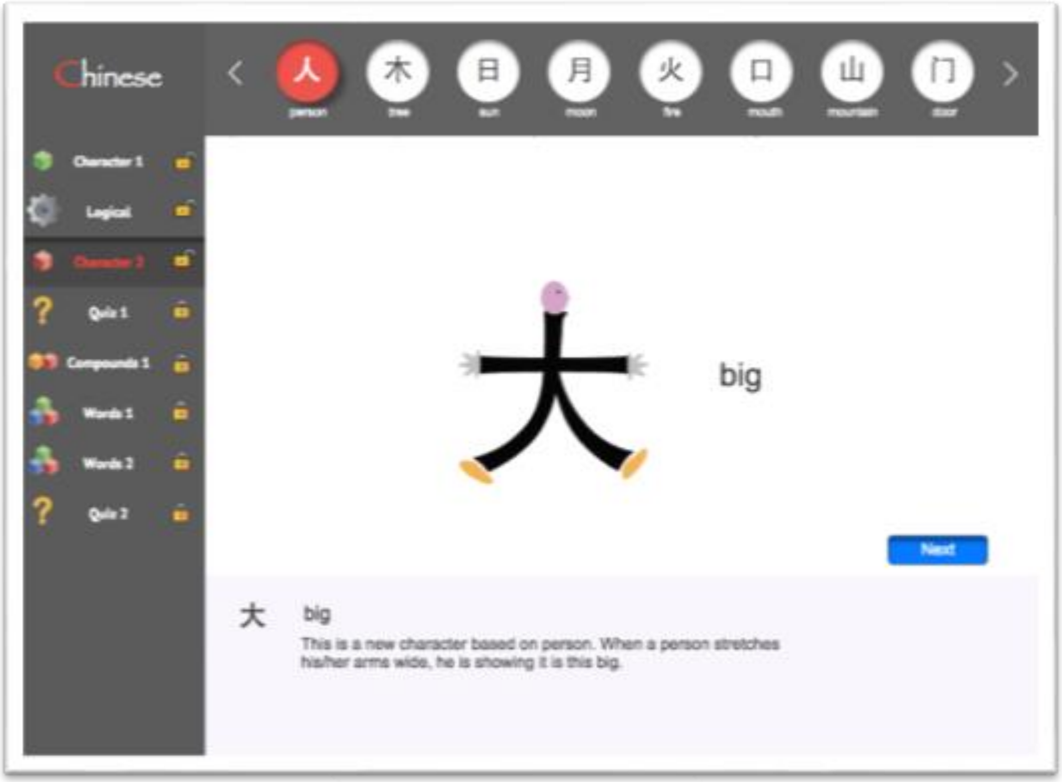

Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
门
door
Character 1
Logical
Character 2
Quiz 1
Words 1
Compounds 1
Words 2
Quiz 2
Compound 2
Side Menu
Then follow the sequence to unlock each learning step for new things.
Next
Skip the rest of the tour
erson


Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
门
door
Character 1
Logical
Character 2
Quiz 1
Words 1
Compounds 1
Words 2
Quiz 2
Compound 2
Annotation Panel
The annotation panel will pop up when needed for more explanation.
Go to course page
Take the tour again
人
person
This character looks like a person walking.

COURSES
---PERSON COURSE



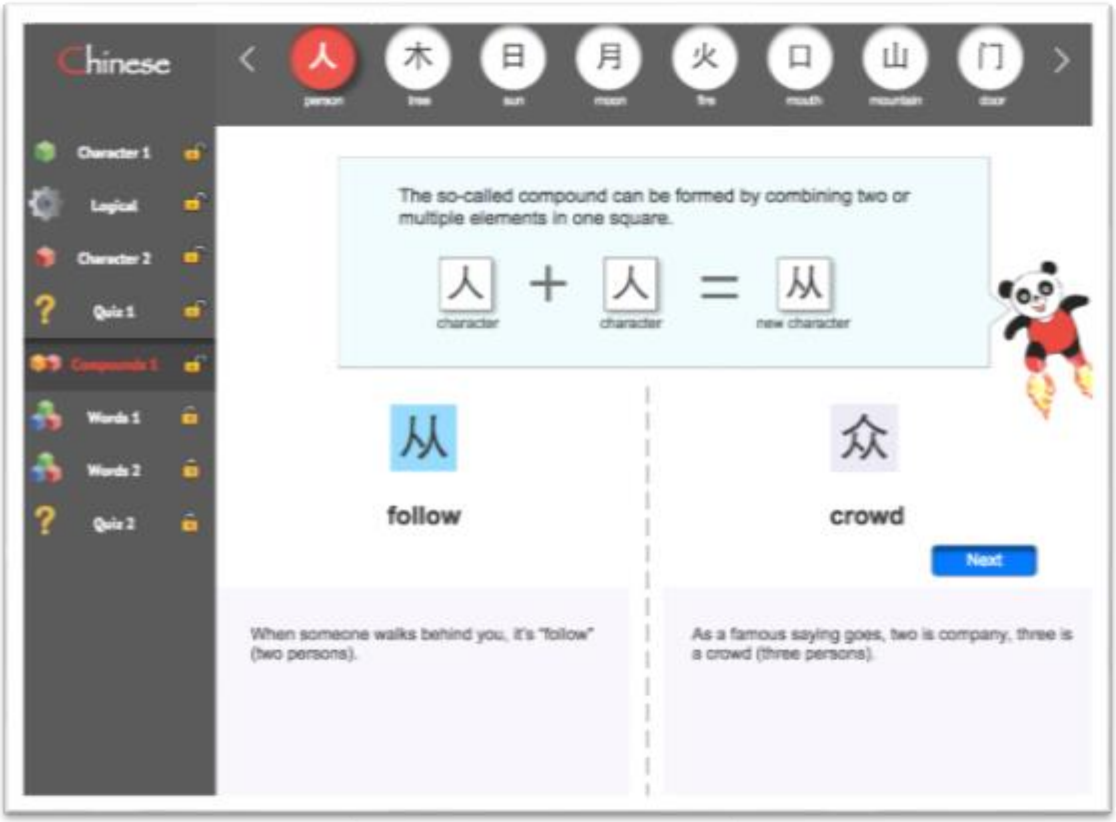



Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
山
mountain
门
door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
big
Next
大
big
This is a new character based on person. When a person stretches his/her arms wide, he is showing it is this big.

