Graduate School Form 9
(Revised 6/03)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis Acceptance

This is to certify that the thesis prepared

By YU ZHU

Entitled

Effects of Voiced-Pronunciation and Stroke Sequence Animation on Production of Characters by Beginners of Chinese as a Foreign Language

Complies with University regulations and meets the standards of the Graduate School for originality and quality

For the degree of Doctor of Philosophy

Signed by the final examining committee:

_____________ Chair 04-15-2005

_____________ 4/15/2005

_____________ 4/15/2005

George ______ /AG per phone 04/15/2005

Approved by: _____________ 11/9/05
Head of the Graduate Program Date

This thesis ☐ is ☒ is not to be regarded as confidential. _____________
Major Professor

Format Approved by:

_____________ or _____________
Chair, Final Examining Committee Department Thesis Format Advisor

EFFECTS OF VOICED-PRONUNCIATION AND STROKE SEQUENCE ANIMATION ON PRODUCTION OF CHARACTERS BY BEGINNERS OF CHINESE AS A FOREIGN LANGUAGE

A Thesis

Submitted to the Faculty

of

Purdue University

by

Yu Zhu

In Partial Fulfillment of the

Requirements for the Degree
of

Doctor of Philosophy

December 2005

To my mentors

ACKNOWLEDGMENTS

Blessed with unfailing support and assistance from my research committee, family, and friends, I managed to complete this study. Here I would like to take the chance to express my whole-hearted thanks to them all.

I would like to express my deepest gratitude to my major professor, Dr. Alan Garfinkel, for all his advices, helps, and encouragements in making this study accomplished. Dr. Garfinkel has always been ready to offer me his feedback and insights timely and to the point. His guidance has helped me to intensify my research focus from time to time. Furthermore, Dr. Garfinkel and his dear wife Sonya took the extra work of editing my thesis word by word. Their reading of the dissertation many times lent to various changes that lead to a higher readability. Their dedicated attitudes affected me deeply and made me felt an even pressing obligation to do my absolutely best for the study. Dr. Garfinkel and his dear wife Sonya are one of the nicest couples I have ever met. I admire Dr. Garfinkel not only as my professor but also as a great person and a dear friend.

To Dr. Wei Hong, who served on my committee and as my teaching supervisor at Purdue University, without her helps, the study would never be possible. As an expert in the field, she continuously provided valuable advices in forming ideas for this study. For nearly every step, I got valuable clues and detailed suggestions from Dr. Hong. I also appreciated the chance to work with her to finally have this study published in a core journal of the field. Dr. Hong challenged me to be more explicit all the time. Her ideas, criticisms, and suggestions helped to shape both this dissertation and the journal paper based on the dissertation. Under her guidance, I was able to grow up as a researcher and a practitioner in the field.

I am also appreciative of the insights, expertise, and practical guidance of my two other committee members: Dr. Charles A. Elster, and Dr. George Font. With their expertise in research methodology, I was able to improve my study to great extent. They also helped me to develop a research agenda right after completion of this study.

To Shanwen, my dearest wife, life partner, and best friend, without whose unconditional care, love, and support, I can hardly imagine how to survive all the frustrations and hardships.

My deepest thanks also extend to the support of my family. Dr. Mei Zhu, my dear sister and Fei Chi, my brother-in-law, whose wisdom, love, and generous spirits have touched me more than they can possibly know.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Zhu, Yu. Ph.D., Purdue University, December, 2005. Effects of Voiced-Pronunciation and Stroke Sequence Animation on Production of Characters by Beginners of Chinese as a Foreign Language. Major Professor: Alan Garfinkel.

In the field of teaching Chinese as a foreign language (TCFL), many beginning learners were attracted by the unique Chinese writing system, but meanwhile, they also experienced considerable frustrations in production of learned characters. The literature has focused on character recognition, and suggested ways to improve recognition. But so far, there have been very few studies concerned production of characters. Furthermore, although learners' needs and concerns were undoubtedly felt, limited by classroom resources, Chinese teachers normally could do very few to address character production problems during teaching sessions. Thus, learners typically had to struggle lonely through the process of character production. Fortunately, with more and more sophisticated computer-based multimedia instructional materials becoming available in the field, learners of today have access to various out-of-classroom resources on character memorization and production that were unavailable to many earlier learners. Studies on the effects of these electronic materials on learners' character production, however, remain scarce.

This study was designed to examine the effects of voiced-pronunciation and stroke sequence animation, two frequently selected inputs in Chinese multimedia flashcards, on production of characters by beginners of Chinese as a foreign language. One hundred students, enrolled in an elementary Chinese course at a large Mid-western university, participated. Students were divided into four treatment groups. Four different formats of computer-based multimedia flashcards of the same vocabulary list were developed for this study. Each group was required to memorize and produce afterwards the characters on the assigned flashcards. The instrument was administered as one pre-test and two post-tests to examine the individual and interaction effects of the two concerned media inputs.

Findings of this study were that the group that viewed characters displayed with both voice and animation (WVWA) performed significantly worse, while the group with voice only (WVNA) scored much better than the rest of the experimental groups. Explanations and discussions were provided from the perspective of dual-coding theory (Paivio, 1986; Clark & Paivio, 1991). Further investigations on the effect of stroke sequence animation are needed to determine, for example, whether this medium can assist learners' mastery of proper stroke sequences.

CHAPTER 1. STATEMENT OF THE PROBLEM

1.1. Introduction

For the majority of beginning learners of Chinese as a Foreign Language (or CFL beginners henceforth), learning to write Chinese characters is believed to be one of the most difficult tasks. As X.-L. Li observed (1996), "high dropout rate in Chinese language learning has much to do with its writing components known as characters." (p. 77). However, this most difficult part of learning Chinese is unavoidable because to be literate in Chinese, one has to master at least 3,500 high frequency hanzi (Norman, 1996). This may also be the underlying reason why *Practical Chinese Reader* (Beijing Language Institute, 1983), the most popular CFL textbook in the past 20 years, requires learners to master about 600 Chinese characters during the first year of study, which is comparable to the number mastered by native Chinese primary school students (S.-H. C. Wang, 1998; see also Serruys, 1962; Leong, 1973). In all philosophies of CFL teaching and learning, memorization of Chinese character undeniably contributes to success. Practically speaking, whether to memorize Chinese characters or not is not the question. It is what Chinese teachers should do to facilitate and transfer learners' memorization that matters. This thesis centers its focus on the effects of fundamental multimedia variables on

Chinese character (i.e., hanzi) memorization of CFL beginners. The rationale will be briefly introduced as follows and further discussed in the literature review.

1.2. Problem Statement

CFL beginners usually struggle to memorize hanzi. For example, S.-H. C. Wang (1998) documented that virtually all (around 93.33 percent) CFL beginners in her study "strive to memorize" when learning a new hanzi. Over 80 percent of these beginners self-reported that their experience of studying hanzi is time-consuming. Furthermore, one third of them found their hanzi learning experience frustrating. Ke echoed this finding in his empirical study (Ke, 1996) and found that CFL learners tend to perform significantly worse in production tasks than in recognition ones. Similarly, Usuki (2000) reported that Westerners found it difficult to produce proper kanji (i.e., Japanese scripts borrowed from the Chinese writing system).

The causes of such struggle and frustration, however, are seldom systematically explored. S.-H. C. Wang (1998) and Usuki (2000) are examples of the few exceptions. In her study, Wang introduced contradictory phenomena familiar to many Chinese teachers and learners. On one hand, CFL beginners feel that learning and producing hanzi is painful. On the other hand, over 80 percent of them would like to study hanzi by themselves instead of with classmates. On one hand, the teacher acknowledges and feels apologetic that her students are struggling with hanzi. On the other hand, she believes "studying and memorizing hanzi should be an individual activity" (Wang, 1998, p. 76) and leaves learning hanzi a task for her students to accomplish outside the classroom. Ma (2000, p. 1) observed "the most interesting and challenging aspect of studying Chinese is writing Chinese characters. Unfortunately, the learning of Chinese characters receives only marginal attention in a typical classroom." Ma further pointed out that CFL learners' struggles with hanzi are at least partially due to the lack of systematic introduction of Chinese hanzi morphology in textbooks and the time pressure of a CFL classroom. Usuki (2001) reported that teaching kanji in Japanese as a foreign language classes is also viewed by many teachers as a task students should do alone, despite the fact that learners of Japanese found this to be very intimidating. Wang argued that CFL beginners' experience of learning hanzi should not be such a lonely, frustrating experience, but fail to advise the reader how to improve this situation in the limited class time available for learning hanzi. Learning hanzi, although important, is only one of the skills covered in CFL syllabi. Since both CFL teachers and learners seem to assume that learning hanzi should be an individual effort outside the classroom, a realistic solution to the problem could be the development of resources that assist CFL learners in independent study.

Well-designed multimedia flashcards are one example of such tools thanks to easier accessibility to Computer Assisted Language Learning (or CALL, henceforth) resources. Recent multimedia flashcards in CFL are computer-based, systematically displaying target words and mnemonics through simultaneous presentation of multiple media such as text, graphics, voiced pronunciation, and animation or video.

Multimedia flashcards of Chinese vocabulary have, as do many other CALL programs and products, potential advantages that are hard to realize in conventional classroom teaching situations. Such advantages include easy accessibility without time and space restraints, standardized inputs, learner-controlled instructional sequences, and so forth. However, these general advantages have not necessarily led to more effective hanzi learning outcomes than has conventional classroom teaching. Further investigation into the outcomes of multimedia flashcard usage is necessary for both theoretical and practical reasons. Identification of these factors will enrich learning theory, especially CALL theories. Practically, these findings will help instructional multimedia designers and consumers.

Unfortunately, currently available literature on multimedia flashcards in CFL is not helpful in the sense that it usually includes only specific product descriptions, program reports, and/or evaluations, which contribute little information on the specific factors leading to successful hanzi learning (Lam, 2001). Some studies even do not provide any testing for the effects of such products. For example, Robert (2002) observed that although knowledge of Chinese radicals has always been considered to facilitate hanzi learning, classroom teaching environments usually do not allow time to thoroughly teach them. To solve the problem, Robert created a computer-assisted hanzi radical learning program and simply claimed expectations of enhanced motivations, as well as hanzi memory, without performing any testing. A few studies have identified beneficial variables that may enhance learning. For example, Lu, Webb, Krus, and Fox (1999) found an effect of instructional sequence on character recognition. The sparse literature on multimedia flashcards is focused on the effects of such fundamental media factors as voiced pronunciation, stroke sequence animation, and so forth.

A related problem rests in the concerns of current CFL hanzi learning studies. Despite the fact that hanzi production is very difficult for many CFL beginners, studies in CFL hanzi learning have instead focused on hanzi recognition (e.g., Lu et al. 1999; Mori & Nagy, 1999; Sergent & Everson, 1992). This has resulted in a surprisingly sparse body of literature on this crucial issue. Consistent with findings that recognition and production are two distinct aspects of vocabulary learning, Clariana and Lee (2001) found that production tasks are more effective, although not to a statistically significant degree, than recognition tasks in vocabulary learning. This finding lends further credence to investigations on vocabulary learning concerning word/character production.

In summary, current hanzi learning practice and studies reveal the need to more thoroughly explore the effects of multimedia flashcards on the production of hanzi.

1.3. Purpose of the Study

Given the fact that most computer-based CFL vocabulary flashcards integrate various media in their design, and given the problem that very few studies have specifically examined how fundamental media variables (such as auditory and/or imagery inputs) affect learners' performance on character production tasks, this study will examine the effects of some widely applied multimedia factors on CFL learners' production of Chinese characters. Specifically, the question that this study asks is:

> *What are the individual and collective effects of voiced- pronunciation and stroke sequence animation in flashcards on beginning CFL learners' production of Chinese characters?*

The rationale for choosing voiced-pronunciation and stroke sequence animation as variables to be examined is stated as follows.

As argued by Usuki (2000), one of the difficulties most CFL beginners struggle with is the pronunciation of hanzi. By providing CFL beginners with the sound of hanzi, the majority of current multimedia flashcards do address this problem. Learners can not only see the pinyin (Romanized Chinese pronunciation notations) of a hanzi but also can actually hear a native speaker pronounces it. Furthermore, this aid can be accessed as many times as desired. In the early 1970s, studies (Kappel, Harford, Burns, & Anderson, 1975; Tell & Ferguson, 1974) were conducted to compare active vocalization versus passive vocalization (hearing the sound) while memorizing vocabulary in the target foreign language. The findings showed that participants who were required to do active vocalization had lower retention scores than those who were exposed to passive vocalization, which suggested that for long-term retention, hearing vocabulary pronounced is superior to active vocalization of the vocabulary “when the information in echoic memory has dissipated” (Tell & Ferguson). Digital voiced-pronunciation of new words in flashcards resembles the passive vocalization situation in the above studies. Thus, by examining the effect of this type of input, the hypothesis that such a multimedia input can improve CFL beginners’ hanzi memorization can be tested.

Another difficulty reported by many CFL beginners is the visual complexity of hanzi. This makes it difficult for learners to discern a character’s structural components or to reproduce the proper character shape as Usuki (2000) reported. To address this difficulty, animations of target characters stroke by stroke and part by part are widely included in current CFL flashcard designs. This stroke animation resembles finger writing on a computer screen. The positive effects of finger writing on kanji learning has been reported by Sasaki (1987) who reviewed several experiments examining the cognitive function of finger writing, an activity widely used by native Japanese and thought to be helpful in kanji learning. By investigating the effect of such an input, the hypothesis that multimedia input will enhance CFL beginners' memory of hanzi can be studied.

Despite the fact that voiced pronunciation and stroke sequence animation were used in most CFL multimedia flashcards, there is a lack of evidence regarding their efficacy. Taking into consideration the resources needed for developing these learning tools, a careful examination prior to their development is definitely justified.

1.4. Significance of the Study

Addressing production task performance will contribute literature on CFL hanzi learning as well as address many CFL beginners' learning difficulties.

The findings of this study will also enrich our understanding of hanzi learning in CFL by identifying fundamental media factors that enhance learners' performance of character production tasks. This, in turn, can guide CFL teachers to better instructional media selection in complex CALL situations.

The literature of educational psychology regarding media effects on learning predicts that simultaneous auditory and textual inputs or concurrent imagery and auditory inputs will enhance memory (e.g., Clark & Paivio, 1991; Moreno & Mayer, 2002). However, the effects of media on learning in more complex, less controlled, hence more realistic situations such as multimedia flashcards in CFL remain unclear in currently available literature. Identifying the effects of voiced pronunciation and stroke sequence animation in computer-based flashcards, this study will also contribute to the body of literature on language learning in educational psychology.

1.5. Overview of Chapters

This study has five chapters. Chapter One introduces the problem and sets direction for research. The rationale and significance of the study is also briefly explained. Chapter Two is a review of literature constructs a theoretical framework for the study. A literature review of Dual Coding Theory (DCT) and Split-Attention Effect (SAE) are the two major topics, since both concern the effects of media or media combinations on learning. Studies on CFL hanzi acquisition are briefly discussed to support better ways to design. Research hypotheses based on the literature review are presented in Chapter Three, which also supplies the methodology of this study. Topics in this chapter include a description of sampling method, subjects, experimental situation, instructional tools, dependent and independent variables, and instruments. Validity issues in this study are also discussed in this chapter. Chapter Four focuses on data analysis where analysis tools, procedures and results are reported. Chapter Five draws conclusions based on the data analysis, and discusses the theoretical and practical implications of the study, and proposes directions for further study.

CHAPTER 2. REVIEW OF LITERATURE

2.1. Introduction: Rationale for Learning Hanzi

Everson (1988a) provided empirical evidence that American students can read faster and understand passages better when they are presented in pinyin (Romanized notations of Chinese pronunciation) rather than in hanzi. And as mentioned in Chapter One, CFL learners usually struggle painfully when learning hanzi. So, why bother to teach hanzi to CFL learners anyway? Why not just teach pinyin? An intuitive answer to these questions is that Chinese do not use pinyin as a means of communication in writing. For example, shoppers in a supermarket who do not know hanzi will have great difficulty. This is not like Japanese which uses hanzi (called Kanji in Japanese) but also allows extensive use of hiragana and katakana (which denotes word pronunciation like pinyin in Chinese) in its writing system. But why don't the Chinese use pinyin in their writing? To better understand this, some knowledge of the characteristics of the Chinese writing system and the history of the development of modern hanzi is necessary.

Chinese lacks sound-symbol correspondence. The written symbols do not represent pronunciations as well as many other writing systems (Yang, 2000; see also Koda, 1996). There are large numbers of homophones (pronounced exactly the same, including rhyme and tone) in Chinese. For example, without a context, by pronouncing the pinyin of "zhū," even native speakers can not whether tell whether it refers to 诸 , 蛛, 株, 珠, 朱, 铢, 洙, 诛, 侏, or 茱 . Although all these ten hanzi are pronounced exactly the same, each of them has a unique definition and use. Furthermore, tones differentiate possible rhymes in Chinese. In other words, the same rhyme with different tones may refer to different hanzi. According to Chinese linguist, Zi-Jie An (Qin, 2000), as many as 1,672 hanzi share the same rhyme with over a hundred other hanzi. Laychuk (1983) has noticed that Chinese has significantly fewer phonetic possibilities as compared with Indo-European languages and called this an advantage. But given the burden required for subtle hearing and accurate pronunciation and the easy confusion that it may bring to oral communication, it is arguable whether this is actually an advantage or a disadvantage. As Allen (1992) reported, in oral communications with clearly a defined context, native Chinese still need to clarify their speech from time to time by referring to the shapes of hanzi because of large numbers of homophones in Chinese. For example, if a Chinese pronounces his or her last name as Zhāng, usually a speech will follow to clarify the components of this Zhāng as 弓and 长or 立and 早, because 张and 章both can be used as a last name and be pronounced exactly the same as Zhāng. Therefore, even to make authentic oral communications successful, CFL learners need to know the shapes of hanzi. For written communications, it is even clearer why one needs to know how to produce hanzi. If CFL learners do not know any hanzi shapes, they can not read a text. If CFL learners do not know how to produce hanzi, they can not express themselves by writing. Even with the help of Romanized computer input method, one still needs to know shape of the target hanzi to choose it from the number of homophones that the machine will present.

Actually, in the 1950s, some linguists once successfully persuaded the Chinese government to simplify and reform the Chinese writing system, and the final goal of this reform was to replace hanzi with pinyin. The first attempt to simplify hanzi was made in 1956 and was successful. It resulted in simplified Chinese characters, now used officially on mainland China and in Singapore. The second simplification attempt in 1977, however, received very negative feedback and was abandoned in 1986. Thus, the final goal of replacing hanzi was proved to be unrealistic (Qin, 2000). To illustrate the absurdity of the final goal for Chinese writing system reform, Yuanren Chao, one of the most famous Chinese linguists, purposefully compiled an 83-word story which is logical and meaningful in hanzi but nonsensical in pinyin (Chao, 1992). In pinyin, it became 83 "yi" sounds, and readers have no idea what these "yi" referred to. After computers became popular in the 1980s, hanzi faced another chance of being replaced by pinyin. First, hanzi required much more hard disk space to display than alphabetic writing systems, which was impossible for many computers at the time. Second, input of hanzi was much slower than for many other languages. With the development of more sophisticated computer hardware, displaying hanzi is no longer a problem (Zhong Hua Zi Jing, 2004). Since the WBZX (五笔字型) method, created by Yong Min Wang, allowed one to input a hanzi by hitting four keys without looking at the keyboard, the input speed of hanzi became at least comparable to English (Qin, 2000).

