華語文教學研究 14.1:21-48, 2017

Identifying the Difficulty Index of Chinese Characters Based on CFL Beginners' Performance[1]

Chiu-Hung CHEN
Mills College

Ke PENG
Western Kentucky University

Abstract

This exploratory study investigated whether characters should be introduced according to the difficulty of their internal structures or according to a frequency index adopted by most introductory textbooks. The following question was addressed: How does each of the following six properties of a character—type, formation, internal structures, components, number of strokes and frequency—affect CFL learners' performance? The results of this exploratory indicate that stroke numbers and character frequency are the only two significant predictors defining the difficulty level of characters based on CFL beginners' actual performance data.

Keywords: Chinese as a foreign language, Chinese character acquisition

[1] We would like to thank two anonymous reviewers for their valuable comments and suggestions.

1. Introduction

According to the U.S. State Department's Foreign Service Institute, the National Foreign Language Center, and the Center for Applied Linguistics (Jackson and Malone 2010), it takes approximately 88 weeks or 2200 class hours for English speakers to attain the S3 level in Mandarin Chinese (henceforth, Chinese), general professional proficiency, the equivalent of the Superior Level of the American Council on the Teaching of Foreign Languages (ACTFL, henceforth) standards.[2] This contrasts dramatically with other languages which take 575-600 class hours (Spanish, French) to 1100 class hours (Russian, Tagalog) to attain the same proficiency. They attribute the difficulty of learning Chinese mainly to the character-based writing system. Similar conclusions are reached by DeFrancis (1984), Everson (1988), and Walker (1989), among others. Numerous studies have investigated the nature of character learning in both L1 and L2 acquisition (Shu and Anderson 1999; Chen et al. 2002; McBride-Chang et al. 2005; Cui 2008), the effects of character component awareness (Ke 1998; Shen and Ke 2007; Shen 2010; Chen et al. 2013), linguistic/cognitive processing of characters (Shen 2004, 2005; Everson 1988, 1998), the impact of delaying character learning (Packard 1990; Everson 1998; McGinnis 1999; Dew 2005; Allen 2008; Cui 2008; Zhao 2011; Yeh et al. 2012), and the acquisition order of characters (Cui 1997; Yin 2006, 2007; Fan 2010).

This small-scale descriptive study is concerned with the last line of inquiry, namely, whether characters should be introduced according to the difficulty of their internal structures (Cui 1997; Yin 2006, 2007) or according to a frequency index adopted by most introductory textbooks (Fan 2010). With the latter approach, the learner typically encounters a mix of simple and complicated characters drawn from a frequency index which does not allow for a gradual learning process. The more gradual process advocated by Yin (2006, 2007), takes account of characters' structural properties as well as their difficulty. However, these competing views both arise from the perspective of instructors who are native speakers. None of the researchers has established that their ratings of character difficulty are consistent with the actual performance of learners who take Chinese as a foreign language (CFL,

[2] Source: http://fsitraining.state.gov course catalogue: p. 156.

henceforth). Since this is a descriptive study, it must be considered as an exploratory investigation of character difficulty defined by students' performance. The generalizability of this study thus may be limited. The contribution of this study lays the groundwork that will lead more understanding of character learning and of the character types and patterns that learners find most challenging.

2. Literature Review

We first introduce Chinese characters formation and discuss character instruction for college adult beginners. Then we review research pertinent to the present study.

According to *Shuō Wén Jiě Zì* (An Analysis of Explanation of Characters) written by Xu Shen during the Han Dynasty (206 BC -220 AD), there are six types of characters formation: pictographic, ideographic, indicative/semantic compounds, semantic-phonetic compounds, phonetic-loan characters, and explanatory characters. Pictographic characters (*xiàngxíngzi*) making up four percent of the characters in *Shuō Wén Jiě Zì* were the earliest form of characters[3] and were depictions of material objects, such as 日 *(sun)* and 山 *(mountain)*. Ideographic characters (*zhǐshìzi*) (1.3%) were formed to indicate more abstract concepts, such as 上 and 下, created as a pair to express the meanings *up* and *down*. Indicative/semantic compounds (*huìyìzi*) (12.3%) were created by combining two or more characters. Their meanings were based on the meanings of the combined characters. For instance, the character 明 *(bright)* was formed by two characters 日 *(sun)* and 月 *(moon)*. Phonetic-loan characters (*jiǎjièzi*) (1.2%) originally represented one concept but were later borrowed to express another meaning that was unrelated but had similar pronunciation. For instance, 花 *(flower)* was borrowed to express the meaning of *spend*. Explanatory characters (*zhuǎnzhùzi*) (0.07%) was a set of characters that share the same radicals (components) and meanings, such as 老 *(old)* and 考 *(old)*.[4] The last and largest category is semantic-phonetic compounds (*xíngshēngzi*), which made up 81.2% of the characters in *Shuō Wén Jiě Zì*, as well as 81% of the 7,000 most frequently used characters according to Li and Kang (1993) (cited in Ye 2011, p. 12).

[3] The percentage of each type of character in *Shuō Wén Iiě Zì* is from Norman (1988).
[4] 考 has several meanings, e.g. old, take a test, examine, etc.

Semantic-phonetic characters consist of two components, one suggestive of the meaning and the other a clue to the pronunciation. For instance, the character *媽* (mother, pronounced as *mā*) is composed of the semantic component *女* (meaning female), and the phonetic component *馬 (mǎ)*. With the majority of characters in this category, one might be tempted to consider Chinese a phonetic language. However, the phonetic components provide reliable cues for pronunciation in only 26% of cases (Fan et al. 1984, cited in Shen and Ke 2007, p. 98).

Following the ACTFL's Performance Guidelines, many standards-based Chinese language programs expect students to simultaneously demonstrate competence in three modes of communication: interpretive skills (listening and reading), interpersonal skills, and presentation skills (speaking and writing). Reading and writing characters is, therefore, an integral part of CFL curriculum.