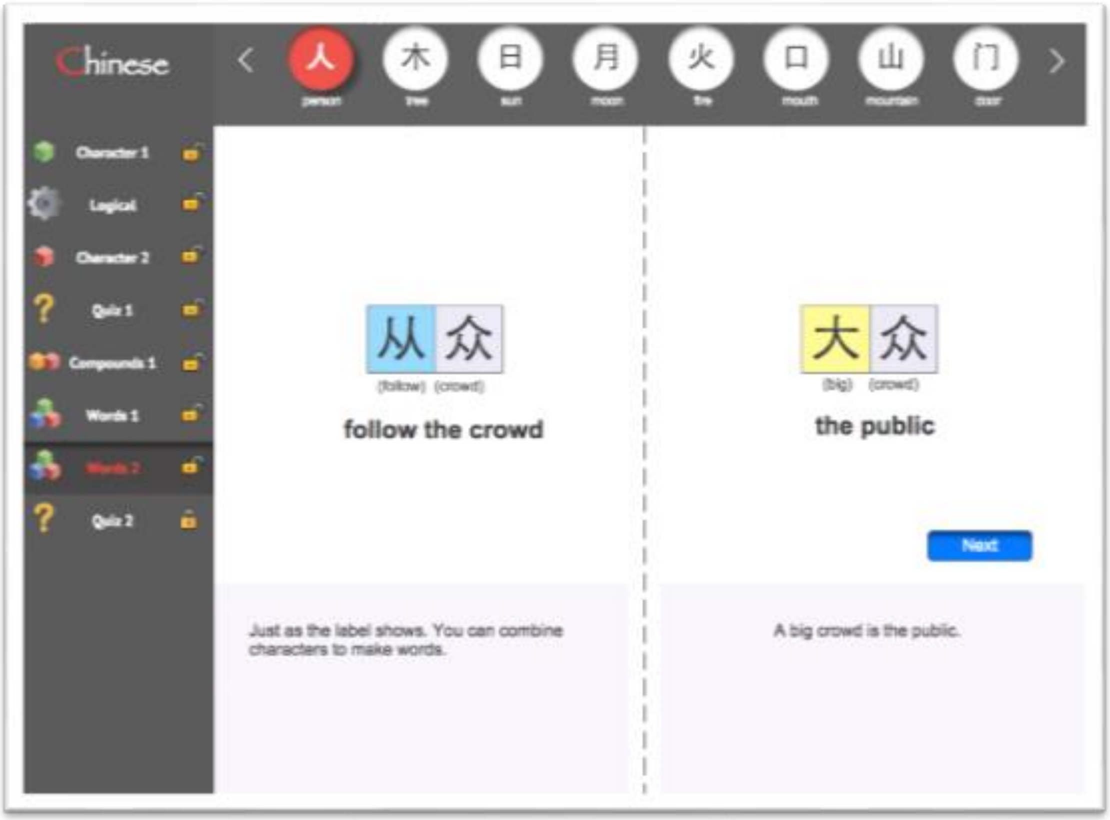

Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
山
mountain
门
door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
Select translation of "big"
大
人
小
check

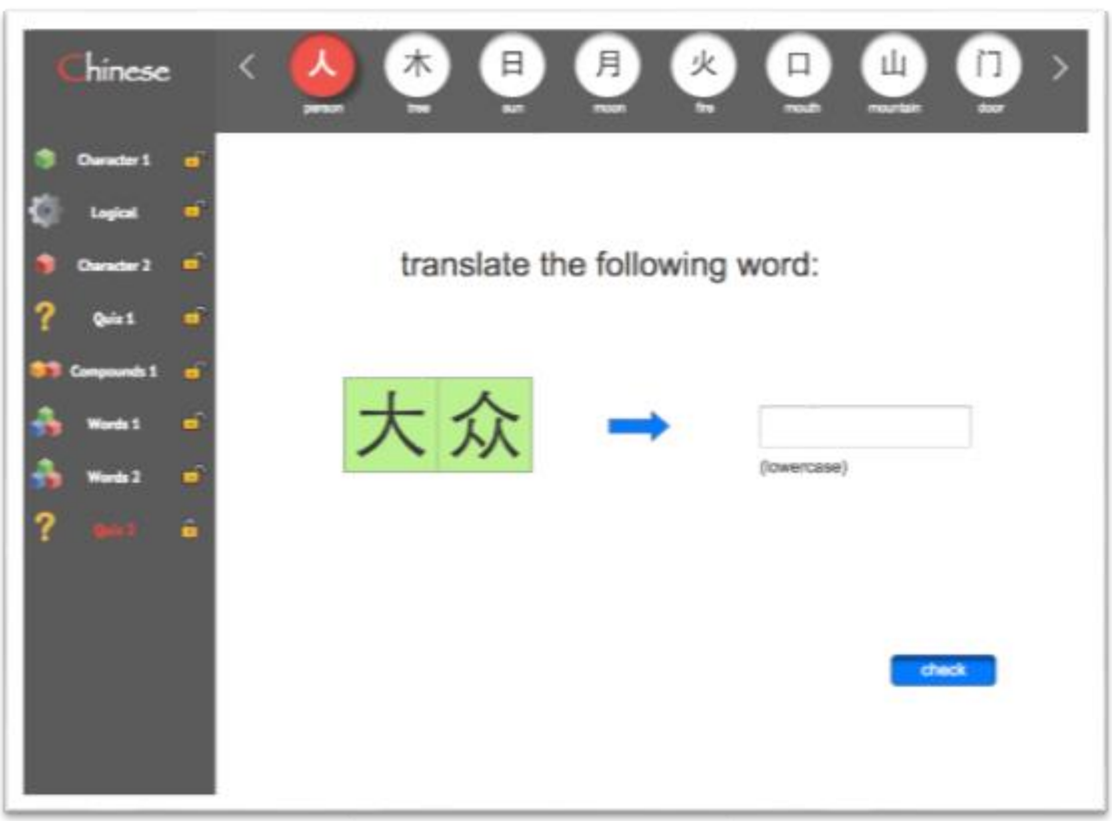

Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
山
mountain
门
door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
Select translation of "big"
大
人
小
You are correct!
Next


Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
山
mountain
门
door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
Select translation of "big"
大
人
小
Correct solution:
Next
Learn this character again
大 big
When a person stretches his/her arms wide, he is showing it is this big.


Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
山
mountain
门
door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
The so-called compound can be formed by combining two or multiple elements in one square.
人 + 人 = 从
character
character
new character
从
follow
众
crowd
Next
When someone walks behind you, it's "follow" (two persons).
As a famous saying goes, two is company, three is a crowd (three persons).


Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
山
mountain
门
door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
Combining two or multiple characters together to form words.
大 + 人 = 大人
character
character
word
人人
everyone
大人
adult
Next
When doubling person, it means everyone.
A big person is "adult".


Chinese
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
从众
follow the crowd
大众
the public
Next
Just as the label shows. You can combine characters to make words.
A big crowd is the public.

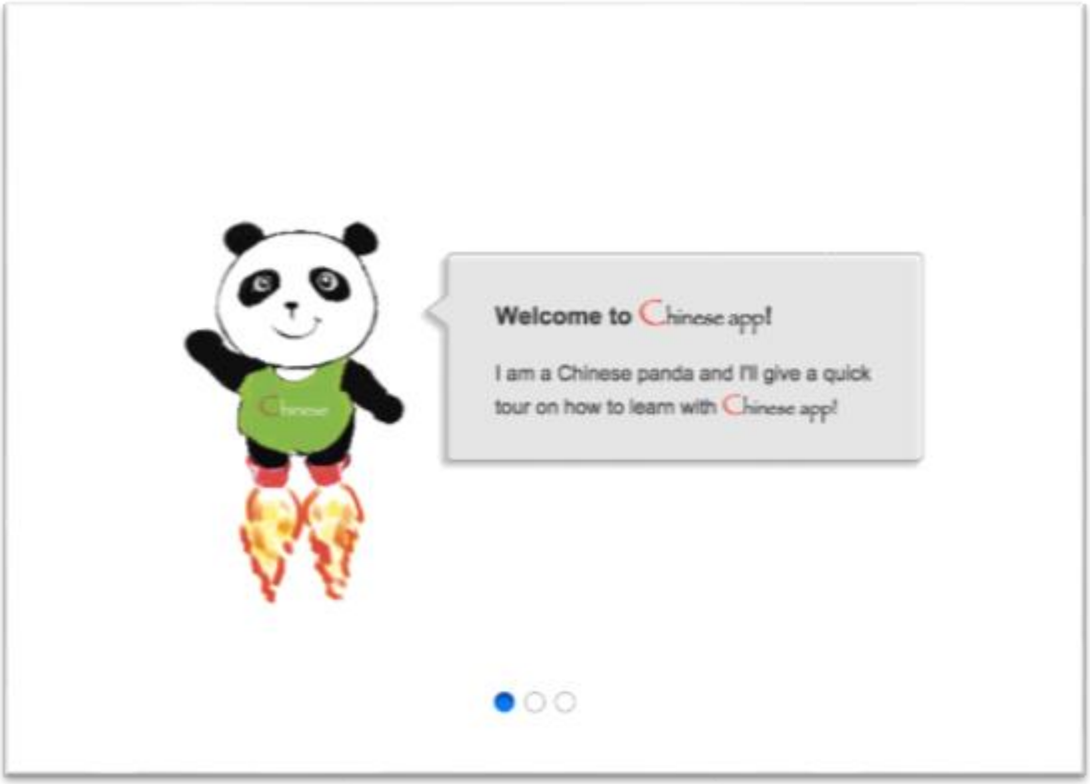

Chinese
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
translate the following word:
大众
(lowercase)
check

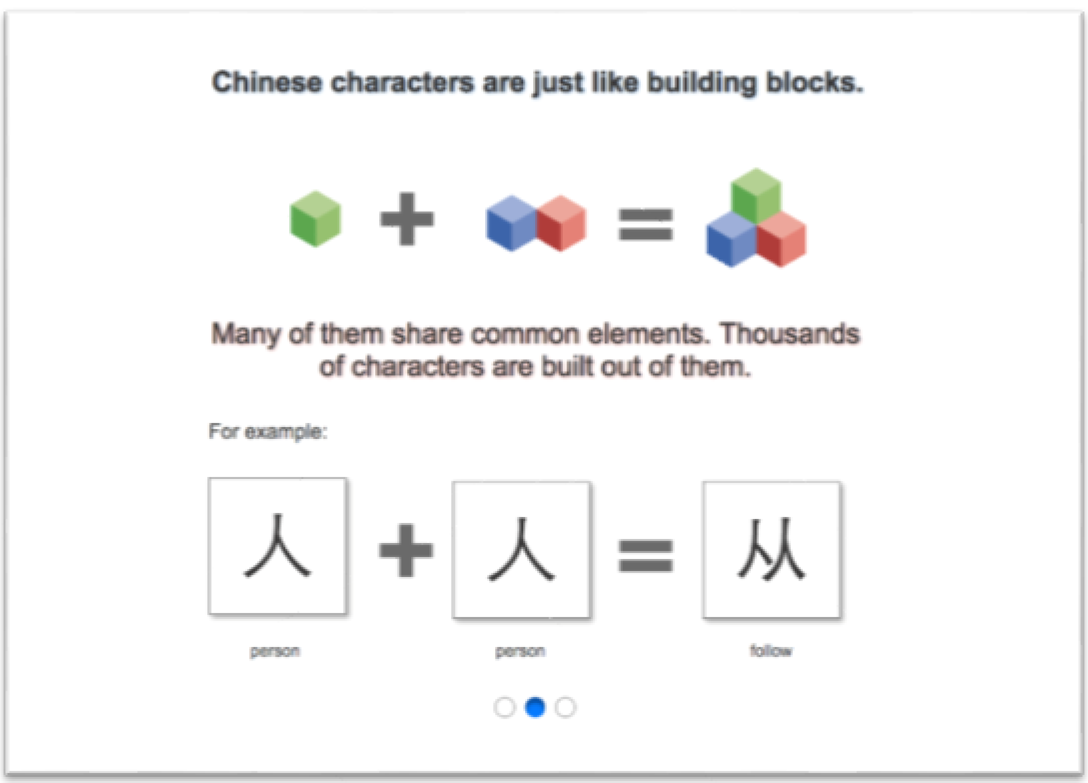

Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
山
mountain
门
door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
translate the following word:
大众
public
(lowercase)
You are correct!
Go to next course