In summary, hanzi will not be replaced by pinyin in the near future. Knowing how to actually produce hanzi or at least knowing the shapes of hanzi is necessary not only in written but also in oral communication. Learning the shapes of hanzi is a challenge CFL learners have to face. The problem for CFL educators is to find out and apply effective approaches for teaching hanzi.

2.2. Existing Studies on Hanzi Production

There are a number of studies on hanzi recognition or recognition tasks as measurements of learning. However, studies on hanzi production are much fewer. An extensive search for related academic journal papers and theses discovered only Ching and Ching (1975), Mickel (1980), Liu (1983), Ke (1996; 1998a; 1998b), X.-L. Li (1996), S.-H. C. Wang (1998), Yin (2003), and W. Li (2004). Some of these studies were mentioned in Chapter One (S.-H. C. Wang; Ke, 1996). A brief review of others will be presented in this section. However, this review of literature does not include descriptions or evaluations of CALL programs. Since available literature only rated the effectiveness of these programs as a whole without identifying effective factors and without using statistical inferences, there is no way to relate such findings to the current study.

Ke (1998a) and Yin (2003) observed CFL learners' hanzi learning strategies. Ke found that non-heritage learners have similar hanzi learning strategies as do heritage learners. However, "more non-heritage learners considered practicing character writing more effective than reading a character text for character learning." "More non-heritage learners considered associating new characters with characters they are already familiar with in terms of graphic structure to be more effective than associating them with sounds" (Ke, p. 102). Ke also found that although CFL beginners perceived "the learning and using of knowledge on graphic structure and character component" in hanzi learning as valuable, most of them still simply memorize a hanzi as a whole, by repetitively producing it. Yin found that of the three essential elements (i.e., shape, phonology, and semantics) of hanzi, shape production is the most difficult to them. He also found the most frequently used strategy to memorize hanzi is to write them over and over. Ke explained that because CFL beginners have mastered only a few hanzi, they are limited in their ability to use hanzi dissembling strategy when memorizing them. Writing hanzi over and over, though extensively used by CFL learners (especially beginners, see Ke 1998a; De Francis, 1984), has been found to be as not as effective as mnemonic methods based on information-processing theories (e.g. S. Li, 1997; W. Li, 2004).

The "semantic analysis method" discussed in S. Li (1997) links hanzi's form to its meaning. Such information is used as mnemonics in his study. For prompt and delayed post-tests, mnemonic groups taught with the semantic-analysis method performed much better in both recognition and production of hanzi than did groups using the rote memory method. S. Li also pointed out that this method of learning hanzi is, to an extent, similar to the way in which Westerners learn words in their native languages. Furthermore, S. Li concluded that this method of teaching hanzi works well for hanzi production, especially when one considers its similarities to the "semantic analysis method." The "grapheme combination method" discussed in W. Li (2004) emphases taking advantage of graphemic information. He also found that the group treated with "grapheme combination method" was superior in Chinese reading tests. The underlying rationale is that complex hanzi are usually constructed with several simple but meaningful graphemes. By learning the graphemes and their combination rules, learners have more opportunity to better understand the target hanzi, review previously learned hanzi with common graphemes, and build up new hanzi in related schema.

Ke (1998b) found differing perceptions of the importance of sound in hanzi learning as compared with Everson (1998b). While sound was found to be a very important factor in hanzi recognition tasks in Everson's study, Ke found that most learners (64% ~ 88%) self-reported that sound was not valuable. Producing was seen as more effective than reading when learning hanzi; meaning was seen as more effective than sound when producing hanzi, and associating new hanzi with familiar ones using graphic structure was seen as more effective than using sound. Ke doubted that these opposing findings could be due to the different pedagogies used in the two studies. He also called for data other than self-report to be used in further investigations on this topic.

Ke (1998b) studied the effects of language background on CFL learners' hanzi acquisition. Counter-intuitively, CFL learners' language background (heritage or non-heritage) was not found to affect learners' performance significantly on either hanzi production or recognition tasks. Ke conjectured that this could be because all the learners in the study were in the early hanzi "accumulation stage" (see Ke, 1996b), when heritage learners have no significant advantages in hanzi production and recognition than do non-heritage learners.

CFL learners self-reported that producing hanzi accurately is their most difficult task (Yin, 2003). When generating factors that distinguish one hanzi from another, Ching & Ching (1975, p. 22-24) considered "manipulations of strokes" as the primary one, since similar but slightly different strokes (e.g. "犬 versus 尤"), number of strokes (e.g. "一 versus 二"), position of strokes (e.g. "犬 versus 太"), length of strokes (e.g. "目versus 且 "), etc. all directly contribute to distinguish hanzi one from another. The position of components is also a noticeable factor for distinguishing compound hanzi (e.g. "杏versus

呆" and "叨versus 召"). As a result, the authors emphasized that CFL beginners need to form "good habits" and be given as thorough as possible instruction by CFL teachers, including issues associated with stroke production (legibility, position, size, and length) and spacing (especially for compound hanzi). Meanwhile, learners should also be taught the sound and meaning of a hanzi when producing it so as to derive its correct applications.

Advocating early exposure of CFL beginners to hanzi with a conceptual learning approach, Liu (1983) argued that mastery of recurring radicals of compound hanzi provides invariant meaning cues which can effectively enhance hanzi learning. She further suggested that phonetic compound hanzi should be taught first as they integrate the three elements (orthographies, phonetics, and semantics) of Chinese characters. She advocates conceptual learning based on a better understanding of hanzi component construction rules. The idea of exposing CFL learners to hanzi early is also echoed in Mickel (1980). He asserts that by helping students in their first year of study learn a solid foundation of hanzi origin, script types, and component analysis etc. and by helping them to better understand the logic underlying the Chinese writing system (origin, basic strokes and stroke order, radical and phonetic component rules, etc.), CFL beginning learners can overcome their anxieties about hanzi learning which benefit their subsequent studies.

In summary, a brief review of existing literature on hanzi education confirmed the necessity of exposing CFL beginners to hanzi early and revealed the importance of instructing them in the recurring phonological components and/or radicals that compose a hanzi. A mastery of basic strokes and correct stroke orders are considered by many researchers to be important for the correct production of hanzi. The pronunciation of hanzi also receives considerable attention in studies of hanzi learning. However, the effects of stroke order and/or sound on hanzi form memorization is still not clear and has not been carefully explored.

The review of literature will continue with studies concerning topics such as the inherent features of hanzi and information processing theories on memory in CALL conditions. The following reviews include studies using native Chinese as subjects in order to determine the factors which influence the recognition of hanzi.

2.3. Hanzi Morphological Characteristics

Hanzi morphological characteristics here refer to hanzi's most prominent features in structure and the procedure of production. In structure, most hanzi are phonetic compounds which combines phonological components and radicals. In the procedure of production, native Chinese put much emphasis on stroke sequence.

2.3.1. Structural Features and Educational Implications

It is acknowledged that although hanzi are categorized into six different categories (i.e. six script types, or 六 书 / liù shū in Chinese), the phonetic compound script (i.e. 形 声 / xíng shēng in Chinese) comprises over 70 percent of hanzi used today. The most prominent feature of xíng shēng hanzi is that they usually consist of two components, with one component implying meaning and the other approximating pronunciation. The meaning indicators are also known as radicals. Herein, pronunciation indicators will be called phonological components. Since radicals and their corresponding meanings are relatively stable and traceable across time, thanks to the available historical documents, they thus receive more attention than phonological components. A phonological component within a hanzi was created at the time that the specific hanzi was created. Although the correspondence between a phonological component and its actual pronunciation might have been very strong originally, it could be much weaker or even seem irrelevant today because of phonological changes over time. This is one of the most important reasons that phonological components are frequently ignored, intentionally or unintentionally, in hanzi teaching. However, it is still arguable that although some phonological components may be pronounced differently today, many more phonological components still retain their original pronunciations or at least the original or similar finals. Even for those hanzi whose modern pronunciations are unrepresented by their phonological components, usually called irregular hanzi, it is still good to tell learners that these phonological components once represented the actual pronunciation of characters so as to allow them to understand hanzi structures better. Actually, according to a study examining the 2,570 hanzi taught in Chinese primary schools, Chinese has a writing system with high learnability, because the few irregular hanzi which are taught first also have high frequency and are written with only a few strokes. Although target hanzi tend to have less frequency and many more strokes, they are mostly regular. This means one can use morphological knowledge to predict their meaning and pronunciation more accurately (Shu, Chen, Anderson, Wu, & Xuan, 2003). Moreover, being able to predict pronunciation with morphological knowledge will also benefit reading ability because it can help the learner better guess both the pronunciation and meaning of an unknown hanzi. A comparative study (Ku & Anderson, 2003) suggests for Chinese-speaking children, as for English-speaking children, morphological awareness develops according to grade level and is closely correlated with reading ability.

2.3.1.1. Phonological encoding evidenced in mental processing of hanzi

It is almost a consensus in the field that the phonological information of words is activated in mental processing, even for logographic writing systems such as Chinese or Japanese. For example, Saito, Masuda, and Kawakami (1998) reported four experiments which consistently indicated that phonological information will be activated in Kanji matching tasks even when pronunciation is not involved. However, there are different views of at what level, phonological component or whole-character, phonetic information of a logographic character is activated.

Using native Chinese college students as subjects, Tan, Hoosain, and Peng (1995) found that high-frequency, "vague-meaning" hanzi stimuli are first processed phonologically, although clear evidence as to whether semantic or phonological activation is first accessed for high-frequency "precise-meaning" stimuli was not found. Tan et al. further argued that pre-semantically activated phonological decoding occurs not at the component-level (i.e. phonological component) but at the whole-character level. This is also consistent with findings in earlier studies (see also Fang, Horng, & Tzeng, 1986; Hue, 1992; Peng & Yang, 1994; Wu, (Wu, 1994)). As Tan, Hoosian, and Siok (1996, p. 876) explained, the so called "character-as-a-whole-to-sound-as-a-whole association" is made possible because: (1) a very limited number of different hanzi are in daily use. For example, as Tan and Perfetti (1998) showed, 4,574 hanzi constitute a whole corpus of 1.8 million Chinese characters; (2) possible pronunciations in Chinese are relatively few as compared with many other languages.

Other researchers, however, found evidence that the phonetic information of hanzi or Kanji are processed at the component level (for example, Seidenberg, 1985; Saito et al., 1998). By manipulating stimuli in a series of experiments, Saito et al. illustrated in detail how information from both phonological components and radicals are interactively activated in Kanji matching tasks. They found that recognition of Kanji depends on the activation of information in components (radical and phonological), and that the figurative activation of components is a prerequisite for further detailed phonological processing.

If phonological components in hanzi facilitate the mental processing of hanzi, it is reasonable to expect that knowing phonological components (such as identification of phonological components, relationship between phonetic information of phonological components and the whole-character pronunciation, etc.) will help one learn hanzi more efficiently. The following studies provide empirical evidence supporting such an expectation.

Chen et al. (2003) argued that a morphological analysis of compound hanzi (mostly phonetic compounds) can communicate to Chinese children the underlying phonological components which can be used to learn how to pronounce target hanzi.

Anderson, Li, Ku, Shu, and Wu (2003) investigated the usage of phonological components in hanzi and found that native learners can better predict the pronunciation of hanzi containing full or partial phonetic information that provided in phonological components than in those without such components.

At the first glance, the findings that phonological components are not determinants of phonological processing (e.g., Tan et al., 1996) may seem to contradict advocates of taking advantage of phonological components in hanzi learning (e.g. Anderson et al., 2003). Further examination, however, reveals that the seemingly opposite findings are just two sides of the same coin. Even though pronunciation of high-frequency hanzi (usually not phonetic compounds) is processed at the whole-character level by skilled native adults, pronunciation of relatively low-frequency hanzi (usually phonetic compounds) can still be successfully predicted with information contained within phonological components by native child learners. Hanzi as phonetic compounds are found to be vocalized more quickly than those which do not contain phonological components only when they are all low-frequency (Seidenberg, 1985). Leong, Cheng, and Mulcahy (1987) found consistent evidence that low-ability native readers differed from high-ability readers in hanzi vocalization latency and/or accuracy mainly for low-frequency and complex (having more strokes) stimuli. A hypothesis based on all these findings is that for CFL learners, as well as native learners had better rely on rote memory when learning pronunciation of non-phonetic-compound hanzi; they should use their knowledge of phonological components to learn phonetic compounds.

In summary, the literature on the mental processing of hanzi suggests that teaching CFL learners the phonological components of phonetic compound hanzi (especially for those with relatively low frequency) will benefit their hanzi learning.

2.3.1.2. Radicals help semantic processing

Laychuk (1983) declared that given the complexity of Chinese orthography, teaching CFL learners the etymological explanations of hanzi forms can enhance hanzi memorization and retention. McBride-Chang, Shu, Zhou, Wat, and Wagner (2003) showed a significant effect of morphological awareness on native Chinese learners' recognition of hanzi.

Li (1996) argued that although a thorough study of Chinese morphology is neither possible nor necessary for CFL learners, a basic knowledge of hanzi structure will definitely help them understand hanzi instead of just viewing them as random combinations of dots and lines. It is also true, as she pointed out that understanding the phonetic and semantic components of hanzi is also one of the tools helps native Chinese memorize thousands of characters. Her pilot study of 48 first-year CFL learners further confirms the argument. In the study, interviewees generally reported that with a brief explanation of how each target hanzi is developed and structured, learning of hanzi becomes easier and retention more stable. Similar findings are also found in Duan and Cuvo (1996) where six adult learners learned faster and retained more when they were given instructions on the meaning of each component of a compound hanzi than when they were just asked to trace the character and write its translation.

Edwards (1997) criticizes textbooks which emphasize phonetic learning to increase learners' Chinese vocabulary, and advocates a more efficient approach in which the ideographic nature of vocabulary is analyzed and used to help CFL learners build up vocabulary more quickly by using characters the learners have already committed to memory.

In summary, there is persuasive evidence that giving CFL learners a brief morphological analysis of components for target hanzi leads to bring more efficient and effective learning than when such information is not provided. Because this kind of analysis helps learners link a new character to related hanzi which share radical or phonetic components they have already memorized, they are better prepared to learn related hanzi in the future.

2.3.2. Stroke Sequence and Educational Implications

According to Sasaki (1987), when most Japanese were prompted with a word memory task, whether in kanji/hanzi or in English, they performed finger writing. Students from non-kanji/hanzi backgrounds tended not to do so. Sasaki (1987, p. 146) hypothesized that "For Japanese, kanji form representation may be an integral part of the stroke action made in writing Kanji characters" and that finger writing provides an action-based representation that can be used as an external aid for mental processes. This can be viewed as an illustration of what Piaget (1977, p. 30) argued:

> To know, therefore, is to produce or reproduce the object dynamically; but to reproduce it is necessary to know how to produce and this is why knowledge derives from the entire action.

Sasaki (1987) suggested that when learners who have a kanji/hanzi background experience learning character by pen writing and finger writing from an early age, their recall of characters will be facilitated by finger actions.

One concept in Chinese literacy that is closely related to the concept of finger writing is stroke sequence. Foreign learners find it hard to understand why a customary stroke sequence should be followed for a hanzi. In contrast, native Chinese learners seldom doubt the justification of stroke sequence. For example, Ma (2000) developed a self-paced tutorial for CFL learners studying Chinese orthography which incorporates stroke and stroke sequence. Ching (1983) explained why CFL learners ought to learn Chinese calligraphy. The author argued that those who learn stroke count and stroke order will be able to produce hanzi more correctly.

According to Law, Ki, Chung, Ko, and Lam (1998), correctly producing stroke sequence of hanzi (especially compound characters consisting of at least two components) requires the ability to (1) properly deconstruct a compound characters into hierarchical parts, (2) produce a part in correct stroke order, and (3) legal individual strokes. The importance of stroke sequence has been emphasized in Chinese literacy education ever since Kai Shu became the standard hanzi font over a thousand years ago. However, the effects of stroke sequence on hanzi learning have not been well studied so far. Law et al. found contradictory facts. On one hand, almost all the native, Chinese-speaking children studied reported that stroke sequence is important and was emphasized by their teachers; on the other hand, half of them could not even properly produce the most familiar high-frequency hanzi like 媽 (read as mā, meaning mother) in the correct stroke sequence. With further analysis, they found the most frequent mistakes those children made is illegal individual strokes followed by incorrect stroke order. The authors argued that given the fact that there are only eight basic strokes (with composite strokes made up of these basic ones), producing legal strokes should not be a problem if children are given sufficient training and practice. The rationale for teaching proper stroke sequence, according to the authors, is that (1) learners can produce hanzi more quickly and accurately; (2) stroke counting and deconstruction of a target hanzi are necessary for looking it up in a dictionary if its pronunciation is unknown in advance; (3) improper deconstruction of a hanzi usually leads to misunderstanding; and (4) some minor mistakes may produce totally different characters. Hierarchical deconstruction of hanzi is not seen as a serious problem for native children. However, given CFL learners' limited knowledge of hanzi and limited time assigned to practice writing, it is reasonable to assume that CFL learners will make more mistakes than native speakers. Knowing the basic rules for stroke sequence and the components of compound hanzi are expected to be helpful.

Producers of instructional multimedia for teaching CFL have longed desired to incorporate stroke sequence principles into relative product designs. Early attempts started in the late 1980s and early 1990s. Ng and Wu (1990) advocated the benefits of stroke sequence instruction and discussed a "modified keyboard for Chinese character entry," which could be used to enter hanzi into a computer by strokes and proper stroke sequence. Ng and Wu argued that this kind of keyboard might help CFL learners better understand hanzi structure, so as to facilitate the learning of characters. Current computer technology makes this much easier to realize than ever, and one of the interests of this study is to determine the effects of stroke sequence instruction on hanzi learning in multimedia conditions.

2.3.3. Section Summary

The review of literature in this section confirms the importance of a knowledge of radicals and phonological components in hanzi learning. Such knowledge is found to be an effective tool in helping both CFL and native Chinese-speaking children to better recognize hanzi. Finger writing hanzi is believed to help memorization but seems not to work well for learners without a kanji background. Another interesting finding is that although mastery of basic strokes and correct stroke sequence has been emphasized by many Chinese researchers and teachers, native Chinese-speaking children generally performed quite unsatisfactorily on tasks of producing proper strokes and correct stroke sequence, even for those hanzi very familiar to them. This finding invites further investigations on stroke sequence instruction and its effect on hanzi learning.

2.4. Information Processing Theory and Memory

Media when used in instruction have a great potential to affect learning. Dual-coding theory (Paivio, 1971, 1986), which relates regarding the interaction between modality and information processing, is an empirical theory on the dynamic networks of verbal and nonverbal representations that underlie human behavior and experience. It also has wide applications on the development of instructional multimedia. The split-attention effect within the paradigm of dual-coding theory also sheds light on learning in multimedia situations.