Recent studies indicate that most Chinese programs at the college level in both the United States and China teach characters at or near the beginning of the first semester (Yin 2006; Zhao 2011; Ye 2011). Once characters are introduced, they are integrated into the four language skills (listening, speaking, reading and writing) with little to no consideration of how students actually learn them. Instead, characters are taught based on students' knowledge of spoken Chinese. For instance, when greetings such as *nǐ hǎo (hello)* and *nín guì xìng (what is your honorable last name)* are introduced, learners are usually expected to learn the associated characters in the same lesson. As a result, the topics covered in a particular course dictate which characters are introduced and their order. This is less problematic with phonetic languages, such as English and Spanish, in which a word's spelling has information about, if not a one-to-one correspondence with its pronunciation. However, the Chinese writing system is logographic with every character, representing a single syllable and word. For instance, the character for *fast* is written and pronounced *快 kuài.*[5] A syllable may be represented by multiple characters each with its own meaning (homophones). Homophones for *kuài*, for example, are *塊 (piece/dollar), 筷 (chopstick).*

[5] The symbol on top of the letter *a (à)* represents the fourth tone.

Previous studies in both SLA and CFL literature have shown that frequency plays an important role in word retention. In the SLA literature, researchers show that the acquisition order of grammatical structures is closely related to input frequency (Wagner-Gough and Hatch 1975; Hatch and Wagner-Gough 1976; Larson-Freeman 1976; N. Ellis 1996, 2002; Goldschneider and DeKeyser 2001). Similar results are also found in vocabulary acquisition, as stated by Ellis and Shintani (2014). The more frequently the learner sees a character, the easier it gets. In the CFL literature, Shen (2005) also found a significant difference in character retention between high frequency words and low frequency ones when investigating the relationship between linguistic complexity and reading comprehension.

To understand the selection of Chinese characters in CFL textbooks for beginners, Fan (2011) analyzed ten textbooks published between 1961 and 2008. The study is informed by the theoretical assumption in which frequency of characters and their structural components, as well as frequency types of structural components play an essential role for CFL learners to acquire the underlying structure of Chinese characters. Fan's results indicate that all 10 textbooks contain a high percentage of high frequency characters based on a comparison with the frequency lists from three frequency dictionaries, Yahoo and Google search counts, and the HSK proficiency level list.[6] With regard to structural components, Fan finds that less than 17% of the characters have semantic or phonetic components that provide reliable cues to their meaning or pronunciation. While it is common to include high frequency vocabulary in language textbooks, Fan's findings imply that most CFL learners are learning characters that do not follow any particular progression, which might impede learning (Yin 2006, 2007). If only semantic-phonetic characters with reliable cues are taught, it would be impossible for true beginners to establish immediate connections between semantic and phonetic components. In order to form the mental associations that help them achieve some degree of automaticity, CFL learners need to know a large number of characters. However, research has not yet established what that number is.

[6] HSK (Hànyǔ Shuǐpíng Kǎoshi) is a standardized test of Chinese language proficiency for non- native speakers.

To help beginners who are required to learn characters together with other skills, researchers have provided empirical evidence showing the positive effects of developing orthographic awareness. Knowledge or radicals (Shen 2000) and the skills to apply that knowledge (Shen and Ke 2007) are positively correlated with learner performances in both recognition and production tasks. Nevertheless, given the difficulty of reading and writing Chinese characters and the higher course drop rate attributed to that difficulty, researchers question whether teaching Chinese characters along with their oral counterparts in the most efficient way to learn Chinese (Yin 2006, 2007; Cui 2008; Ye 2011; Ross et al. 2012). Cui (2008) notes that, when Chinese children begin to learn characters in school, they have already developed oral fluency. Moreover, at the beginning, the emphasis is on recognizing characters gradually, rather than writing them. Characters are acquired gradually based on their relevance. The characters introduced earliest have simple forms, concrete meanings, and are relevant to children's daily lives.

Yin (2006, 2007) also argues that characters should be taught based on their complexity, as well as structural properties. He suggests that the order of introduction be based on four factors. First, characters with few strokes should be taught before those with many. Second, pictographic characters, based on depictions of natural objects, should precede non-pictographic characters. Third, single-component characters should be taught before characters with multiple components (e.g. semantic-phonetic characters). Finally, semantic compounds should be taught before semantic-phonetic compounds because beginners can more easily connect the character's meaning with semantic components than with phonetic components. To teach characters in such a fashion, Yin suggests a separate writing class should be provided during the first semester together with a speaking class that focuses only on pronunciation and oral practice.

Like the frequency index approach, the view that character order should be governed by difficulty level arises from the perspective of instructors and native speakers in mind. However, there is no support that their assessments of character difficulty correspond well with CFL learners' actual performance. Discrepancies between learners' and teachers' view of character difficulty might lead to less effective teaching. The present study addresses the following question: How does

each of the following six properties of a character—type, formation, internal structures, components, number of strokes and frequency—affect CFL learners' acquisition of characters?

The premise behind this research questions is that analyzing of students' actual performances is the best way to measure character difficulty from the learners' perspective. Correlating students' recognition and production of a character with its various attributes will unveil the nature of character learning at the beginner level.

3. Method

A total of fifty-four first-year students of Chinese as a foreign language (CFL) from two American universities participated in the study over the course of one year (N=54). By the time the data were collected, it was towards the end of the second semester, the participants were given the following recognition tests as review practice in three consecutive classes[7]. A total of seventy-six characters were tested (n=76).

The research study is three-folded: firstly, how to evaluate CFL learners' performance by using item analysis, namely, by analyzing the difficulty level and discrimination index of each character (n=76); secondly, how character properties, including character types, internal structures, components, number of strokes and frequency, affect the difficulty level of each character demonstrated by CFL learners' performance; and, thirdly, to examine whether or not there would be any interaction between character difficulty and task types (e.g. contextual vs. non-contextual tasks).

