International Journal of Disability, Development and Education

ISSN: 1034-912X (Print) 1465-346X (Online) Journal homepage: http://www.tandfonline.com/loi/cijd20

Supporting Orthographic Learning at the Beginning Stage of Learning to Read Chinese as a Second Language

Li-Yun Chang, Yi Xu, Charles A. Perfetti, Juan Zhang & Hsueh-Chih Chen

To cite this article: Li-Yun Chang, Yi Xu, Charles A. Perfetti, Juan Zhang & Hsueh-Chih Chen (2014) Supporting Orthographic Learning at the Beginning Stage of Learning to Read Chinese as a Second Language, International Journal of Disability, Development and Education, 61:3, 288-305, DOI: 10.1080/1034912X.2014.934016

To link to this article: http://dx.doi.org/10.1080/1034912X.2014.934016

Published online: 03 Sep 2014.

International Journal of Disability, Development and Education, 2014
Vol. 61, No. 3, 288–305, http://dx.doi.org/10.1080/1034912X.2014.934016

Supporting Orthographic Learning at the Beginning Stage of Learning to Read Chinese as a Second Language

Li-Yun Chang[a]*, Yi Xu[b], Charles A. Perfetti[a]*, Juan Zhang[c] and Hsueh-Chih Chen[d]

[a]*Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA, USA;* [b]*Department of East Asian Languages and Literatures, University of Pittsburgh, Pittsburgh, PA, USA;* [c]*Faculty of Education, University of Macau, Macau, China;* [d]*Department of Educational Psychology and Counseling, National Taiwan Normal University, Taipei, Taiwan*

Learning to read a second language (L2) is especially challenging when a target L2 requires learning new graphic forms. Learning Chinese, which consists of thousands of characters composed of hundreds of basic writing units, presents such a challenge of orthographic learning for adult English speakers at the beginning stages of learning. In this study, we use an *in vivo* classroom design to extend previous research on how to support orthographic learning. First, we test the hypothesis that learning characters is enhanced by a grouped sequence of characters that share sub-character graphic components. Second, we examine the effects of four encoding methods that have been investigated in laboratory studies—handwriting, visual chunking, passive reading, and stroke-reporting. The results demonstrate that the grouped approach facilitated character production compared with the distributed approach and that visual-chunking outperformed the other three encoding methods under the grouped sequence. We propose that learning via visual chunking with characters grouped by the same chunks enhances the Chinese orthographic representations of beginning L2 learners.

Keywords: characters-distributed instruction; characters-in-group instruction; character scoring scheme; Chinese as foreign/second-language learning; encoding methods; handwriting; *in vivo* experiment; orthographic learning; reading; visual chunking

Introduction

Reading is fundamentally learning to associate written forms with pronunciations and meanings. Among the three constituents of lexical representations—orthography, phonology, and semantics (Perfetti, Liu, & Tan, 2005)—the orthographic constituent plays a prominent role in reading as the initial step of visual word recognition. Although robust orthographic representations are universally important for reading, the visual perceptual properties differ across written languages (Pelli, Burns, Farell, & Moore-Page, 2006), producing corresponding differences in learning the graphic forms of a writing system. For example, Nag (2007) reports that in reading Kannada, which is an alphasyllabic language with hundreds of symbols and more complex visual-spatial features than alphabetic languages, children take longer to master the orthographic forms. Such properties of written language become an issue in second-language (L2) learning, which is learning to map novel orthographic forms to a L2 as it is being learned. This is especially true when the L2 writing system contrasts with the first-language writing system in its visual forms and its mapping principles.

*Corresponding authors. Email: lic72@pitt.edu (L.-Y Chang); perfetti@pitt.edu (C. A. Perfetti)

Chinese thus presents orthographic learning challenges to readers of English and other alphabetic languages. Studies of both Chinese in foreign language (CFL) and Chinese as second-language (CSL) contexts report that recognising Chinese characters is the most challenging reading task for English speakers (Everson, 1998). A Chinese character is composed of strokes interwoven in patterns to form chunk(s)[1] in a square-like form. However, what make Chinese visual orthography strikingly different from other languages are its large number of orthographic units (i.e., stroke, chunk, radical, and character) and their complex combinations (DeFrancis, 1989). Five basic strokes (i.e., horizontal stroke, vertical stroke, slant stroke, point stroke, and angled stroke) can yield up to 44 additional variations of stroke shapes (Wang, 2011). These strokes combine to form 439 chunks. Most of these chunks must follow certain positional constraints (i.e., spatial layouts) within the two-dimensional square to form compound characters, which accounts for 80% of the character types of 7000 frequent characters (Chinese Language Committee, 2009).

Despite the crucial role that visual complexity plays in establishing orthographic representations in Chinese reading, little attention has been given to how to support the orthographic representations of CFL/CSL learners. According to a review that summarises the history of CFL/CSL teaching in regular classroom settings, most instructors focus on improving learners' listening and speaking language skills, while leaving reading and writing as the learners' responsibility to practice outside the classroom (Zhong, 1990). Recently, a growing body of laboratory research has investigated orthographic learning of adult CFL/CSL learners directly (Cao et al., 2013a, 2013b; Guan, Liu, Chan, Ye, & Perfetti, 2011; Xu, Chang, Zhang, & Perfetti, 2013). These studies demonstrated that various encoding methods—handwriting, stroke animation, and visual chunking—on character forms can strengthen orthographic constituents and their connections to semantics and phonology. For instance, writing characters from memory after a brief exposure led to better learning of the link between the character and its meaning, while typing the pinyin (an alphabetic spelling) helped to learn the link between the character and its pronunciation. Integrating handwriting and pinyin typing facilitated orthographic recognition and strengthened both orthography–semantics and orthography–phonology links (Guan et al., 2011).

To explain these effects, the researchers hypothesised that handwriting draws learners' attention to the decomposed structures of characters and establishes motor memory traces that support recognition through production knowledge. This explanation of the writing-on-reading effect was supported in a functional magnetic resonance imaging study (Cao et al., 2013a), which found greater activations in bilateral superior parietal lobules and lingual gyri—brain areas that support visual word processing and motor processes—when Chinese characters were learned via handwriting compared with pinyin typing.