2.4.1. Dual-Coding Theory

2.4.1.1. Basic Assumptions

Dual-Coding Theory (henceforth DCT) assumes associative and referential associations/networks within and between the verbal and nonverbal systems which code mental representations (see Fig.1, as in Clark & Paivio, 1991, p. 152). Associative networks within the verbal and nonverbal/imaginational representations are basic mental structures, and activations of these representations are fundamental mental processes (Paivio, 1971, 1986; Paivio & Begg, 1981; Clark & Paivio, 1987). Referential connections between the two systems, as Clark and Paivio asserted (1991, p. 153), "join corresponding verbal and imaginational codes and potentially allow such operations as imaging to words and naming to pictures."

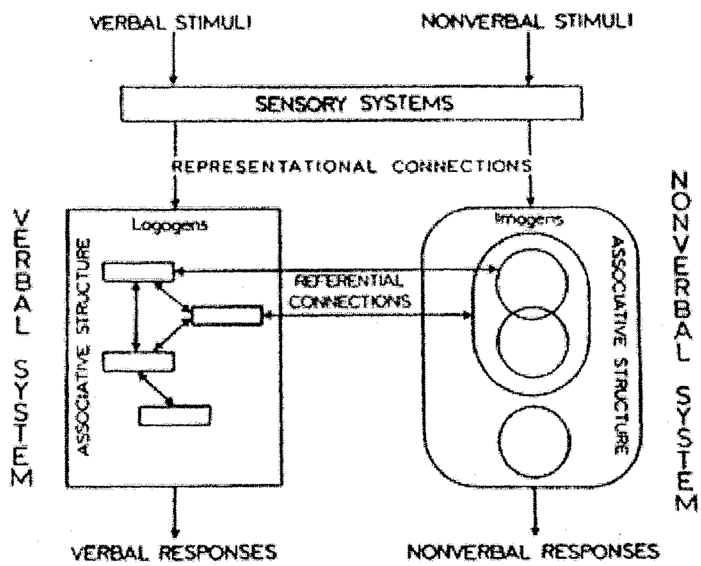


Figure 2.1. Schematic depiction of the structure of verbal and nonverbal symbolic systems.

Note. Figure 2.1 shows the representational units and their referential (between system) and associative (within system) interconnections as well as connections to input and output systems. The referentially unconnected units correspond to abstract-word logogens and "nameless" imagens, respectively. From *Mental representations: A dual-coding approach,* 67, by A. Paivio, 1986, New York: Oxford University Press. Reprinted by permission (See Appendix B).

From the perspective of DCT, imagery processes can facilitate the memorization memory of learning materials because of "elaborative" and/or "integrative" procedures (Clark & Paivio, 1991, p. 165-166). Elaborative processing refers to the additive effects of generating imagery coding from verbal coding (Paivio, 1975; Paivio & Lambert, 1981; Vaid, 1988), and integrative processing refers to the "redintegration" phenomenon of recalling an entire representation via partial cues (Begg, 1973; Bower, 1970; Paivio, 1969).

2.4.1.2. Mental Codes & Sensory Modality

As DCT concerns both verbal and non-verbal coding, it is applicable to studies on multimedia learning. The relationship between mental codes and sensory modality as theorized in DCT is illustrated by the following table (Sadoski & Paivio, 2001, p. 45):

Table 2.1.

Orthogonal conceptual relation between symbolic systems and sensorimotor systems

	Symbolic Systems	
Sensorimotor	Verbal	Nonverbal
Visual	Visual words	Visual objects
Auditory	Auditory words	Environmental sounds
Haptic	Writing patterns	"Feel" of objects
Taste	—	Taste memories
Smell	—	Olfactory memories

Note. Table 2.1is with examples of types of modality-specific information represented in each subsystem. From *Mental representations: A dual-coding approach,* 57, by A. Paivio, 1986, New York: Oxford University Press. Reprinted by permission (See Appendix B).

2.4.1.3. Relevance to this Study

This study concerns the effects of factors in multimedia flashcards on hanzi production from memory. Factors of interests such as sound, animation, and text can be viewed as both visual and auditory, in terms of sense modality, or as verbal and nonverbal codes. Since DCT provides a model depicting how these codes dynamically interact, it can serve as a theoretical framework for the study. More specifically, it is reasonable to treat explanation text (i.e., texts using character definition and mnemonics, including both English and Chinese) in flashcards as verbal codes through the visual modality, and pronunciation of target hanzi to be memorized as verbal codes in the auditory modality. The nature of target hanzi as well as animations of target hanzi, however, is arguable. For CFL beginners from non-Chinese backgrounds, these inputs are hypothesized to be similar to non-verbal information. Hanzi never seen before may be viewed as a new "graph" rather than as meaningful verbal information. CFL beginners from a Chinese background may still be able to view new inputs as verbal information because of the influence of cultural experience.

2.4.2. Theory of the Split-Attention Effect

2.4.2.1. Brief Accounts of the Split-Attention Effect

The split-Attention Effect (SAE) seemingly contradicts but actually conforms to DCT. The basic assumptions of SAE are that: (1) modality has an effect on information processing; (2) within a modality, the capacity for processing information is limited, and, if more than one source of input is presented simultaneously, SAE will take place; and (3) when the above inputs are from different modalities with one input in each modality, SAE will not happen and learning results are facilitated.

2.4.2.2. Relationship between SAE & DCT

Consistent experimental results are found to support SAE in some multimedia learning situations. Visual text and animation presented simultaneously (hereafter: AT treatment) are more likely to restrict the cognitive resources available for effective learning (as measured in both knowledge retention and transfer) than auditory verbal information, and visual text presented concurrently (hereafter: AN treatment) because of the presence of split-attention effects. From a DCT approach, Mayer and Moreno interpret the AT treatment condition as a learning environment where verbal text and visual animation are presented in the learners' visual channel. This tends to cause the attention split because of human's limited cognitive resources available in a particular modality. The AN treatment condition represents learning environments where auditory verbal information and visual animation are presented in two different modalities (i.e. visual and acoustic) and are more likely to cause referential connections between the modalities to facilitate learning. In Moreno and Mayer (Moreno, 1999), the advantage of dual modalities presentation even holds for sequential presentation in addition to the concurrent one. Moreno and Mayer (2000) tested the effect of background music and sound in multimedia learning. They found that background music or sounds negatively affected performance on retention and transfer tests, which is consistent with the limited capacity of the working memory hypothesis assumed by DCT.

As Moreno and Mayer (2002) discussed, there are conflicting findings on redundant presentations of information as the cause of a positive DCT effect (see also Broadbent, 1956; Hede, 1980; Martin, 1980; Montali & Lewandowski, 1996; Penney, 1989; Treisman & Davies, 1973) versus a negative redundant signals effect (see also

Colquhoun, 1975; Halpern 1974; Kinchla, 1974; Kobus et al., 1986; Lewandowski & Kobus, 1989; Nickerson, 1973), or the so-called SAE effect (see also Chandler & Sweller, 1992; Kalyuga, Chandler, & Sweller, 1999; Mousavi, Low, & Sweller, 1995; Smith, Miller, Grossman, & Valeri-Gold, 1994). Moreno and Mayer conducted studies to investigate "whether and under what conditions the addition of on-screen text would facilitate the learning of a narrated scientific multimedia explanation" (p. 156). The previous contradiction is reconciled by new findings that are consistent with DCT. The new findings favor the redundancy of on-screen text only when competing visual components are not used. According to DCT, since visual working memory is limited, SAE will impair learning when on-screen text and animation are simultaneously provided. Kalyuga et al. successfully ameliorated split-attention effects while adding redundancy to graphic presentations by changing text to manageable narration or by coloring codes to make referential connections easily accessible. Both approaches appear to reduce cognitive load and allow more resources for referential processing, thus benefiting learning.

In summary, SAE is consistent with DCT in that both (1) admit limited capacity for information processing within one modality (visual or auditory) or representation (verbal or non-verbal) and (2) acknowledge the advantages of using corresponding redundant information from different modalities (visual and auditory or verbal or nonverbal). The major difference is that DCT focuses on codes and SAE focuses on modality. Another difference is that DCT emphasizes the positive effect of connections between verbal and non-verbal coding, while SAE devotes primary attention to negative effects taking place within a modality.

2.4.2.3. Relevance of SAE to the Study

As the study is under multimedia conditions involving more than one modality, SAE theory will shed light on the hypothesis of the study and the explanation of findings as DCT will do.

2.4.3. Section Summary

The proposed study concerns the effects of multimedia in Computer Assisted Language Learning on the learning outcomes. Dual-coding theory and the split-attention effect give the study a theoretical framework. The dual-coding (verbal and nonverbal) approach is usually found to be beneficial in facilitating the mental processing of information. However, given the limited capacity of our information processing system (visual or acoustic), the split-attention effect suggests avoiding the presentation of more than one medium in a modality at a time.

Although many studies on individual and/or interactive effects of sounds, graphics, animations, and texts are conducted in the field of educational psychology, most test no more than two media types at a time. These laboratory type studies are well controlled, but, without actual testing, we can not generalize their findings to more complex multimedia learning environments such as hanzi shape memorization using multimedia flashcards which integrating text (Chinese and English), hanzi stroke sequence animation, and voiced pronunciation.

2.5. Keyword Mnemonics as a Variables Affecting Memory

Mnemonics are taught techniques for memorizing target characters in foreign language education. Imagery and keywords are mainstreams of the mnemonic technique. Whereas imagery mnemonics are generally found to be effective for memorizing more concrete characters/words, they still have relatively limited use, because there are many abstract concepts that are hard to depict in graphs. The ineffectiveness of imagery mnemonics is reported in studies of hanzi memorization such as Wang and Thomas (1992). On the other hand, some studies (e.g., Lu et al., 1999) do provide evidence of the effectiveness of imagery mnemonics in hanzi recognition tasks. The target hanzi in these studies, however, are mainly pictographs which can be easily explained with help of graphs, but these are in the minority. The dominant type of hanzi in current use is phonetic compound hanzi. Therefore, in the literature review on mnemonics, the keyword mnemonics method is emphasized.

The keyword mnemonic method in foreign vocabulary learning is defined in Pressley, Levin, and Delaney (1982) as a two-phase procedure for remembering unfamiliar foreign words, where learners first establish a firm association between a target foreign word and an acoustically similar mother tone word (the keyword), which then "encodes a meaningful interaction between the keyword and the foreign word's definition" (p.61).

The keyword mnemonic method is consistently found to be an effective approach to L2 vocabulary memory in many studies (Pressley et al. 1982; see also Ott, Butler, Blake, & Ball, 1973; Johnston, 1974; Atkinson, 1975; Atkinson & Raugh, 1975; Raugh & Atkinson, 1975).

Some studies examined if and how the effectiveness of mnemonics is affected by variables such as learner proficiency, the strategy of mnemonics use, and the ease of mnemonics generation, etc. For example, Delaney (1978) investigated whether learners' proficiency will affect effectiveness of keyword mnemonic usage. He found that when low fluency learners are given the option of using either imagery or keyword mnemonics, they will perform as poorly as peers using no mnemonics at all. Cohen and Aphek (1980) studied how different strategies of using keyword mnemonics affected the performance of learners of Hebrew as a second language on recognition tasks. They found that previously formed keyword associations benefit recognition better than either changing keywords afterwards or using no keywords at all. Hall (1988) examined student-generated keyword mnemonics. He found that when keywords could be easily generated, learners did not perform significantly better on recognition tests, but they performed significantly worse when keywords generations could not be easily made.

The cited studies have influenced the design of this study. To help learners memorize new target words, mnemonic keywords should be supplied initially and continued throughout the instructional presentation. The participants, of course, are all CFL beginners who have relatively low fluency and are not expected to have sophisticated skills in generating effective keywords by themselves.

Although most mnemonic studies use definition recall/recognition tasks, there is a few which use spelling/production as post-test tasks. A review by Pressley et al. (1980, p. 73) found that for production tasks, even "when keywords and foreign word syllables were orthographically identical," there is still "no negative consequences on incidental spelling were observed." This finding, if true, encourages the use of orthographically identical or similar keywords as mnemonics for production tasks. Since the task used in this study is learners' orthographical production, it can be reasonably expected that using hanzi or partial Chinese hanzi (i.e. components) as mnemonics will not negatively affect their performance. Another reason for using hanzi or hanzi components as keywords is that it is hard to build up memory cues between logographic characters like hanzi and alphabet characters like English without analyzing the structural features of hanzi. In other words, even in cases when English is used for keywords, they will still have to be built up on a phonological or semantic analysis of the target hanzi.

2.6. Chapter Summary and Research Questions

Based on this review of literature, it is clear that: (1) teaching phonologic components and radicals to CFL learners will facilitate their hanzi memory, especially for phonetic compound hanzi; (2) knowing stroke sequence helps CFL learners produce hanzi from memory; (3) using inputs from dual-modality or inputs as dual-coding will enhance memory; (4) teaching keyword mnemonics using English as well as Chinese will tend to improve learners' memory of hanzi; and (5) testing short-term memory in L2 vocabulary learning is meaningful.

Given these findings, it is reasonable to predict that multimedia hanzi flashcards designed to include all the above mentioned factors will enhance the production of hanzi from memory more than multimedia flashcards which have none of them. For example, current multimedia hanzi flashcards involve radical explanations, mnemonics, embedded pronunciations, and stroke sequence animations. Although it is clear that flashcards with these features are more desirable than flashcards with just target hanzi and English definitions, it is not known which combinations of the above features will produce the best results. An exploration in this direction will benefit CFL education. More importantly, since former studies on DCT and SAE usually involve no more than two types of inputs at a time (e.g., graphs and texts in the visual modality, or background and foreground sounds in the auditory modality, or two modalities but each with only one input type), the effects of more complex multimedia combinations are never discussed. An exploration of more complicated multimedia combinations will thus enrich our knowledge on learning theories such as DCT and SAE.

Specifically, this study concerns the examination of the individual and interactive effects of digital voiced pronunciation and stroke sequence animation on CFL beginners' production of hanzi from memory.

The primary research questions are: (1) does voiced pronunciation have an effect on CFL beginners' production of hanzi from memory? (2) does stroke sequence animation have an effect on CFL beginners' production of hanzi from memory? (3) does using both voiced pronunciation and stroke sequence animation create an interactive effect? A less critical question is that is, can a difference be found in terms of the effect of stroke animation between learners from a Chinese background and those from a non-Chinese background?

CHAPTER 3. METHODOLOGY

3.1. Situation

This study took place at a large public Midwestern university, which offers Chinese Level II in spring semester each year. Chinese Level II is a four-credit comprehensive course in elementary Mandarin, designed to develop basic proficiency in listening, speaking, reading, and writing for foreign language learners. Like other levels, the Chinese Level II course includes computer lab classes. Six computer labs were scheduled for the spring 2004 semester. All of these computer lab classes met on Monday, Wednesday, or Friday. Generally, after two classroom sessions, a computer lab class was scheduled for learners to practice writing on the computer using vocabulary and sentence patterns they had just learned. This experiment took place in the fifth computer lab class. In spring 2004, 106 students were enrolled the course. All these 106 learners had either taken Chinese Level I course or had tested out of that course. The majority of these learners were students in the College of Engineering or the College of Science. No learner took the course as a requirement for a major. Only a small percentage (less than 5%) of the learners took this course as a requirement for a minor.

About 40% of the learners were native speakers of English, about 40% were native speakers of Asian languages other than Chinese, and the remaining 20% or so were Chinese heritage learners, whose first language is not Chinese but know a limited amount of Chinese through learning from family members. All the learners were 18-30 years old. About 2/3 were males and the rest were females.

The Chinese Level II had six sessions which were taught by two course coordinators and four teaching assistants. All the instructors were native Chinese, two from Taiwan and the rest from the People's Republic of China. The four assistants teaching the Monday, Wednesday, and Friday classes (including computer lab classes) were referred to as Y, J, C, and Z (initials of their first name). The teaching schedule of Chinese Level II, prepared by the coordinators before the semester began, restricted the subjects to be covered or activities to be done for each class. All the teaching assistants were required to teach accordingly. The following table summarizes the instruction schedule for these four assistants.

Table 3.1.

Teaching assistants and teaching schedules

T. A.	Time (M/W/F)	Session
Y	9:30 a.m.—10:20 a.m.	1
J	9:30 a.m.—10:20 a.m.	1
C	10:30 a.m.—11:20 a.m.	1
C	1:30 a.m.—2:20 p.m.	1
Z	10:30 a.m.—11:20 a.m.	1
Z	1:30 p.m.—2:20 p.m.	1

3.2. Research Design

3.2.1. Independent and Dependent Variables

The independent variables involved in this study were voiced pronunciations and stroke sequence animations of hanzi. Control and experimental groups were established for each variable.

The dependent variable of concern was participants' scores on the hanzi production-from-memory post-test. The pre-test score of each participant was treated as a covariable to eliminate prior treatment differences (a widely used approach to correct errors caused by non-randomized sampling) through statistical measures that considered the diverse language backgrounds of the learners.

3.2.2. The Two-Factor Factorial Design

To test the effects of voiced pronunciation, stroke sequence animation, and their interaction, a two-factor factorial design was applied. The full treatment group (WVWA) used multimedia flashcards with target hanzi, English definition, pinyin (Romanized notation of pronunciation), mnemonics (with phonological and radical information), voiced pronunciation, and stroke sequence animation. The control group (NVNA) used the same material without voiced pronunciation or stroke sequence animation. The partial treatment groups were the WVNA group, which used the same material as the full treatment group but without animation, and the NVWA group, which used the same material as the full treatment group, but without voiced pronunciation.

3.2.3. The Null-Hypothesis

According to the theoretical framework of DCT and SAE, it was reasonable to expect presence of SAE and DCT in the four different treatments. However, since the four treatments appealed to different modalities as well as to different types of inputs within the visual modality, it was hard to predict their effects on learning results. Therefore, given the complicated situation, a null hypothesis was made: there will be neither main (i.e. voiced pronunciation effect, stroke animation effect) nor interactive effects of voiced pronunciation of hanzi and stroke sequence animations. More specifically, the null hypotheses addressing the research question were as follows:

Null hypothesis 1: The performance of the WVWA group and the WVNA group will not be significantly different on the task of hanzi production from that of the NVWA group and the NVNA group. In other words, there will be no effect of voiced pronunciation.

Null hypothesis 2: The performance of the WVWA group and the NVWA group will not be significantly different on the task of hanzi production from that of the WVNA group and the NVNA group. In other words, there will be no effect of stroke sequence animation.

Null hypothesis 3: The performance of the WVWA group and the NVNA group will not be significantly different on the task of hanzi production from that of the NVWA group and the WVNA group. In other words, the voiced pronunciation effect will be consistent regardless of the levels of stroke sequence animation and vice versa.

3.3. Sampling Techniques

3.3.1. Sample Size and Test Power

As Montgomery (2001) pointed out, "Selection of an appropriate sample size is one of the most important aspects of any experimental design problem." (p. 40). A calculation of sample size can allow experimenter know in advance how large a minimum sample size should be in order to achieve a desired level of test power. The more powerful a test is, the higher the probability that it will be able to detect the mean difference between treatment groups.

In this study, the test power was set at 0.90, which meant that about 90% of the time, the tests to be performed will detect this possible significant effect of stroke sequence animation, voiced pronunciation, and interaction between voiced pronunciation and stroke sequence animation.

