

When Should Characters Be Introduced to Novice-Level Chinese in a Blended Learning Setting?

初级汉语混合式学习环境中的汉字教学时机

Sihui (Echo) Ke¹

Stacy Dubravac

University of Kentucky

University of Kentucky

[ORCID: #0000-0003-2665-0637](https://orcid.org/0000-0003-2665-0637)

[ORCID: #0000-0001-8790-4579](https://orcid.org/0000-0001-8790-4579)

Abstract: After many decades, the debate continues over when characters should be introduced to college-level Chinese programs in the U.S. This study aimed to assess the impact of early instruction versus delayed instruction of characters on oral and written competency development. A time-series design was used to track two novice-level cases in a blended learning setting for one semester at an American university. One case received the early introduction of character instruction, the other received the delayed instruction. The two participants completed eight oral and written tests after treatment, three repeated summative assessment tasks, as well as pre- and post-intervention attitude questionnaire survey. The findings suggest that the early instruction case performed better than the delayed instruction case in both oral and written tests. Additionally, the delayed instruction case changed attitudes toward the timing of character instruction. Implications for Chinese as a foreign language education are discussed.

Keywords: USA, Postsecondary education, Foreign language teaching, Teaching methods

摘要:近几十年来,关于何时将汉字引入美国大学汉语课堂的问题仍存在争议。该项研究旨在评估汉字教学起始时间差异是否对学习者口语和书面语能力发展有影响。笔者运用时间系列设计和案例分析,跟踪美国大学混合学习环境中的两名初级汉语学习者一个学期。一个案例同时开始汉字和口语教学;另一个案例延迟汉字教学,学生先习得口语后才开始学习汉字。两名被试在研究过程中完成了八个不同的单元测试,重复完成三次综合能力评估测试,在教学实验实施前后都填写了汉字学习态度问卷调查。研究结果表明:同时学习汉字和口语的被试在口语和书面语测试的表现都优于延迟汉字学习的被试。此外,延迟汉字学习的被试改变了其对汉字教学时机的看法。笔者结合研究结果讨论该研究对汉语教学的启示。

关键词:美国、大学教育、外语教学、教学方法

¹ Correspondence concerning this article should be addressed to Sihui (Echo) Ke (email: Sihui.Ke@uky.edu) at the University of Kentucky.

1. Introduction

This exploratory study aimed to address a prolonged debate in teaching Chinese as a foreign language: when should characters be introduced? (e.g., DeFrancis, 1966; Krashen, 2017; Ye, 2013). The Chinese writing system is distinct from the alphabetic writing system. It is a morphosyllabary, with one character mapped onto a morpheme at the syllable level (e.g. 雪 carries the pronunciation of *xuě* and means ‘snow’; 人 is pronounced as *rén* and means person; and 雪人 means ‘snowman’). Some researchers and educators argue that it is easier to learn to read and write after a solid background for speaking and listening has been established (e.g., Packard, 1990); others posit that it is important to get students familiar with characters as early as possible (Knell & West, 2017). However, major gaps remain open in this research area: first, prior studies varied in their operationalizations of the “time lag/delay” in introducing characters (ranging from three weeks to one semester). Second, empirical studies that address this issue have focused on face-to-face traditional classroom settings. Little evidence is available for classroom setting with technology facilitation and flexible schedule (e.g., online Chinese classes).

This research set out to provide additional empirical evidence for this heated debate. Specially, it compared the effects of delaying character instruction versus no delay in novice-level Chinese in a blended learning setting in an American university. Time-series case design was adopted to track learners’ language development over one academic semester. Rather than providing a definite answer to this complex issue, the findings of this longitudinal study aim to provide some specific implications for character instruction in Chinese as a foreign language with consideration of learner attitude in a blended-learning environment.

2. Literature Review

2.1 To Delay or Not to Delay Character Instruction: Theoretical and Practical Rationales

A delay in character instruction is referred to as “a time lag between the time the course starts and the time the characters are introduced” (Packard, p.167). Packard (1990) also pointed out the lag generally lasts from one to four weeks. To our knowledge, DeFrancis (1966) might be among the first to discuss the timing of character instruction. He compared two approaches, that is, the Chao approach and the IFEL¹ approach. The former allowed for flexibility as to when to initiate character learning, ranging from Lesson One to as late as Lesson Nine in the first semester. The latter mandated a time lag in teaching and learning characters. Students at the IFEL first learned conversation texts, then learned characters later. The characters would not go beyond the conversation texts. To date, the majority of the Chinese programs adopts a no-time-lag approach, and begin to teach characters at the beginning or near the beginning in the first semester (see a survey with 124 postsecondary institutions by Ye, 2013). However, researchers and educators are far from a consensus about the best timing for character instruction.