Following these results, Xu et al. (2013) compared the effectiveness of passive reading, handwriting, and computerised animation software presentation. The animation was a video showing how a character was written with correct stroke order and direction, resembling a situation of implicit writing. While this study found that writing and animation were more effective than passive reading, as indicated in the response times to make lexical decisions on characters, there was no clear pattern in the character-to-meaning or character-to-pronunciation links.

An alternative encoding method that has been investigated is visual chunking, which directs learners' attention to decomposed spatial structures and chunks (see Figure 1 for an encoding method of visual chunking). On average, a chunk has five strokes (Chen et al., 2011). Remembering the chunks of a character, which on average has 15 strokes,

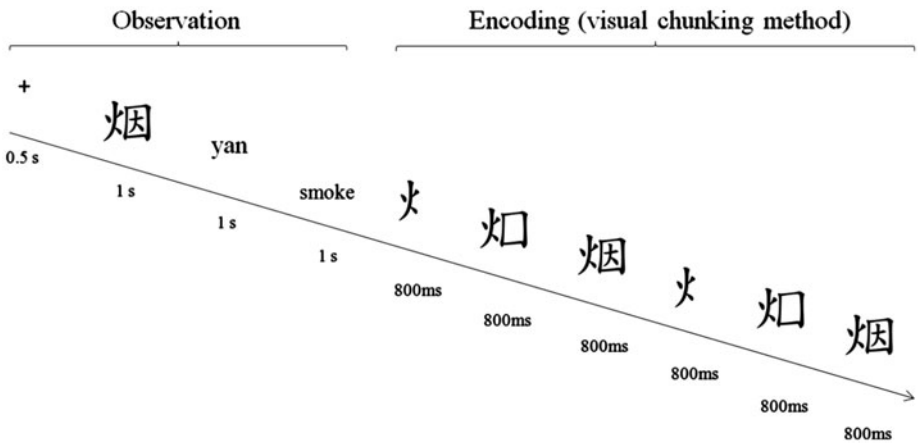

Figure 1. Learning procedure for one sample character with three chunks using the visual chunking encoding method.

would therefore greatly reduce visual memory load. In an event-related potential study, Cao et al. (2013b) found that, although a visual chunking condition did not outperform handwriting and passive reading in behavioural data, chunking during learning later elicited greater amplitudes in the N170, an event-related potential indicator of orthographic recognition.

Collectively, these studies suggest that handwriting supports learners' orthographic learning, while the effects of stroke animation and visual chunking are less robust and need further investigation. Moreover, these studies were carried out in laboratories, which poses the question of whether the effects observed under well-controlled experimental conditions are robust enough to be observed in actual classrooms. For example, although handwriting practice improved orthographic recognition, the practice time average from those studies is about six minutes per character, a practice time unlikely to be implemented in traditional language classrooms. Hence, one goal of the present study is to examine whether the effects found in the laboratory studies can be generalised to real classroom settings, which have shorter practice time and a noisier environment. Such settings, with real students in actual classrooms, allow *in vivo* studies of learning (Koedinger, Aleven, Roll, & Baker, 2009; Koedinger, Corbett, & Perfetti, 2012).

Another instructional factor in orthographic learning is how graphic forms are introduced. A critical review of teaching methods for Chinese children highlights a contrast between meaning-centred and character-centred approaches (Lam, 2011). The meaning-centred approach assumes that communication is the purpose of learning to read and, accordingly, introducing characters in the sequence in which they are introduced in school textbooks allows learners to read these books (Wan, 1991). This is distributed learning, with no design to introduce together characters that share graphic forms. This distributed method dominates current practice for both first-language and L2 learning (Lam, 2011). In contrast, the character-centred approach introduces characters that share the same chunk together in a group. For instance, characters 饱, 蚀, 饶, 馅, 馋, and 馒 all share the same chunk 饣. Thus, learners are able to observe the shared features across characters and develop orthographic awareness for the specific chunks and the layouts of the characters. With this orthographic knowledge of chunks, learners can distinguish orthographically similar characters and infer the meanings or pronunciations of novel characters also containing the chunk (Wan, 1991). The distributed approach thus emphasises the pragmatic aspect of word usages as reflected in texts, whereas the grouped approach focuses on the constituent forms of orthographic units.

Although some studies appear to support the grouped approach (Su, 2006) while others appear to support the distributed approach (Si, 2001), there are no well-controlled comparisons of the two approaches. To establish valid comparisons between the two, one needs a set of meaningful texts that present characters sharing the same chunks in each unit, and a different set of meaningful materials in which those identical characters are introduced in separate units, so that each unit does not contain characters that share the same chunks. Moreover, comparisons in classrooms face the challenge of comparing different teachers and different students, creating confounds with the design of instruction. One solution is to hold the instructor factor consistent while varying the introduction sequence—grouped or distributed—of characters in the learning materials. This is the approach we use in this *in vivo* study.

The current study is the first attempt to investigate beginning L2 learners' orthographic learning as a function of these two presentation sequences, as well as the first study to examine the effect of character encoding procedures in an *in vivo* setting. The study compared the effects of character introduction designs—namely, grouped versus distributed—and four encoding procedures—handwriting, visual chunking, passive reading, and stroke-reporting—on first-year CFL learners' orthographic learning. On the assumption that drawing attention to form strengthens orthographic representation, we hypothesised that grouped characters, compared with distributed, would lead to better orthographic learning. Based on the laboratory results of character encoding methods, we expected that handwriting of characters would have a larger general effect on character recognition than the other three encoding methods. However, we expected that visual chunking would especially facilitate orthographic learning when characters were introduced in groups, providing additional attention to orthographic forms that would pay off in later orthographic recognition.