To calculate a sample size, the parameters needed are: alpha (α), desired difference (D), power level (1-β), standard deviation (σ^2) of response variable, levels of treatments, and the value of phi-square (ϕ^2).

Alpha is the type I error rate defined as the probability of rejecting a null hypothesis when it is actually true. Alpha is usually set as either 0.05 or 0.01, and here 0.05 was used. Desired difference is the minimum significant mean difference on performance between any two treatment groups. Desired difference is arbitrarily determined based on experience. Here, it was reasonable to set the value of desired difference as 3, 4, or 5, because, based on experience and CFL classroom practice, between-group mean differences smaller than 3 points of a test may be thought as insignificant. In other words, only mean differences greater than 3 points were considered to be meaningful or significant. Those less than three were treated as insignificant even if the statistical test found them to be significant. As stated before, the test power used in this study was set at 0.90. Standard deviation of the response variable or in other words, the mean difference in performance between treatment groups, was found to be 5.63 points. This was based on previous Chinese Level II course data revealing the distribution of the CFL learners' scores on the hanzi-production-from-memory task using different kinds of flashcards as instruction materials similar to this study. Levels of treatments in this case were with or without treatments, or 1 and 0 levels, for each of the two concerned factors (voiced pronunciation and stroke sequence animation). Phi is a parameter corresponding to the desired difference depending on all the previously mentioned parameters. A calculation of Phi makes the application of the Operating Characteristic Curve (a graph used to determine test power) more convenient. The calculation formulas for Phi in two-way fixed effect factorial model are as follows (Montgomery, 2001, p. 189):

$$\text{Factor A: } \phi^2 = n\, b\, D_A^2 \,/\, 2a\sigma^2$$

$$\text{Factor B: } \phi^2 = n\, a\, D_B^2 \,/\, 2b\sigma^2$$

$$\text{Interaction: } \phi^2 = n\, D_{AB}^2 \,/\, 2[(a\text{-}1)\,(b\text{-}1)+1]\,\sigma^2$$

In this case, voiced pronunciation and stroke sequence animation were represented by factors A and B respectively. The following table summarizes relationships between sample size (subjects per treatment group) and several candidates of desired differences (D) owing to the two factors of concern (supposing on equal number of subjects per treatment group) under the condition of setting the alpha value at 0.05.

Table 3.2.

Relationships between desired difference, number of subjects, and test power of the effects of voiced pronunciation and stroke sequence animation

D	Subjects per group (n)	ϕ^2	ϕ	V1	V2	β	Power: 1- β
3	20	2.84	1.69	1	76	0.37	0.63
	30	4.26	2.06	1	116	0.21	0.79
4	20	5.05	2.25	1	76	0.13	0.87
	25	6.31	2.51	1	96	0.06	0.94
5	18	7.09	2.66	1	68	0.045	0.955
	22	8.68	2.95	1	84	0.015	0.985

Note. V1 and V2 refer respectively to the degree of freedom of the numerator and the denominator in calculating the parameter ϕ^2.

The table shows that when a statistically significant difference of 3 points is used with a sample size of 30 subjects, there will be a weak test power of 0.79. This means that when the experiments are repeated over and over, about 2 times out of 10, the tests will yield no significant difference when there actually is one. More importantly, only 106 students registered for the Chinese Level II course in spring 2004, and thus the maximum available number of subjects per group was 26, which will yield an even weaker power. When the desired detectable difference increases to 4, however, the minimum number of subjects needed is only about 22 per group to get a fairly high power of 0.90, which means that, 90% of the time, the difference will be significant.

Similarly, detecting the interaction effect, the relationship between sample size and desired difference can be summarized by the following table.

Table 3.3.

Relationships between desired difference, number of subjects, and test power of the interaction effect of voiced pronunciation and stroke sequence animation

D	Subjects per group (n)	ϕ^2	ϕ	V1	V2	β	Power: 1-β
3	20	1.41	1.19	1	76	0.50	0.50
	30	2.13	1.46	1	116	0.48	0.52
4	20	2.78	1.67	1	76	0.40	0.60
	25	3.48	1.87	1	96	0.25	0.75
5	18	3.92	1.98	1	68	0.23	0.77
	22	4.79	2.19	1	84	0.15	0.85

Note. V1 and V2 refer respectively to the degree of freedom of the numerator and the denominator in calculating the parameter ϕ^2.

Because the main effect is more important to test than the interaction effect, a sample size of 22 per treatment group will yield an acceptable test power of 0.85 when the mean difference owing to interaction is 5 points.

In summary, the power and sample size calculation demonstrated that a minimum of 22 participants per treatment group was necessary for this study to detect a significant difference in participants' post-test mean scores (at least 4 points) owing to either the effect of stroke order demonstration or voice input. The test is capable of detecting such a difference 90% of the time if the experiment is going to be repeated over and over. This required a total of 88 subjects, which was not difficult given the situation of this study.

3.3.2. Convenience Sampling Method

Based on the above calculations, the sampling method chosen for this study was convenience sampling. The target sample consisted of Chinese Level II students. Furthermore, regular class sessions were treated as experimental groups for this study.

The rationale for using convenience sampling is explained as follows:

Suitable size of available participants — given situation of this study, it was reasonable to expect at least 80 participants, which allowed a satisfactory test power (around 90%) for detecting a minimum mean difference of 4 points on the post-test;

Representativeness of the sampling — the demographic information of the participants in this study was quite comparable to some other studies of CFL education (e.g., Wang, 1998; Ke, 1996; Li, 1996 etc.).

Similarity to normal classroom instruction — the regular classes used in this study resembled normal classroom instructional situations. Participants therefore were in a situation very familiar to them. This helped to reduce the feeling of taking part in an experiment.

3.4. Recruitment of Subjects

This study was classified as "exempt" by the Institutional Review Board on Human Subjects.

In the spring semester of 2004, the fifth computer lab class was chosen to conduct the experiment. During the next regular classroom meeting immediately after that lab class, the subjects learned how to use the target words they had just memorized in the lab. This was to ensure: (1) a higher participation rate. Because the flashcard instructions taught the required hanzi for the Chinese Level II course, the participation rate was expected to be relatively high. The participation rate of 94.3% (100 /106) verified this assumption; (2) participants generally paid more attention to the flashcard instruction material because the content relevance was high; (3) a better control of pre-knowledge, because the target hanzi had never been taught before.

At the beginning of the lab class, all students in the six sessions were told about the study, used the assigned flashcards, and took the pre-test and post-test anonymously. Students were encouraged to participate by being informed the purpose of the study and the importance of their participation, but they were also told that their participation was voluntary and would not affect their final grade in any way. Since the data were collected anonymous, students knew that their confidentiality was protected.

Examination of the data showed that of the 100 observations, five were classified as invalid by experimenter (students who did not strictly follow lab instructions), six were considered to be unusual observations (students whose hanzi production pre-tests scored higher than their post-tests), four contained missing data (such as missing pre-test post-test responses), and two observations were treated as outliers (refer to Chapter Four). Thus, only 83 observations were valid for this study. The valid data rate was 83% (the ratio of valid observations to total observations).

The subjects were from different language backgrounds, dominated by native speakers of Asian languages other than Chinese (42.7%) and native speakers of western languages (40%). The vast majority speakers of western languages in this study were native speakers of English. Chinese heritage learners constituted the rest subjects (17.3%).

3.4.1. Assignment of Flashcards Treatments

The six sessions of the Chinese Level II course were divided into four experimental groups. To better contrast the mean scores of the WVWA group (which used flashcards with both voiced pronunciation and stroke sequence animation) with that of the NVNA group (which used flashcards with neither voiced pronunciation nor stroke sequence animation), four randomly chosen class sessions were used with two randomly chosen as WVWA groups and the other two as NVNA groups. One of the remaining two sessions was the WVNA group, and the other was the NVWA group.

NVNA, WVNA, NVWA, and WVWA were represented with the numbers 1, 2, 3, and 4 respectively. Six identical white paper slips were prepared with the numbers 1, 2, 3, or 4 written on them. Two slips were marked number 1, and two were marked number 4. The four teaching assistants did not know in advance how those numbers represented the experimental conditions. They were each told to draw one of the six slips. The following table summarizes the results of the drawing.

Table 3.4.

Results of the treatment assignment drawing

T. A.	Time (M/W/F)	Treatment
Y	9:30 a.m.—10:20 a.m.	1
Z	10:30 a.m.—11:20 a.m.	1
C	10:30 a.m.—11:20 a.m.	2
C	1:30 a.m.—2:20 p.m.	3
J	9:30 a.m.—10:20 a.m.	4
Z	1:30 p.m.—2:20 p.m.	4

The number of valid observations for the NVNA (treatment 1), the WVNA (treatment 2), the NVWA (treatment 3), and the WVWA (treatment 4) was 27, 9, 15, 32 respectively.

3.5. Materials

3.5.1. The Target hanzi

For this study, the 24 target hanzi to be memorized by subjects were chosen from the Lesson 28 vocabulary list in *Practical Chinese Reader I* (Beijing Language Institute, 1983) (henceforth PCR). Only hanzi never taught before in the vocabulary list were selected as the target hanzi.

3.5.2. Description of the Flashcards

Four different formats of computer-based flashcards of the target hanzi were developed by the researcher.

The flashcards were designed in such a way that each screen presented only one target vocabulary item with its corresponding hanzi (one or two characters), pinyin, English definition, and mnemonic note. However, the use of availability of corresponding voiced pronunciation and stroke sequence animation were varied across the treatment groups.

1. For the NVNA group, neither voiced pronunciation nor stroke sequence animation were presented.
2. For the WVNA group, voiced pronunciation was presented but stroke sequence animation was not.
3. For the NVWA group, stroke sequence animation was presented but voiced pronunciation was not.
4. For the WVWA group, both voiced pronunciation and stroke sequence animation were presented.

The software used to develop these flashcards included Microsoft Paint®, Macromedia Flash®, and Macromedia Director®.

The dimensions of each single screen of the flashcards were 14.5 cm x 11 cm. The flashcard background was white. Each card had a 0.5 point solid black line border to provide a contrast between the card and the screen background. The screen background was also white. During the learning process, each of the flashcards appeared at the center of the screen. Each target hanzi was demonstrated on the upper half of a card in a True Type MS Hei style font, size 72. The radical (if there was one) of the target hanzi was highlighted in blue. All other verbal information, such as pinyin, English, or Chinese characters, was in black. Both the pinyin and the English definition of the target hanzi were right below the character. The capitalized English definition (True Type Times New Roman style font, size 14) was provided in the parentheses that followed the pinyin of the character. Only the first letter of the pinyin was capitalized. The pinyin was also presented in a True Type Times New Roman font, size 14.

The lower half of the card provided mnemonic cues for the target hanzi. The mnemonic was presented in English (True Type Times New Roman style font, size 14), with the definition of the character capitalized. The radical (if any) was presented in True type MS Hei style font size 14 with an English explanation. Morphological cues for pronunciation (if any) were also presented in English and pinyin (True Type Times New Roman style font, size 14).

There were navigation buttons ("next" and "back") at the bottom of each page. The pace of presentation of the cards was subject controlled. Learners could navigate forward and backward by clicking the corresponding buttons.

Two examples of NVNA flashcard follow, one for a one-character word, and the other for a two-character word. The actual flashcards were larger than the examples given.

箱 xiāng (suitcase) Mnemonics: The suitcase is made of bamboo. 竹is a radical meaning "bamboo." 相is a character pronounced "xiāng."	公平 gōngpíng (fair, fairness) Mnemonics: The two characters are well balanced in their shapes. The first character means "openness," and the second one means "balance" or "impartial," thus the word means "fair" or "fairness."

Figure 3.1. Examples of No Voice No Animation (NVNA) flashcards.

For the WVNA type of flashcard, all settings were exactly the same as described above except that there was a simultaneous digital voiced pronunciation for the target hanzi (pronounced by adult native Chinese speakers with a male voice for odd pages and a female voice for even pages). The pronunciation sounded only once.

For the NVWA type of flashcard, all settings were the same as for the NVNA except that the static character in the upper part of the card was replaced by a movie created using Macromedia Flash®. The movie showed the correct stroke sequence of the target hanzi. Two to three frames constituted a completed Chinese stroke depending on the physical length of the particular stroke. The display rate of the movie was 6 frames per second. The movie started with the first stroke and ended with the last one.

For the WVWA type of flashcard, all settings were the same as those for the NVWA, except that voiced pronunciation (as described in the WVNA type) was heard once simultaneously with the movie.

It should be noted that the sound and/or stroke order animation was played only once. If subjects wanted to hear the sound and/or see the animation multiple times, they needed to refresh the screen page by revisiting it.

3.5.3. The Experimental Equipments

Two neighboring computer laboratories assigned for the Chinese Level II course were used as experiment sites. In each lab, there were 30 Optiplex GX260 Dell® personal computers (CPU 2.8 GHz), each with an UltraSharp 1702 FPV 17-inch flat panel LCD monitor. Each participant in the WVNA and WVWA groups was also provided with a headset for listening to the voiced pronunciations.

Each of the four different sets of flashcards was saved as an execute file on a World Wide Web five minutes prior to each lab meeting. No subjects knew about these websites in advance.The following table shows the allocation of websites for the four treatments. During the experiment, the subjects visited the website assigned. When they clicked the link of an assigned flashcards type, the file was executed and shown as a set of flashcards.

Table 3.5.

Times, locations, and website URLs of experimental materials

Time & T.A.	lab	Treatment	Website
9:30 J	289	WVWA	HTTP://WEB.ICS.PURDUE.EDU/~CHEN155
9:30 Y	283	WVWA	HTTP://WEB.ICS.PURDUE.EDU/~ZHU1
10:30 C	289	WVNA	HTTP://WEB.ICS.PURDUE.EDU/~CHEN155
10:30 Z	283	NVNA	HTTP://WEB.ICS.PURDUE.EDU/~ZHU1
1:30 C	289	NVWA	HTTP://WEB.ICS.PURDUE.EDU/~CHEN155
1:30 Z	283	WVWA	HTTP://WEB.ICS.PURDUE.EDU/~ZHU1

3.6. Data Collection Techniques

3.6.1. Instruments

3.6.1.1. Demographic Information

A very brief section (Appendix C, Part I) was included at the beginning of the instrument to collect participants' information on language backgrounds (i.e. mother tongue as English, Chinese, or other Asian language), and flashcard preference (no voice/ no animation, with voice/ no animation, no voice/ with animation, or with voice/ with animation).

3.6.1.2. Pre-test

In case the participants were differed in prior-knowledge of the target hanzi, a pre-test (Appendix C, Part II) was administered. The pre-test listed English definitions for the target hanzi and asked the subjects to write any of the target hanzi that they already knew. The goal of the pre-test was to measure participants' pre-knowledge of production of the target hanzi so as to eliminate the pre-knowledge effect on the post-test scores. The way to eliminate such an effect was by including the pre-test score as a covariable in the two-way ANOVA model.

There were 24 target hanzi, and each correctly produced hanzi counted as one point. The possible range for pre-test scores was from 0 to 24.

3.6.1.3. Production Post-test

The production post-test (Appendix C, Part III) was the instrument used to test the participants' short-term memory of the target hanzi presented in the flashcards. The English definition of each target vocabulary was listed in the same order as presented in the flashcards. Participants needed to produce corresponding hanzi for each English definition. Each correctly produced hanzi was counted as one point. The possible range for the production post-test score was from 0 to 24 points.

3.6.1.4. Recognition Post-test

After viewing the flashcards assigned, the participants were asked to take the recognition post-test (Appendix C part IV). The 24 target hanzi were listed, and participants were required to write down the English definition of each word together with the Chinese pinyin. The order of the hanzi tested was exactly the same as presented in the vocabulary list of lesson 28 in PCR and also the same as presented in the flashcard presentation. Although this fixed word order of presentation may affect learners' post-test performance (as suggested by Lu et al. 1999), such an effect (if any) was applied equally to each treatment group.

Each correctly marked pinyin or English definition was counted as one point. As there were 24 target hanzi (19 vocabulary words), the possible range for pinyin marking was 0 to 24 points, and the possible range for the English definition test was from 0 to 19 points.

3.6.1.5. Content Validity

According to the commonly acknowledged definition of test validity, which "refers to the degree in which our test or other measuring device is truly measuring what we intended it to measure" (AllPsych and Heffner Media Group Inc, 2004), it was reasonable to assume a high validity for the instrument used in this study because the goal of the post-test was to measure participants' memory of shapes for each of the target characters, and that was also exactly what the participants were directed to do.

3.6.2. Procedures

All the Chinese Level II learners attended the pre-scheduled computer lab class on the date of experiment as they usually did. Upon entering the lab, they sat at the computer of their choice.

Once the class began, the instructor gave the Guide for Lab Activity (see Appendix D) to each learner. The guide informed learners of the purpose and learning content of the experiment, the voluntary participation principle, and the experimental procedure.

After clarifying any questions about the guide, the instructors posted the pre-assigned website URL for each class. Next, each subject was given two minutes to take the production pre-test for the 24 target hanzi. Subjects were then informed to memorize within 30 minutes as much as possible of the information given about the target hanzi, including the pinyin, the English definition and the orthography by viewing the assigned flashcards. The subject then took the production post-test for 5 minutes. Finally, they took the recognition post-test for another 5 minutes.

The subjects were not identified. The production pre-test and post-test papers were attached to each other to ensure that scores from a subject could be compared for the data analysis.

3.7. Data Analysis Techniques

The data collected were very brief descriptive data on the subjects' native language, flashcard media preference, and the quantitative observations of subjects' pre-test and post-test scores.

Descriptive data were totaled to reveal tendencies which were most likely to facilitate meaningful interpretations of the quantitative data.

Quantitative data were tested statistically determine if the individual and/or collective effects for factors of concern were significant.

SAS statistical programming was used as the data analysis tool for this study. The steps of the analysis were as follows: (1) set up model; (2) adjust pre-test scores for nonequivalent groups; (3) check assumptions of two-way ANOVA model by plotting QQ-Plot and residuals; (3) identify outlier(s); (4) perform type III two-way ANOVA analysis; and (5) do pair-wise comparisons of least square means.

3.8. Validity Issues

3.8.1. Internal Validity Issues

3.8.1.1. History Threat and Controls

Because the six groups included in this study met in six different sessions and were taught by four different instructors, there was a potential threat of unplanned variable. The instructors conducting the experiment might have behaved differently, and the subjects in different sessions might have encountered different events which might have affected their learning process or performance on the test.

To control this, the lab guidance, which was the same as what the subjects would see in the experiment, was given to each of the four instructors one week prior to the experiment. The researcher and the instructors met to clarify any questions about the guidance three days prior to the experiment. The four instructors were also informed to supervise participants' behavior during the experiment to prevent any irrelevant events that might affect participants' study and/or test performance. Furthermore, since the laboratories were in two adjacent rooms, the researcher was able to observe each laboratory and talk with the instructors while the experiment was carrying on to make sure that no disturbing events happened.

3.8.1.2. Testing Threat and Controls

The content of the pre-test and the production post-test were exactly the same. Both included the same new words in Lesson 28 of the textbook PCR. This could have caused testing variable because some subjects might have had prior knowledge of the content of the post-test.