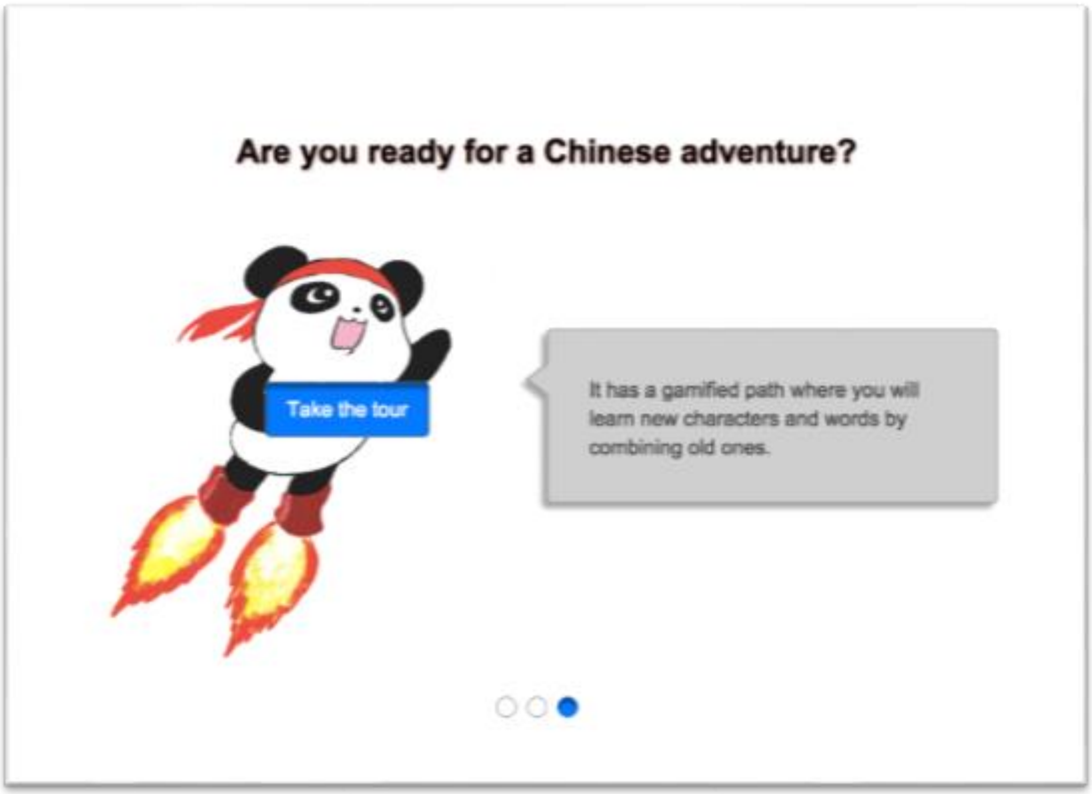

Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
山
mountain
门
door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
translate the following word:
大众
(lowercase)
Correct solution:
大众 the public
Go to next course
Learn this character again

VERSION B (PLAIN, fewer pedagogical interactions)