Methods

Participants

Forty-eight monolingual English speakers (18 males; mean age 18.96 years) enrolled in first-year Chinese language classes at the University of Pittsburgh participated in this study. They received credits and payment for participating. Prior to the experiment, the participants had received eight weeks of classroom Chinese instruction. During these eight weeks, they were taught general rules of stroke orders, knowledge of pinyin, and had acquired about 180 characters. To ensure that the Chinese proficiency levels of participants between the two conditions were equivalent, we used the participants' Chinese overall course score—which includes their listening, speaking, reading, writing performances, and vocabulary knowledge—to match their proficiency. Data from an additional seven participants whose course score did not match with the others were excluded from the analysis. There were 24 participants in the grouped and distributed conditions, respectively; no Chinese course score difference was found, $t(46) = -1.50$, $p = 0.14$.

Design

A 2 × 4 mixed design was used with sequence (group vs. distributed) as a between-participants factor and encoding (handwriting, visual chunking, passive reading, and stroke-reporting) as a within-participant factor.

Stimuli

Forty-eight simplified Chinese characters were selected from the participants' Chinese textbooks and a Chinese Orthography Database (Chen et al., 2011). None of these characters had been taught in class before the experiment. These characters were derived from eight semantic radicals—女, 饣, 日, 火, 木, 心, 贝, and 钅—with six characters in one radical group. (See Appendix 1 for the 48 characters organised into eight sets in the grouped and the distributed conditions, respectively.) Across the eight radical groups, characters were matched by number of strokes (average: 10.11 strokes), number of chunks (average: 2.98 chunks), spatial layouts (left-right, or top-down), and frequency of English translations (Brysbaert & New, 2009). The key manipulation of the learning materials is the distribution of the 48 characters. For the grouped condition, six characters within one radical group compose a text; for the distributed condition, six characters from different radical groups compose a text. This manipulation led to eight texts for grouped and distributed conditions, respectively (see Appendix 2 for sample texts for each condition). The average length of texts was 46 characters and there was no length difference between the two conditions, $t(14) = -1.77$, $p = 0.10$.

Procedure

The experiment consisted of a pre-test session, a four-day learning session, a post-test session, and a two-week delayed test session. Throughout these sessions, both conditions were taught by an instructor who regularly lectured these students in their regular language curriculum. All sessions were conducted in the participants' regular class hours and in a classroom equipped with 24 personal computers.

Pre-test

Prior to learning, participants were asked to write pinyin with tone and meaning for any of the 48 characters that they recognised. No participant had correct responses on any of the 48 characters for pinyin or for meaning.

Learning Stage

For four consecutive days, participants learned two texts containing 12 characters. In each learning day, the instructor first guided the participants to read the new words containing these characters, and led them to comprehend each text by comprehension questions. Next, the participants continued to learn the 12 characters on computers with an average of 30 seconds for each character. They used one encoding method for each learning day. A Latin square arrangement was used to balance the encoding order (see Table 1), such that after four days all participants had used each encoding method, but on different days for different characters.

Table 1. Latin square arrangement used for balancing the encoding order.

	Day 1	Day 2	Day 3	Day 4
Participant 1	R	C	S	W
Participant 2	W	R	C	S
Participant 3	S	W	R	C
Participant 4	C	S	W	R

Notes: C = visual chunking, R = passive reading, S = stroke reporting, W = handwriting.
This design applied to both participants in the characters-in-group and the characters-distributed conditions, while the sets of characters in each cell differed for participants in different conditions.

The learning procedure for each character was divided into a 3000 ms observation time and an approximately 6000 ms practice time. In the observation time, for the first 1000 ms, the participants saw a character in the centre of a computer screen; for the second 1000 ms, they saw that character's pinyin with its tone and listened to a recording of its pronunciation; for the final 1000 ms, they saw the English translation of the character. In the practice time, the participants focused on encoding the character form with an assigned method. The period of practice time varied based on the number of chunks of each character (800 ms per chunk), and this time was constant across encoding conditions. Specifically, in the chunking condition, participants twice saw a computerised presentation of how a character was composed by chunks one at a time. In the reading condition, each character was presented for passive viewing at an exposure duration equated to that of the chunking condition. In the writing condition, participants were asked to write the character from memory as completely as possible for as many times as they could within the given 6000 ms. In the stroke-reporting condition, participants were asked to write down the number of strokes while viewing the character; next, they were asked to write down a particular stroke corresponding to the stroke number shown on the screen. This dual task served as a baseline condition. The entire procedure, including the initial 3000 ms exposure and the 6000 ms practice time, was repeated three consecutive times for each character.

Testing Stage

Immediate Test (Days 1–4) of Character Learning

After every learning stage, the participants' orthographic learning was assessed in a lexical decision task, computerised with E-Prime software (Schneider, Eschman, & Zuccolotto, 2002). In this task, the participants judged whether the stimulus presented on the screen was a real character by pressing one of two buttons as quickly and as accurately as possible. Each stimulus was presented for 1000 ms followed by a blank interval of 3000 ms. The stimuli included three types of characters: 24 familiar characters that did not include any of the eight radicals taught in this experiment, 48 target characters from the learning stage, and 24 novel characters that included the key radicals from the learning stage. For each stimuli type, one-half stayed as real characters and one-half were created as non-characters. For familiar and target characters, the non-characters were created by deleting a stroke or adding a stroke at the beginning, the middle, or the end of characters. For novel characters, the non-characters were created by switching the position of radicals to impermissible positions. The stimuli were divided into four testing sessions for each instructional condition and the different types of stimuli were presented randomly. The participants' responses on the learning materials were analysed.

Post-test (Day 5) of Production

On the fifth day, the participants underwent a rehearsal and then completed a character production task. In the rehearsal, 48 characters were reviewed one at a time with six seconds for one character. For the first three seconds, the character form, pinyin, and meaning were presented for one second each alone; for the last three seconds, all three remained in view together. Participants were instructed to integrate the orthographic, phonetic, and semantic information of each character. They experienced the same introduction order as they received in the learning phase.

After the rehearsal, participants were given 10 minutes to complete a paper-and-pencil character production task, which tested the participants' ability to recall a learned character from memory. In this task, participants were asked to write a character given its written English translation and were encouraged to write any orthographic features of the character that they remembered. The order of 48 test items on the test sheets was counterbalanced.