The theoretical bases supporting the delay or no delay in character instruction differed between the delay and no delay camps. The delay camp drew on the important role of oral language in reading development based on child first language reading research (e.g., Unger, Lorish, Noda & Wada, 1993). Unger et al. related second language acquisition with child first language literacy development, “[t]he fact that children normally learn to understand and produce speech before acquiring the skills of literacy shows that literacy is not necessary for normal language acquisition” (p.29). It should be noted that Unger et al.’s framework was built for Japanese curricula for American high school and college students originally, but has been cited by Chinese scholars to support a time delay (e.g., Ye, 2013). Unger et al. also consider two types of students, those who

prioritize oral communication skills and those who have long-term learning goals and will learn listening, speaking, reading and writing. They thought that, for both groups, “a solid foundation in speaking is the best insurance that they will make steady progress in reading” (p. 21).

The no delay camp argued that the separate presentations of pinyin, character meaning, and characters might result in a higher level of *cognitive load* in Chinese learners. For instance, Chung (2007) examined the more effective order of presenting characters, pinyin and English translation for Australian high school learners of Chinese, and found that word retention occurred more rapidly when a character was presented first and then its pinyin and color-coded English translation. The transfer facilitation model specific to L2 reading (Koda, 2005) provides additional justification for no delay. The transfer facilitation model argues that adult L2 readers differ qualitatively from child L1 readers in that adult L2 readers can transfer previously acquired literacy experiences to L2 reading, and that the transfer effect might be moderated by L1-L2 linguistic distance.

As to practical rationales, the camp that supports a delay has proposed three major reasons: first, considering the alphabetic background of American students, it is beneficial for them to command the pinyin system first. Pinyin is the romanization of the Chinese characters based on their pronunciation. The oft-cited research is Everson (1988), who found that beginning-level American university learners of Chinese read texts in pinyin faster than texts written in characters. Moreover, their comprehension accuracy rates were higher. A second reason is related to teaching implementation. Walker (1984) held that a delay provides opportunities for students to review vocabulary and grammar learned previously. Finally, students’ affective factors are considered (Zhao, Guo, & Dynia, 2013). The logic is that learning characters, a distinct orthography, at the onset of exposure will overwhelm the students, cause anxiety, and affect their motivation to learn Chinese.

Counterarguments have been proposed by those who support no delay in character instruction (e.g., Knell & West, 2017). First, regarding beginning-level American university learners' processing of character-based and pinyin-based texts, Light (1976) provided a different interpretation of Everson (1988). Light found that more proficient learners were faster at reading texts written in characters whereas less proficient learners were faster at reading texts in Romanized pinyin. Interestingly, participants in Light's study thought they were better readers of characters. In a recent study by Zhang, Lin, Zhang and Choi (2019), they compared two conditions (character only versus character plus pinyin) and their effects on non-heritage learners and heritage learners' access to word pronunciation in Chinese. It was found that in the no-pinyin (character only) condition, the two groups showed no significant score difference, whereas in the character plus pinyin condition, heritage learners significantly outperformed non-heritage learners. The pinyin/no pinyin conditions were less marked for non-heritage foreign language learners of Chinese. The second reason is also related to curriculum design. Knell and West (2017) proposed that reading and writing have the potential to increase students' Chinese language performance in general. Recent research that implemented a reading-to-learn approach with intermediate- and advanced-level Chinese learners in an American university also indicates that learning to read and write in Chinese can promote lexical learning and presentational communication skills in Chinese learners (e.g., Author, 2018; Zhang & Koda, 2018). Lastly, in relation to students' attitudes toward character learning, Ye (2013) surveyed 907 college students' beliefs in character instruction and found that the majority (about 77%) prefer character instruction near or at the beginning of the first semester. In a recent study that compared two instructional methods (no delay versus a delay in character instruction) in American high school learners of Chinese, Knell and West (2017) did not find students intimidated by characters, which was at least true for the no delay group.

In the next section, we try to consolidate direct evidence that compared the two approaches toward character instruction (i.e., Knell & West, 2017; Packard, 1990; Ye, 2013).

2.2.Previous Research Evidence: Packard (1990), Knell & West (2017) and Ye (2013)

To our knowledge, the aforementioned studies have generated direct evidence regarding the debate on the best timing of teaching characters. Packard (1990) is often cited by those who support a delay in character instruction. He compared the performance of two university-level learner groups (delay versus no delay) and found that there was no significance difference in reading and writing test performance between the two groups, but the delay group achieved better scores in fluency and phonological discrimination tasks. In contrast, Knell and West (2017) partially replicated the study in American high school students and found that there was no significance differences in oral interview and fluency assessment between the two groups, yet the no delay group performed significantly better in reading and writing tests. Moreover, Knell and West observed no significant difference between the two groups in terms of students' attitudes toward the timing of character instruction. Ye (2013)'s study did not involve any testing or group comparison. Rather, she did a large-scale survey with students and instructors from 124 postsecondary institutions and asked about their beliefs about the best timing to teach and learn characters. The following discussion evaluates the design and evidence of these three studies.