However, since all the vocabulary were supposed to be new words for the subjects, and since they were required to produce the hanzi, just knowing the content of the post-test in advance was unlikely to improve their performance on the production post-test.

Since participants were very likely to vary in prior-knowledge of the target hanzi because of such factors as diverse language backgrounds and so on, a pre-test of the hanzi to be memorized was used to measure the prior-knowledge. In the data analysis, the pre-test score was treated as a covariable so as to measure the treatment effects while eliminating knowledge of prior differences.

3.8.1.3. Selection Threat and Controls

The subjects in this study were all volunteers which might have caused selection variables. However, the high participation rate (94.3%) or the valid observations divided by the total number of registered students (78.3%) reduced the impact to a large degree.

Since the experimental groups were formed without randomization, in other words, intact class sections were used instead of groups with randomly selected subjects, equivalent groups prior to the experiment can not be assumed. Thus the experimental situation became a quasi-experimental situation. To reduce the unequal groups effect to a maximal degree, the non-equivalent groups analysis techniques (Trochim, 2004) was applied.

3.8.2. External Validity Issues

Any generalizations based on this study should be made with caution. This is because the characteristics of the subjects may not be the same as the students in other situations.

To help readers make the decision if appropriate generalizations could be made or not in their situations, the language background and media preference characteristics of the subjects in this study will be described in detail in Chapter Four.

Another noticeable external validity threat is that although the vocabulary list used to generate target hanzi in this study was considered to be randomly chosen from the textbook PCR, the textbook itself however was not randomly chosen from all the Chinese language textbooks on the market. Similarly, no assumption can be made that hanzi covered in the textbook were randomly chosen from all available Chinese hanzi. The textbook was purposefully chosen because it is the current most popular textbook for teaching CFL both in the People's Republic of China and across the world. This fact makes the choice meaningful, but one should be aware of these non-random factors in order to make more appropriate generalizations.

3.9. Reliability Issue

The reliability issue is critical because measures inherently involve random measurement errors. By estimating test reliability and making adjustment accordingly, one can improve a measurement and make it closer to its true score. There are different reliability estimators. Those often used include inter-rater reliability, test-retest reliability, parallel forms reliability, and internal consistency reliability. As test-retest and parallel forms reliability involve either test and retest or two sets of instruments, they are not suitable for the design of this study. Emphasis was given to inter-rater reliability and internal consistency reliability.

In this study, the main concern was the hanzi production instrument. The hanzi production test was administered twice as pre-test and post-test. The pre-test measurement error was critical and affected the analysis results because of the quasi-experimental design used in this study (please refer to Chapter Four for details). The post-test measurement error was much less influential since the data analysis used the ANCOVA model which is by nature a regression analysis as discussed in Trochim (2004):

> In regression analysis we fit the line relative to the vertical displacements of the points. Post-test measurement error affects the vertical dimension, and, if the errors are random, we would get as many residuals pushing up as down and the slope of the line would, on average, remain the same as

> in the null case. There would, in this post-test measurement error case, be more variability of data around the regression line, but the line would be located in the same place as in the no error case.

Thus, only pre-test reliability was assessed, and the measurement error in the pre-test was fixed according to estimates of inter-rater reliability and/or internal consistency reliability.

3.9.1. Internal Consistency Reliability

The internal consistency reliability is "used to assess the consistency of results across items in a test," according to Trochim (2004). The purpose of calculating internal consistency reliability is to see how consistent the test items within a measure are. In other words, are the items in a test measuring the same construct? How closely are the items related to each other?

There are different approaches to estimate internal consistency reliability, such as average inter-item correlation, average item-total correlation, split-half reliability, and Cronbach's Alpha (Trochim, 2004). Cronbach's Alpha is used in the reliability calculation because "Cronbach's Alpha is mathematically equivalent to the average of all possible split-half estimates" (Trochim), is widely used, and can be easily achieved by using computing software.

3.9.2. Inter-Rater Reliability

Inter-rater reliability was also applied when the pre-test error was estimated. The grading of hanzi production unavoidably involved subjective judgments. The accuracy of the test grading was open to question. Using two graders and checking the inter-rater reliability allowed this study to reduce possible serious measurement errors.

The researcher of this study and the principal coordinator of the Chinese language program under investigation served as the two graders. The following table summarizes basic information about these two graders:

Table 3.6.

Qualifications of raters involved in inter-rater reliability calculation

	The Coordinator	The Researcher
Gender	Female	Male
Highest Degree Earned	Ph.D.	M.Phil.
Native Language	Mandarin	Mandarin
Years of CFL teaching	10	3

A stratified sampling of the pre-test papers was used to estimate the inter-rater reliability. Twenty-one subjects' responses were selected at random for the coordinator to grade. The samples consisted of 7, 2, 4, 8 random samples from the NVNA, the WVNA, the NVWA, and the WVWA treatment group respectively. An inter-rater reliability was calculated for both the pre-test and post-test of hanzi production. More detailed information on reliability estimates is available in Chapter Four.

3.10. Summary of the Chapter

In this chapter, more detailed information on situation of the experiment was provided, followed by the specification of null hypotheses based on the review of literature in Chapter Two. Sampling issues were then discussed to show how the sample size was determined for this study and how the participants were recruited. Information was given about how the experimental groups were formed. Descriptions of experimental materials and equipments were given in detail. Finally data collection and analysis techniques were briefly introduced, followed by a discussion of validity and reliability issues.

In the Chapter Four, data analysis procedures and results will be discussed in detail.

CHAPTER 4. ANALYSIS OF DATA

4.1. Descriptions of the Demographic Data

Demographic data on the subjects' language backgrounds and preferences for flashcard types were collected. The language background question asked the subjects to select their level of home exposure to Chinese from one of the following: (a) some knowledge of Chinese other than Mandarin, (b) knowledge of Asian languages other than Chinese, or (c) no home exposure to Chinese. Among the 83 valid answers for the two-way ANCOVA, there were 73 responses to the question. Knowledge of Chinese other than Mandarin was reported by 15.1%, 39.7% were native speakers of other Asian languages, and 45.2% were native speakers of English who had no home exposure to Chinese. The learners who had some knowledge of Chinese were heritage speakers born in America. They know a little Chinese, mostly oral, from their parents. The Asian languages other than Chinese included (in decreasing number) Malaysian, Korean, and Japanese. Many of the Indonesian and Malaysian subjects were descendants of immigrants who speak Chinese. Therefore, it was reasonable to assume that they had a home exposure to Chinese comparable to that of the heritage speakers.

The flashcard preference question asked the subjects to choose their preferred flashcard type from (a) voice only, (b) animation only, (c) both, or (d) neither. As in their previous multimedia flashcards experience, the subjects were exposed to one of the four types about two weeks before the experiment. The previously exposed flashcards were also developed by the researcher. The target hanzi were from the vocabulary list of another lesson in the textbook *Practical Chinese Readers I*. The designs shared most of the features of the flashcards used in this study, except that, in the latter case, no example sentence was given for the target vocabulary, the radicals were highlighted in blue, and the stroke sequence animation stopped after one exposure. The subjects might have explored other computer-based multimedia flashcards beyond these classroom exposures, but no such information was collected.

The assignments of the previous exposure and the experiment treatment are summarized in the following table, together with a summary of the subjects' responses to the flashcard preference question.

Table 4.1.

Subjects' flashcard experiences and preferences according to class sessions

Class session	Flashcard Experience	Flashcard Assigned	Preference information			
			NVNA	WVNA	NVWA	WVWA
C1	NVNA	WVNA	22.2%	0%	0%	77.8%
C2	NVWA	NVWA	8.3%	16.7%	25%	50%
J	N/A	WVWA	5.26%	5.26%	5.26%	84.21%
Y	WVWA	NVNA	10%	10%	0%	80%
Z1	NVWA	NVNA	20%	0%	0%	80%
Z2	WVWA	WVWA	27.3%	9.1%	0%	63.6%
Average			15.51%	6.84%	5.04%	72.61%

From the above table, it is clear that, although the subjects had diverse preferences, most of them liked the flashcards with both voiced pronunciation and stroke sequence animation. The next most favored type was that with none of the multimedia inputs. The following table shows the preference level for each flashcard type across the four treatment groups.

Table 4.2.

Flashcard preferences for each experimental group

	Preference			
Group	NVNA	WVNA	NVWA	WVWA
NVNA	15%	5%	0%	80%
WVNA	22.2%	0%	0%	77.8%
NVWA	8.3%	16.7%	25%	50%
WVWA	16.28%	7.18%	2.63%	73.91%

In conclusion, the summary of demographic information found that the subjects in this study were diverse in both language background and flashcard type preference.

4.2. Analysis of the Quantitative Data

4.2.1. Model Selection

Terms included in the model of the two-way ANCOVA (analyzing data collected from Post-test 1) were voiced pronunciation, stroke sequence animation, and their interaction, since the study set out to be a 2-factor-factorial design. However, it was necessary to consider the following factors that could have affected the Post-test 1 scores.

4.2.1.1. Prior-knowledge

Since the subjects were assigned randomly to treatment groups of class sessions established by the university, the subjects might have had relatively diverse prior knowledge of the target hanzi. Therefore, the Post-test 1 mean difference between the treatment groups might have been caused by the prior-knowledge difference as well as by the treatment factors. To control the prior-knowledge effect, the Pre-test was administered to measure the prior-knowledge, and the pre-test score was treated as a covariable in the data analysis.

4.2.1.2. Test Reliability

Test reliability values in this study told how reliable the test measurements were and also served the purpose of eliminating biased estimates for the effects of the factors of concern (voiced pronunciation, stroke sequence animation, and their interaction).

The SAS output summarizing Cronbach's Alpha for the Pre-test follows. Raw variables' Alpha (i.e., $\alpha \approx 0.973$) was chosen because all the pre-test items used the same scale of points, that is, 1 for a correct response and 0 for a wrong one.

Table 4.3.

Values of Cronbach Coefficient Alpha

Cronbach Coefficient Alpha	
Variables	Alpha
Raw	0.972817
Standardized	0.975970

The inter-rater reliability (as described in Chapter Three) was calculated from 21 randomly chosen test papers (see Appendix E for details). As long as a produced hanzi was recognizable (with perhaps the stroke direction changed or with one or two strokes missing), the researcher treated it as a correct. But the Chinese course coordinator graded the tests more strictly. Stroke direction changes or missing stroke(s) were considered to be wrong responses. The resulting instances of mismatched grading were only 1 item and 14 items for the pre-test and post-test, respectively. Thus, the estimates of overall inter-rater reliability for the pre-test and post-test were 0.998 and 0.972, respectively. Given the high values of test-reliability estimates, it was concluded that both the pre-test and the post-test were reliable instruments.

For non-equivalent groups design, according to Trochim (2004), not only should the Pre-test be included in the model, but also should be adjusted to exclude measurement errors as much as possible. Trochim further argues that post-test measurement error will not result in biased estimates for treatment effects because of "the criterion that is used in regression analysis to fit the line." Specifically, he suggests the following adjustment to be performed for each pre-test observation:

$$X_{adj} = \overline{X} + R\,(X - \overline{X})$$

Where: X_{adj} = adjusted pre-test value, $\overline{X}$ = mean value of the original pre-test for a treatment group, X= original pre-test value, and R= original pre-test reliability.

As for the reliability value for the adjustment, Trochim suggests using Cronbach's Alpha or any other reliability which is lower. For this study, the Cronbach's Alpha for the original pre-test was around 0.973, and the inter-rater reliability for the pre-test was 0.998. So, Cronbach's Alpha as 0.973 was chosen to adjust the pre-test scores.

4.2.1.3. Prior-knowledge and Treatment Interaction

Because the pre-test measured the subjects' hanzi production performance for the target hanzi, and the instructional treatment of stroke sequence animation concerned the same factor, it was reasonable to assume an interaction between these two terms. Thus, an interaction term of adjusted pre-test and stroke sequence animation was also included in the model.

In summary, the terms included in the model to test the effects of digital voiced pronunciation and the stroke sequence animation on the hanzi production test were adjusted pre-test, stroke sequence animation, voiced pronunciation, interaction of stroke sequence animation and voiced pronunciation, interaction of stroke sequence animation and adjusted pre-test.

4.2.2. Outlier Identification

To identify the outliers, a preliminary running of the collected whole dataset yielded residual values (the difference between the value of an observation and its predicted value according to the model) and the value of root MSE as follows.

Table 4.4.

SAS output for Root MSE value

R-Square	Coeff Var	Root MSE	Post-test Mean
0.494874	36.29700	5.213957	14.36471

As discussed in Montgomery (2001), outlier candidates are observations whose standardized residual is either larger than 2 or smaller than -2. The standardized residual is the ratio of a residual to value of the root MSE, and the root MSE in the study was 5.21 as shown in the above table. Examining the SAS output for residuals yielded the following three outlier candidates.

Table 4.5.

Outlier candidates

Observation No.	Residual	Standardized Residual
13	10.73	2.06
41	-11.78	-2.26
67	12.76	2.45

Since there was no information on why these three subjects behaved so unexpectedly as compared with their group peers, it was decided to keep as many of them in the data as possible. Thus, only the largest positive and negative candidates were treated as outliers in this study (i.e., observations no. 41 and 67), and keep observation no. 13 was kept in the dataset for model checking and data analysis.

4.2.3. Model Adequacy Checking

According to Kleinbaum, Kupper, Muller, and Nizam (1998), to analyze data using the two-way ANCOVA model, three assumptions should be valid: (1) each observation should be from a normally distributed population, (2) each observation should have the same population variance, and (3) all observations should be statistically independent of one another. To simplify the expressions, the study used normality, constancy, and independence referring to the three assumptions, respectively.

A good way to check these assumptions is to draw the diagnostic plots using residual information (Montgomery, 2001). Following is a brief report of adequacy checking of the model. All the checking procedures used the previously discussed model, not including the two outliers.

4.2.3.1. Checking of Normality

The SAS output of "goodness-of-fit tests for normal distribution," as shown in the following table, suggests that the normal distribution should be rejected at Alpha equal to 0.05, but not if Alpha is set as 0.01. Given further considerations that the groups have different numbers of subjects, and each group is relatively small, conclusion is drawn that that the normal distribution assumption was violated but still valid for data analysis.

Table 4.6.

Goodness-of-Fit tests for normal distribution

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.11053812	Pr > D	0.014
Cramer-von Mises	W-Sq	0.16684702	Pr > W-Sq	0.015
Anderson-Darling	A-Sq	0.90181272	Pr > A-Sq	0.022

The Figure 4.1 helps to visualize residual distributions.

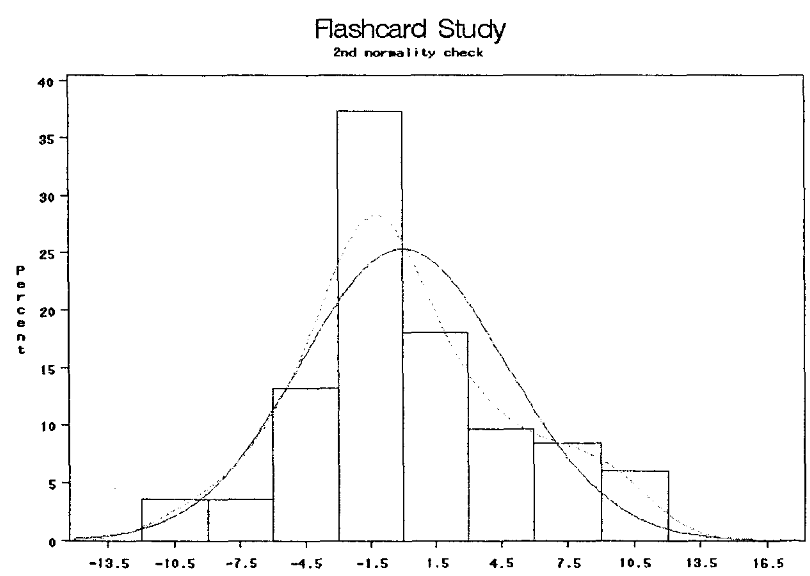


Figure 4.1. Histogram plot for Post-test 1 model adequacy checking.

Note. The solid red line denotes the theoretical distribution. The dotted red line illustrates the observed distribution.

The "normal quantile plot" clearly shows that the residuals generally fit into the theoretical regression line, i.e., the solid red line in the Figure 2, and the residuals almost all fit proportionally to the normal quantiles. However, the rising tail still suggests a violation of the normality assumption.

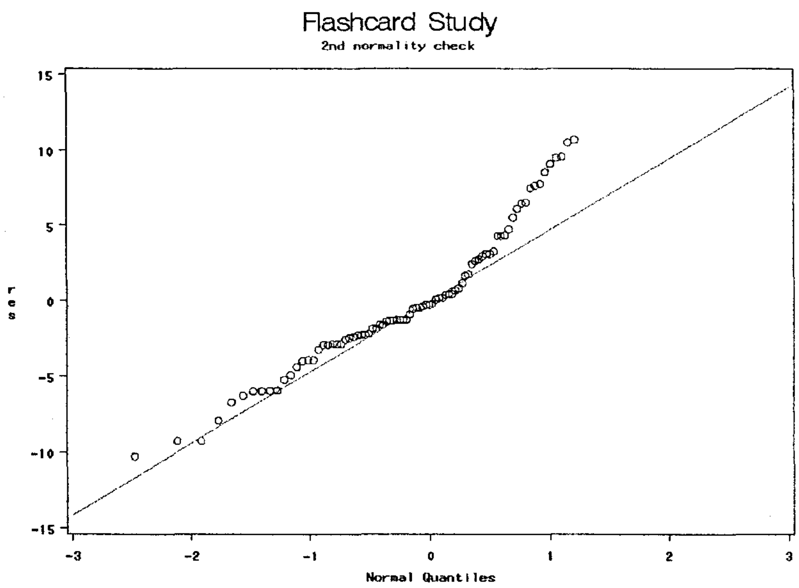


Figure 4.2. QQ-plot for Post-test 1model adequacy checking

Since both the histogram (Figure 1) and the QQ-plot (Figure 2) showed clear deviations from their corresponding theoretical distribution patterns, the researcher needed to find out if data transformation was necessary by calculating Box-Cox transformation lambda. If the best lambda and the convenient lambda are both calculated as a value other than 1, data transformation is needed before conducting the two-way ANCOVA. The data transformation procedure makes the two-way ANCOVA that followed legible, but it is arguable whether or not one should force the data to become normal when it actually is not. Fortunately, as shown in the Table 1, the best lambda is high at 0.8, and the convenient lambda is given as 1, which means there was no need to transform the data collected before analysis.

Table 4.7.

Transformation information for Box-Cox

Transformation Information for BoxCox(post-test)				
Lambda		R-Square	Log Like	
0.6		0.09	-158.624	*
0.8		0.09	-158.237	<
1.0	+	0.09	-159.038	*

< - Best Lambda
* - Confidence Interval
\+ - Convenient Lambda

In summary, the normality assumption was found violated, but not severely, in the normality diagnostic plots and in the goodness-of-fit tests. The Box-Cox lambda further verified no need to perform data transformation to satisfy the normality requirement before the analysis.

