SCIENTIFIC STUDIES OF READING, *8*(4), 357–379

The Implicit and Explicit Learning of Orthographic Structure and Function of a New Writing System

Min Wang
University of Maryland, College Park

Ying Liu and Charles A. Perfetti
University of Pittsburgh

Two experiments were carried out to examine how adult readers of English learn to acquire the orthographic structure and function of Chinese characters selected from reading material in their first-semester college course in Chinese. The first experiment, an online lexical decision task, demonstrated that the learners quickly acquired knowledge about the orthographic structure of the characters. With only limited vocabulary and without explicit instruction, they were sensitive to the characters' curriculum frequency and internal compositional features. Overall, these results were consistent with our previous findings (Wang, Perfetti, & Liu, 2003). The second experiment further tested the learners' implicit and explicit learning of the orthographic component function of Chinese characters. In an offline unknown character identification task, these learners showed difficulty in making use of the functional cue of the known semantic radical without any probing. However, with probing they identified visually the majority of target semantic radicals. In addition, they demonstrated implicit knowledge of the meaning cue of the semantic radical by making more appropriate meaning inferences on high-frequency radicals than on low-frequency ones. After receiving explicit instruction, the students extracted significantly more meaning information from known semantic radicals, particularly low-frequency ones, compared to their performance prior to such instruction.

In learning Chinese, adult speakers of English face several challenges. Aside from the extra effort required by adult learners of a second language, compared with

Requests for reprints should be sent to Min Wang, Department of Human Development, University of Maryland, 3304P Benjamin Building, College Park, MD 20742. E-mail: minwang@umd.edu

children, there is the challenge of learning a new writing system that contrasts sharply with their native alphabetic writing system. These contrasting features indeed lead to some differences in how reading works in Chinese compared with English (Leong, 1997; Perfetti, Liu, & Tan, 2002).

Learning to read Chinese involves mapping between spoken forms and print forms of the language just like learning to read other languages in the world. However, in Chinese, the learner faces a rich and complex orthographic structure expressed in a nonlinear form. The basic unit, the character, maps onto a syllabic morpheme and is pronounced as a syllable including onset, rime, and tone. Each character consists of interwoven strokes. There are 24 basic strokes: 一, 丨, 丿, and 丶 are the most common. Strokes are combined to form component radicals, for example, 工 and 扌. The combination of strokes has certain positional constraints. Among all radicals, some are single characters on their own. For example, 工 has a meaning of 'work' and is pronounced /gong/1 (the number 1 indicates the tone of the pronunciation). Some are not real characters on their own (e.g., 扌 and 氵) and have no pronunciations. However, they may sometimes provide meaning cues for the whole characters. Two categories based on the characters' structural complexity are present: simple characters and compound characters. Simple characters are those that have one single radical such as 工.

Of special importance for the learner are the more common compound characters, those that contain two or more distinct radical components, for example, 扛 (meaning 'carry with shoulder,' pronounced as /kang/2). The configuration of the radicals in compound characters is normally either a left–right or top–bottom structure. Each component of a compound character has its own fixed position, or legal position, either left, right, top, or bottom. For example, 扌 always occurs at the left position of a character, as in 打, 拍, 挑, and 拣 (meaning 'hit,' 'pat,' 'choose,' and 'pick' and pronounced /da/3, /pai/1, /tiao/1, /jian/3, respectively). Presence of this radical on the right side of the character would violate its positional constraint, and the character would be considered illegal.