Packard's (1990) research is among the oft-cited literature to support a delay in character instruction. The design of his study was to implement a three-week time lag in American university students of elementary-level Chinese in one academic year, and compared the oral and written language performance between the time lag group and the no time lag group. However, the instruction was carried out by two different instructors. Also, the methods used to teach characters were unclear. Moreover, contact hours of oral and written Chinese in the two groups were not

specified. Knell and West (2017)'s study could be viewed as a semi-replication study of Packard's. The research method was similar, that is, comparing the oral and written performance in mid-term and final tests between two instructional groups (termed as "early introduction" and "delayed introduction" in their study). Knell and West's study included three notable differences: (1) the population was American high school students; (2) the time lag was longer (i.e., one semester vs. 3 weeks); (3) the same instructor carried out the intervention. Finally, Knell and West provided more details regarding teaching methods and also specified the contact hours for both learner groups. The teaching approach was traditional and drill-based, as students were asked to practice stroke order and to copy characters. Krashen (2017) evaluated Knell and West's study design and inferred that the performance difference in written language assessment could be due to the longer exposure to reading and writing in the early introduction group.

Another difference between Knell and West's (2017) study and Packard's (1990) study is that the former conducted an attitude questionnaire survey. However, they did not find any notable difference between the two instructional groups. To our knowledge, Ye's (2013) survey may be the most comprehensive and provide more meaningful insights of American learners' attitudes toward character teaching and learning². The majority of students preferred to learn characters at or near the beginning of the first semester, and they believed that learning characters is an essential part of learning Chinese. Participants also posited that an early start makes it less difficult in the long run to learn characters. For those who preferred a delay (about 23% of the informants), their rationales are consistent with the viewpoints of some researchers and educators (e.g., McGinnis, 1990; Unger et al., 1993; Zhao et al., 2013). They held that (1) learning to speak and listen is more important than learning to read and write; (2) learning to speak and listen sets a foundation for reading and writing; and (3) learning all four skills simultaneously increases students' cognitive

load and decreases their interest in learning Chinese. Interestingly, in a follow-up interview, when the informants were presented with the rationales for a delay, those who preferred no delay changed their attitudes.

In spite of the lingering debate over the timing of character instruction in Chinese as a foreign language, the persistent dearth of empirical research evidence leaves instructors with no reliable implications for real world practices. Existing research findings vary (e.g., Knell & West, 2017; Packard, 1990; Ye, 2013). Other inconsistencies include the operational definition of time delay, instructional methods, and testing methods. The common research design has been to compare changes in two instructional groups via assessment at two time points only. Current researcher lacks longitudinal data tracking of learners' dynamic development of oral and written language competencies. Finally, there is very little attention given to how learners' affective factors might influence the interpretation of findings.

3. The Present Study

The current research aimed to provide additional evidence regarding the debate over the timing of character instruction in Chinese as a foreign language. Based on time-series case design (for a review, see Mellow, 2013), data were gathered from two participant cases over one semester in an American university in a lab-based blended learning setting. Two questions guided the research:

- (1) Are there any intraindividual and interindividual differences between early instruction (EI) and delayed instruction (DI) learners in their oral and written skills development over one semester of learning?
- (2) Is there any difference between EI and DI learners' attitudes toward the timing of introducing characters?

4. Methods

4.1. Participants

Two participants who had had no previous exposure to (Mandarin) Chinese language were recruited for this research. Both were English-speaking students from an American research university in the Midwest. Each participant received a certificate by completing online Chinese course requirements via Coursera, and was compensated with a \$10 Amazon gift card for each face-to-face meeting with the instructor. Each participant spent approximately four hours learning each week, including weekly online self-learning and a face-to-face meeting with the instructor in a language lab.

According to Winke (2013), learner-related factors play important roles in predicting Chinese learners' performance. In her study with 96 English-speaking advanced-level learners of Chinese in an American university, aptitude, motivation and learning strategies predicted about 11% of the variance in learners' Chinese listening, speaking and reading skills. The current study focused on learner individual differences by reporting (1) their motivation, (2) aptitude and (3) attitudes at two time points (i.e., pre- and post-intervention). Participants' motivation was determined by a questionnaire survey constructed after Kormos and Dörnyei (2004) and Winke (2013). The questionnaire included twenty items with a five-point Likert scale. As shown in Table 1, both participants were highly motivated learners because their responses maintained an average score at or above four point five out of five. Participants' language aptitudes were tested with the Pimsleur Language Aptitude Battery (PLAB) (Pimsleur, Reed & Stansfield, 2004). Both participants were considered to have *excellent* aptitudes (with grades above 74 out of 111 points). Notably, they both excelled in subtest of auditory ability. Lastly, their attitudes toward character

instruction were surveyed with items adopted from Ye (2013). Changes in their attitudes (if any) are reported in the Results section.

Table 1.

Participant Background

Participant	Gender	Age	L1	L2s	Attitude	Aptitude	Motivation	Time spent weekly (hrs)
Early instruction (EI)	Female	19	English	French & Spanish	No delay	88 (49 out of 50 for auditory ability)	4.50 out of 5	3
Delayed instruction (EI)	Male	20	English	German, Spanish & French	Delay	91 (50 out of 50 for auditory ability)	4.85 out of 5	3

Note. L1, first language. L2s, additional languages.