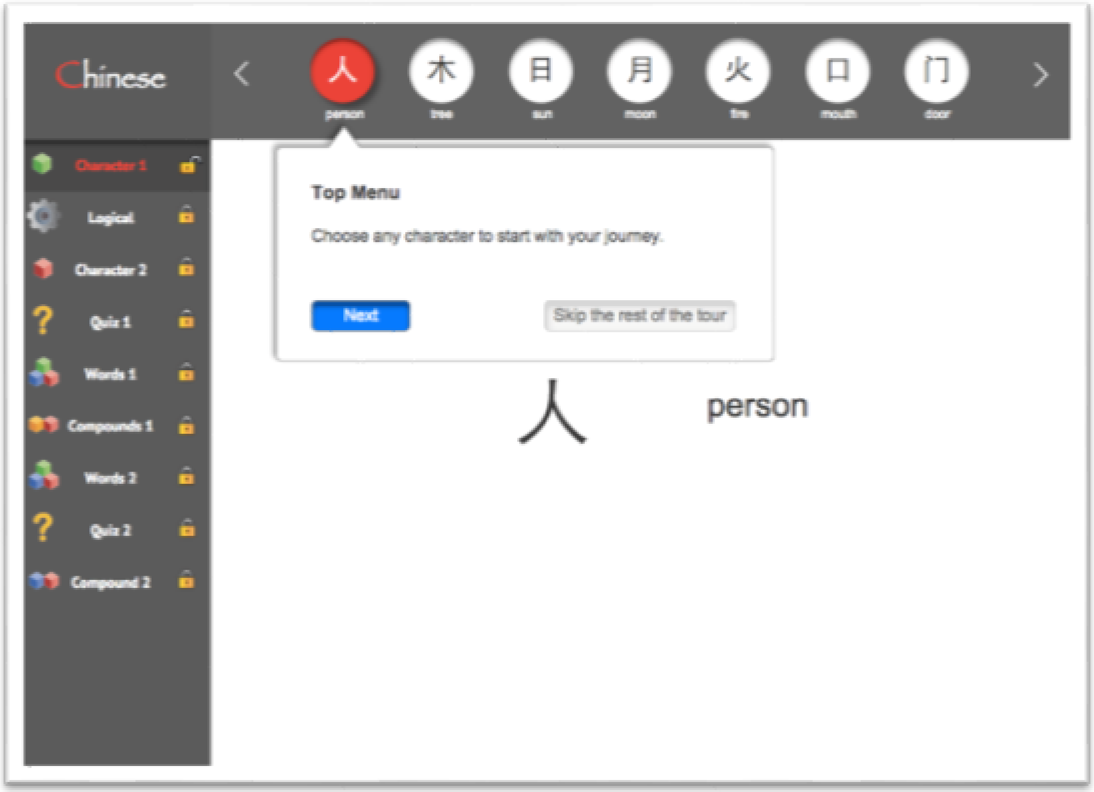
INTRODUCTION PAGES

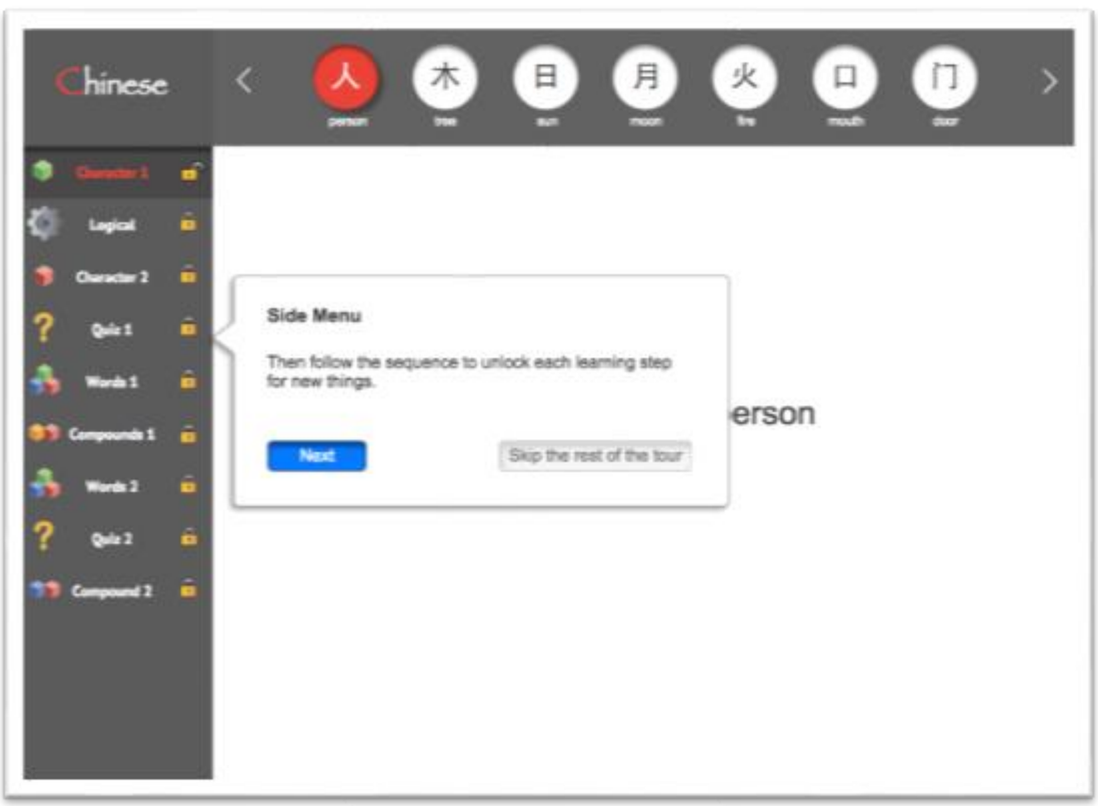


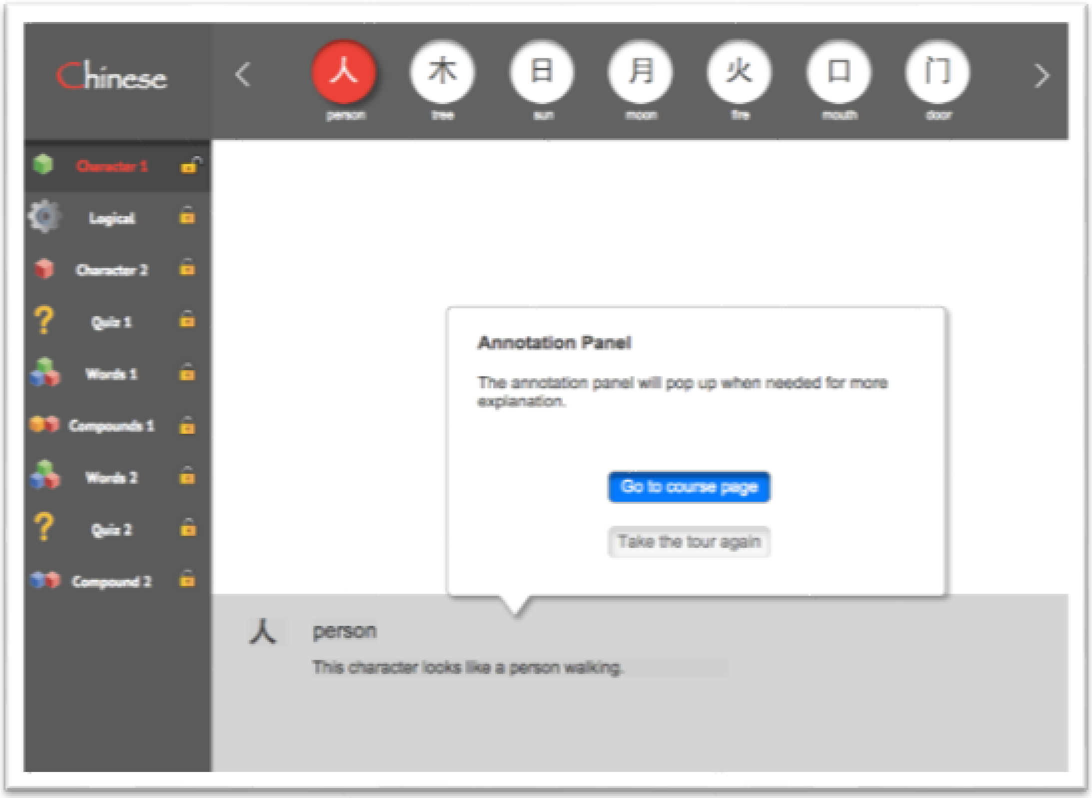



Are you ready for a Chinese adventure?
Take the tour
It has a gamified path where you will learn new characters and words by combining old ones.