One test provides context for the target characters and the other does not. The participants were required to provide English translation for the characters as in the examples below:

(1) Recognition test with context:
zhè 本[8] shū shì zhōng wén shū
qǐng 坐， nǐ xiǎngdiǎn shén me

[7] Please see sample test items in the Appendix.
[8] Simplified characters were adopted in the tests.

wǒ bù 想 qù gōng zuò

xià kè yǐ hòu , 回 jiā ba

(2) Recognition test without context:

本: _____ (English translation)

Table 1 below illustrates how the characters in the above sentences are analyzed based on the property variables.

Table 1: Character Property Variables

	本	坐	想	回
Type	dútǐzì 'Independently-formed Character'	hétǐzì 'Compound Character'	hétǐzì 'Compound Character'	hétǐzì 'Compound Character'
Formation	zhǐshìzì Self-explanatory Character	huìyìzì Associative Compound Character	xíngshēngzì Semantic-phonetic Compound	xiàngxíngzì Pictographic Character
Internal Structure	n/a	top-bottom	top-bottom	outside-inside
Phonological or Semantic Component[9]	n/a	indirect	clear indication	n/a
Stroke Number	5	7	13	6
Frequency	2 (times)	3	5	3

Each test consisted of 76 target characters (n=76). In the recognition test with context, some sentences contained one target character, and some more than one. In the recognition test without context, the target characters were listed randomly. These characters were introduced to the students in both groups in the first year. The scoring

[9] Components include semantic cues for huìyìzì and phonological hints for xíngshēngzì. Semantic cues include direct meaning association, indirect meaning association, and indiscernible meaning association. Phonological hints include clear phonological cues, partial phonological cues, and no phonological cues. Please see sections 4.6 and 4.7 for more details.

of the tests is that each character is worth one point for correct translations. The maximum number of points is 76 for each test. Each right answer was coded "1", wrong or partially wrong answers were coded "0". For the calculation of item statistics, missing values were treated as wrong answers.

4. Results

4.1 Reliability Analyses

Before comparisons were made between contextualized and non-contextualized tasks, we conducted reliability analyses to validate the consistency of each task. Cronbach's α was used to inspect the internal consistency of the tasks. When alpha equals 0, the true score is not measured and there is only an error component. When alpha equals 1.0, all items measure only the true score, and there is no error component. In general, a lenient cut-off of 0.60 is common in exploratory studies, a Cronbach's α of 0.7 is an acceptable level of agreement, and by convention, a value of 0.80 is a good scale for many researchers. The Cronbach's α of the contextualized task was .982, and the non-contextualized task yielded an α value of .906. This is an ideal situation, as both tasks analyzed in this study were proven to be reliable and consistent.

4.2 Item Analysis

Having confidence in the reliability of the tasks, statistical item analysis was performed to investigate the actual difficulty in character learning from the learners' perspective. In this study, the character recognition task was treated as a test, and each character was treated as an individual item. The purpose of doing so is to identify the difficulty level of each item or character, and to see how they affected students' overall performance. Character items were assessed according to two indexes: the difficulty level (*p*, i.e. prevalence of correct responses) and the discrimination index (D-index, *r*, ability to differentiate learners on their ability to recognize characters), which is the corrected item-total correlation in SPSS.

The item difficulty index (*p*) refers to the ratio of the total number of correct responses to the total number of responses, expressed as a percentage. It is calculated by the formula P=R/T, where R is the number of correct responses and T is the total

number of responses (i.e. correct + incorrect + blank responses). The higher the index value, the lower the difficulty of the character, and vice versa. Both tasks involved in this study were administered to students at the end of the semester as cumulative assessments, therefore, items that have a *p* value of 0.75 or higher were easy, items with a *p* value between 0.60 and 0.75 were medium, items with a *p* value of 0.40 ~ 0.60 were difficult, and items with a *p* value below 0.40 were very difficult. Figure 1 illustrates the distribution of difficulty levels in both contextualized and non-contextualized recognition tasks. Table 2 categorizes the tested characters into four difficulty levels based on students' performance. Why do CFL beginners find some characters more difficult than others? Is there any way to explain this phenomenon? We will approach this challenge in the next sections by examining the properties of the characters, including character types, internal structures, components, stroke numbers and frequency.

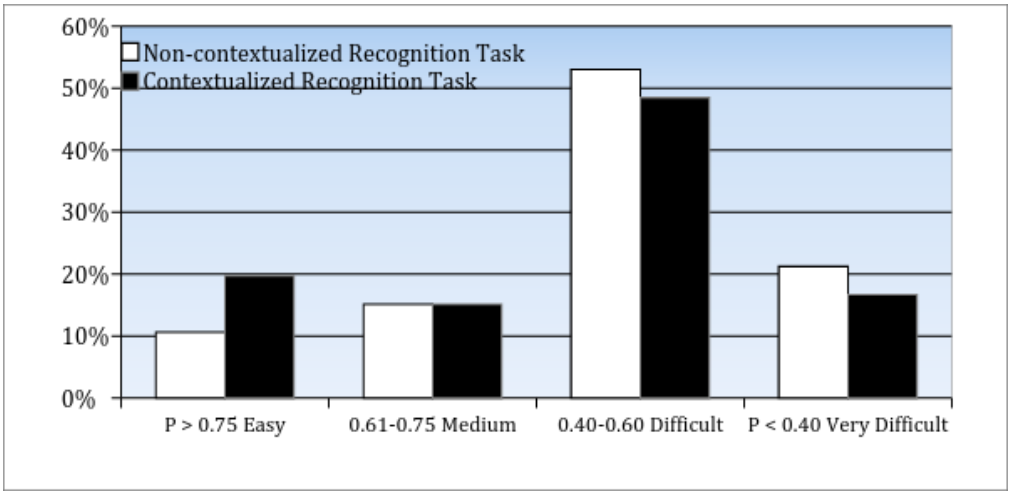

Figure 1: Distribution of Difficulty Levels Based on Students' Performance

Table 2: Characters Categorized by Difficulty Levels Based on Students' Performance