The majority of Chinese characters are compound characters (more than 80%) according to Zhu (1988). Radical components of compound characters are normally categorized into two types: semantic radicals (often labeled as radicals) and phonetic radicals (often labeled as phonetics). Semantic radicals are those components in characters that provide meaning information for the whole character. For example, 氵, the water radical, provides an important meaning cue for the characters 江 'river' (/jiang/1) and 海 'sea' (/hai/3), although they are pronounced differently. Phonetic radicals are those that provide sound cues for the whole character. For example, the radical 青 'green' is pronounced as /qing/1, and the two characters 请 'invite' (/qing/3) and 晴 'sunny' (/qing/2) sharing the radical 青 are pronounced similarly to 青 and differ only in tone. However, they have very different meanings. There are approximately 200 semantic radicals in Chinese. Many of them cannot stand alone as characters; therefore, they do not have pronunciations associated with them. Each semantic radical can form on average about 20 compound characters. It is important to point out that most Chinese semantic radicals provide useful meaning information to the whole character (these radicals are often called transparent semantic radicals). Although there are some cases where the semantic radical does not predict the meaning of the whole character, the relation between semantic radicals and whole character meaning is more reliable than the relation between phonetic radicals and whole character pronunciation (see Feldman & Soik, 1999, for a review).

Recent research has shown that orthographic information including the radical and its position is explicitly represented in a Chinese reader's lexicon not only for adult skilled readers (e.g., Peng, Li, & Yang, 1997; Taft, Zhu, & Peng, 1999) but also for young school children (e.g., Peng et al., 1997; Shu & Anderson, 1999). Furthermore, adult alphabetic learners of Chinese are also sensitive to this radical information. In a long-term study of learning to read Chinese by American college students, Wang, Perfetti, and Liu (2003) found the learners were sensitive to the structural complexity of characters. In a lexical decision task, they accepted simple characters faster and more accurately than compound characters. They were also sensitive to the compositional relationship of the character: They rejected noncharacters containing illegal radical forms faster and more accurately than those containing legal radical forms in illegal positions, which in turn were rejected faster and more accurately than those containing legal radical forms in legal positions.

In this article, we report results from a second cohort from this same population. Our main aim is to address the extent to which learners acquire the function of character components as well as the orthographic structure of the characters. It is quite possible that learners' rapid acquisition of orthographic structure is primarily a matter of perceptual learning that excludes function. On the other hand, it is likely that motivated learners acquire implicitly the functional values of components—for example, that one component often gives a cue to meaning—just as they implicitly learn the stroke forms of legal radicals and legal characters. To examine this question, we first had to establish that learners indeed acquired the orthographic structure of characters, as Wang et al. (2003) found. Thus the first of two studies described next is a replication of Wang et al. with a new cohort of student learners. The second study then examines the functionality question.

This functionality of components (radicals) is important for native speakers of Chinese. Both phonetic and semantic radicals have been shown to play an important role in learning to read Chinese characters (e.g., Chen, 1993; Chen & Allport, 1995; Fang, Horng, & Tzeng, 1986; Seidenberg, 1985; Zhu, 1987). Shu and Anderson (1997) showed that by Grade 3, Chinese children who are good readers are able to use the information in semantic radicals as assistance in learning and remembering new characters. This is especially true when the characters are less frequent or less familiar. Chan and Nunes (1998) used a creative spelling task to test

Chinese children's use of semantic and phonetic radicals to infer meaning and generate a pronunciation of the character. The children were encouraged to help a Chinese boy create new characters to name unfamiliar objects by selecting certain stroke patterns, either the semantic or the phonetic radicals. The results showed that children develop understanding and use of semantic radical components of the characters earlier than phonetic radicals. From age 6, the children could systematically employ information from semantic radicals (providing that they were frequent ones). Only after age 9 were children able to use phonetic information provided by phonetic radicals to read the invented characters. However, Ho and Bryant (1997) demonstrated that even Grade 1 Chinese children showed better performance in reading regular compound characters in which the phonetic components provided sound cues for the whole characters than in reading irregular characters. Taken together, these results suggest that utilizing the radical component function is fundamental in learning to read Chinese and children may develop sensitivity to semantic radicals earlier than phonetic ones. However, phonetic information may be used in early primary grades. In our study presented here, we were interested in investigating whether and how students learning to read Chinese as a second language in a college classroom develop this sensitivity.