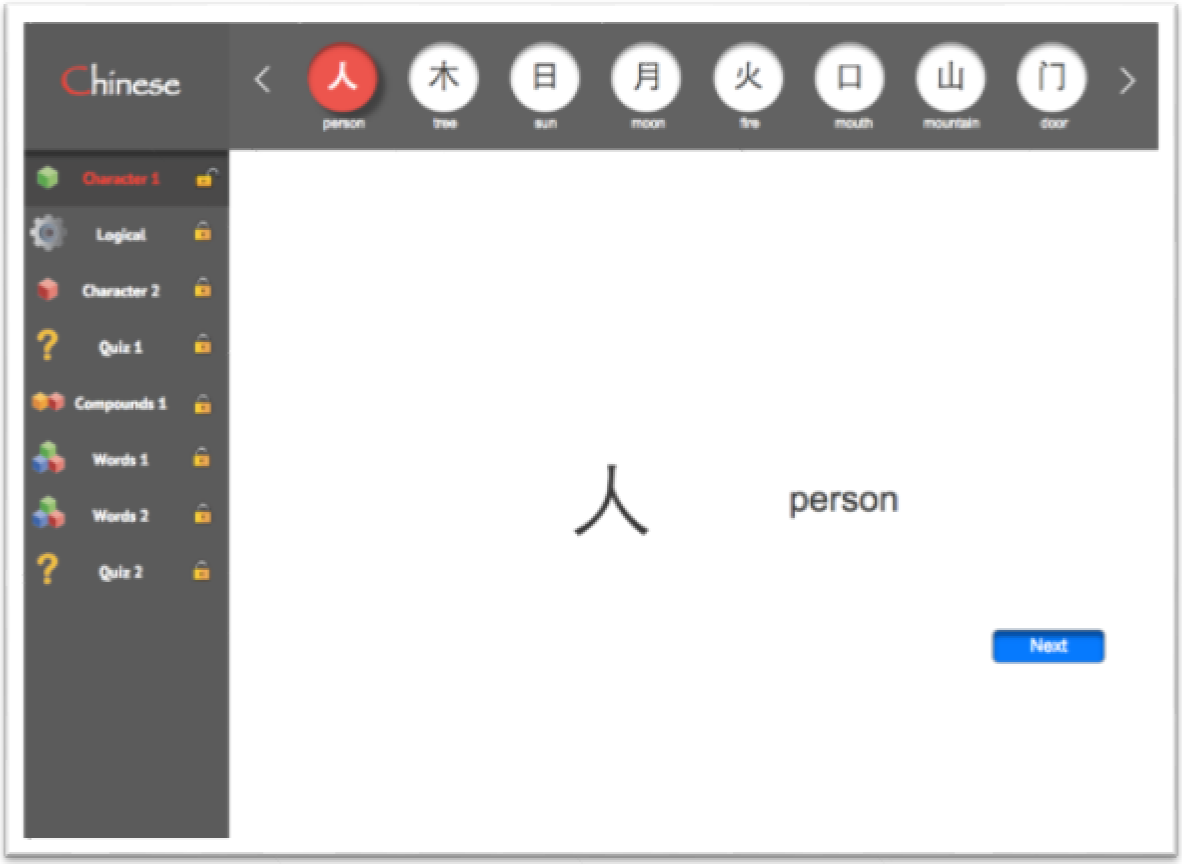
TOUR PAGES

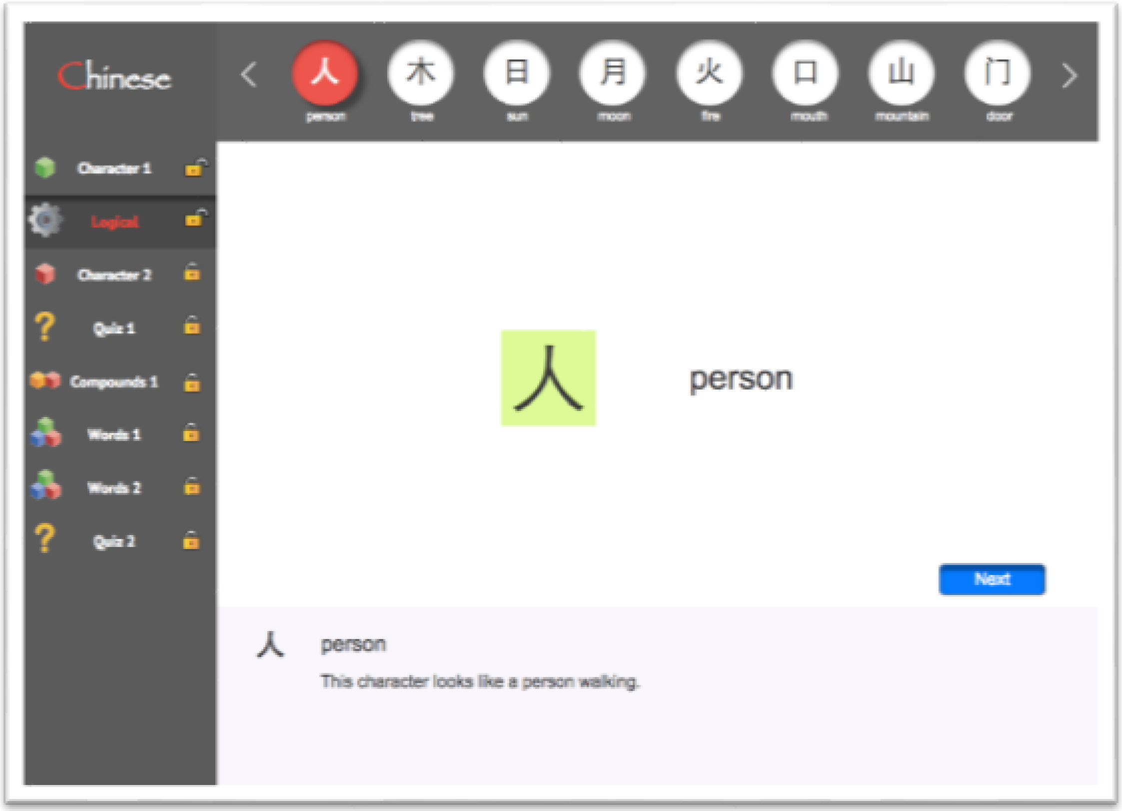

Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
门
door
Character 1
Logical
Character 2
Quiz 1
Words 1
Compounds 1
Words 2
Quiz 2
Compound 2
Top Menu
Choose any character to start with your journey.
Next
Skip the rest of the tour
人
person