Easy Characters	Medium Difficulty Characters	Difficult Characters	Very Difficult Characters
書點和電媽國文美學朋友圖館爸吃說中	去以家本問飯没師有生	晚打話做來想坐起半還要難看車少男今年下對回多會分叫日可天您大老小謝號	再道知上誰給現在請就都怎吧太先

In addition to difficulty level, another critical notion for classical item analysis is the item discrimination power or D-index (*r*). In classical test theory, *r* measures how well each item tests the difference between the percentage of students in the upper group who gave the correct response, and the percentage of those in the lower group who gave the correct response. The higher the discrimination index, the better the item can differentiate between high and low test scorers (Mery, Newby, and Peng, 2011). The objective of the current study is not to determine which characters have higher discrimination power to distinguish high-achieving CFL beginners from low ones, but to explicate which characters are more difficult than others from the learners' perspective. That being said, however, some characters tend to serve as a more effective measurement when time becomes a constraint. In other words, if CFL educators do not have enough time to test students' knowledge of all of the seventy-six characters, which ones should be given priority? Table 3 provides a recommendation to such situations. The characters that should be avoided may be so difficult that neither the lower group nor the upper group made the correct responses, but it may also be the case that they are so easy that both groups tend to give the correct responses.

Table 3: Characters Categorized By Discrimination Power Based on Students' Performance

Characters to Avoid for Testing Low Discrimination Power	**Possible Characters for Testing Medium Discrimination Power**	**Characters Recommended for Testing High Discrimination Power**
中說爸吃坐做號上吧還太話再就給大要誰日您在文	書點和電媽國美學朋友圖館以本飯没師有生來起少今對多分老現	怎謝天家小車下可請打想看年晚會都道叫問先知去回半男難

4.3 Types of Characters (Dútǐzì Versus Hétǐzì)

In the item analyses above, it is noted that some characters are more difficult than others based on learners' performance data. What gives rise to the different difficulty indexes? Using the difficulty indexes of each character in the two tasks, from this section on we will examine whether the following property variables, namely, character types, internal structures, character components, stroke numbers and frequency, can explain the variances in difficulty.

Data were analyzed using a mixed-design ANOVA with a within-subjects factor of task types (contextualized and non-contextualized recognition) and a between-subject factor of character type (*dútǔzì* 'independently-formed character', *hétǐzì* 'compound character')[10]. The descriptive statistics demonstrated that students performed slightly better with *dútǔzì* (Contextualized Recognition Task: M= 0.57, SD= 0.038) than with *hétǐzì* (M=0.535, SD=0.023) regardless of the task type. There was a significant main effect of task type, $F(1, 62) = 7.006$, $p = .01$. This effect indicates that the difficulty levels of characters vary between the contextualized and the non-contextualized recognition tasks. The interaction between task types and

[10] Independently-formed characters refer to characters that cannot be further divided into any meaningful unit (cf. morpheme). Pictographic and ideographic characters are considered independently-formed characters. In contrast, compound characters are composed of two or more meaningful units as commonly seen in semantic-phonetic compounds, in which one component provides the hint as to the meaning and the other provides the hint as to the pronunciation.

character types was not significant, F (1, 62) = 1.933, p = .169, nor was the main effect of character types (*dútǔzì i* and *hétǐzì*), F (1, 62) =. 482, p = .490. In other words, there were no significant differences in the difficulty indexes between *dútǔzì* 'independent characters' and *hétǐzì* 'compound characters'.

4.4 Formation of Characters (Xiàngxíngzì, Zhǐshìzì, Huìyìzì, and Xíngshēngzì)

A mixed-design ANOVA was conducted with task types (contextualized recognition, non-contextualized recognition) as the within-subjects factor, and formation of characters (*xiàngxíngzì* 'photographic characters', *zhǐshìzì* 'indicative or self-explanatory characters', *huìyìzì* 'associative compound characters', and *xíngshēngzì* 'semantic-phonetic compound characters') as the between-subjects factor. The descriptive statistics showed that regardless of task type, students performed best with *huìyìzì* (Contextualized Recognition Task: M= 0.59, SD= 0.13) in comparison with the other three types of characters (*xiàngxíngzì*: M= 0.57, SD=0.20*; xíngshēngzì:* M=0.55, SD=0.19; *zhǐshìzì*: M=0.49, SD= 0.14). The data revealed no significant main effects of task types, F(1, 62) =2.293, p = .135, or main effects of character types, F(3, 62)= .654, p= .583. The interaction between task types and character types was not significant either, F (3, 62) = .966, p = .415. The formation of characters, specifically, *xiàngxíngzì* 'photographic characters', *zhǐshìzì* 'indicative or self-explanatory characters', *huìyìzì* 'associative compound characters' and *xíngshēngzì* 'semantic-phonetic compound characters', was not helpful in capturing the differences in difficulty levels of the tested characters.

4.5 Internal Structure of Characters

Since the attempt to utilize character types or formations to interpret the variances in the difficulty indexes of characters, the focus now shifts to the investigation of whether the internal structure of characters, i.e. top-bottom structure, left-right structure, and outside-inside structure, can help us to understand the difficulty indexes of Chinese characters for CLF learners[11]. As in the previous section, data were analyzed using a mixed-design ANOVA with a within-subjects factor of

[11] Examples of top-bottom structure: 早, left-right structure: 好, and outside-inside structure: 回.

task types (contextualized recognition, non-contextualized recognition) and a between-subject factor of a character's internal structures (top-bottom, left-right, and outside-inside). The descriptive statistics demonstrated that students found the outside-inside structure (Contextualized recognition task: M= 0.65, SD =0.13) easier than the top-bottom structure (M= 0.59, SD = 0.14) and the left-right structure (M = 0.54, SD = 0.17). There was a significant main effect of task type, and the difficulty level was lowered with the use of contextualized tasks, F(1, 45) =12.859, p = .001. The interaction between task types and the internal structures of characters failed to reach statistical significance, F (2, 45) =1.090, p = .169. Nor was there any statistical significance in the main effect of character internal structures, F(2,45) = .638, p = .533.