4.2. Instructional Design

The present research was carried out by a junior researcher who is a native speaker and a senior researcher who is a nonnative speaker of Chinese. The intervention was carried out in a blended learning setting over 14 weeks in one semester, delivered by the junior researcher who has taught Chinese as a foreign language for about eight years in the U.S.

The target proficiency level toward the end of the intervention was novice mid to novice high according to ACTFL proficiency guidelines (Swender, Conrad & Vicars, 2012). The teaching materials were theme-based, covering nine topics (greetings, time and date, shopping, family,

dining, directions, making friends, making phone calls, and location). The learners were expected to learn about 252 single- and multi-character new words.

For the early instruction case, characters were taught in the first week; whereas for the delayed instruction case, characters were not taught until the fourth week. In the early introduction case, characters were taught mainly through implicit methods. The students were exposed to characters in vocabulary and text learning and exercises. The instructor did not provide any explicit instruction in the first four weeks. Beginning in Week Five, the instructor started to introduce the strategy of using radical awareness or morphological awareness, that is using familiar phonetic/semantic radical in a character and familiar word parts in a two-character word to guess new word sounds and meanings. The schedule for the two cases is illustrated in Table 2. Both cases had about 20 hours exposure to oral and written Chinese by the end of Week Five.

From Week Six forward, the instruction was the same for the two cases: each week, the learners first spent about three hours on a Chinese course titled “Chinese for Beginners” and administered by Shanghai Transportation University via Coursera, then they joined a one-hour individual face-to-face meeting with the same instructor. During the online independent learning section, learners would practice new vocabulary items, listen to audios and watch videos, practice conversations, read short texts in Chinese and culture notes, and complete quizzes to check their understanding. During the face-to-face meetings, the instructor helped the learner to review and practice conversations, answered questions raised by the learner, and administered a weekly unit test.

Table 2.

Instructional Schedule

Weekly schedule	Early instruction	Delayed instruction
1	Course introduction pinyin & characters, Lesson One Hello	Course introduction Lesson One & Lesson Two (researcher-developed file in pinyin only)
2	Lesson Two What's the time?	Lesson Three & Lesson Four (researcher-developed file in pinyin only)
3	Lesson Three What do you want to buy?	Lesson Five (researcher-developed file in pinyin only)
4	Lesson Four Family	Review Lessons One to Three online and learn characters
5	Lesson Five What would you like to eat?	Review Lessons Four & Five online and learn characters
6	Midterm test	Midterm test
7	Lesson Six Directions	Lesson Six Directions
8	Lesson Seven Phone number	Lesson Seven Phone number
9	Lesson Eight What's up	Lesson Eight What's up

10	Lesson Nine Where's my cellphone?	Lesson Nine Where's my cellphone?
11	Holiday (no class)	Holiday (no class)
12	Review	Review
13	Final test	Final test
14	Delayed post-test	Delayed post-test

4.3. Assessment Instruments

Summative assessments were conducted at three time points, including mid-term, final, and delayed post tests. The delayed post-test was administered one month after the intervention. The assessment battery was adopted from Knell and West (2017) that tested participants' oral and written language competencies in Chinese respectively. Assessments consisted of two tasks (i.e., oral interview, and character and word recognition). In addition to the summative assessments, weekly unit tests in the same format were administered to both participants, for a total of eight tests per participant. Eight observations met the minimum requirements for time-series design (Mellow, Reeder & Forster, 1996). Task instructions and sample items are described below.

There were two sections in the oral interview accuracy and fluency task. In the first section, each participant was asked to answer 10 questions (e.g., *Xiànzài jǐdiǎn jǐfēn?* [What time is it now?]). In the second section s/he was provided with pictures and responded to 10 questions in relation to the pictures. For example, the researcher showed the participant a picture of the national flag of the United Kingdom and asked "*Tā cóng nǎlǐ lái?* [Where is he from?]"). The maximum

score possible for accuracy was 20. A fluency scale was generated on a five-point scale (after Knell and West, 2017) (e.g., 1, Has great difficulty producing smooth expressions and often hesitates when speaking; 5, almost always speaks smoothly with little or no hesitation or breaks.)

There were two sections in the character and word recognition task, with 16 items in the first section and 10 items in the second section. In the first section, the researcher read aloud a single character or a two-character word, and the participant were given a written multiple choice response from which they selected the correct answer. In the second section, the format was the same, but prompts were short sentences, with no more than 10 characters in each sentence.

It should be noted that it was impossible to replicate Knell and West's (2013) study by adopting their instruments directly. Knell and West used traditional characters in their study, whereas simplified characters were taught to the participants in this research. Also, since the teaching materials were different, the researcher had to adjust the contents of several items in Knell and West's assessment instruments. Lastly, no writing skills were tested in the research because the two participants were not motivated to write Chinese characters but rather to read and type the characters.