To set the stage for the studies, we must characterize the nature of the college classrooms we studied. We carefully examined the textbook used, observed the lessons, and interviewed the instructors about their teaching methods. We found that students were taught neither the orthographic structure of Chinese characters nor the function of the character components. That is, the instructors did not explicitly teach the decomposition of characters into radicals and did not teach the positional constraints of radicals. The instructors also did not draw attention to the function of radicals as cues to pronunciation and meaning. The rationale for these instructional practices was a concern that because of the fairly low reliability of these cues, instruction on radical components would cause overgeneralization among beginning Chinese learners. Students learned to read and write the characters, mostly through copying and memorization, in sentences and written conversational context. Students were not tested on their knowledge of the characters' internal structure. Because Chinese was not used outside of the classroom, the curriculum was the major source of exposure to the new system. This unique learning environment provided an opportunity for examining implicit learning of some critical properties of a new writing system.

Learning to read via implicit procedures has been documented in the literature for alphabetic writing systems. Thompson and his colleagues (Fletcher-Flinn & Thompson, 2000; Thompson, Cottrell, & Fletcher-Flinn, 1996; Thompson, Fletcher-Flinn, & Cottrell, 1999) suggested that there are two knowledge sources in acquiring letter–phoneme correspondences in English reading acquisition. One source is implicit learning, involving children's self discovery of the letter–phoneme relationship as well as induction from their accumulated print experiences.

For example, children may acquire the association between the orthographic representation of the letter *-t* and the /t/ sound through various positional contexts of the right boundary of the visual-orthographic representation of words such as *get*, *cat*, *went*, and *got*. The other knowledge source is explicit learning, involving explicit instruction of these correspondences. Thompson and his colleagues successfully demonstrated implicit learning for very young children in learning to read, especially when an explicit instruction source is not available. Our notion of implicit and explicit learning is consistent with these two knowledge sources, and the purpose of our study is to show how alphabetic skilled readers acquire implicit knowledge of orthographic structure and function in learning to read Chinese.

EXPERIMENT 1

Wang et al. (2003) suggested that learning the structure of characters can occur despite the lack of instruction we described previously. Experiment 1 was a replication of their study with a new cohort of students, who would then serve in the next experiment on implicit learning of functionality. Thus, student learners in their 1st year of Chinese carried out lexical decisions on Chinese characters and noncharacters designed to expose the structure of characters.

Method

Participants. There were 15 participants in the study. Ten were attending the 1st-year Chinese language program at the University of Pittsburgh (Pitt), and 5 were attending the 1st-year Chinese language program at Carnegie Mellon University (CMU). None of them had formal Chinese learning experience. Fourteen students had English as their first language, and the remaining participant had German as a first language. This group of participants constituted a new cohort of students from the same programs at Pitt and CMU where we recruited our participants in Wang et al. (2003).

The Chinese language program at Pitt consisted of a spoken language course and a reading–writing course. There were ten 50-min sessions per week for a student who was registered for both the oral and the reading–writing course: 7 sessions devoted to oral language practice and 3 to reading–writing. The curriculum content was designed to provide students with cumulative speaking, reading, and writing experience. Both simplified and traditional versions of the characters were taught. The number of characters introduced in the first term was 263. Pinyin (a Roman alphabetic system) was taught to assist in reading Chinese characters. The CMU Chinese program had an integrated listening, speaking, reading, and writing curriculum. The characters were taught both individually and in the text context. As with the Pitt program, both simplified and traditional versions of the characters were taught. There were five 50-min sessions per week. The number of characters introduced in the first term was 230. The two curricula had130 characters in common. All students were tested at the end of their first term.