4.6 Semantic Clues

The descriptive statistics demonstrated that with either task type, direct meaning association (Contextualized Recognition Task: M = 0.67, SD = 0.21) provided more cues than indirect meaning association (M = 0.61, SD = 0.14) in helping students perform better. A two-way ANOVA with task types (contextualized recognition, non-contextualized recognition) as the within-subjects factor and semantic clues (direct meaning association, indirect meaning association, and indiscernible meaning association)[12] as the between-subjects factors yielded a significant main effect of task type, F(1, 46) = 21.131, p < .001, and a significant interaction between task types and semantic clues, F(2,46) = 4.052, p = .024. The remaining main effect of semantic clues was not statistically significant.

In order to examine the simple effects of semantic clues on the difficulty indexes of characters, two one-way, between-subjects ANOVA were conducted in contextualized and non-contextualized recognition tasks, respectively. When the characters were tested in contexts, there was a significant effect of semantic clues on the character difficulty indexes in direct meaning association, indirect meaning association, and indiscernible meaning association, F(2, 46) = 5.133, p = .010. Post

[12] Examples of direct meaning association: 看 'to see' whose radical 目 means 'eye'; indirect meaning association: 吃 'to eat' whose radical 口 means 'mouth'; indiscernible meaning association: 晚 'late' whose radical 日 'sun' does not provide any clue as to the meaning of the character.

hoc comparisons using the Tukey HSD test indicated that the mean difficulty index for the direct meaning association condition (M = .667, SD = .21) was significantly lower than the indiscernible meaning association condition (M = . 486, SD = .14), *p* = .044. Similarly, the indirect meaning association condition (M = .606, SD = .14) was also significantly different from the indiscernible meaning association condition, *p* = .025. However, in non-contextualized recognition tasks, there was no statistically significant effect of semantic clues on the character difficulty indexes in any the three conditions. The results validated the contribution of semantic clues to character recognition only in contextualized tasks, not in simple recognition tasks where there was no context.

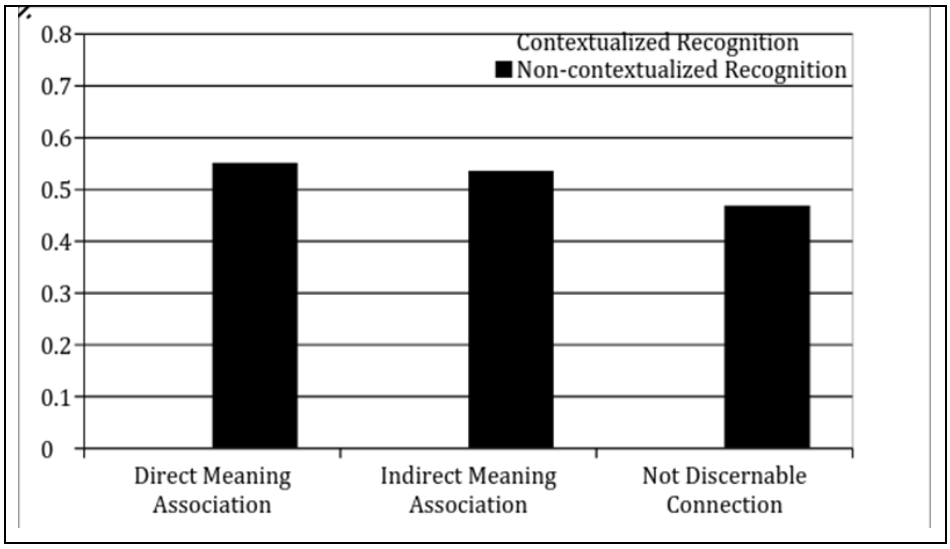

Figure 2: The Effect of Semantic Cues on Character Recognition

4.7 Phonological Hints

The descriptive statistics demonstrated that clear phonological cues were more helpful (Contextualized Recognition Task: M = 0.56, SD = 0.24) than no phonological cues (M = 0.54, SD = 0.19) and partial phonological cues (M =0.54, SD =0.18) in helping students perform better regardless of the task type. A two-way ANOVA with task types (contextualized recognition, non-contextualized recognition) as the within-subjects factor and phonological hints (clear phonological cues, partial

phonological cues, and no phonological cues)[13] as the between-subjects factors only generated a significant main effect of task type, $F(1, 18) = 10.006$, $p = .005$. Neither the main effect of phonological hints, $F(2, 18) = .051$, $p = .950$, nor the interaction between task types and phonological hints, $F(2, 18) = .145$, $p = .866$, were statistically significant.

4.8 Number of Stroke

The complexity of the tested characters ranges from 4 strokes to 13 strokes. All the characters were divided into three groups to ensure similar number of characters in each group: Group 1 consisting of characters with 5 or fewer strokes, Group 2 of characters with 6-8 strokes, and Group 3 with 9-13 strokes. The descriptive statistics demonstrated that Group 1 characters were much easier (Contextualized Recognition Task: M = 0.60, SD = 0.14) than characters with more strokes in Group 2 (M = 0.59, SD = 0.18) and Group 3 (M =0.50, SD =0.17) regardless of the task type. The main effect of stroke number groups, however, was not statistically significant: $F (2, 60) = 2.326$, $p = .107$. Simple linear regression analysis was used to test if the number of strokes significantly predicted the difficulty indexes of Chinese characters for CFL beginners in the contextualized recognition task. The result of the regression was statistically significant. The number of strokes can only predict 7.6% of the variance ($R^2 = .076$, $F(1, 60) = 4.85$, $p < .05$). The findings suggest that stroke numbers of a character significantly predicted the difficulty index of the character ($\beta = -.019$, $p < .05$). The larger the stroke number is, the more difficult the character becomes for CFL beginners. Although the number of strokes is a significant predictor, the effect size is considerably small.