4.4. Data Analysis Procedures

Mixed methods were used to analyze the quantitative and qualitative data gathered in the research (Creswell, 2015). Quantitative data generated from the time-series case design and descriptive statistics are reported first, with oral language competence and written language competence as the dependent variables. No inferential nonparametric statistical design has been undertaken due to the numbers of cases and observations associated with each case in this research. Rather, we followed previous time-series research approaches seen in in Applied Linguistics (e.g., Mellow et

al., 1996) and reported changes in means and changes in trends by visual plotting (for reviews of time-series designs in Applied Linguistics research, see Mellow, 2013). Qualitative data regarding participants' change in attitudes toward character teaching and learning included participants' oral and written responses to interviews and surveys. These comments were analyzed to monitor participants' attitude changes and what led to such changes (if any).

5. Results

The first goal of this study was to investigate any intraindividual and interindividual differences between early instruction (EI) and delayed instruction (DI) learners in the development of their oral and written skills over one semester of learning. Summative assessment results are shown in Table 3. Regarding oral interview accuracy, the EI learner improved between mid-term and final tests, yet her accuracy rate decreased sharply in the delayed post-test one month after the intervention. The DI learner started strongly in oral interview, with an 80.00% accuracy rate, yet the accuracy rate continued to drop (to 60.00% in the final test and to 38.33% in the delayed post-test). As to oral interview fluency, both learners' performance was rated between the ranges of two and three (five being the highest) across three time points. Finally, the EI learner did well in the character and word recognition task consistently and maintained a 95.24% accurate rate in the delayed post-test whereas the DI learner's accuracy rate was notably lower (73.81%).

Table 3.

Summative Assessment Results at Three Time Pints

Task	EI	DI

	Mid-term	Final	Delayed	Mid-term	Final	Delayed
Oral interview accuracy rate	68.33%	75.00%	33.33%	81.67%	60.00%	38.33%
Oral interview fluency	2.88	2.25	2.44	2.75	1.80	2.72
Character and word recognition accuracy rate	76.19%	97.61%	95.24%	42.86%	71.43%	73.81%

The participants also completed eight weekly tests, whose results served to track the development of their oral and written language competencies. The results are illustrated in Table 4 and Figures 1-3. The EI learner's developmental trajectory for the first three unit tests followed a U-shape turn. She started strongly in tasks measured by the three different indices; the score decreased on the second test, and increased on the third test. For each of the next five tests her performance was stable regarding oral interview accuracy and fluency; she performed better in the character and word recognition tests. The DI learners' performance in the first three unit tests was surprisingly low, despite that no characters were taught during that time. He caught up in the final two tests, with higher accuracy rates (both being 70.00%) than the EI learner (both being 60.00%). However, he scored lower on the fluency tests than the EI learner. Also, his performance in the character and word recognition overall was poor, with six test scores below chance level (33.33%). The EI learner clearly outperformed the DI learner with regard to oral interview fluency and character and word recognition accuracy. Her performance in oral interview accuracy tests was also more stable and consistent than that of the DI learner.

Table 4. Results of Weekly Unit Tests

Oral Interview Accuracy Rate								
Unit test	1	2	3	4	5	6	7	8
EI	83.30%	26.70%	55.00%	78.30%	51.70%	46.70%	60.00%	60.00%
DI	41.67%	36.67%	30.00%	35.00%	46.67%	35.00%	70.00%	70.00%
Oral Interview Fluency								
Unit test	1	2	3	4	5	6	7	8
EI	1.50	1.00	2.75	2.25	1.75	3.00	2.50	2.00
DI	1.50	2.00	1.00	1.00	2.50	1.25	1.25	1.25
Character and Word Recognition Accurate Rate								
Unit test	1	2	3	4	5	6	7	8
EI	100.00%	71.43%	89.29%	89.29%	100.00%	57.14%	78.57%	92.86%
DI	42.86%	0.00%	14.29%	28.57%	14.29%	35.71%	0.00%	0.00%