Materials and designs. There were 160 items—80 characters and 80 noncharacters. The characters and noncharacters were the same as in Wang et al. (2003). The complete set of materials of this experiment for the Pitt students is in the Appendix. The CMU students had a different set of real characters selected from their own curriculum; there were 21 characters in common between the Pitt and CUM set. Both groups used the same noncharacters. Characters were presented in the simplified (modern) style. Two variables were manipulated for the real characters: One was the frequency of the character defined as the number of appearances of the character in the textbooks of Pitt or CMU Chinese curriculum (high or low frequency). The mean count for high-frequency characters used in this task was 44.85 for Pitt students and 15.55 for CMU students. The mean for low-frequency characters was 9.68 for Pitt students and 4.8 for CMU students. The other variable was the structural complexity of the character, defined as whether the character is composed of a single radical or more than one radical (simple or compound character). The average number of radicals for compound characters was 2 for Pitt students and 2.28 for CMU students. The combination of these two variables resulted in four conditions: (a) high-frequency simple (HS) characters (e.g., 小), (b) high-frequency compound (HC) characters (e.g., 她), (c) low-frequency simple (LS) characters (e.g., 生), and (d) low-frequency compound (LC) characters (e.g., 机). Twenty characters of each type were selected from the curriculum. Stroke numbers were matched for HS and LS characters and for HC and LC characters to control for visual complexity when considering the frequency effect. Among HC characters, there were three characters containing phonetic radicals that provided pronunciation cues for the whole characters and four characters containing semantic radicals that provided some meaning cues for the whole characters. Among the LC compound characters, there were five characters containing phonetic radicals and five containing semantic radicals. There were no systematic differences between the HC and LC characters in terms of their phonetic regularity or semantic transparency. Due to the limited number of characters acquired in the first term of the two Chinese programs, very few of the simple characters used in this experiment appeared in the compound characters encountered by the students in their curricula. Therefore, we considered only the frequency of these simple characters when they stand alone. A combined frequency of the simple characters and their occurrence in compound characters should be taken into consideration in future studies with students having a larger reading vocabulary.

The noncharacters were constructed varying the legality of the radical forms and the legality of the radical positions. There were four conditions: (a) legal radicals in legal positions (LR-LP; e.g., 泪), (b) legal radicals in illegal positions (LR-ILP; e.g., [illegible]), (c) illegal radicals (ILR; e.g., [illegible]), and (d) visual symbols, (e.g., $, *). The ILR were generated by adding, deleting, or moving a stroke from one location to another within a legal radical. The visual symbols were taken as a baseline condition. The order of all the items was randomized for each participant.

Procedure. The experiment was implemented using E-Prime (Psychology Software Inc., Pittsburgh). The participant saw a fixation sign (+) first for 500 msec, followed by a target character until a response was given. The participant was instructed to press the left mouse button to indicate Yes and the right one to indicate No to make the lexical decision as quickly and accurately as possible.

Results and Discussion

The percentage of data loss as a result of data screening in each experimental condition was similar to Wang et al. (2003). These outliers were 6.7%, 5.3%, 4.7%, 5.7%, 2.0%, 4.0%, 4.7%, and 3.0% for HC, HS, LC, and LS characters and for LR-LP, LR-ILP, ILR, and visual symbols, respectively. The trimmed data underwent a log transformation to increase the normality and homoscedasticity of the reaction times (RTs) for the participants. The logRTs were taken as the dependent variable for the subsequent analyses.

Real characters: Yes responses. The decision time analysis was based on correct responses. The frequency variable consisted of two levels (high vs. low), and the structural complexity had two levels (simple vs. compound characters). The group (Pitt vs. CMU students) was treated as a between-participants variable. A Frequency × Structural Complexity × Group repeated measures analysis of variance (ANOVA) showed the following main results: The two student groups did not differ, $F(1, 13) = .33$, $p > .5$. The interactions between the group and the two stimulus variables (frequency and structural complexity) were not significant either, both Fs < 3.5, ps $> .05$. The data pooling the two groups are shown in Figure 1. The main effect of frequency was significant, $F(1, 13) = 19.02$, $MSE = .0014$, $p < .01$. The main effect of structural complexity of the characters was also significant, $F(1, 13) = 27.01$, $MSE = .0009$, $p < .001$. An interaction between character frequency and structural complexity was somewhat less reliable, $F(1, 13) = 4.12$, $MSE = .0011$, $p = .06$. Post hoc pairwise comparisons using the Bonferroni adjustment method for multiple comparisons indicated that RTs for HC and HS characters were not significantly different from each other, $p > .3$. However, RTs for HC characters were shorter than for LC characters, HS characters shorter than LS characters, and LS characters shorter than LC characters, all ps $< .05$.