4.9 Frequency

Frequency in this study refers to the number of times that the target character appears in the textbook. Since the two student samples came from two institutions

[13] The examples of clear phonological cues: the pronunciation of 媽 ‘mā’ whose pronunciation is same as the sound radical 馬 ‘mǎ’ except tones (first tone vs. third tone); partial phonological cues: 現 ‘xiàn’ which shared the same vowels as its sound radical 見 ‘jiàn’; no phonological cues: 都 ‘dōu’ in which the pronunciation of both components are not associated with ‘dōu’.

employing different textbooks, the frequency for each character varies between the two groups. An average of the two character frequencies was calculated. The results of the simple linear regression suggest that a significant proportion of the total variation in the character difficulty indexes was predicted by the frequency of the character. In other words, the number of occurrences in the textbook is a reliable predictor of the difficulty index of the character, $F(1, 65) = 4.09$, $p < .05$, explaining 6% of the variance. Additionally, we find the frequency of the character in the textbook statistically significantly predicted its difficulty index ($\beta = .005$, $p < .05$). In other words, the more frequently the character appears in the textbook, the less difficult it gets. But yet again, the effect size is considerably small, which marks the finding less meaningful.

5. Discussion

The research question the present study addresses is how character properties, including character types, internal structures, components, number of strokes and frequency, affect CFL learner's performance. The study yields several findings.

First, two difficulty indexes are helpful to character teaching based on the item analysis: the difficulty level and the discrimination index. The means of students' performance in character assessments differentiate easy characters from and difficulty ones. What students find easy or difficult characters may be different from the teacher's perception. For instance, the character 謝 which is generally considered difficult because of its structure (three components with a larger stroke number). However, according to students' actual performance, it is less difficult than the character 再 which is a single-component character with fewer strokes. With regard to the discrimination index, it indicates that some characters have more discriminatory power than others. For instance, the character 中 is not an ideal test item because both highly proficient and less proficient students are able to recognize it. In contrast, the character 怎 which has high discrimination power serves as a better option for character assessment. In other words, when it comes to evaluate students' mastery of characters, teachers cannot afford and do not need to test all the characters. Rather, they can be selective and effective when time and resources are limited.

Second, the statistical results indicate that number of strokes and frequency are the only two significant predictors defining the difficulty level of characters. Though the results cannot be necessarily meaningful in implication due to the small sample size and effect size, frequency does play an important role in word retention in both SLA and CFL literature as discussed in the literature review section. Recall that Shen's (2005) study, high frequency words have better retention rate compared with low frequency ones. However, unlike Shen's study in which word frequency was defined according to *Modern Chinese Frequency Dictionary* (1986), the present study used occurrences of characters in the textbook to determine frequency rates. We suggest that, based on the results of the study, high frequency characters should not be the only consideration for vocabulary selection. The number of occurrences or repetition of characters that affect the level of character difficulty should also be taken into consideration.

With the exception of the number of strokes, the other factors which CFL educators generally view as important or relevant to character teaching and learning, such properties of characters as types, internal structures, semantic and phonetic components, fail to consistently predict the students' actual performance of a character. One explanation is that novice CFL learners have not yet developed sensitivity towards semantic cues in relation to the character meaning, and sensitivity to phonetic cues in relation to the character pronunciation, among other factors. This is not surprising as only 17% of the characters have reliable semantic and phonetic cues (Fan 2011). There may be a threshold for the number of characters that the CFL beginner must acquire before s/he can establish such sensitivity to either semantic or phonetic cues.

Another reason for the unexpected results is because educators typically assume that CFL learners acquire characters the same way as native speakers do. In fact, empirical evidence from CFL character acquisition has been presented to challenge the way that characters are taught. For instance, researchers have raised concerns about the prevailing approach in which characters are taught together with other skills from the beginning (Yeh et al. 2012; Chen and Peng 2013). One major objection to this approach is that the CFL beginner does not have oral fluency when they begin to learn characters. For CFL beginners, writing characters is a completely new task, in

addition to the development of interpersonal and presentational skills. This is in contrast with native speakers who already have a good, if not perfect, command of the oral language at the time of formal character learning. The fundamental learning task for native speakers is to map the oral form with the written script when they first begin to learn characters. However, CFL beginners cannot establish immediate connections between semantic components and phonetic components as easily as their counterparts (Shu and Anderson 1999). In his survey study on character learning strategies, McGinnis (1999) found that CFL beginners used rote repetition more than other strategies, and they did not apply knowledge of radicals to character learning as expected. It is important to note that we do not intend to deny the benefits that learning character properties could bring to the CFL learner, as argued by Yin (2006, 2007), Cui (2008), among others. Nevertheless, previous studies such as Shen and Ke (2007) have cautioned that even though character properties such as radical knowledge are important in character learning, there are limitations in using semantic radicals to infer the meaning of a character, particularly when the character is morphologically opaque (also see Shen 2000). Therefore, we suggest that, based on the results of the current study, frequency should also be stressed equally, if not more, as the other factors.

The last finding reveals that the difficulty of a character varies according to the type of task used to elicit the CFL learners' performance in recognizing characters. Main effects of task type have been found significant among most of the factors discussed in Section 4. Contextualized tasks yield higher recognition performance and are more rigorous in their ability to capture the actual recognition and production of the participant. This result implies that despite the fact that all the participants in this study are CFL beginners, they are already adept at using context to help them recall or produce targeted characters. Researchers thus should be aware of the differences caused by assessment tasks. The outcome of each individual assessment task provides a limited representation of word recognition or reading comprehension (Brantmeier, 2012). Character knowledge ultimately contributes to the reading comprehension of Chinese.