Figure 1. Oral Interview Accuracy

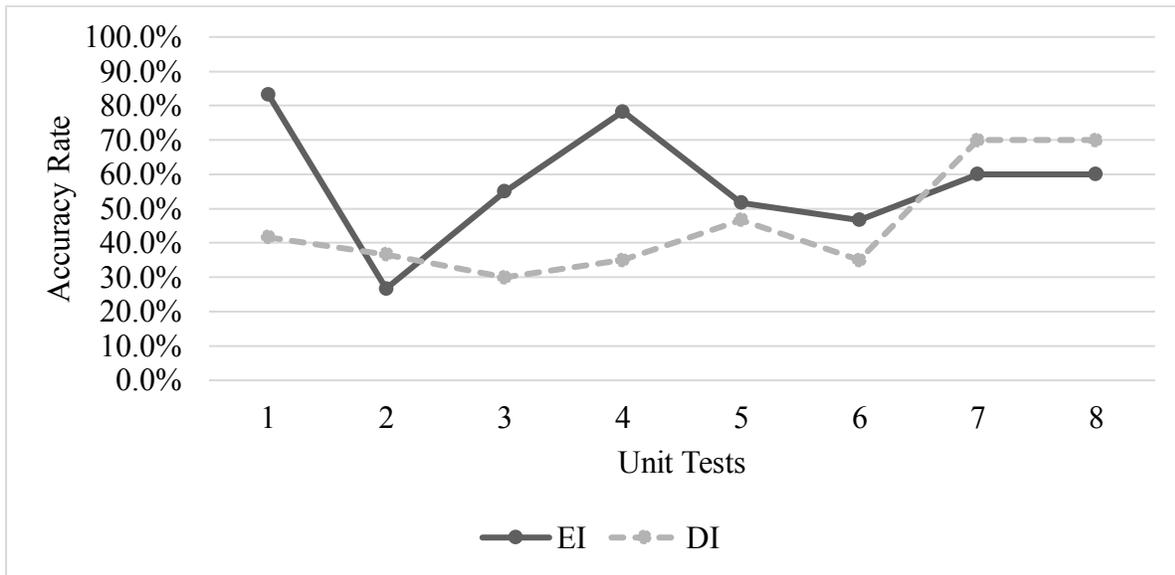


Figure 2. Oral Interview Fluency

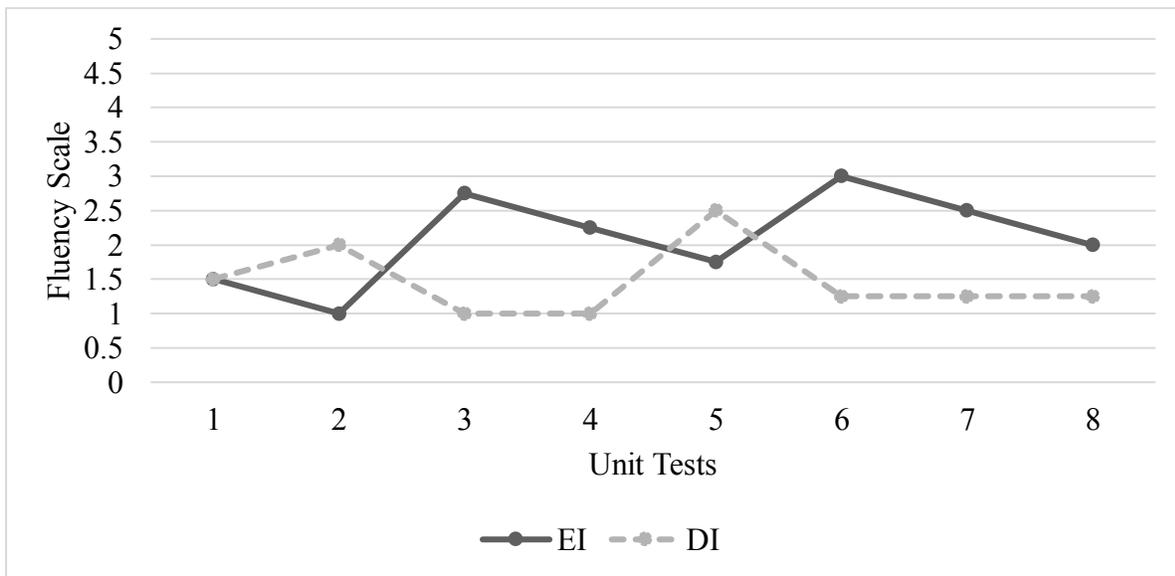
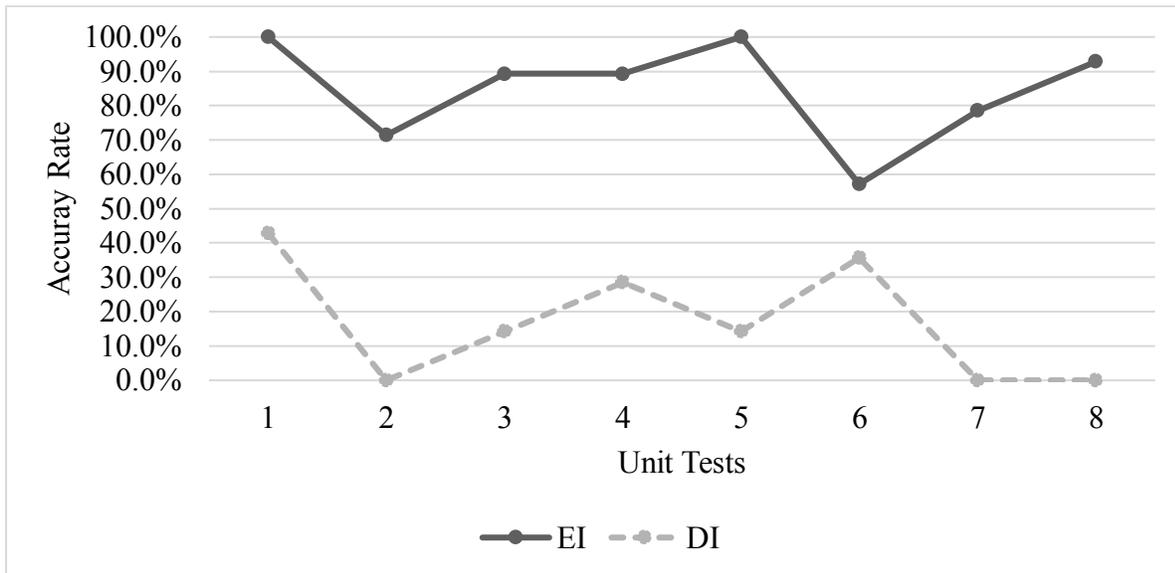