Delayed Test (Two Weeks) of Production

After two weeks, participants carried out, without rehearsal, the character production task of the post-test session. The order of test items was counterbalanced.

Character Scoring Scheme

Given that Chinese orthography consists of four units—stroke, chunk, layout, and character—participants' writing responses were scored according to the following schemes (see Appendix 3 for each scoring formula and authentic examples). All scores were calculated based on the proportion of correct responses. The advantage of using the proportion is to provide a continuous scale for measurement and to accommodate all types of orthographic units from simple to complex. For scoring the character responses, one researcher coded the entire set of responses and a native Chinese speaker followed the scoring scheme to code one-half of the responses. In what follows, we introduce each scoring scheme and then report its inter-rater reliability.

Stroke Scoring. A stroke is a basic writing pattern that can be finished without a pen leaving the paper (Chinese Language Committee, 1997). There are five basic strokes that generate up to 44 additional derived strokes (Wang, 2011). Stroke scoring is concerned with the correct reproduction of the stroke's shape, regardless of the compositional relations of strokes with one another. For scoring a character with the stroke scheme, the denominator is the character's total number of strokes and the numerator is the number of correct strokes in the response. Incorrect responses are defined as adding, deleting, or changing strokes and result in no score. The score reflects the participants' representation concerning the lowest-level building blocks of a character form. The inter-rater reliability of the stroke scoring was 0.996.

Chunk Scoring. A chunk is a basic orthographic unit that is composed of more than one stroke. The number of chunks varies based on the sample size of characters; sampled from 6097 frequent characters, there are 439 chunks (Chen et al., 2011). The composition of chunks follows certain positional constraints; a chunk that is in an impermissible position leads to a non-character. Chunk scoring is concerned with not only the orthographic form of chunks but also their positional information. For scoring a character with the chunk scheme, the denominator is the character's total number of chunks and the numerator is the number of correct chunks appearing in their correct positions. Correct forms of chunks are described by the Chinese Orthography Database (Chen et al., 2011). Incorrect responses are defined as adding, deleting, or changing chunks, and having the correct chunk form appearing in an impermissible position. In scoring, a single and correct chunk form, without any positional information, would count as one correct response. The score reflects the participants' representation concerning intermediate-level units and their permissible positions of a character. The inter-rater reliability of the chunk scoring was 0.950.

Layout Scoring. A layout is a positional constraint that can form a hierarchy of spatial information within a square. There are six basic layouts: single, half-split, two-quarter circled, three-quarter circled, enclosed, and converging (Chen et al., 2011). These layouts can combine with each other to form a composite structure. Layout scoring is concerned with the correct spatial hierarchy within a square, disregarding whether the composed units are correctly written. For scoring a character with the layout scheme, the denominator is the number of sub-layouts defined by the Chinese Orthography Database and the numerator is the number of correct sub-layouts composed by the written responses. Incorrect responses are defined as sub-layouts that do not fit those of the target character. For example, a response with a top-down layout for a left–right character would be counted as zero. The score reflects the participants' representation concerning spatial arrangements of any units within a character. The inter-rater reliability of the layout scoring was 0.995.

Character Scoring. Character scoring recognises only completely correct characters as correct. For scoring a character with the character scheme, the denominator is one and the numerator is either one (i.e., a fully correct written form) or zero (i.e., any responses different from the correct character). The inter-rater reliability of the character scoring was 1.

These various scoring schemes are sensitive to degrees of correctness and completeness of the learner's representation of Chinese characters. They allow a range of measures that can capture partial learning as well as complete learning.

Results

First, we report the influence of sequence and encoding on orthographic representation in the learning phase. For the lexical decision task, we performed a 2×4 mixed analysis of variance with sequence as a between-participants variable and encoding as a within-participant variable. We analysed accuracy instead of reaction time because our participants were beginning learners of Chinese. We found no effects of sequence, $F(1,46) = 0.42$, $p = 0.52$, $\eta_p^2 = 0.01$, or encoding, $F(3,138) = 1.62$, $p = 0.19$, $\eta_p^2 = 0.03$, and no interaction of these factors, $F < 1$.

Next, we performed a 2 (sequence) × 4 (encoding) repeated-measures multivariate analysis of variance on all character production tasks. When the difference reached significance, pairwise comparisons with Bonferroni adjustment are reported. Table 2 presents the descriptive statistics for all dependent measures across testing times for the two groups of participants. Below we report the results measured by the four scoring schemes first at post-test and then the delayed test.

Table 2. Means (standard errors) of accuracy for each instructional group on the lexical decision task and the character production task at different testing points.

		Characters-in-group				Characters-distributed			
		R	C	S	W	R	C	S	W
Lexical immediate decision		0.72	0.73	0.76	0.71	0.70	0.68	0.74	0.69
		(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)
Character production	Post	0.15	0.26	0.14	0.11	0.13	0.15	0.12	0.17
(character scoring)		(0.03)	(0.04)	(0.03)	(0.02)	(0.03)	(0.04)	(0.03)	(0.02)
	Delayed	0.07	0.09	0.07	0.09	0.06	0.07	0.05	0.06
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Character production	Post	0.33	0.37	0.25	0.22	0.25	0.26	0.19	0.30
(layout scoring)		(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
	Delayed	0.22	0.17	0.12	0.16	0.15	0.11	0.09	0.11
		(0.04)	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.02)
Character production	Post	0.30	0.38	0.24	0.21	0.25	0.25	0.22	0.28
(chunk scoring)		(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)
	Delayed	0.17	0.17	0.13	0.17	0.14	0.11	0.09	0.11
		(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Character production	Post	0.38	0.43	0.29	0.26	0.29	0.28	0.24	0.32
(stroke scoring)		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
	Delayed	0.25	0.21	0.18	0.20	0.19	0.15	0.12	0.14
		(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)

Notes: C = visual chunking, R = passive reading, S = stroke reporting, W = handwriting.

Post-test: Four scoring schemes

In the post-test, we found in the grouped condition that visual chunking consistently led to superior results in learners' orthographic production in all the four scoring schemes. In the distributed condition, writing was beneficial relative to the stroke-counting baseline, but only when we used the layout scheme.