4.2.3.2. Checking of Independence

The data in this study were collected across different time sessions (i.e., 9:30-10:20 am, 10:30-11:20 am, and 1:30-2:20 pm). According to Montgomery (2001), it is necessary to check if the plot of residuals versus time shows any clear structure. Runs of positive and/or negative residuals in the plot will give strong evidence for one to reject the independence assumption.

Following is the independence diagnostic plot for this study. Numbers 1, 2, and 3 in the horizontal axis represent the 9:30-10:20 am, 10:30-11:20 am, and 1:30-2:30 pm sessions, respectively. For each time session, it was clear that positive and negative residuals were balanced fairly evenly. Thus, it was reasonable to conclude that there was no evidence of violation of the independence assumption for the collected data.

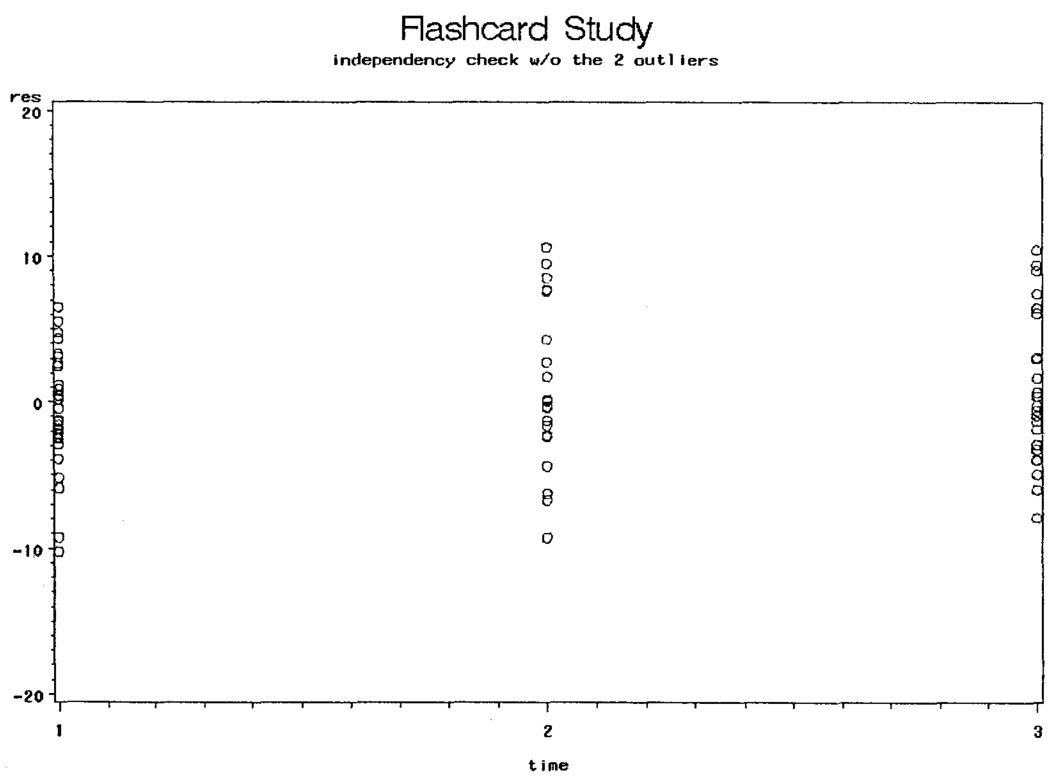


Figure 4.3. Residuals versus times plot for Posttest 1 model adequacy checking

4.2.3.3. Checking of Constancy

The two-way ANCOVA, diagnostic plots for the constancy assumption referred to the plot of residuals versus predicted values and residuals versus each of the two main factors. For the plot of residuals versus predicted values, it was desirable to see the residuals distributed evenly above and below the 0 value of the vertical axis for each predicted value. The Figure below is satisfactory except for the slight skew when the predicted value is close to 9. To draw a conclusion about constancy, it was necessary to study the other two plots (i.e., residuals versus each factor). If the distance between residuals was consistent across different levels of a factor, it was concluded that the constancy assumption was not found violated.

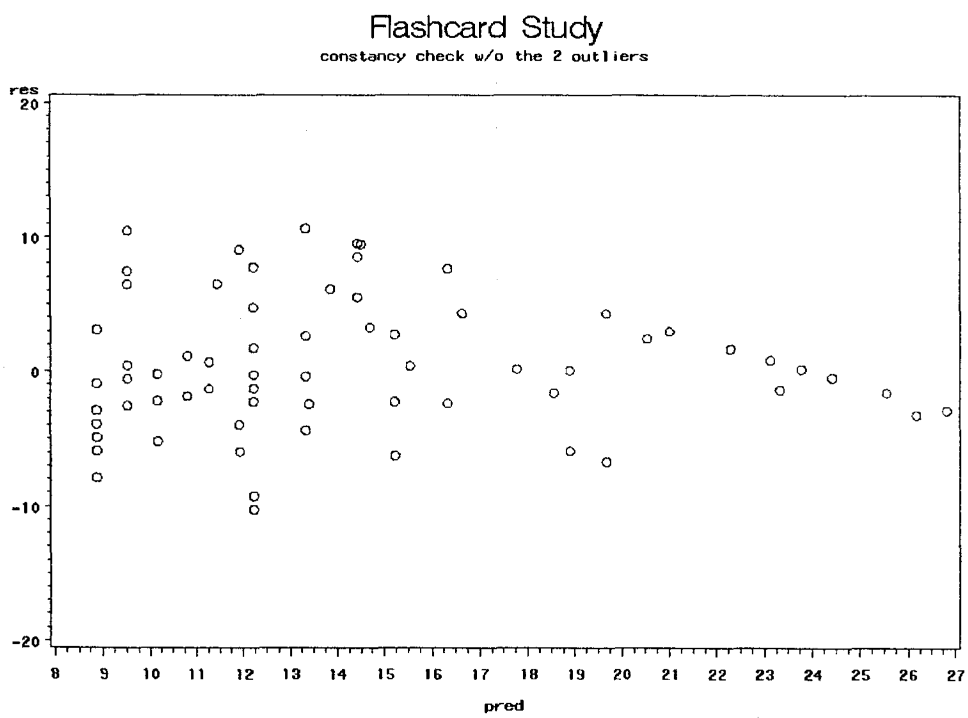


Figure 4.4. Residuals versus predictors plot for Post-test 1 model adequacy checking

In this study, 0 and 1 ("with" and "without") levels were used for both voiced pronunciation and stroke sequence animation factors. The plots are shown below. An examination of these two plots indicated that for the factor of digital voiced pronunciation, the between-residual-distance was fairly consistent. The plot for the stroke sequence animation, however, showed a somewhat more dense distribution of residuals at the "with" level than at the "without" level. However, given the consideration that the difference across the two levels was not significant, it was concluded that the constancy assumption was satisfied.

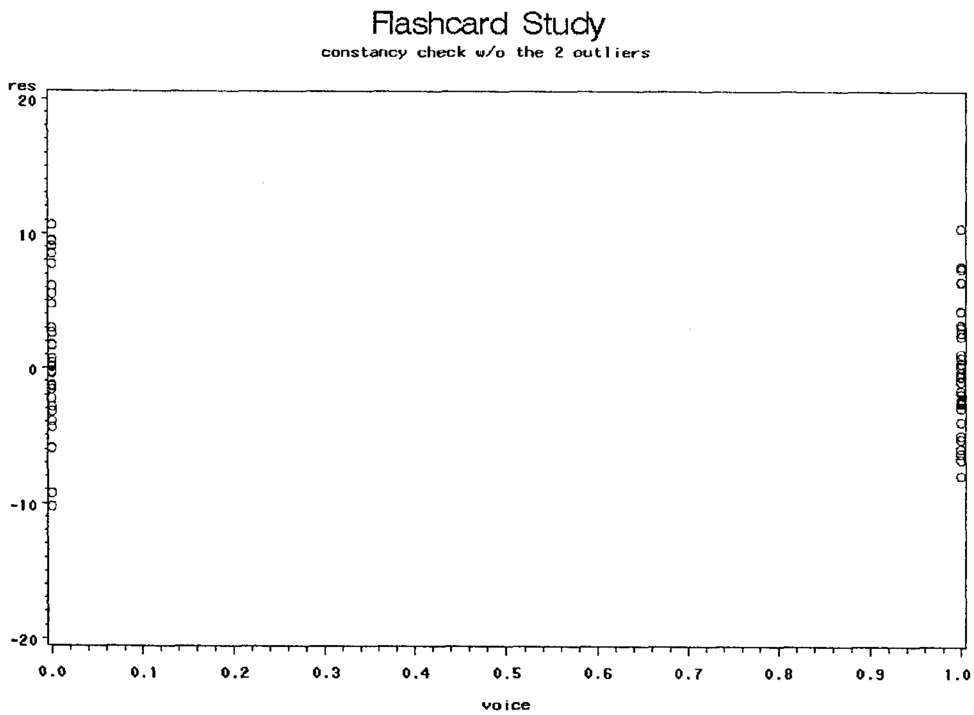


Figure 4.5. Residuals versus voiced pronunciation plot for Post-test 1 model adequacy checking

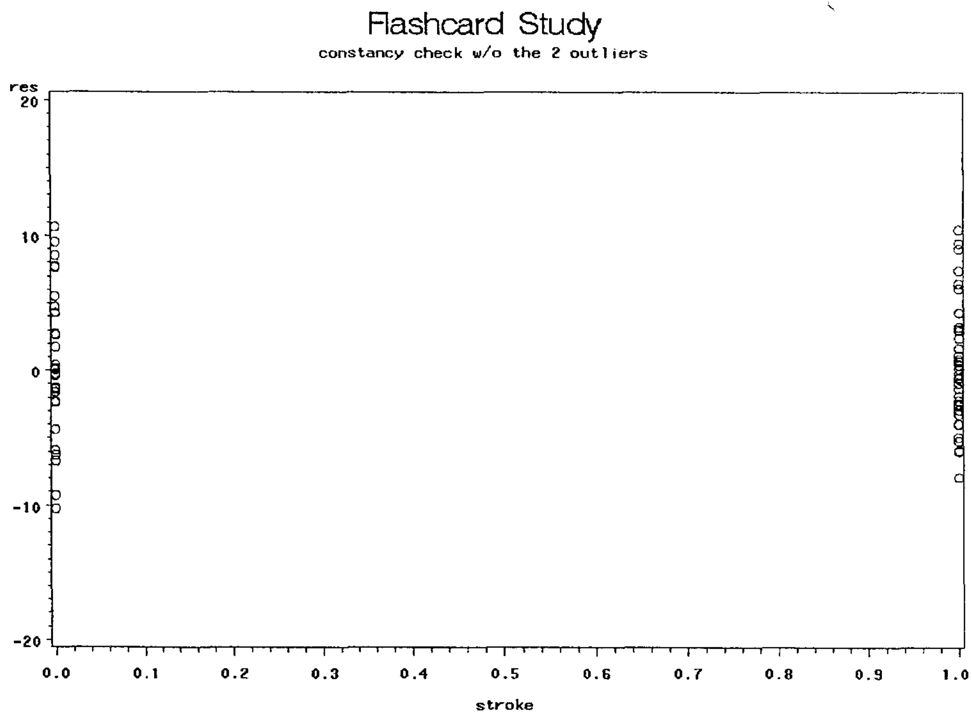


Figure 4.6. Residuals versus stroke sequence animation plot for Posttest 1 model adequacy checking

In summary, the adequacy checking did not find evidence of severe violation of any of the three model assumptions. Thus, it was possible to proceed with the two-way ANCOVA.

4.2.4. The Two-Way ANCOVA

The ANCOVA table shows significant effects of stroke sequence animation ($p=0.01$) and voice and stroke interaction ($p=0.03$).

Table 4.8.

SAS Type III output Posttest 1 2-way ANCOVA test

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Adjusted pre-test	1	834.9982657	834.9982657	35.26	<.0001
Voice	1	1.8361565	1.8361565	0.08	0.7814
Stroke	1	153.3671122	153.3671122	6.48	0.0129
voice*stroke	1	113.5587663	113.5587663	4.80	0.0316
Adjusted pre-test*stroke	1	57.7597392	57.7597392	2.44	0.1225
Error	77	1823.495	23.68175		

The LS means table and comparison matrix (as shown below) suggest that stroke sequence animation had a significant negative effect. More specifically, the adjusted Post-test 1 mean score of the NVNA group (mean=17.3) was found to be significantly higher ($p \approx 0.002$) than that of the WVWA group (mean=11.8). The adjusted Post-test 1 mean score of the WVNA group (mean=20.3) was significantly higher ($p \approx 0.03$) than that of the NVWA group (mean=14.1). The adjusted Post-test 1 mean score of the WVNA group (mean=20.3) was also significantly higher ($p \approx 0.0003$) than that of the WVWA group (mean=11.8).

Table 4.9.

Tukey-Kramer adjustment for least squares means

Index	Group	LS Mean
1	NVNA	17.314
2	NVWA	14.118
3	WVNA	20.325
4	WVWA	11.786

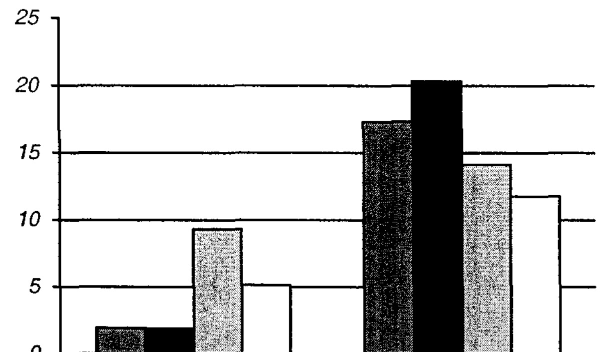
Table 4.10.

Pair-wise comparison matrix for least squares means

i/j	1	2	3	4
1		T=1.788845 P=0.2865	T=-1.60751 P=0.3805	t=3.736415 p=0.0020
2	T=-1.78885 P=0.2865		T=-2.78568 P=0.0333	t=1.491414 p=0.4476
3	T=1.607512 P=0.3805	T=2.785682 P=0.0333		t=4.29001 p=0.0003
4	T=-3.736415 P=0.0020	T=-1.491414 P=0.4476	T=-4.29001 P=0.0003	

Note. In this table, "t" stands for the t-test statistic for differences between paired means. The shadowed p-value is significant at the Alpha=0.05 level. The "i" and "j" are indices of a treatment group under comparison.

The graphic representation of adjusted Pre-test and Post-test 1 means for each group is also provided as follows:



	Pre-test	Post-test1
NVNA	1.9259279	17.3135352
WVNA	1.8837889	20.32459
NVWA	9.3333243	14.1177649
WVWA	5.1544883	11.7860832

Figure 4.7. Adjusted Pretest and Post-test 1 means for each group

Note. The maximum possible score for Post-test 1 was 24.

4.2.5. The Post-hoc Analysis

The two-way ANCOVA served the primary purpose of this study well, that is to examine the effects of voiced pronunciation and stroke sequence animation on character production. However, it was still interesting to explore how the experimental setup affected recognition and pinyin test scores. It was also interesting to find out, given the same treatment, how language background affected Post-test 1.

4.2.5.1. The one-way ANOVA for Post-test 2

The analysis showed that treatment (NVNA, WVNA, NVWA, or WVWA) as a factor significantly ($p \approx 0.01$) affected pinyin test performance. More specifically, a pair-wise comparison indicated that the WVNA group scored significantly higher ($p<0.05$) than both the NVNA group and the WVWA group with a mean difference of 7.7 points and 8.1 points respectively.

The analysis also found that the treatment factor did not significantly ($p \approx 0.49$) affect the recognition test performance at all. A pair-wise comparison did not indicate any between-treatment mean difference significant at the $p<0.05$ level.

It should be noted, however, that since no data were collected on pinyin and recognition pre-tests, it is unknown if prior-knowledge affected the test results and, if so, to what extent.

4.2.5.2. A further analysis for Post-test 1

Although language background, when included in the ANOVA model yielded a highly insignificant p-value of 0.77 and did not have an overall significant effect on Post-test 1 across treatments, it may still have had a significant effect within a given treatment group.

A one-way ANOVA (with pre-test score as covariable) shows that within each treatment group, language background did not significantly affect production Post-test 1, except for the NVWA group. There were three subjects in this group who did not release information on their home exposure to Chinese, and therefore their data were excluded for this analysis. Analysis indicated that, within the NVWA group, language background did have an extremely significant effect ($p<0.0001$). Eight subjects in this group were native speakers of Asian languages (other than Chinese), while the remaining four subjects in this group were native speakers of English. The native speakers of Asian languages performed significantly better on Post-test1 than the English speakers (with a mean difference as high as 14 points).

CHAPTER 5. CONCLUSION OF THE STUDY

5.1. Discussion of ANCOVA Findings

In discussing two-way ANCOVA findings, the interaction effect of the two main factors should be examined first. If the interaction effect is found to be significant, meaning that the effect of one main factor is not consistent with the different levels of the other main factor, pair-wise comparisons of the treatment means should be made to determine the best and worst treatments. If the interaction effect is not found to be significant, then one can safely discuss the main factor effect without performing the pair-wise comparisons.

5.1.1. The Interaction Effect

The study found a significant effect of stroke sequence animation and a significant interaction effect of voice and stroke sequence animation. More specifically, the significant interactive effect refers to the positive effect of voiced pronunciation in the treatment without-stroke-sequence animation, which became negative in the with-stroke-sequence animation treatment. This can be illustrated in the pair-wise comparisons, as discussed in the next section.

5.1.2. The Pair-wise Comparisons

The WVNA treatment group had the highest Post-test 1 mean score, because the advantages of referential connections were made possible through dual-modality processing of information (i.e., both auditory voice and visual texts). The absence of stroke sequence animations also reduced visual loads as compared with the other treatments. The hanzi production performance of this group was significantly better than both the WVWA and the NVWA groups (with $p \approx 0.03$ and $p \approx 0.0003$ respectively).

It is interesting to find that the WVWA treatment group had the lowest character Post-test 1 score, rather than the NVWA group. Based on DCT, one might assume that the presence of voiced pronunciation in the WVWA group would reduce, to some degree, the disadvantages of SAE caused by the overloading of visual information. Thus, the WVWA group would perform better than the NVWA group on Post-test 1. However, the findings showed that the NVWA group outperformed the WVWA group by 2.4 points at an insignificant probability level of $p=0.45$. On one hand, given the observed highly insignificant level for the difference between these two groups, one could reasonably suspect that the observed superiority of NVWA is not sufficiently reliable. On the other hand, if the finding is true, one can conclude that when SAE took place (caused by overwhelming inputs in the visual modality), the extra audio input harmed rather than benefited subjects' performance.

According to SAE, a much more serious negative effect on the WVWA treatment than on the NVNA one might be expected, because of the overwhelming visual load caused by using stroke sequence animations in the former treatment. However, based on DCT, one could also expect that for the WVWA group, the additional audio input would benefit the subjects. Thus, significant difference was not expected between the mean scores of the WVWA and NVNA groups. However, it was found that the WVWA treatment was significantly outperformed ($p \approx 0.002$) by the NVNA on Post-test 1. One plausible explanation for this discrepancy is that the negative split attention effect may be much more significant than the positive dual-coding processing effect. This explanation can be supported by ANCOVA findings on Post-test 1. The effect of stroke sequence animation factor was found to be statistically significant ($p \approx 0.01$), while the effect of voiced pronunciation factor was found to be insignificant ($p \approx 0.78$).