Figure 3. Character and Word Recognition



The second goal of the research is to examine any differences between EI and DI learners' attitudes regarding the timing of introducing characters. Participants oral and written responses pre- and post-intervention were compared to determine any changes in attitude.

The EI learner did not show a change in attitude toward the timing of learning Chinese characters (she wanted to learn characters from the beginning). She maintained that there were two “excellent” reasons: “characters are an important part of the Chinese language. There are many homophones (i.e., the same pinyin can have many different meanings) in Chinese.” In her final reflection, she responded,

“The characters are different, and a better way of learning is studying them right away. But they are different enough that maybe slowly learning them would be good.”

As to the DI learner, before the intervention cycle, he indicated that he preferred learning characters in the middle of the first semester. A good reason for the delay was that “it is too much for students to learn all four skills from the very beginning”. But he also acknowledged three

excellent reasons to learn characters at the beginning: “characters are an important part of the Chinese language. Students eventually have to learn characters anyway. There are many homophones (i.e., the same pinyin can have many different meanings) in Chinese.” After the intervention, his attitude shifted. He wished that he had learned characters from the beginning and provided his emotive reasoning in written response:

“Because I didn’t learn them right away, I feel like that’s a big factor into why I have trouble reading them.”

These findings suggest that there were both intraindividual and interindividual differences between the EI and DI learners in the development of their oral and written competencies over one semester in a blended learning setting. At the intraindividual level, the EI learner’s oral and written language competencies underwent consistent development while showing moderate improvement in both oral and written language competencies. In contrast, the DI learner’s developmental trajectory was chaotic and development appeared sporadic while demonstrating improvement in oral language competence only. His reading accuracy rate was below chance level.

At the interindividual level, there was no notable difference between the two learners’ performance in oral interview accuracy rates; the EI learner outperformed the DI learner in oral interview fluency and character and word recognition accuracy. As to learners’ attitudes toward the timing of introducing characters, the EI learner and she retained her preference for learning characters at the onset of instruction. The DI learner, however, pivoted substantially during the study; he shifted his initial preference for delayed character instruction toward a belief that it is more beneficial to learn character at the beginning of instruction. Both learners provided a linguistic reason supporting an early introduction of characters: there are many homophones in Chinese, which carry the same sounds but correspond to different characters and meanings.

6. Discussion, Conclusions, Limitations and Implications

The findings of this research with American university novice-level learners of Chinese in a blended learning indicate that early introduction of characters appears more beneficial for learners' development of both oral and written language competencies, which echoes Knell and West (2013)'s conclusions, yet contradicts Packard's (1990) findings. In addition, both learners indicated a preference for early introduction of characters in the post-intervention survey. The primary goal of this study was not to posit a definitive answer to the debate about the best timing of character instruction. We are simply interested in exploring the influence of individual differences and the potential interaction with intervention that could have led to our findings. The two cases were similar in terms of their age, previous language learning background, language aptitude and (high) motivation. Their contact hours in oral and written Chinese were manipulated and controlled to ensure equal numbers of learning hours in each case. The notable difference is their preference toward the timing of character instruction before intervention. The EI learner preferred no delay, while the DI learner preferred a delay. The findings of this research might have been different if the two cases had been switched and they were not assigned according to their preferred instructional approach.

In the context of online learning or blended learning of Chinese as a foreign language in a flexible manner, with learners of high-level language aptitude and motivation, teaching characters at the beginning seems to be more beneficial for learners to develop both oral and written language competencies.

It is noteworthy that the research design of this study is different from prior studies. Time-series case design and delayed post testing were adopted in this research to track intraindividual and interindividual differences over one semester's learning, whereas aggregative summative

assessment scores based on different two interventional groups (i.e., a delay versus no delay) were used in previous research to indicate changes between pre- and immediate post-tests due to different types of interventions. In future research, at least four participant cases and eight observations pre- and post-intervention for at least three time points are needed to for baseline data collection and inferential statistical analysis (Mellow et al., 1996; Mellow, 2013).

Future research will also need to compare the effects of the different operationalizations of the time delay (e.g., three to four weeks versus one semester over one year's formal learning) and character instruction. As pointed out by DeFrancis (1966), a major error in Chinese reading program is the obsession with how many characters a learner should know. Rather, instructional attention should be paid to fostering students' ability to use known word parts to learn novel characters and words. Also, there is a need to reconceptualize the role of reading in promoting content and language learning since existing Chinese programs in North America are heavily reading-based anyway (see a reading-to-learn approach in Author, 2018; Zhang & Koda, 2018).