A $2 \times 2 \times 2$ (frequency × structural complexity × group) repeated measures ANOVA was performed on the accuracy data. Again, the group effect and the in-

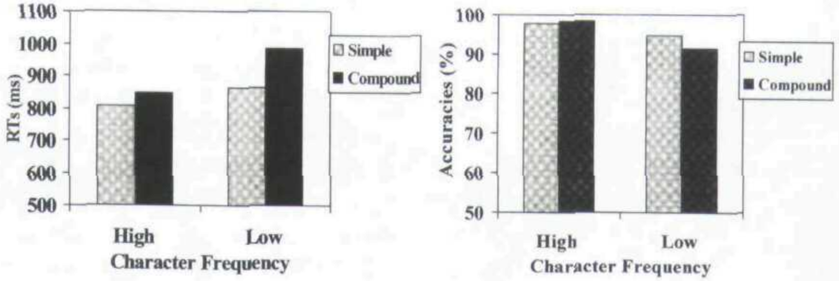

FIGURE 1 Reaction times (RTs) and accuracies of real characters in lexical decision task

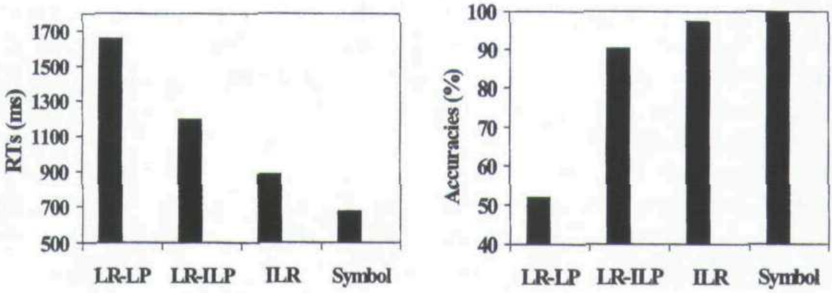

FIGURE 2 Reaction times (RTs) and accuracies of noncharacters in lexical decision task. LR-LP = legal radicals in legal positions; LR-ILP = legal radicals in illegal positions; ILR = illegal radicals.

teraction between the group and main effects were not significant. The two-group pooled data are shown in Figure 2. The main effect of frequency was significant; participants responded more accurately on high-frequency characters than low-frequency characters, $F(1, 13) = 8.12$, $MSE = .0033$, $p < .05$. The main effect of structural complexity was not statistically significant, $F(1, 13) = 3.22$, $MSE = .0013$, $p = .096$. There was a significant interaction between frequency and structural complexity, $F(1, 13) = 6.24$, $MSE = .0016$, $p < .05$. Post hoc pairwise comparisons using the Bonferroni adjustment method for multiple comparisons indicated that only accuracies for HC characters were significantly higher than those for LC characters, $p < .05$. All the other pairwise comparisons were not, all ps > .05.

Noncharacters: No responses. The group effect again was not significant. Therefore, the data for the two groups were pooled as shown in Figure 3. The participants made the fastest rejection on visual symbols and the slowest rejection on the noncharacters containing LR-LP. Pairwise comparisons using the Bonferroni adjustment method for multiple comparisons showed that RTs for LR-LP noncharacters were slower than those for LR-ILP noncharacters, $p < .001$. RTs for LR-ILP noncharacters were slower than those for ILR noncharacters, $p <. 001$. Finally, RTs for ILR noncharacters were slower than visual symbols, $p < .001$.

The accuracy results were consistent with the RT data. The participants were close to 100% correct at judging visual symbols; characters with the LR-LP caused the most errors for the learners. Pairwise comparisons using the Bonferroni adjustment method for multiple comparisons showed that accuracies for LR-LP noncharacters were lower than those for LR-ILP noncharacters; accuracies for LR-ILP noncharacters were lower than those for ILR noncharacters, all ps < .05. Only the accuracies for ILR noncharacters were not significantly lower than those for visual symbols, $p = .08$.