Using the character scoring scheme, we found that chunking led to the highest level of character production in the grouped condition. The sequence × encoding interaction was significant, Pillai's Trace = 0.26, $F(3,44) = 5.02$, $p < 0.01$, $\eta_p^2 = 0.26$, with no significant difference between sequences, $p = 0.46$, and a significant difference among encoding methods, Pillai's Trace = 0.22, $F(3,44) = 4.07$, $p = 0.011$, $\eta_p{}^2 = 0.22$. A significant effect of encoding methods was present in the grouped condition, Pillai's Trace = 0.32, $F(3,44) = 6.84$, $p < 0.01$, $\eta_p{}^2 = 0.32$, but not the distributed condition ($p = 0.09$). In the grouped condition, chunking led to higher accuracy than passive reading, stroke-reporting, or handwriting ($p = 0.008$, $p = 0.005$, $p < 0.001$, respectively); no other effects were found.

Using the layout scoring scheme, encoding effects were present for both the grouped condition, Pillai's Trace = 0.26, $F(3,44) = 5.09$, $p < 0.01$, $\eta_p{}^2 = 0.26$, and the distributed condition, Pillai's Trace = 0.18, $F(3,44) = 3.18$, $p = 0.03$, $\eta_p{}^2 = 0.18$. Chunking again was most effective in the grouped condition, differing significantly from stroke-reporting and handwriting ($p = 0.02$, $p = 0.01$, respectively); in the distributed condition, writing was most effective, although statistically it differed significantly ($p = 0.02$) only from the stroke-counting, which was least effective.

Using the chunk scoring scheme, chunking was again more beneficial than stroke-reporting and handwriting, whereas other pairs of comparisons were not significant. This conclusion follows from the statistical analyses that showed no main effect of sequence, but a main effect of encoding, Pillai's Trace = 0.19, $F(3,44) = 3.44$, $p = 0.03$, $\eta_p^2 = 0.19$, and a significant sequence × encoding interaction, Pillai's Trace = 0.26, $F(3,44) = 5.02$, $p < 0.01$, $\eta_p^2 = 0.26$. This interaction reflected an encoding effect for the grouped condition, Pillai's Trace = 0.32, $F(3,44) = 6.93$, $p < 0.01$, $\eta_p^2 = 0.32$, but not the distributed condition, Pillai's Trace = 0.09, $F(3,44) = 1.53$, $p = 0.22$, $\eta_p^2 = 0.09$.

Using the stroke scoring scheme, we again witnessed the advantage of chunking under the grouped condition. The statistical analyses showed an encoding effect, Pillai's Trace = 0.22, $F(3,44) = 4.21$, $p = 0.01$, $\eta_p^2 = 0.22$, no sequence effect ($p = 0.22$), and a significant sequence × encoding interaction, Pillai's Trace = 0.23, $F(3,44) = 4.42$, $p < 0.01$, $\eta_p^2 = 0.23$. Again, the encoding effect was restricted to the grouped condition, Pillai's Trace = 0.32, $F(3,44) = 6.78$, $p < 0.01$, $\eta_p^2 = 0.32$, where chunking outperformed stroke-reporting and writing ($p = 0.003$, $p = 0.001$, respectively); no other pairs of comparison were found.

Delayed Test: Four scoring schemes

In the delayed test, we observed that the grouped sequence led to better memory retrieval than the distributed sequence in the layout and the stroke scoring schemes, while we did not find other differences.

Using the character scoring scheme, there were no significant effects of sequence ($p = 0.20$), encoding method ($p = 0.64$), or their interaction ($p = 0.88$). Using the layout scoring scheme, the interaction was not significant ($p = 0.88$). However, there was a significant difference between sequences, $F(1,46) = 4.30$, $p = 0.04$, $\eta_p^2 = 0.09$, with the grouped condition having higher accuracy than the distributed condition ($p = 0.04$). There was also a significant difference among encoding methods, Pillai's Trace = 0.22, $F(3,44) = 4.18$, $p = 0.01$, $\eta_p^2 = 0.22$, with passive reading outperforming stroke-reporting ($p = 0.02$), while no other difference among encoding methods was found. Using the chunk scoring scheme, we found the same results as using the character scoring scheme: there was no interaction ($p = 0.86$); neither the main effect of sequence ($p = 0.07$) nor the main effect of encoding method ($p = 0.13$) reached significance. In contrast, using the stroke scoring scheme, we observed the beneficial effect of grouped sequence we found with the layout scoring scheme. The interaction was not significant ($p = 0.99$), while the difference of sequences reached significance, $F(1,46) = 4.09$, $p = 0.04$, $\eta_p^2 = 0.08$, with the grouped condition outperforming the distributed condition ($p = 0.04$). The difference among encoding methods was also significant, Pillai's Trace = 0.17, $F(3,44) = 3.08$, $p = 0.04$, $\eta_p^2 = 0.17$, but marginal comparisons with Bonferroni adjustment did not find any significant difference in any pairs.

Discussion

The study represents the first simultaneous test of multiple instructional methods for supporting orthographic learning of Chinese characters by beginning CFL learners. Each of these methods draws the learner's attention to the form of the character, but varies in how this is done. An important feature of the study is its *in vivo* design; we conducted a laboratory-style multi-condition experiment in real but well-controlled classroom settings, a strategy that helps narrow the gap between the results of learning research and educational practice (Koedinger et al., 2012).

By combining two methods of sequencing the characters to be learned with four methods for the learners to study the characters, we first demonstrate in the grouped sequence that visual chunking strengthens learners' orthographic memory better than the other three encoding methods. Further, we found that a grouped sequence enhances character production relative to a distributed sequence, as seen by better recall in the layout and the stroke scoring schemes in the delayed test. Collectively, the findings indicate that grouped sequence and visual chunking encoding help with CFL beginning learners' memory of orthographic information in character learning.