Finally, one alternative approach to address the issue of character instruction is to validate empirically the cognitive load of reading pinyin only versus reading pinyin and characters simultaneously in learners of Chinese as a foreign language. One oft-cited rationale for a delay in character instruction is that learning the sounds and orthography simultaneously will increase learners' cognitive load. However, emergent research evidence suggests the alternative (e.g., Chung, 2007; Zhang et al., 2019). With the advance of technology, more research with learners of Chinese as a foreign language in an online or blended learning environment is needed to test the cognitive load hypothesis. It should be noted that no follow-up in-depth interview about learners' experience with blended learning was included in the present study. Researchers should consider including more comprehensive interviews in future inquiries.

Notes

1 IFEL, the Institute of Far Eastern Languages at Yale.

2 Ye (2013) also surveyed Chinese teachers' beliefs in character instruction. Given the scope of this study, the discussion is focused on learners' beliefs only.

Acknowledgements

We are grateful for the support of ACTFL 2018's Research Priorities Initiative and the feedback from colleagues who attended the ACTFL/ Middlebury Research Forum 2019. Special thanks go to Dr. Margaret E. Malone and Ms. Amanda Kwak. Also, we greatly appreciate Dr. E. Knell's sharing of assessment instruments. Finally, we would like to thank the editor and anonymous reviewers for their feedback and comments that help improve this manuscript.

References

- Chung, K. K. (2007). Presentation factors in the learning of Chinese characters: The order and position of Hanyu pinyin and English translations. *Educational Psychology, 27*(1), 1-20.
- Creswell, J.W. (2015). *A concise introduction to mixed methods research, 4th ed.* California: Sage Publications, Inc.
- DeFrancis, J. (1966). Why Johnny can't read Chinese. *Journal of the Chinese Language Teachers Association, 1*(1), 1-20.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition.* Mahwah, NJ: Lawrence Erlbaum.
- Everson, M. E. (1988). Speed and comprehension in reading Chinese: Romanization vs. characters revisited. *Journal of the Chinese Language Teachers Association, 23*(2), 1-15.
- Knell, E., & West, H. I. (2017). To delay or not to delay: The timing of Chinese character instruction for secondary learners. *Foreign Language Annals, 50*(3), 519-532.
- Kormos, J. & Dörnyei, Z. (2004). The interaction of linguistic and motivational variables in second language task performance. *Zeitschrift für Interkulturellen Fremdsprachenunterricht [Online]*,9(2).
<http://www.ualberta.ca/~german/ejournal/kormos2.htm>
- Krashen, S. (2017). Early vs. Delayed Introduction of Chinese Characters in Beginning Mandarin Classes: Is the “Early Advantage” Due to More Instructional Hours?. *Foreign Language Annals, 50*(4), 829-830.

- Light, T. (1976). Comparative reading speeds with Romanized and character texts. *Journal of the Chinese Language Teachers Association*, 11(1), 1-10.
- Mellow, J. D. (2012). Time series. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1-5). Online: Blackwell. DOI: 10.1002/9781405198431.wbeal1216
- Mellow, J. D., Reeder, K., & Forster, E. (1996). Using time-series research designs to investigate the effects of instruction on SLA. *Studies in Second Language Acquisition*, 18(3), 325-350.
- McGinnis, S. (1999). Student goals and approaches. *Mapping the Course of the Chinese Language Field*, 3, 151-168.
- Packard, J. L. (1990). Effects of time lag in the introduction of characters into the Chinese language curriculum. *Modern Language Journal*, 74(2), 167-175.
- Pimsleur, P., Reed, D. J. & Stansfield, C. W. (2004). *Pimsleur Language Aptitude Battery: Manual, 2004 Edition*. N. Bethesda, MD: Second Language Testing, Inc.
- Swender, E., Conrad, D. J., & Vicars, R. (2012). *ACTFL proficiency guidelines 2012*. American Council on the Teaching of Foreign Languages.
- Noda, M., Wada, Y., & Unger, J. M. (1993). *A Framework for Introductory Japanese Language Curricula in American High Schools and Colleges*. Johns Hopkins University.
- Walker, G. (1984). Literacy and 'reading' in a Chinese language program. *Journal of the Chinese Language Teachers Association*, 19(1), 67-84.
- Winke, P. (2013). An investigation into second language aptitude for advanced Chinese language learning. *The Modern Language Journal*, 97(1), 109-130.

Ye, L. (2013). Shall we delay teaching characters in teaching Chinese as a foreign language?. *Foreign Language Annals*, 46(4), 610-627.

Zhang, H. S., & Koda, K. (2018). Enhancing lexical sophistication through IC skills training in low-intermediate Chinese as a foreign language in the US. In K. Koda & J. Yamashita (eds.), *Reading to Learn in a Foreign Language: An Integrated Approach to Foreign Language Instruction and Assessment (pp.180-190)*. New York: Routledge