The results replicated those of Wang et al. (2003). Thus, the effect of frequency on lexical decisions is reliable across the two different participant cohorts, indicating that lexical decisions are a meaningful measure to apply to second-language word form learning. The effect of structural complexity is also replicated and was more reliable in decision times than in accuracies. The participants made significantly faster decisions on simple characters than compound characters. This result may suggest that the radical components are represented in reading Chinese characters. However, when making such a claim we need to be cautious about potential confounding factors such as the possible differences in visual complexity (e.g., the number of strokes) between simple and compound characters. The interaction between frequency and structural complexity was significant for accuracies and marginally significant for RT data ($p = .06$). Frequency effects were larger for compound characters than for simple characters. This interaction is analogous to a well-known Frequency × Regularity interaction in reading literature across different writing systems (e.g., Fang, Horng, & Tzeng, 1986; Seidenberg, 1985). The results for noncharacters were very much consistent with our previous findings. These results again pointed to the ability of beginning Chinese learners to detect the legality of the forms and positions of the radicals in the character. They further indicated that the learners could detect illegal forms of the radicals faster than illegal positions of the radicals. We reiterate how important it is that this learning occurred without explicit instruction.

The results in Experiment 1 suggest that this cohort of students acquire implicit structural knowledge of Chinese characters. In Experiment 2 we take up the question of implicit learning of function. First, do beginning Chinese learners acquire implicit learning of radical component function without explicit instruction? If they do not, can explicit instruction in the laboratory produce this functional learning?

EXPERIMENT 2

In this experiment, we focused on testing Chinese learners' implicit and explicit learning of semantic radicals. The students were asked to identify novel Chinese characters to which they had no exposure in their curriculum. These unknown characters contain semantic radicals that the students have encountered in their Chinese curriculum. Explicit instructions were then provided to teach the function of the semantic orthographic components. We were interested in demonstrating an effect of explicit learning of the function of semantic radicals for the whole characters. The effect of frequency of curriculum exposure of the radicals was also examined.

The semantic radical is the component radical that provides a meaning cue for the whole character. Learning orthographic functional regularities imposes higher cognitive demands on the learner than learning perceptual features of the writing system, especially in a case where no instruction is provided. The learner is required to process structural information perceptually first, then extract the functional cue from repeated exposures to the component, and finally make appropriate inferences for the newly encountered print based on the extracted regularity. As these beginning learners have very limited vocabulary, the limited exposure to component radicals might not be sufficient for learners to establish the association between the radical unit and the meaning of the characters. Also important, their curriculum did not provide instruction on the function of the semantic radical. Therefore, implicit learning of the orthographic component function might not be feasible for beginning Chinese learners. We predict that the Chinese learners might fail to utilize the semantic radical information in the unknown characters for their meaning task implicitly. They might be able to identify visually the radical components in the unknown characters with certain probing.

The frequency of the radical component should have an effect on their performance. The more exposure to the target radicals the students have in learned compound characters, the higher the probability they would recognize these radicals visually in unknown characters. Furthermore, if frequently encountered radicals also consistently associate with the same meaning cues across compound characters, the students would be more ready to draw inferences about a meaning relation between the high-frequency radical and an unknown character, especially when some probing is available. With explicit instruction, the students would learn these orthographic functional regularities quickly. An alternative hypothesis would be that learners might be able to acquire the functionality of radical components implicitly just as they implicitly learned the internal structural regularities of the character.

Method

Participants. The same 15 participants from Experiment 1 participated in this experiment.

Curriculum analysis. A computerized curriculum file was created , which included all of the characters inwhich the textbook learned up to the end of the first term when students were recruited for the experiment. Extra possible characters learned via assignments or supplementary materials were not taken into consideration. The numbers of the radicals and strokes of each character were tallied. About 76% of the characters in the curriculum were compound characters. Sixteen semantic radicals occurred in 263 characters. Each radical's type frequency—the number of different characters in it occurred—was tallied. This frequency information was used to select radicals for the subsequent experiment. Semantic consistency was also tallied for each radical. We calculated the semantic consistency as the number of characters in which the radical's meaning was congruent with the character's meaning over the number of total characters containing the radical.