5.1.3. Effect of Language Background or Flashcard Preference as a Factor

To find out if language background or flashcard type preference can be a factor significantly affecting the subjects' performance on the hanzi production test, the two factors were included, one at a time, in a one-way ANCOVA model with terms of adjusted pre-test and treatment. The results indicated that the effects of language background ($p \approx 0.80$) and flashcard type preference ($p \approx 0.60$) were not found significant.

The unfound significance of the effect of language background is consistent with the findings in Ke (1998b), which also report that CFL learners' language background (heritage or non-heritage) was not found to affect their performance significantly on either hanzi production or recognition tasks. Ke's conjecture was that this finding might be because all the learners in his study were in the early hanzi "accumulation stage." This can also be applied to the present study, as all the subjects were also CFL beginners.

The unfound significance of the effect of flashcard type preference may be due to the homogenous media preference of the subjects as a whole. Overall, more than 70%t of the subjects preferred flashcards with both voiced pronunciation and stroke sequence animation. At least half of the subjects in each group preferred this particular flashcard type.

The study appears to speak unfavorably about multi-media effects, because the positive effect of voiced pronunciation was not found to be significant, and the negative effect of stroke sequence animation was found to be significant. Post-hoc analysis, however, helps one understand more precisely the effects under examination, and helps one to believe that these multimedia inputs served their purpose.

5.2. Discussion of Post-hoc Analysis Findings

The post-hoc analysis for the pinyin production performance as in Post-test 2, showed that the WVNA group scored the highest among all the four groups and performed significantly better than the NVNA and WVWA groups. This implies that the voiced pronunciation did significantly improve subjects' pinyin mastery for new vocabulary. It should be noted that this conclusion is made under the assumption that there was no prior-knowledge difference on the pinyin of the target hanzi across the treatment groups. Here, it is understood that the assumption is true because the target hanzi had not been taught before. However, only when a pinyin pre-test is administered, can this assumption actually be tested. As pinyin production was not the primary concern of this study, no pinyin pre-test was given to the subjects. Thus, further investigation is needed to verify this finding.

The post-hoc analysis also reveals the following important findings. Although language background was found to be insignificant across different treatment groups, it was found to be extremely significant ($p<0.0001$) within the NVWA group. This group happened to include only two types of language background, Asian languages other than Chinese and English. The native speakers of Asian languages performed significantly better, by an average of 14 points, than the native speakers of English. It is reasonable to conclude that the Asian languages native speakers benefited much more from the stroke sequence animation treatment than did the native speakers of English.

This post-hoc analysis implies that the effect of stroke sequence animation may closely depend on learners' language backgrounds. While it may improve the performance of learners with a particular language background, it may impair rather than benefit learners who do not have such a background. Such a finding is consistent with that of previous studies. For example, Sasaki (1987) found that stroke by stroke finger writing, a widely observed behavior among native Japanese and Chinese speakers, may help native speakers of Japanese to perceive the images of Kanji graphemes more clearly. It may also aid anagrammatical performance (i.e., manipulating kanji components to form new kanji). Sasaki also found that native speakers of Chinese or Japanese even extend finger writing to English learning tasks. This is unlikely in the learning strategy of native speakers of English. Subjects whose native language does not have Kanji or pictographs usually are not aware of the function of finger writing. Law et al. (1998) found that the importance of stroke sequence is fully understood even by Chinese elementary school children. They further propose that improper stroke order may lead to difficulty in hanzi production memory tasks.

Considering the post-hoc analysis and the relevant literature, this study concludes that the negative effect of stroke sequence animation found in this study is at least partially due to the language background of the subjects. A further examination of the test papers also revealed other information that might explain why stroke sequence animation may not have the expected positive effect. On the test papers, some subjects voluntarily wrote notes. For example, one subject wrote, "easier if we could write the characters also." This subject and another subject actually used pre-test space to practice writing (their scores were not used in the data analysis, because, according to the experiment's instructions, they could memorize only by viewing the flashcards, not by writing). These two subjects were an Asian language speaker and a Chinese heritage speaker, respectively. On one hand, this observation is consistent with Sasaki's findings that the Chinese and/or Japanese prefer to use finger writing when memorizing hanzi/kanji. On the other hand, it implies that if the hanzi/kanji background subjects are provided opportunities to write after the stroke sequence animation, their post-test performance may be even better. Although finger writing is not the primary concern of this study, it seems relevant to the application of stroke sequence animation.

It is also to be noted that the negative effect of stroke sequence animation was only found in situations where learners were already overwhelmed by visual modality inputs. Furthermore, it was only found in the hanzi production from memory test. For other tests, for example, a test concerning proper stroke sequence production, it is unlikely such a negative effect would be found.

5.3. Implications

The findings of this study have both theoretical and practical implications. The animations applied in previous studies on DCT or SAE were always conventional nonverbal representations. The stroke sequence animations of Chinese characters that were used in this study, however, were semi-verbal in nature. They were, in part, non-verbal because they had most of the properties of non-verbal representations, as discussed by Clark and Paivio (1991). They were, in part, also inherently verbal representations, because each frame of the animations consisted of only strokes of Chinese characters, which definitely convey verbal information. Therefore, it is reasonable to assume an extension of DCT that for situations where verbal information is presented in a non-verbal way, such as the animations used in this study, DCT and SAE both hold. Furthermore, based on the findings of this study, one can also set forth a hypothesis for situations when the dual-coding effect and SAE are both present. Under such conditions, one can expect the negative SAE to be much more significant than the positive dual-modality effect. However, this assumption needs further evidence.

The practical implications are straightforward. Since this experiment implies that multimedia do not necessarily increase the effectiveness of character memorization and production, it is essential for instructional multimedia designers to ensure effectiveness before undertaking major developments of costly multimedia projects. For educators, this study suggests that more attention to learner characteristics is needed when they incorporating multimedia instructional materials in their lesson plans.

Finally, with an R-square value around 0.54, which means that about 54% of the variations in Post-test 1 scores can be explained by variations of factors in the model, the model used for this study is acceptable as a social science study. This acceptable, but some-what low, R-square value was caused mainly by the research question. That is, the focus of this study was on the effect of multimedia. Individual factors such as learning aptitude, spatial ability, and so on were not taken into account. It is certainly desirable to find out all the factors that can better explain the Post-test 1 scores. Such findings are helpful in identifying all factors, not limited to multimedia factors, which can enhance students' Chinese character memorization. Further investigations in this direction are needed.

5.4. A Brief Summary of the Study

This study analyzed 83 valid data collected at a large, Midwestern public university to examine the effects of voiced pronunciation and stroke sequence animation, inputs usually found in typical computer-based multimedia Chinese character flashcards. The following table summarizes the hypotheses and findings (both primary and post-hoc analysis) of this study.

Table 5.1.

Test hypotheses and conclusions

Null Hypotheses	Conclusions
No voiced pronunciation effect.	Do not reject (p=0.78)
No stroke sequence animation effect.	Reject (p=0.01). A significant negative effect was found in this experimental setting.
No interaction effect between voiced pronunciation and stroke sequence animation.	Reject (p=0.03). Pair-wise comparisons were performed, and the WVNA treatment was found to be the best on the hanzi production test.
Pinyin test mean scores will be equal across treatments.	Reject (p=0.01). The WVNA treatment also resulted in the best pinyin test score.
Recognition test mean scores will be equal across treatments	Do not reject (p=0.49).
Language background will have no effect across treatments.	Do not reject (p=0.77).
Subjects with different language backgrounds will have identical mean scores within a treatment.	Reject for the NVWA group (p<0.0001). Native speakers of Asian languages performed much better than native speakers of English.

Note. In this study, every alternative hypothesis for its corresponding null hypotheses was set as a 2-sided one.

Further investigations on the effect of stroke sequence animation will be needed to find out, for example, whether this media input can assist learners' mastery of proper stroke sequences. The conclusions reached in this study on multimedia effects in foreign language education need additional testing so that better products can be developed and appropriate classroom materials can be selected.

LIST OF REFERENCES

LIST OF REFERENCES

Allen, J. R. (1992). I will speak, therefore, of a graph: A Chinese metalanguage. *Language in Society, 21*(2), 189-206.

AllPsych and Heffner Media Group Inc. (2004). *Research methods: Chapter 7.* Retrieved Feburary, 27, 2005, from http://allpsych.com/researchmethods/validityreliability.html

Anderson, R. C., Li, W. L., Ku, Y. M., Shu, H., & Wu, N. N. (2003). Use of partial information in learning to read Chinese characters. *Journal of Educational Psychology, 95*(1), 52-57.

Atkinson, R. C. (1975). Mnemotechnics in second-language learning. *American Psychologist, 30*, 821-828.

Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory, 104,* 126-133.

Begg, I. (1973). Imagery and integration in the recall of words. *Canadian Journal of Psychology, 27*, 159-167.

Beijing Language Institute. (1983). *Practical Chinese reader.* Beijing, China: Beijing Language Institute Press.

Bower, G. H. (1970). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior, 9*, 529-533.

Broadbent, D. E. (1956). Successive responses to simultaneous stimuli. *Quarterly Journal of Experimental Psychology, 8*, 145-152.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*, 233-246.

Chao, Y. R. (1992). *Yu yan wen ti*. Taipei, Taiwan: Commercial.

Chen, X., Shu, H., Wu, N. N., & Anderson, R. C. (2003). Stages in learning to pronounce Chinese characters. *Psychology in the Schools, 40*(1), 115-124.

Ching, E. (1983). Why a calligraphy course for language students? Paper presented at the Annual Meeting of the American Council on the Teaching of Foreign Languages, San Francisco, CA.

Ching, N., & Ching, E. (1975). Teaching the writing of Chinese characters. *Journal of the Chinese Language Teachers Association, 10*(1), 20-24.

Clariana, R. B. (2001). The effects of recognition and recall study tasks with feedback in a computer-based vocabulary lesson. *Educational Technology Research & Development, 49*(3), 23-36.

Clark, J. M., & Paivio, A. (1987). A dual coding perspective on encoding processes. In M. A. McDaniel, & Pressley, M. (Ed.), *Imagery and related mnemonic processes: Theories, individual differences, and applications* (pp. 5-33). New York: Springer-Verlag.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149-210.

Cohen, A. D., & Aphek, E. (1980). Retention of second-language vocabulary over time: Investigating the role of mnemonic associations. *System, 8*, 221-235.

Colquhoun, W. P. (1975). Evaluation of auditory, visual, and dual-mode displays for prolonged sonar monitoring in repeated sessions. *Human Factors, 17*, 425-437.

De Francis, J. (1984). *The Chinese language: Fact and fantasy.* Honolulu, HI: University of Hawaii Press.

Delaney, H. D. (1978). Interaction of individual differences with visual and verbal elaboration instructions. *Journal of Educational Psychology, 70*, 306-318.

Duan, D. W., & Cuvo, A. J. (1996). Comparison of prototype and rote instruction of English names for Chinese visual characters. *Journal of Applied Behavior Analysis, 29*(1), 125-127.

Edwards, L. (1997). Vocabulary expansion in modern standard Chinese. *Babel: Australia, 32*(3), 24-25, 29.

Everson, M. E. (1988a). Speed and comprehension in reading Chinese: Romanization vs. characters revisited. *Journal of the Chinese Language Teachers Association, 23*(2), 1-15.

Everson, M. E. (1988b). Word recognition among learners of Chinese as a foreign language: Investigation the relationship between naming and knowing. *Modern Language Journal, 82*(2), 194-204.

Fang, S. P., Horng, R. Y., & Tzeng, O. J. L. (1986). Consistency effects in the Chinese character and pseudo-character naming tasks. In H. S. R. Kao, & Hoosain, R. (Ed.), *Linguistics, psychology, and the Chinese language* (pp. 11-22). Hong Kong: University of Hong Kong Press.

Hall, J. W. (1988). On the utility of the keyword mnemonic for vocabulary learning. *Journal of Educational Psychology, 80*(4), 554-562.

Halpern, J. (1974). Learning to utilize information presented over two sensory channels. *Perception & Psychophysics, 16*(2), 321-328.

Hede, A. J. (1980). Dichotic and bisensory grouping effects. *Journal of Experimental Psychology, 32*, 295-306.

Hsu, H.-M., & Gao, L. (2002). Computer-Mediated Materials for Chinese Character Learning. *CALICO Journal, 19*(3), 533-536.

Hue, C. W. (1992). Recognition processes in characternaming. In H. C. Chen, & Tzeng, O. J. L. (Ed.), *Language Processing in Chinese* (pp. 93-107). Amsterdam, North-Holland: Elsevier Science.

Johnston, C. W. (1974). *An experimental study of two methods of vocabulary development.* Unpublished doctoral dissertation, University of Colorado.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied cognitive psychology, 13*, 351-371.

Kappel, S., Harford, M., Burns, V. D., & Anderson, N. S. (1975). Effects of vocalization on short term memory for words. *Journal of Experimental Psychology, 101*, 314-317.

Ke, C. R. (1996a). An empirical study on the relationship between Chinese character recognition and production. *Modern Language Journal, 80*(3), 340-350.

Ke, C. R. (1996b). *A model for Chinese orthographic awareness developmental stages.* Unpublished manuscript.

Ke, C. R. (1998a). Effects of strategies on the learning of Chinese characters among foreign language students. *Journal of the Chinese Language Teachers Association, 33*(2), 93-112.

Ke, C. R. (1998b). Effects of language background on the learning of Chinese characters among foreign language students. *Foreign Language Annals, 31*(1), 91-100.

Kinchla, R. A. (1974). Detecting target elements in multielement arrays: A confusability model. *Perception and Psychophysics, 15*, 149-158.

Kleinbaum, D. G., Kupper, L.L., Muller, K.E., & Nizam, A. (1998). *Applied Regression Analysis and Other Multivariable Methods*. Pacific Grove, CA: Duxbury.

Kobus, D. A., Russotti, J., Schlichting, C., Haskell, G., Carpenter, S., & Wojtowicz, J. (1986). Multimodal detection and recognition performance of sonar operators. *Human Factors, 28*, 23-29.

Koda, K. (1996). L2 Word Recognition Research: A Critical Review. *Modern Language Journal, 80*(4), 450-460.

Ku, Y. M., & Anderson, R. C. (2003). Development of morphological awareness in Chinese and English. *Reading and Writing: An Interdisciplinary Journal, 16*, 399-422.

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*(1), 115-128.

Law, N., Ki, W. W., Chung, A. L. S., Ko, P. Y., & Lam, H. C. (1998). Children's stroke sequence errors in writing Chinese characters. *Reading and Writing: An Interdisciplinary Journal, 10*(3-5), 267-292.

Laychuk, J. L. (1983). *The use of etymology and phonetic symbols (zhuyin fuhao) in teaching first year Chinese*. Paper presented at the Annual Meeting of the American Council on the Teaching of Foreign Languages, San Francisco, CA.

Leong, C. K. (1973). Hong Kong. In J. Downing (Ed.), *Comparative reading: Cross-national studies of behavior in reading and writing*. New York: Plenum.

Leong, C. K., Cheng, P. W., & Mulcahy, R. (1987). Automatic processing of morphemic orthography by mature readers. *Language and Speech, 30*(2), 181-197.

Lewandowski, L. J., & Kobus, D. A. (1989). Bimodal information processing in sonar performance. *Human Performance, 2*, 73-84.

Lewandowski, L. J., & Kobus, D. A. (1993). The effects of redundancy in bimodal word processing. *Human Performance, 6*, 229-239.

Li, S. (1997). *The semantic analysis method for teaching Chinese characters.* Unpublished doctoral dissertation, State University of New York at Albany.

Li, W. (2004). *The 'grapheme combination method': Teaching and learning Chinese characters through associative links.* Unpublished doctoral dissertation, George Mason University.

Li, X.-L. (1996). HyperCharacters: A pilot study in computerized learning of Chinese characters. *CALICO Journal, 14*(1), 77-94.

Liu, I. (1983). The learning of characters: A conceptual learning approach. *Journal of the Chinese Language Teachers Association, 18*(2), 65-76.

Lu, M. Y., Webb, J. M., Krus, D. J., & Fox, L. S. (1999). Using order analytic instructional hierarchies of mnemonics to facilitate learning Chinese and Japanese "Kanji" characters. *Journal of Experimental Education, 67*(4), 293-311.

Ma, J. H. S. (2000). *Keys to Chinese character writing: Step-by-step directions to writing characters quickly and easily.* (Eric Document Reproduction Service No. ED450600).

Martin, M. (1980). Attention to words in different modalities: Four-channel presentation with physical and semantic selection. *Acta Psychologica, 44*, 99-115.

McBride-Chang, C., Shu, H., Zhou, A. B., Wat, C. P., & Wagner, R. K. (2003). Morphological awareness uniquely predicts young children's Chinese character recognition. *Journal of Educational Psychology, 95*(4), 743-751.

Mickel, S. (1980). Teaching the Chinese writing system. *Journal of the Chinese Language Teachers Association, 15*(1), 91-98.

Montali, J., & Lewandowski, L. (1996). Bimodal reading: Benefits of a talking computer for average and less skilled readers. *Journal of Learning Disabilities, 29*, 271-279.

Montgomery, D. C. (2001). *Design and analysis of experiments* (5^{th} ed.). New York: John Wiley & Sons.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358-368.

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*(1), 117-125.

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94*(1), 156-163.

Mori, Y., & Nagy, W. (1999). Integration of Information from Context and Word Elements In Interpreting Novel Kanji Compounds. *Reading Research Quarterly, 34*, 80-101.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319-334.

Ng, Y. H., & Wu, W. H. (1990). Learning to write Chinese from first principles. *Computer and Education, 15*(1-3), 119-126.

Nickerson, R. S. (1973). Intersensory facilitation of reaction time: Energy summation or preparation enhancement? *Psychological Review, 80*(6), 489-509.

Norman, J. (1996). Learning Chinese in the 1990s. *ADFL Bulletin, 27*(2), 4-8.

Ott, C. E., Butler, D. C., Blake, R. S., & Ball, J. P. (1973). The effect of interactive-image elaboration on the acquisition of foreign language vocabulary. *Language Learning, 23*, 197-206.

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review, 76*, 241-263.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart, and Winston.

Paivio, A. (1975). Coding distinctions and repetition effects in memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 179-214). New York: Academic.

Paivio, A. (1986). *Mental representations: A dual-coding approach.* New York: Oxford University Press.

Paivio, A., & Begg, I. (1981). *Psychology of language*. Englewood Cliffs, NJ: Prentice Hall.

Paivio, A., & Lambert, W. (1981). Dual coding and bilingual memory. *Journal of Verbal Learning and Verbal Behavior, 20*, 532-539.

Peng, D.-L., & Yang, H. (1994). *The psychological research into naming Chinese characters.* Paper presented at the Workshop on Cognitive Processing of Asian Languages, University of New South Wales, Sydney, Australia.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17*(4), 398-422.

Piaget, J. (1977). The role of action in the development of thinking. In W. F. O. J. M. Gallagher (Ed.), *Knowledge and development* (Vol. 1, pp. 17-42). New York: Plenum.