The beneficial effect of the grouped sequence in comparison with the distributed sequence can be attributed to the emphasis of recurring radicals in the grouped condition. Xu, Chang, and Perfetti (in press) showed this effectiveness of presenting repeated radicals close together to strengthen the form–meaning and form–sound links and to further bootstrap the generalisation of radical knowledge. These results are in line with previous studies. Prior research on first-language orthographic learning (Chang & Han, 2004) reported that the grouped approach increased second-grade elementary school students' orthographic knowledge more than the distributed approach after eight-week courses. One study on adult L2 orthographic learning (Chen et al., 2013) also suggested that chunk-derived character learning enriched CSL learners' orthographic knowledge of taught chunks after three-week courses.

Our study went beyond previous studies in implementing an *in vivo* procedure in which a single instructor followed a standardised instructional plan to teach two sets of texts, and in which the encoding methods were presented in balanced orders, thus removing typical confounds that occur in instructional experiments between instructors' characteristics and instructional methods.

Although our results revealed the advantage of the grouped sequence in the delayed test, no differences were observed in the lexical decision task in the immediate test or the production task in the post-test. This delayed effect may suggest an underlying mechanism of learning and retention—instructional effects might be delayed because they boost long-term memory more than initial learning (Bloom & Shuell, 1981; Cao et al., 2013b). Our results suggest that learning characters presented in a grouped sequence can strengthen the orthographic constituents over a period of two weeks.

As for the comparative effectiveness of encoding methods, we did not find any effect in the lexical decision task; nevertheless, we consistently found the advantage of visual chunking over the other three methods in the post-test. The results contrast with the writing benefit reported in previous studies (Cao et al., 2013a; Guan et al., 2011; Xu et al., 2013). Note that prior research always introduced materials sharing the same chunks in a distributed fashion. In the grouped condition of this study, however, we found visual chunking to be the most effective encoding method that resulted in more accurate recall through the emphasis of orthographic chunks. This finding is actually in line with previous developmental research, which reports that Chinese-speaking children are capable of applying visual chunking strategies in character production and that their perceptual ability on processing chunks is associated with reading proficiency (Pak et al., 2005). Moreover, there are teaching implications from our significant results of the visual chunking encoding. They suggest that when characters are taught in a curriculum with enough emphasis on their shared chunks, the learners' visual–orthographic representation of characters can be strengthened because attention is drawn to such decomposed structures and smaller components of a character (Cao et al., 2013b). In other words, learners' familiarity and acquisition of chunks can be used as a crutch in character learning.

On the other hand, for characters introduced in a distributed sequence, results showed that the handwriting was only effective based on the layout scoring scheme.

Compared with previous studies, this advantage under the distributed sequence was relatively small. One possible explanation is that, in this *in vivo* design, the practice time for each character was restricted to only 30 seconds to represent the practical situations in a classroom, as compared with an average of 360 seconds in previous studies (Cao et al., 2013a, 2013b; Guan et al., 2011; Xu et al., 2013). It is also likely that previous studies instructed their participants to write identical characters on consecutive days, which allowed encoded memory to be consolidated after sleep (Walker & Stickgold, 2006). In this study, the participants only experienced handwriting once and they received a memory recall task right after rehearsal. This one-time practice may not leave enough time for memory consolidation, and for handwriting to be effective requires time to practice.

Finally, it is also worth noting that the four scoring schemes of writing responses revealed different patterns of instructional effect. Specifically, the layout and the stroke scoring schemes, not the character or chunk schemes, show that the grouped condition outperformed the distributed condition in the delayed test. These patterns suggest that scoring based on layouts and strokes might better reflect learners' orthographic representation, especially when the learners are new to the Chinese orthographic system. For instance, for a two-chunk character, a novice may write two partial chunks and put them in the correct layout. If the scoring scheme is strict (i.e., character or chunk scheme), the novice would receive zero credit for his or her effort. However, one should recognise that this novice's performance is qualitatively different from another novice who could not produce any strokes. Thus, we suggest that using a combination of the four scores can help instructors and researchers better evaluate a novice's recall accuracy on different levels of orthographic information—from strokes to spatial layouts, then to chunks, and finally to the whole characters. This comprehensive measure reflects the extent to which the learners can recall and reproduce the character form and which level of orthographic knowledge the leaner still needs to gain.

To conclude, this study examined how instructions focused on orthographic form support the learning of Chinese characters at the beginning of study by CFL learners. The contribution of this study is threefold. Theoretically, it addresses the debate about the effectiveness of learning sequence—that is, whether grouped versus distributed presentation of characters sharing a radical is more effective—further demonstrating that a grouped arrangement is more effective for the recall and production of characters, although not necessarily for their recognition. Methodologically, the study provides a comprehensive character scoring system to reveal distinctive information of learners' acquired orthographic representations. Practically, this study examined instructional strategies for encoding characters, finding a synergy effect of visual chunking with a grouped sequence as well as a small but significant effect of handwriting with a distributed sequence. Overall, our results demonstrate that, for beginning CFL/CSL learners in authentic classroom settings, visual chunking encoding for learning characters shares the same chunk results in improved recall, and a grouped sequence supports learners to establish more robust orthographic representations than a distributed sequence.

Acknowledgements

The authors are grateful to Ms Fan Fan for conducting the classes, Lidia Zacharczuk, Jon-Michel Seman, and the staff in the Language Lab of the University of Pittsburgh for their assistance with various tasks, and all participating students for their cooperation.

Funding

This research was funded from the National Science Foundation (grant #SBE-0836012) through the Pittsburgh Science of Learning Center and from the "Aim for the Top University Project" of National Taiwan Normal University. No restrictions have been imposed on free access to, or publication of, the research data. Opinions reflect those of the author(s) and do not necessarily reflect those of the funding agency(ies). The author (s) had no financial or other conflicts of interest.