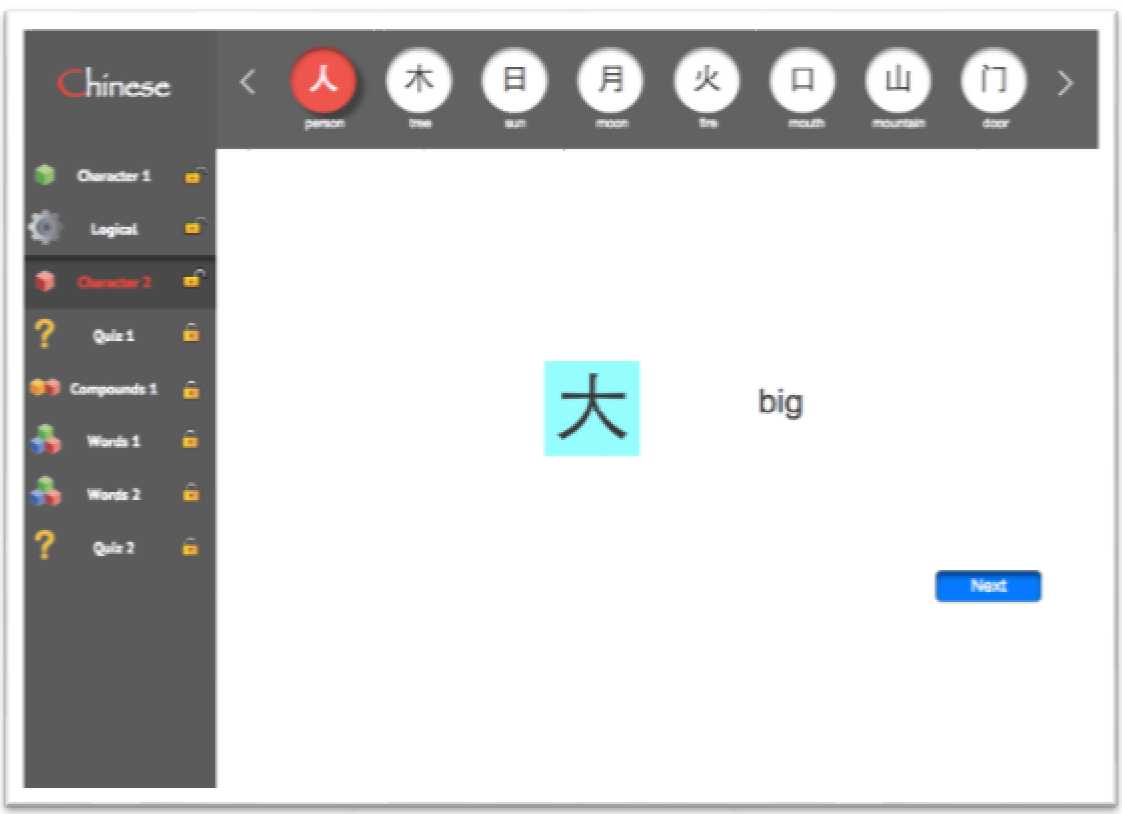

Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
门
door
Character 1
Logical
Character 2
Quiz 1
Words 1
Compounds 1
Words 2
Quiz 2
Compound 2
Side Menu
Then follow the sequence to unlock each learning step for new things.
Next
Skip the rest of the tour
erson


Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
门
door
Character 1
Logical
Character 2
Quiz 1
Words 1
Compounds 1
Words 2
Quiz 2
Compound 2
Annotation Panel
The annotation panel will pop up when needed for more explanation.
Go to course page
Take the tour again
人
person
This character looks like a person walking.