Materials and design. A list of 18 unknown characters was selected for the experiment from both the Pitt and CMU programs. The participants had not learned these characters from their curricula. Fourteen of the characters had a left–right structure. Four items had a top–bottom structure. These unknown characters are semantic compounds with the semantic radicals either on the left parts or on the bottom parts of the characters. The target semantic radicals provided transparent meaning cues for the whole characters. For example, the unknown character 湿 'wet' contains the known semantic radical 氵 'water' at the left side of the character. The unknown character 怒 'angry' contains the known semantic radical 心 'heart' at the bottom of the character. The students had encountered these semantic radicals in their curriculum. The semantic consistency of these target radicals was high with a mean of .83. The right or top parts of the characters may or may not provide pronunciation cues. However, for those that do provide pronunciation cues, they were selected to be unknown to the students to control for the influence of these phonetic radicals on our target semantic ones.

The frequency of the semantic radicals was manipulated: six high- and six low-type-frequency semantic radicals were selected from their curriculum. The means of the high- and low-frequency radicals were 8.33 and 3.00 for Pitt students, and 7.33 and 2.50 for CMU students, respectively. There exists a difference in terms of the degree of transparency of the semantic radicals in Chinese characters. According to Shu, Chen, Anderson, Wu, and Xuan (2003), semantic transparency refers to the contribution of a semantic radical to the meaning of a compound character. The more meaning information a semantic radical contributes, the more transparent the compound character. To control for semantic transparency when considering the effect of frequency of the radicals, we designed a questionnaire to investigate the semantic transparency of the high-frequency and low-frequency radical groups. Twelve native Chinese readers were asked to rate how much meaning information the target semantic radical provides for the whole character. We used a 5-point scale ranging 1 (*complete amount of information*), 2 (*a large* *amount of information*), 3 (*some amount of information*), 4 (*a little amount of information*), and 5 (*not at all*). The mean score of semantic transparency was 2.14 (SD = 0.55) for the high-frequency group and 2.10 (SD = 0.42) for the low-frequency group. Therefore, the high-frequency and low-frequency radical groups did not differ from each other in terms of their semantic transparency, p >.1.

Three of the six radicals in each frequency group had two trials, and the other three had only a single trial. In other words, each of these three radicals appeared in two different characters. See Table 1 for the design and complete set of materials. The two frequency groups were matched in terms of their radical numbers and stroke numbers. The order of the characters was randomized and was kept the same for all participants. Of 18 characters, 8 were common to the Pitt and CMU curricula. Among the 8 shared unknown characters, 5 contained high-frequency semantic radicals and 3 contained low-frequency ones.

Procedure. Participants were tested individually, and the entire experiment was videotaped via a Web cam. The experiment consisted of three steps.

• Step 1: Unknown character identification with no probing. The participants were asked to identify the unknown characters without any probe. They were encouraged to take a guess. The instructions were as follows: "I am going to show you a list of characters you haven't seen before. You are encouraged to take a guess at each of them. Please write down their pronunciations in Pinyin and meanings in English translation, and then tell me how you guess."

• Step 2: Identification with probing. The participants were probed with a series of questions. The purpose of these questions was to encourage them to look at the characters more carefully and to identify the familiar radicals by marking them. They were also encouraged to take a guess at any meaning cues that the familiar radicals might suggest. The participants were instructed as follows: "Now look at the character carefully; have you seen any part of this character before? Please circle it so that I can see it. Does it tell you something? Take a guess."