6. Conclusion

We conclude that the difficulty of a characters is, and should be, defined by learners to better understand the second language acquisition process of character learning. The results of this study indicates that instructors' personal definition of character difficulty is an unreliable guide to students' identification of it in actual performance. The discrepancies between their views are where breakthrough progress may occur in character teaching and learning. Pedagogically, when knowledge of character is evaluated in formative and summative assessments, we suggest that CFL educators test characters with high discrimination power for effective use of time and resources. We also propose that in conducting empirical research, measurements should be contextualized tasks not only for collecting data, but also to create a learning experience for participants. Last but not the least, number of strokes and frequency, the only two factors which impact the CFL beginner's performance in character recognition, should be taken into consideration together with other factors when planning curriculum for first-year CFL learners.

We would also like to acknowledge the limitations of this study. Due to the small number of characters covered in the first-year CFL curriculum, the small sample size in combination with the ambitious number of factors under investigation, many of the results are statistically insignificant. The significant but small effect size also plays restrictions in meaningful interpretation and implication. It is our intention to lay the groundwork for follow-up research in the future that identifies the timing or the threshold for CFL beginners (e.g. the minimum number of characters required to know) to become sensitive to the phonetic and semantic cues which contributes to character recognition with high automaticity, and ultimately, Chinese literacy.

References

Allen, Joseph. R. 2008. Why learning to write Chinese is a waste of time: A modest proposal. *Foreign Language Annals* 41.2: 237-251.

Brantmeier, Cindy. 2012. *Diagnosing Strengths and Weaknesses of Foreign/Second Language Readers*. Presentation delivered at LARC: Pennsylvania State University and San Diego State University.

Chen, Chiu-Hung, and Ke Peng. 2013. The Impact of Instructional Approaches and Task Types on Learner's Character Retention in Chinese as Foreign Language. *Proceedings of International Conference of Association of Teaching Chinese as Second Language (ATCSL 2013)*.

Chen, Hsueh-Chih, Chih-Chun Hsu, Li-Yun Chang, Yu-Chi Lin, Kuo-En Chang, and Yao-Ting Sung. 2013. Using a Radical-derived Character E-learning Platform to Increase Learner Knowledge of Chinese Characters. *Language Learning and Technology* 17.1: 89-106.

Chen, Xuefeng, and Henry S. R. Kao. 2002. Visual-spatial properties and orthographic processing of Chinese characters. *Cognitive Neuroscience Studies of the Chinese Language*, ed. by Henry S. R. Kao, Che-Kan Leong, and Ding-Guo Gao, 175-194. Hong Kong: Hong Kong University Press.

Cui, Yonghua. 1997. Hànzìbùjiàn hé duìwài hànzì jiàoxué. *Yǔyán wénzì yìngyòng* 3:49-62 (In Chinese).

Cui, Yonghua. 2008. Cóng mǔyǔ értóng rènzì kàn duì wài hànzì jiāo xué. *Yǔ yán jiāo xué yǔ yán jiù* 2:17-23 (in Chinese).

DeFrancis, John. 1984. *The Chinese Language: Fact and Fantasy*. Honolulu, Hawaii: University of Hawaii Press.

Dew, James E. 2005. Language is primary, script is secondary: The importance of gaining a strong foundation in the language before devoting major efforts to character recognition. Paper presented at *the International and Interdisciplinary Conference,* Germany: University of Mainz in Germersheim. Accessed online, July 30, 2013 http://www.fb06.uni-mainz.de/chinesisch/Dateien/hanzirenzhi_papers_dew.htm

Ellis, Nick. 2002. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in*

Second Language Acquisition 24:143-188.

Ellis, Rod, and Natsuko Shintani. 2014. *Exploring Language Pedagogy through Second Language Acquisition Research.* New York: Routledge, Taylor and Francis Group.

Everson, Michael E. 1988. Speed and comprehension in reading Chinese: Romanization vs. characters revisited. *Journal of Chinese Language Teachers Association* 23.2:1-15.

Everson, Michael E. 1998. Word recognition among learners of Chinese as a foreign language: Investigating the relationship between naming and knowing. *The Modern Language Journal* 82:194-204.

Fan, Huimei. 2010. *Developing Orthographic Awareness Among Beginning Chinese Language Learners: Investigating the Influence of Beginning Level Textbooks*. Ph.D. dissertation. University of Iowa.

Goldschneider, Jennifer M., and Robert M. DeKeyser. 2001. Explaining the "Natural Order of L2 Morpheme Acquisition" in English: A Meta-Analysis of Multiple Determinants. *Language Learning* 51: 1-50.

Hatch, Evelyn, and Judy Wagner-Gough. 1976. Explaining sequence and variation in second language acquisition. *Papers in Second Language Acquisition (Special Issue, 4)*, ed. by Douglas Brown, 39-57.

Ke, Chuanren. 1998. Effects of strategies on the learning of Chinese characters among foreign language students. *Journal of the Chinese Language Teachers Association* 33.2: 93-112.

Jackson, Frederick H., and Margaret Malone. 2010. *Building the Foreign Language Capacity We Need: Toward A Comprehensive Strategy For A National Language Framework.* Accessed online, August 10, 2013. http://www.cal.org/resource-center/publications/building-foreign-language-capacity

Jiang, Xin. 2009. *Duìwài hànyǔ zìcí yǔ yuè dú xuéxí yán jiù*. (in Chinese). Beijing Language and Culture University Press: Beijing, China.

Larson-Freeman, Diane. 1976. An Explanation for the Morpheme Acquisition Order of Second Language Learners. *Language Learning* 26: 125-134.

McBride-Chang, Catherine, Bonny W.-Y. Chow, Yiping Zhong, Stephen Burgess, and William G. Hayward. 2005. Chinese character acquisition and visual skills in two Chinese scripts. *Reading and Writing* 18: 99-128.