Pressley, M., Levin, J. R., & Delaney, H. D. (1982). The mnemonic keyword method. *Review of Educational Research, 52*(1), 61-91.

Pressley, M., Levin, J. R., Nakamura, G. V., Hope, D. J., Bispo, J., & Toye, A. R. (1980). The keyword method of foreign vocabulary learning: An investigation of its generalizability. *Journal of Applied Psychology, 65*, 635-642.

Qin, Y. S. (2000). *My opinions on Chinese input methods.* Retrieved December, 24, 2004, from http://demo.phdcr.com/~ysqin/my_webs/wz/wbzx.htm

Raugh, M. R., & Atkinson, R. C. (1975). A mnemonic method for learning a second-language vocabulary. *Journal of Educational Psychology, 67*, 1-16.

Robert, A. (2002). *Hanzi Master (V1.3).* Retrieved October, 17, 2005, from http://kamares.ucsd.edu/~arobert/hanzim.html

Sadoski, M., & Paivio, A. (2001). *Imagery and text: A dual coding theory of reading and writing*. Mahwah, NJ: Lawrence Erlbaum.

Saito, H., Masuda, H., & Kawakami, M. (1998). Form and sound similarity effects in kanji recognition. *Reading and Writing: An Interdisciplinary Journal, 10*, 323-357.

Sasaki, M. (1987). Why do Japanese write characters in space? *International Journal of Behavioral Development, 10*(2), 133-149.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition, 19*, 1-30.

Sergent, W. K., & Everson, M. E. (1992). The effects of frequency and density on character recognition speed and accuracy by elementary and advanced L2 readers of Chinese. *Journal of the Chinese Language Teachers Association, 27*(1-2), 29-44.

Serruys, P. L.-M. (1962). *Survey of the Chinese Language Reform and the Anti-Illiteracy Movement in Communist China*. Berkeley: Center for Chinese Studies, University of California.

Shu, H., Chen, X., Anderson, R. C., Wu, N. N., & Xuan, Y. (2003). Properties of school Chinese: Implication for learning to read. *Children Development, 74*(1), 27-47.

Smith, B. D., Miller, C., Grossman, F., & Valeri-Gold, M. (1994). Vocabulary retention: Effects of using spatial imaging on hemispheric-preference thinkers. *The Journal of Research and Development in Education, 27*(4), 244-252.

Tan, L. H., Hoosain, R., & Peng, D. L. (1995). Role of early presemantic phonological code in Chinese character identification. *Journal of Experimental Psychology, 21*, 43-54.

Tan, L. H., Hoosain, R., & Siok, W. W. T. (1996). Activation of phonological codes before access to character meaning in written Chinese. *Journal of Experimental Psychology, 22*(4), 865-882.

Tell, P. M., & Ferguson, A. M. (1974). Influence of Active and Passive Vocalization on Short-Term Recall. *Journal of Experimental Psychology, 102*(2), 347-349.

Treisman, A. M., & Davies, A. (1973). Divided attention to ear and eye. In S. Kornblum (Ed.), *Attention and performance* (Vol. IV, pp. 101-117). New York: Academic.

Trochim, W. M. K. (2004). *The nonequivalent groups design*. Retrieved October 15, 2004, from http://www.socialresearchmethods.net/kb/quasnegd.htm

Usuki, M. (2000). Promoting learner autonomy: learning from the Japanese language learners' perspectives. (Eric Document Reproduction Service No. ED450588).

Usuki, M. (2001). Learner Autonomy: Learning from the Student's Voice. (Eric Document Reproduction Service No. ED452698).

Vaid, J. (1988). Bilingual memory representation: A further test of dual coding theory. *Canadian Journal of Psychology, 42*, 84-90.

Van Aacken, S. (1996). The Efficacy of CALL in Kanji Learning. *On-Call, 10*(2), 2-14.

Wang, A. Y., & Thomas, M. H. (1992). The effect of imagery-based mnemonics on the long-term retention of Chinese characters. *Language Learning, 42*(3), 359-376.

Wang, S.-H. C. (1998). A study on the learning and teaching of hanzi-Chinese characters. *Working Papers in Educational Linguistics, 14*(1), 69-101.

Wu, J.-T., Chou, T.-L., & Liu, I.-M. (1994). The locus of the character/word frequency effect. In H.-W. Chang, Huang, J.-T., Hue, C.-W., & Tzeng, O. J. L. (Ed.), *Advances in the study of Chinese language processing* (Vol. 1, pp. 31-58). Taipei, Taiwan: National Taiwan University Press.

Yang, J. (2000). Orthographic effect on word recognition by learners of Chinese as a foreign language. *Journal of the Chinese Language Teachers Association, 35*(2), 1-18.

Yin, J. J. H. (2003). A survey on strategies that American college students used in memorizing Chinese characters. *Journal of the Chinese Language Teachers Association, 38*(3), 69-90.

ZhongHuaZiJing. (2004). *Han Zi Te Dian*. Retrieved January, 25, 2005, from http://www.zhzj.com.cn/hzzg_td.htm

APPENDICES

Appendix A Waiver of Informed Consent

Revised 11/04/2002 Ref. # 04-208 E

(Approved on March 04, 2004)

Purdue University

Committee On The Use of Human Research Subjects

1. Project Title: The Effects of Voice and Stroke Animation Inputs on Short-term Memory of Written Form of Chinese Characters
2. Principal Investigator: Dr. Alan Garfinkel
3. Informed consent of the subject is one of the fundamental principles of ethical research for human subjects. Informed consent is also mandated by Federal regulations (45 CFR 46) and University policy for research with human subjects. An investigator should seek a waiver of written or verbal informed consent, or required elements thereof, only under compelling circumstances.

Please check below the type of waiver you are requesting:

1. __X_ Waiver of written informed consent because the only record linking the subject and the research would be the consent document and the principal risk would be potential harm resulting from breach of confidentiality.
2. __X_ Waiver of written informed consent because the research presents no more than minimal risk of harm to subjects and involves no procedures for which written consent is normally required outside of the research context.
3. ____ Waiver or modification of specific written informed consent elements, specifically: (insert what elements you request to have waived)

I believe that this request is appropriate because: (check all that apply)

___X_ (1) The research presents no more than minimal risk of harm to subjects.

___X_ (2) The waiver or alteration will not adversely affect the rights and welfare of the subjects.

___X_ (3) The research could not practicably be carried out without the waiver or alteration.

___X_ (4) Whenever appropriate, the subjects will be provided with additional pertinent information after participation.

______________________________ March 4, 2004

Principal Investigator Signature Date

Appendix B Permissions to Use Figure 4-1 and Table 4-2 from "Mental representations: A dual coding approach" by Paivio (1986)

From: Al Paivio [apaivio@uwo.ca]

Sent: Tuesday, September 27, 2005 5:16 PM

To: ZHU

Subject: Re: Reprint Permission Request

Dear Nr Zhu,

You certainly have my permission to use the material but Oxford owns the copyright so you should ask them too. The name I have for that is

Scott Marinaro, Rights Associate
Oxford University Press
198 Madison Avenue
New York, NY 10016-4314
212 726 6037 Permission Phone

Perhaps you've already asked Oxford.

Good luck with your thesis,

Allan Paivio

On Tue, 27 Sep 2005, ZHU wrote:

Dear Professor Paivio,

I'm writing to request for your permissions to use Table 4-2. (Orthogonal conceptual relation between symbolic systems and sensorimotor systems with examples of types of modality-specific information represented in each subsystem) that appeared on page 57 and Figure 4-1. (Schematic depiction of verbal and nonverbal systems) that appeared on page 67 of your book, "Mental Representations: A Dual Coding Approach" published by Oxford University Press (New York) in 1986.

I would appreciate it greatly if your permissions can be granted for me to use both of them (Table 4-2 & Figure 4-1) in my Ph.D. dissertation entitled "Effects of voiced-pronunciation and stroke sequence animation on production of characters by beginners of Chinese as a foreign language" which is to be deposited in November, 2005.

Many thanks for your kind attention and time.

Yours Sincerely,

Yu Zhu
Ph.D. Candidate
Department of Curriculum & Instruction
College of Education
Purdue University
West Lafayette, IN47907
U.S.A.

Inquiry: Permission Requests StatusFrom: Permissions [permissions.us@oup.com]
Sent: Tuesday, October 25, 2005 11:33 AM
To: ZHU
Subject: RE: Inquiry: Permission Requests Status

Dear Yu Zhu,

Contracts were just returned to us yesterday and weren't deliverable. Can you provide us with a better address than below?

Yu Zhu
Purdue University
Nimitz Drive
West Lafayette, IN 47906

Many thanks and best,

Frank Quinn
-----Original Message-----
From: ZHU [mailto:zhu1@purdue.edu]
Sent: Tuesday, October 25, 2005 1:28 AM
To: Permissions
Subject: Inquiry: Permission Requests Status

Dear Sir/Madam,

I'm a doctoral student graduating in December this year. I'm writing to ask the status of my permission requests sent on September 26 via post mail and October 16 online. I had to get your permissions to use the two graphs in my dissertation and deposit it by the due day.

Detail information about my requests is as below:

Permissions requested for me to use Table 4-2 & Figure 4-1 (page 57 & 67 respectively in the book “Mental Representations: A Dual Coding Approach”, ISBN 0-19-503936-X; OUP; 1986) in my Ph.D. dissertation entitled "Effects of voiced-pronunciation and stroke sequence animation on production of characters by beginners of Chinese as a foreign language" which is to be deposited soon.

I’m very grateful to your attention, time, and kind help. I look forward to hearing from you soon.

Yours Sincerely,

Yu Zhu
Ph.D. Candidate
Department of Curriculum & Instruction
College of Education
Purdue University
West Lafayette, IN47907
Tel. (765) 404-8576
Fax: (765) 496-4398

From: "Quinn, Frank" <frank.quinn@oup.com>

To: <zhu1@purdue.edu>

Subject: RE: urgent inquiry: my two permission requests submitted on 09/25/2005

Date: Friday, November 18, 2005 12:12 PM

Dear Yu

I am happy to report to you that both Table 4.2 and Figure 4.1 have been cleared and I will issue you a new contract that includes both gratis. We don't send contracts over email so use this email as a temporary license to submit your paper and the contracts will arrive to you soon.

Thanks and best,

Frank

Appendix C Data Collection Instruments

Part I Demographic Information

For each of the following two statements, please circle the choice that best matches your situation.

Your mother tongue is
A. Chinese B. Asian (other than Chinese) C. Others

You prefer Chinese character flashcards with
A. voice B. stroke animation
C. both D. no voice nor stroke animation

Part II: Pre-test

Please write down the Chinese character(s) for each of the following English words.

to handle
air or angry
a pair of
measure word for hat
suitcase
visa
hat or cap
shoes
slippery or to skate
ice
winter
kick
soccer
team
fair
referee
competition
lose
win

Part III: Post-test 1

to handle
air or angry
a pair of

measure word for hat
suitcase
visa
hat or cap
shoes
slippery or to skate
ice
winter
kick
soccer
team
fair
referee
competition
lose
win

Part IV: Post-test 2

Please write down **Pinyin** and **English Meaning** for each Chinese word given

办
气
双
顶
箱
签证
帽
鞋
滑冰
冬
踢
足球
队
公平
比赛
输
赢

Appendix D Guide for Lab Activity for the Four Experimental Groups

Chinese 102 Computer Lab Activity

03/29/2004

This activity is scheduled to evaluate one of the instructional materials that Chinese program will develop soon in order to help our students memorize Chinese character more efficiently.

You will be asked to memorize the information about new words of Lesson 28, using the material assigned by your teacher (the use of each new word will be explained by your teacher tomorrow). You will then be asked to do a pre-test, two post-test before and after you use the material. Pre-test and post-tests data will be both anonymous.

The time for pre-test is 2 minutes.

The time allocated to study is 30 minutes. Try your best to memorize the shape, the pinyin and the English meaning of each new word.

For post-tests, you have 10 minutes.

Please follow the teacher's instruction for your steps.

As the more data we get, the more precise conclusion we can draw upon, your participation in this activity will be greatly encouraged and appreciated. However, if you do not want to participate, you can simply leave the data collection sheet blank. You decision about participation and your performance on pre-test or post-tests will **NOT** affect your grade on course Chinese 102 in any matter.

Thanks for your attention and time, and hope you can find the material helpful!

Appendix E Inter-rater Reliability Information Sheet

This section above the straight-line is the method and results determining observations used in calculation of the inter-rater pre-test reliability.

The other rater randomly chooses 21 observations (i.e., 25% of the valid observations in this study) coded as the following.

If a randomly chosen code does not represent an observation used for the 2-way ANOVA analysis, the code will be discarded, and she will be told to choose randomly an alternative code.

Please choose and write down 7 codes at random from A1-A16 and D1-D17.

(Result: A1, A4, A15, D5, D7, D10, D17)

Please choose and write down 2 codes at random from F1-F10.

(Result: F9, F10)

Please choose and write down 4 codes at random from E1-E16.

(Result: E1, E3, E8, E13)

Please choose and write down 8 codes at random from B1-B16 and C1-C20.

(Result: B4, B5, B7, B10, C1, C3, C5, C17)

Here are the target *hanzi* for the pre-test. Please note that *hanzi* in parenthesis will not be graded. Full point for this pre-test is 24 with 1 point for each *hanzi*. Grader: please use your own judgment, but no partial credits.

1 办 2 气 3 双 4 顶 5 箱(子) 6 签证 7 帽(子) 8 鞋(子) 9 滑 10 冰 11 冬(天)
12 踢 13 足球 14 队 15 公平 16 裁判 17 比赛 18 输 19 赢

Following is the counts for the number of unmatched inter-rater grading in both pre-test and production post-test. For example, "A1 1, 0" means for the observation coded as A1, there was 1 unmatched grading in the pre-test, and zero unmatched grading in the production post-test.

NVNA

A1 1, 0; A4 0, 1; A15 0, 1; D5 0, 0; D7 0, 2; D10 0, 0; D17 0, 0.

WVNA

F9 0, 2; F10 0, 0.

NVWA

E1 0, 2; E3 0, 1; E8 0, 0; E13 0, 0.

WVWA

B4 0, 0; B5 0, 0; B7 0, 1; B10 0, 2; C1 0, 1; C3 0, 0; C5 0, 1; C17 0, 0.

The pre-test inter-rater reliability is 0.998. The inter-rater reliability for post-test1 is 0.972.

VITA

VITA

Education

Purdue University, West Lafayette, IN, 2005
Doctor of Philosophy in Foreign Language Education & Graduate Certificate in Applied Statistics

The University of Hong Kong, Hong Kong, China, 2001
Master of Philosophy in Chinese Rhetoric

Hebei Normal University, Hebei Province, China, 1995
Bachelor of Arts in Chinese Language and Literature

Hebei University, Hebei Province, China, 1992
Associate of Arts in Chinese Language and Literature

Honors, Awards, and Scholarships

Purdue University, Department of Foreign Languages & Literatures, International Conference Presentation Award 2004
Purdue University, Department of Foreign Languages & Literatures, Finalist for Excellent Teaching Assistant Award 2002
The University of Hong Kong Postgraduate Scholarship 1998-2000
The University of Hong Kong Postgraduate Student Research Award 1998-2000
The University of Hong Kong International Conference Presentation Award 2000
Hebei University Award for Excellence in Scholarship 1992
Hebei University Scholarship 1991-1992

Teaching and Research Experience

Teaching Assistant, Chinese Program, Department of Foreign Languages & Literatures, Purdue University, West Lafayette, IN, 2002-2005

- Teach Chinese courses.
- Grade students' homework and tests.
- Assist in the course related activities organized by the Chinese program.

Research Assistant, Support Centre for Teachers Using Chinese as the Medium of Instruction, The University of Hong Kong, H.K., 1999-2003

- Analyze assigned texts with a systemic functional linguistics approach.
- Assist in preparing instructional materials for teacher training seminars.
- Assist in preparing grant application materials.

Academic Support Experience

Chinese 101 Acting Coordinator, Chinese Program, Department of Foreign Languages & Literatures, Purdue University, West Lafayette, IN, Fall 2003

- Prepare the teaching schedules and syllabus.
- Prepare the examinations and test papers.
- Lead the bi-weekly teaching assistants' meetings.

Course Assistant in the course of Foreign Language Teaching Methods, Department of Curriculum & Instruction, Purdue University, West Lafayette, IN, Fall 2003

- Assist in preparing teaching materials and facilities.
- Lead group discussions in the class.

Student Researcher for the project of the Digital Pack (CD-Rom) of New Practical Chinese Reader, Department of Foreign Languages & Literatures, Purdue University, West Lafayette, IN, Spring & Summer 2003

- Assist in research & grant application.
- Develop template for the CD-Rom.

Instructional Multimedia Developing Experience

Instructional Designer & Script Writer, Instructional TV series project for the textbook, *New Practical Chinese Rea*der (Volume I-IV), Funded by China National Office for Teaching Chinese as a Foreign Language (NOTCFL), May – December, 2005

- Design and write up detailed instructional scripts.
- Plan and describe in detail the multimedia inserts.

Chinese Courses WebCT Developer & Maintainer, Purdue University, West Lafayette, IN, 2002-2005

- Develop voluntarily WebCT for Chinese courses.
- Oversee the Chinese WebCT operations.
- Maintain and update contents as necessary.
- Help colleagues and students using the system.

Assistant, Media Center, Department of Foreign Languages & Literatures, Purdue University, West Lafayette, IN, Spring, 2004

- Assist in developing instructional audio, or AV materials for the faculties.
- Solve the faculties' problems in using instructional multimedia.

Professional Memberships

Chinese Language Teachers Association

International Systemic Functional Linguistics Association

Research Interests

Teaching Chinese as a foreign language

- Computer assisted language learning
- Proficiency assessments
- Classroom pedagogy

Chinese Language Studies

- Rhetoric
- Applications of Systemic Functional Linguistics

Publications

Zhu, Y., & Hong, W. (2005). Effects of digital voice and stroke sequence animation on character memorization of Chinese as foreign language learners. *Journal of the Chinese Language Teachers Association, 40(3),* 49-70.

Zhu, Y. (2000). Adapting S.F. linguistics theory to Chinese advertising text analysis: The why and the how. In the *Booklet of International Systemic Functional Linguistics Association (ISFLA) for the 27th International Systemic Functional Linguistics Congress on 'Traversing Boundaries: SF Theory in New Context',* 61.

Ji, S.C., & Zhu, Y. (2000). Rhetorical characteristics of puns in commercial advertisements. *Journal of Hebei University, 25,* 124-127.

Conference Presentations

Hong, W., & Zhu, Y. (2004). *Effects of Voice and Stroke Order Animation of Electronic Flashcards on Character Memorization of Chinese Language*

Learners. A paper presented at the 4th International Conference on Chinese Language Pedagogy, Kunming, China.

Zhu, Y. (2000). *Adapting S.F. Linguistics Theory to Chinese Advertising Text Analysis: The Why and the How.* A paper presented at the 27th International Systematic Functional Linguistic Congress, Melbourne, Australia.

Zhu, Y. (1999). *Inter-reflection: The Pre-modern Chinese Antithesis on Tea and Chinese Tea Culture.* A paper presented at the 4th International Chinese Rhetoric Association Congress, Changping, China.