Note

1. A chunk, or *bùjiàn* (部件), is a basic orthographic unit composed of more than one stroke. Although chunks may coincide in part with radicals and several studies referred ambiguously to radicals as chunks (for example, Taft & Zhu, 1997, pp. 761–762), they are not the same. Some radicals can be further decomposed into smaller chunks, such as the radical 音, which can be decomposed into chunk 立 and chunk 日. Moreover, radicals usually provide semantic information (i.e., semantic radicals or *bùshǒu*, 部首) or phonetic information (i.e., phonetic radicals), whereas chunks do not necessarily convey functional information. Furthermore, in some cases, a chunk can also be a radical. For example, character 婚 (hūn; marriage) has three chunks (女, 氏, and 日), while chunk 女 (nǚ; female) is also the semantic radical of character 婚. Given that a chunk is an intermediate-level orthographic unit in a character (e.g., the hierarchical model of lexical activation in Taft & Zhu, 1997), assessing the role of chunks in addition to radicals is one aspect in which this study wishes to complement the existing literature. In the present study, we use "chunks" that are defined by the Chinese Orthography Database (Chen, Chang, Chiou, Sung, & Chang, 2011).

References

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research, 74*, 245–248.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 488–496. doi:10.3758/BRM.41.4.977

Cao, F., Vu, M., Chan, H., Lawrence, J., Harris, L., Guan, Q., ... Perfetti, C. A. (2013a). Writing affects the brain network of reading in Chinese: An fMRI study. *Human Brain Mapping*, 1–15. doi: 10.1002/hbm/22017

Cao, F., Rickles, B., Vu, M., Zhu, Z., Chan, D. H. L., Harris, L., ... Perfetti, C. A. (2013b). Early stage visual-orthographic processes predict long-term retention of word form and meaning: A visual encoding training study. *Journal of Neurolinguistics, 26*, 440–461. doi:10.1016/j.jneuroling.2013.01.003

Chang, S. J., & Han, M. J. (2004). The effects of different word recognition teaching methods for lower grade elementary school students. *Educational Review, 22*, 71–88.

Chen, H. C., Chang, L. Y., Chiou, Y. S., Sung, Y. T., & Chang, K. E. (2011). Construction of Chinese orthographic database for Chinese character instruction. *Bulletin of Educational Psychology, 43*, 269–290.

Chen, H. C., Hsu, C. C., Chang, L. Y., Lin, Y. C., Sung, Y. T., & Chang, K. E. (2013). Using radical-deriving e-learning platform to increase Chinese learners' Chinese knowledge. *Language Learning & Technology, 17*, 89–106.

Chinese Language Committee. (1997). *Regulation of BiShun for modern general-purpose Hanzi* (*Xian Dai Han Yu Tong Yong Bi Shun Gui Fan*). Beijing: General Administration of Press and Publication.

Chinese Language Committee. (2009). *Modern Chinese common character list* (*Hanyu Tongyong Zibiao*). Beijing: Commercial Press.

DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems*. Honolulu, HI: University of Hawaii Press.

Everson, M. (1998). Word recognition among learners of Chinese as a foreign language: Investigating the relationship between naming and knowing. *Modern Language Journal, 82*, 194–204. doi:10.1111/j.1540-4781.1998.tb01192.x

Guan, C. Q., Liu, Y., Chan, D. H. L., Ye, F., & Perfetti, C. A. (2011). Writing strengthens orthography and alphabetic-coding strengthens phonology in learning to read Chinese. *Journal of Educational Psychology, 103*, 509–522. doi:10.1037/a0023730

Koedinger, K. R., Aleven, V., Roll, I., & Baker, R. (2009). In vivo experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 383–412). New York, NY: Routledge.

Koedinger, K. R., Corbett, A. C., & Perfetti, C. (2012). The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36*, 757–798. doi:10.1111/j.1551-6709.2012.01245.x

Lam, H. C. (2011). A critical analysis of the various ways for teaching Chinese characters. *Electronic Journal of Foreign Language Teaching, 8*, 51–70.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30*, 7–22. doi:10.1111/j.1467-9817.2006.00329.x

Pak, A. K. H., Cheng-Lai, A., Tso, I. F., Hua, S., Wenling, L., & Anderson, R. C. (2005). Visual chunking skills of Hong Kong children. *Reading & Writing, 18*, 437–454. doi:10.1007/s11145-005-6575-3

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research, 46*, 4646–4674. doi:10.1016/j.visres.2006.04.023

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review, 12*, 43–59. doi:10.1037/0033-295X.112.1.43

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Si, X. (2001). On characters-distributed instructional approach. *Curriculum, Teaching Material, and Method, 2*, 1–5.

Su, C. J. (2006). On grouped-characters instructional approach. *Selected Papers on Scientific Word Recognition Education,* 332–336.

Taft, M., & Zhu, X. (1997). Submorphemic processing in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 761–775. doi:10.1037/0278-7393.23.3.761

Walker, M. P., & Stickgold, R. (2006). Sleep, memory, and plasticity. *Annual Review of Psychology, 5*, 139–166. doi:10.1146/annurev.psych.56.091103.070307

Wan, Y. Y. (1991). Psychological features and instruction for children character learning. In Z. F. Yang & S. R. Gao (Eds.), *Chinese people and Chinese mind—Development and instruction* (pp. 403–448). Taipei, Taiwan: Yuan-Liou Publishing.

Wang, Z. G. (2011). *Questions and answers on modern Chinese character*. Beijing: Peking University Press.

Xu, Y., Chang, L. Y., & Perfetti, C. A. (in press). The effect of radical-based grouping in character learning in Chinese as a foreign language. *The Modern Language Journal*.

Xu, Y., Chang, L.-Y., Zhang, J., & Perfetti, C. A. (2013). Reading, writing, and animation in character learning in Chinese as a foreign language. *Foreign Language Annals, 46*, 423–444. doi:10.1111/flan.12040

Zhong, J. (1990). A report of fifteen years of Chinese language instruction. In Y. Sheng (Ed.), *Selected papers on Chinese as foreign language instruction* (pp. 66–67). Beijing: Beijing Yu Yan Xue Yuan Press.