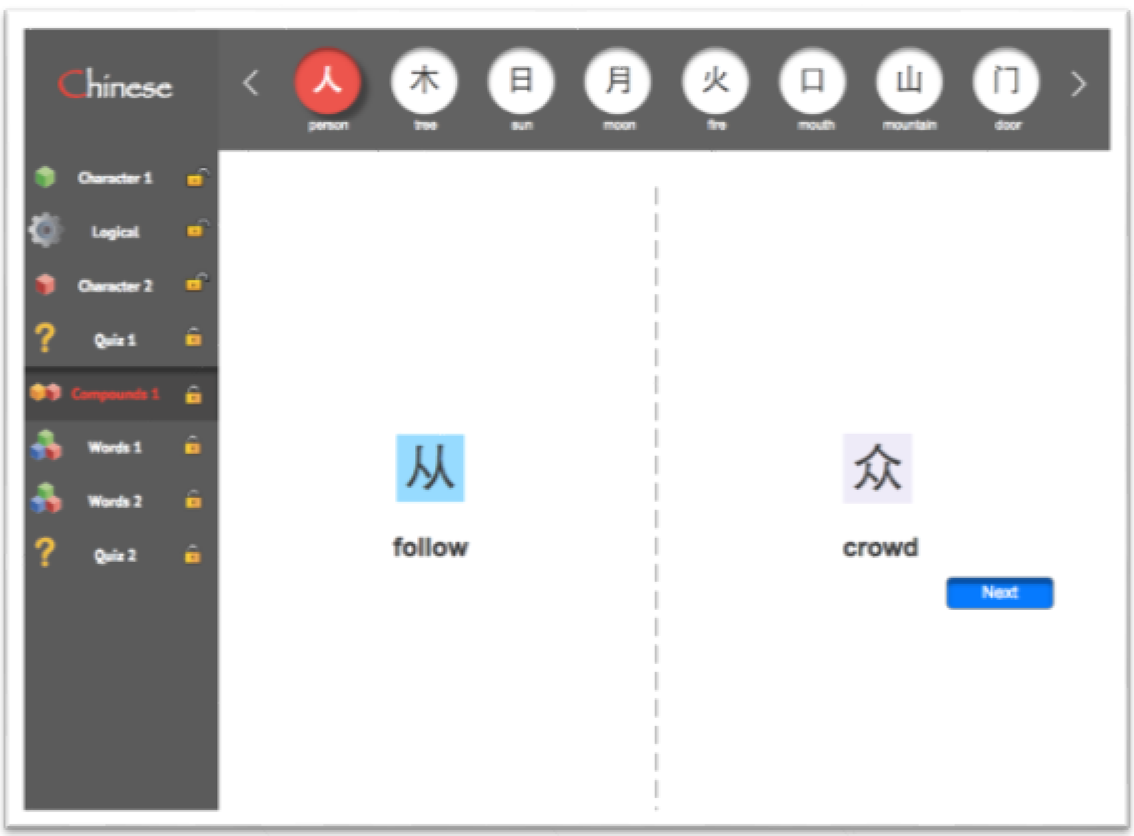
COURSE PAGES
---PERSON COURSE

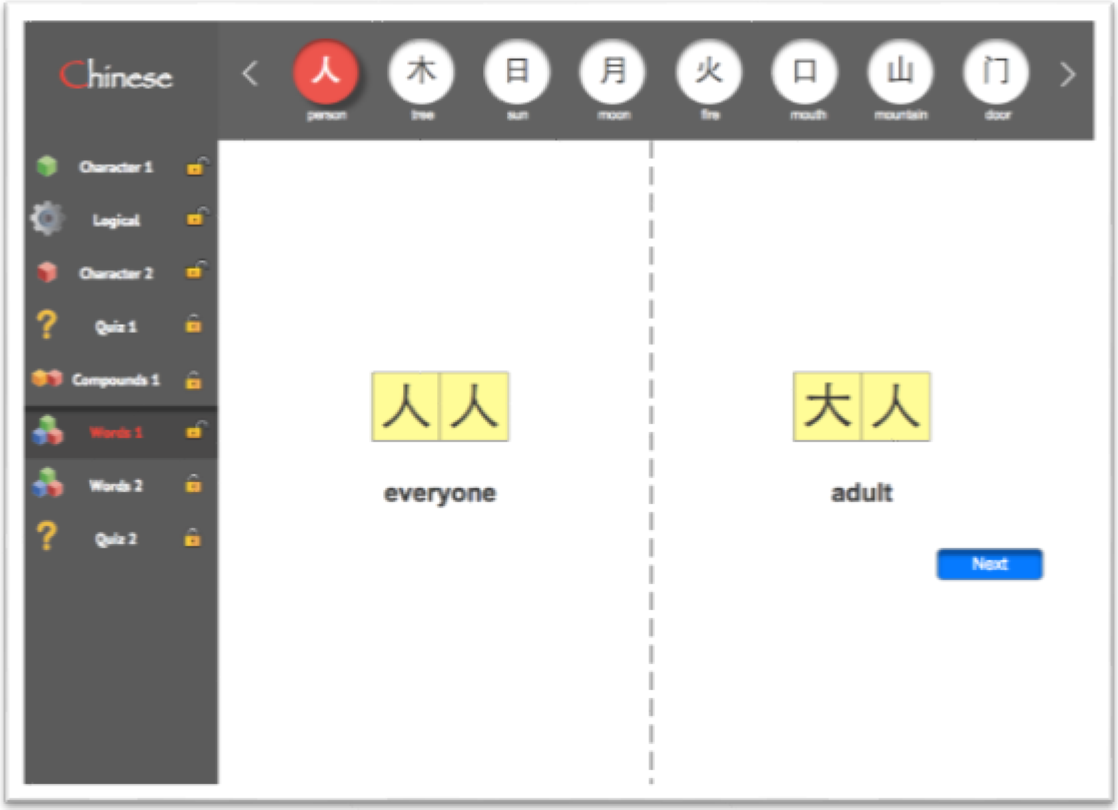


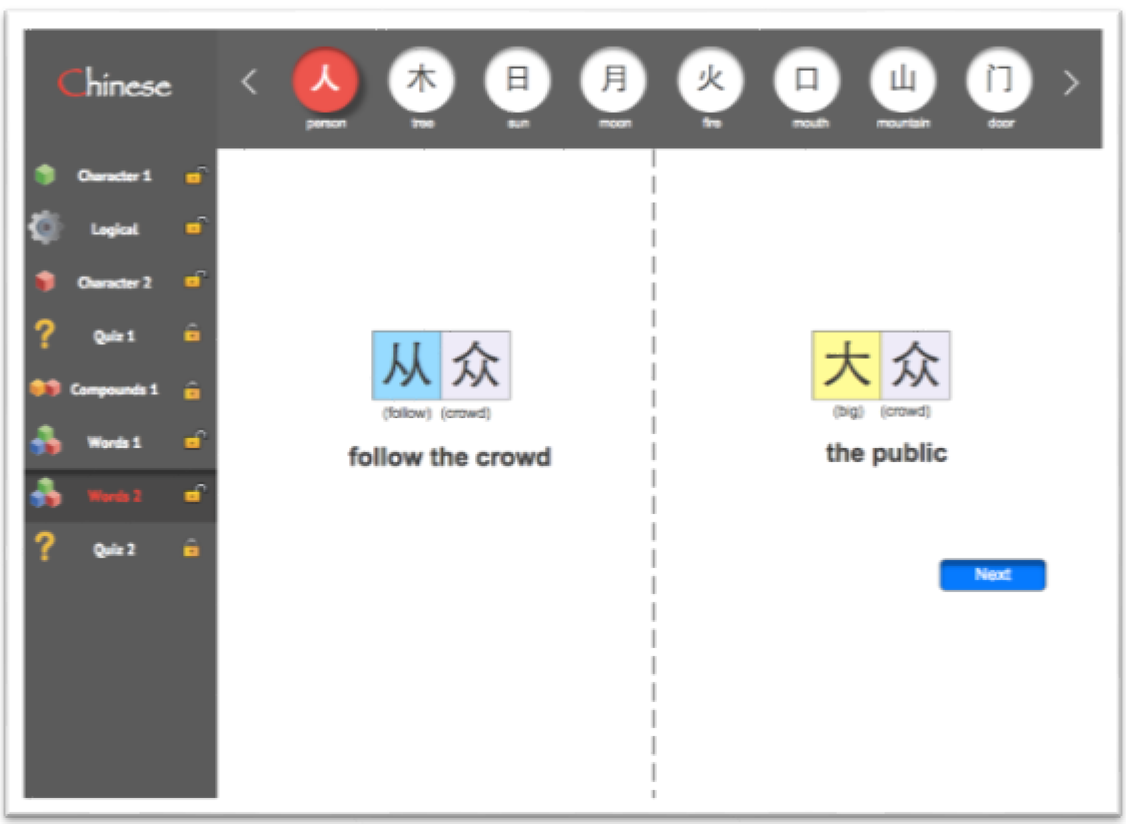

Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
山
mountain
门
door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
人
person
Next
人
person
This character looks like a person walking.


Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
山
mountain
门
door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
大
big
Next


Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
山
mountain
门
door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
Which do you think is the translation of "big"
大
人
小
Next


Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
山
mountain
门
door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
从
follow
众
crowd
Next


Chinese
人 person
木 tree
日 sun
月 moon
火 fire
口 mouth
山 mountain
门 door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
人人
everyone
大人
adult
Next


Chinese
人 person
木 tree
日 sun
月 moon
火 fire
口 mouth
山 mountain
门 door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
Quiz 2
从众
(follow) (crowd)
follow the crowd
大众
(big) (crowd)
the public
Next


Chinese
人
person
木
tree
日
sun
月
moon
火
fire
口
mouth
山
mountain
门
door
Character 1
Logical
Character 2
Quiz 1
Compounds 1
Words 1
Words 2
translate the following word:
大众
(lowercase)
Go to next course