McGinnis, Scott. 1999. Student's goal and approaches. *Mapping the Course of the Chinese Language Field*, ed. by M. Chu, 151-188. Kalamazoo, Michigan: Chinese Language Teachers Association, Inc.

Mery, Yvonne, Jill Newby, and Ke Peng. 2011. Assessing the Reliability and Validity of Locally Developed Information Literacy Test Items. *Reference Services Review* 39.2.

Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.

Packard, Jerome. 1990. Effects of time lag in the introduction of characters into the Chinese language curriculum. *The Modern Language Journal* 74: 167-175.

Shen, Helen. 2000. Radical knowledge and character learning among learners of Chinese as a foreign language, *Yuyan Yanjiu [Linguistic Studies]* June, 85-93.

Shen, Helen. 2004. Level of cognitive processing: effects on character learning among non-native learners of Chinese as a foreign language, *Language and Education* 18: 167-182.

Shen, Helen. 2005. Linguistic complexity and beginning-level L2 Chinese reading. *Journal of the Chinese Language Teachers Association* 43.3: 1-28.

Shen, Helen. 2010. Analysis of Radical Knowledge Development Among Beginning CFL Learners. *Learners of Chinese as a Foreign Language*, ed. by Michael Everson and Helen Shen, 45-65. Honolulu: NFLRC.

Shen, Helen, and Chuanren Ke. 2007. Radical awareness and word acquisition among non-native learners of Chinese, *The Modern Language Journal* 91: 97-111.

Shen, Helen, and Xin Jiang. 2013. Character reading fluency, word segmentation accuracy, and reading comprehension in L2 Chinese. *Reading in a Foreign Language* 25.1: 1-25.

Shu, Hua, and Richard C. Anderson. 1999. Learning to Read Chinese: The Development of Metalinguistic Awareness. *Reading Chinese Script: A Cognitive Analysis*, ed. by J. Wang, H.-C. Chen, R. Radach, and A. Inhoff, 1-18. New Jersey: Lawrence Erlbaum Associates.

Wagner-Gough, Judy, and Evelyn Hatch. 1975. The importance of input in second

language acquisition studies. *Language Learning* 25: 297–308.

Walker, Galal. 1989. Intensive Chinese curriculum: the EASLI model. *Journal of the Chinese Language Teachers Association* 24.2: 43-84.

Ye, Lijuan. 2011. *Teaching and Learning Chinese as A Foreign Language in the United States: To Delay or Not to Delay the Character Introduction.* Ph.D. Dissertation. Georgia State University.

Yeh, Meng, and Claudia Ross. 2012. *Character Retention Study*. Paper presented at the ACTFL Annual Conference, Philadelphia, PA.

Yin, Jinghua. 2006. Proceeding Separately and Striking Together. *Chinese Teaching in the World* 26.1: 116-121.

Yin, Jinghua. 2007. Curriculum design for an entry course on Chinese characters at an American University. *Journal of Chinese Language Teachers Association* 42: 67-75.

Zhao, Jinming. 2011. Chūjí hànyǔ jiāoxué de yǒuxiào tújìng. *Chinese Teaching in the World* 25.3:376-387.

[審查：2016.10.17　修改：2016.10.28　接受：2017.1.11]

陳秋宏
Chiu-Hung CHEN
94613 美國加州奧克蘭麥克阿瑟大道 5000 號　密爾斯大學
Department of Languages and Literatures, Mills College
No.5000, MacArthur Blvd., Oakland, CA, USA, 94613
cchen@mills.edu

彭珂
Ke PENG
42104 美國肯塔基州鮑林格林#31027 College Heights 大道 1906 號
西肯塔基大學
Department of Modern Languages
Western Kentucky University
No.1906 College Heights Blvd., #31027Bowling Green, KY, USA 42104
Ke.peng@wku.edu

Appendix: Sample Test Items

Please Provide the English Translation for the Following Characters.

		Pinyin	English
1	zhè 本 shū shì zhōn g wén shū。		
2	wǒ de fán g jiān hěn 大 。		
3	wǒ 爸爸 shì lǎo shī 。		
4	wǒ hé 朋友 dōu xué zhōn g wén 。		
5	wǒ bù chán g kàn 电 shì 。		
6	wǒ huì zhōn g wén 和 yīn g wén 。		
7	shí diǎn shàn g kè ， shí yì diǎn 下 kè。		
8	qǐn g 坐， nǐ xiǎn g diǎn shén me ？		
9	wǒ 要 hē yì bēi kě lè 。		
10	tā yě shì 美 g uó rén 。		
11	wǒ zǎo sh a n g liù diǎn 就 qǐ chuán g		
12	您 g uì x ìn g ?		
13	wǒ bú qù 图书馆 kàn shū 。		
14	wǒ bù 想 qù g ōn g zuò 。		
15	wǒ xìn g wán g ，叫 wán g xiǎo mín g 。		
16	jīn tiān de g ōn g kè hěn 难 。		
17	wǒ 先 hē chá ， rán hòu chī fàn 。		
18	wǒ 妈妈 bù g ōn g zuò 。		
19	wǒ yǒu liǎn g běn wén xué 书。		
20	今年 shì èr lín g yī sān nián 。		
21	qǐn g 问 ，nǐ shì lǎo shī m a ？		
22	wǒ yǒu yí liàn g dé g uó 车 。		
23	wǒ xiǎn g qù zhōn g 国。		

24	wǒ de shǒu jī 号 shì 2701234567		
25	xià kè yǐ hòu ,回 jiā b a 。		
26	wǒ de shēn g rì shì 1 yuè 1 日 。		
27	现在 shì shí diǎn bàn 。		
28	给 wǒ xiě diàn zǐ yóu jiàn b a 。		
29	wǒ zài g ěi mā m a 打 diàn huà 。		
30	wǒ men yì qǐ qù dǎ qiú 吧 。		
31	“wéi ？wǒ zài tú shū g uǎn ， nǐ yě 来 ba ！”		
32	wǒ bù 知道 tā shì shuí 。		