Appendix 1. Learning materials (48 characters) organised into eight sets, in characters-in-group and characters-distributed presentation sequences, respectively

Material sets in the characters-in-group condition			Material sets in the characters-distributed condition		
Set number	Presentation	Character	Set number	Presentation	Character
	A1	贩		G4	饱
	A2	账		D1	棚
1	A3	败	1	B5	烟
	A4	赔		H1	忽
	A5	贼		C3	钻
	A6	赚		E1	婚
	B1	炉		E4	娃
	B2	烘		F2	旷
2	B3	焰	2	B6	烛
	B4	炮		A3	败
	B5	烟		H5	忍
	B6	烛		C5	钥
	C1	银		D2	桶
	C2	钞		C6	钓
3	C3	钻	3	F3	晒
	C4	锁		A2	账
	C5	钥		H2	患
	C6	钓		G6	饶
	D1	棚		E6	姑
	D2	桶		C1	银
4	D3	梯	4	D6	概
	D4	板		F1	晌
	D5	材		G5	蚀
	D6	概		A4	赔
	E1	婚		B4	炮
	E2	嫁		F6	暗
5	E3	媳	5	C4	锁
	E4	娃		D3	梯
	E5	娇		H3	急
	E6	姑		A5	贼
	F1	晌		E2	嫁
	F2	旷		A1	贩
6	F3	晒	6	C2	钞
	F4	晾		G2	馅
	F5	暄		F4	晾
	F6	暗		B2	烘
	G1	馒		E5	娇
	G2	馅		G3	馋
7	G3	馋	7	H4	忌
	G4	饱		A6	赚
	G5	蚀		D4	板
	G6	饶		B3	焰
	H1	忽		E3	媳
	H2	患		G1	馒
8	H3	急	8	B1	炉
	H4	忌		D5	材
	H5	忍		F5	暄
	H6	愁		H6	愁

Note: For each condition, there are eight sets of characters organised in grouped or distributed fashion. In the Presentation column, the capital letter denotes the chunk and the number denotes the characters. The key manipulation is that, in each set of grouped presentation, all six characters share the same chunk, while, in each set of distributed presentation, no characters share the same chunk.

Appendix 2. Sample texts taught in the characters-in-group condition (G8) and in the characters-distributed condition (D8) with the target character highlighted and underlined

Text G8
王朋喜欢吃鱼，最近他忽然患病了，还要吃鱼。李友急了，说，"你生病了，要忌口，忍一下"。王朋发愁了。李友说，"我做菜，请你吃吧"。
Pinyin
Wáng Péng xǐhuān chī yú, zuìjìn tā hūrán huànbìngle, hái yào chīyú. Lǐ Yǒu jíle, shuō, "Nǐ shēngbìngle, yào jìkǒu, rěn yīxià." Wáng Péng fāchóule, Lǐ Yǒu shuō, "Wǒ zuòcài, qǐng nǐ chī ba."
English Translation
Wang Peng likes to eat fish. Recently, he suddenly contracted an illness but he still wanted to eat fish. Li You became anxious and said, "You are sick. You should avoid eating fish and endure this for a while." Wang Peng became worried. Li You said "Let me cook and treat you."

Text D8
白家的儿媳小月常常做馒头。因为她用火炉和很好的材料，所以很好吃。
白英爱和小月寒暄， 她说：「我喜欢你的馒头，要是没有了，就发愁。」
Pinyin
Báijiā de érxí Xiǎoyuè chángcháng zuò mántou, Yīnwèi tā yòng huǒlú hàn hěnhǎo de cáiliào, suǒyǐ hěnhǎochī. Báiyīngài hàn Xiǎoyuè hánxuān, tāshuō, "Wǒ xǐhuān nǐde mántou, yàoshì méiyǒule, jiù fāchóu."
English Translation
The daughter-in-law of the Bai family, Xiaoyue, makes steamed buns often. Because she uses the stove and very nice materials, the buns are good to eat. When Bai Yingai and Xiaoyue exchanged conventional greetings, she said "I like your steamed buns and I am worried if I don't have any."

Appendix 3. Character scoring scheme for each written response

Orthographic unit		Stroke	Chunk	Layout	Character
Scoring ratio	Numerator	Number of correct strokes	Number of correct chunks	Number of correct sub-layouts composed by the written response	0 = wrong character 1 = correct character
	Denominator	Total number of strokes	Total number of chunks	Number of sub-layouts defined by the Chinese Orthography Database	1
Definition of incorrectness		A. Delete stroke: no score B. Add stroke: no score	A. Delete chunk: no score B. Add chunk: no score C. Correct chunk form but in wrong position: no score	Sub-layouts composed by the written response do not fit those of the target character: no score	Any incorrect stroke or incorrect radical will result in an incorrect character
Note		1. ONLY focus on the shape of correct strokes 2. NOT consider positional correctness of strokes	1. Consider BOTH correct form AND permissible position of a chunk 2. A single and correct chunk form, without any positional information, would count as one correct response since it is written accurately	1. ONLY focus on the spatial structure correctness 2. NOT consider if the form of stroke or radical is correct or not	

Example scoring

Target character (number below is a denominator for each scheme)					Written responses and their properties (number below is a numerator for each response)					Scores for every written response (number below is a scoring ratio for each response)			
	Stroke	Chunk	Layout	Character		Stroke	Chunk	Layout	Character	Stroke	Chunk	Layout	Character
枷	9	3	3	1	枷	9	3	3	1	1	1	1	1
					珈	7	2	3	0	0.78	0.67	1	0
					蜘	8	1	3	0	0.89	0.33	1	0
					楞	9	1	2	0	1	0.33	0.67	0
					枸	9	1	2	0	1	0.33	0.67	0
					架	9	0	2	0	1	0	0.67	0
					茄	6	0	2	0	0.67	0	0.67	0
					加	5	2	2	0	0.56	0.67	0.67	0
					叻	5	0	2	0	0.56	0	0.67	0
					另	5	0	0	0	0.56	0	0	0
					召	5	0	0	0	0.56	0	0	0
					呆	7	0	0	0	0.78	0	0	0
					木	4	1	0	0	0.44	0.33	0	0
					啉	8	0	3	0	0.89	0	1	0
					???	0	0	3	0	0	0	1	0

Note: Participants used ?, ○, □, or ✗ different symbols to represent that they knew certain chunks should be written in certain positions while they were not able to write the correct form of the chunk. These symbols were seen as incorrect chunk forms in scoring with the layout scheme.