• Step 3: Identification with explicit instruction. The students were directed to the target radical in the unknown character and were helped to retrieve information from previously known characters that contain the same target radical. The experimenter wrote down a few learned characters from the curriculum containing the target radical and explained the semantic relation between the target radical and the known characters. Then the students were encouraged to make analogies to the unknown character. In the end, the students were explicitly taught the function of the target radical. The instructions were as follows: "Look at this part [pointing to the target radical in the target character, e.g., 氵]. Does it appear in 海 'sea', 流 'water-flow', and 游 'swim'? What do these characters mean? Does that part tell you something related to 'water'? Now look at this part carefully; I would like to let you know that

TABLE 1
Design and Materials Used in Experiment 2

	Radicals	*Related Meaning*	*Characters*	*Trial Order*	*Meaning*
High-frequency radicals	口	mouth	喝	First trial	drink
			喂	Second trial	feed
	讠	speech	诉	First trial	tell
			谜	Second trial	riddle
	氵	water	湿	First trial	wet
			酒	Second trial	wine
	亻	people	俪		couple
	日	sun	晒		sun-dry
	辶	walk	追		chase
Low-frequency radicals	忄	emotion	忧	First trial	worry
			恼	Second trial	angry
	纟	string-like	线	First trial	thread
			绳	Second trial	rope
	灬	heat	照	First trial	sunshine
			烈	Second trial	hot
	饣	food	饮		drink
	走	walk	超		pass
	心	heart	恩		thankful

this part often tells you something related to 'water'. So this new character 湿 means 'wet' and is pronounced as /shi/1. Try the next character now."

Note that the purpose of the explicit instruction was to draw students' attention to the semantic relationship between the target radical and the whole character. However, in most cases, the semantic radical would not convey the exact meaning of the whole character. The students were made aware of this issue during the explicit instruction. For example, in the case of the 氵 'water' radical, the students were asked to be aware that this radical did not provide the exact meaning of the character 湿 'wet'.

Coding system. The students' unknown character identification performance was coded according to the following scale that ranged from 0 (*don't know*) to 1 (*attempt to make inference, but unsuccessful*; e.g., trying to relate the unknown character 酒 'wine' with 四 'four'). After the probing, the identification performance of the target radicals was coded as follows: 0 (*don't know*), 1 (*identify visu-*

ally only; i.e., circle the familiar radicals), and 2 (*identify visually and make appropriate inference*; e.g., if the student could not only identify 口 'mouth' in 喝 'drink' but also infer that the novel character has a meaning related with 'mouth'). After the explicit instruction, the students' identification performance on the target semantic radicals was coded as follows: 0 (*don't know*), 1 (*identify visually only*), and 2 (*identify visually and make appropriate inference*). Recall that three out of the six radicals in each frequency group had two trials. Because the first trial was used for explicit instruction, only the second trial was coded and used for subsequent analyses. Interrater reliability for the response coding system was established between two independent native Chinese readers. The reliability for coding of character identification was 1.0, for after-probing performance was .96, and for after-instruction performance was .94. Disagreements were resolved through discussions between the raters.

Results and Discussion

Unknown character identification with no probing. The participants could not identify successfully the pronunciations of any of the unknown characters. Because the students' knowledge of the semantic components was the target for this study, we report only the meaning identification performance in the following analyses. The percentage of items that fell into the two coding categories in identifying the unknown characters is shown in Figure 3. The majority of the items (about 83% low-frequency and 86% high-frequency radicals) fell into the "No Answer" category. The participants made an attempt to infer the meanings on a very small percentage of the unknown characters (about 14% low frequency, 10% high frequency) but were not successful.

After probing. The percentage of items that fell into the three coding categories after the experimenter's probing is shown in Figure 4. A repeated measures ANOVA was conducted in which both the radical frequency and response type were treated as within-participants variables. Results showed that there was a significant effect of response type, $F(2, 13) = 14.84$, $p < .001$. Post hoc nonparametric Wilcoxon signed ranks tests showed that for low-frequency radicals the students made significantly more visual identifications of the target semantic radicals than meaning inferences of the semantic radicals ($Z = 3.06$, $p < .01$). For high-frequency radicals, however, the students made more appropriate meaning inferences of the target radicals than visual identifications, although not statistically significant ($Z = 1.749$, $p < .1$). The effect of frequency was not significant ($F < 1$). However, the interaction between frequency and response type was significant, $F(2, 13) = 17.97$, $p < .001$. Post hoc paired sample t tests indicated that the students were able to make more inferences using semantic radicals on high-frequency than low-frequency

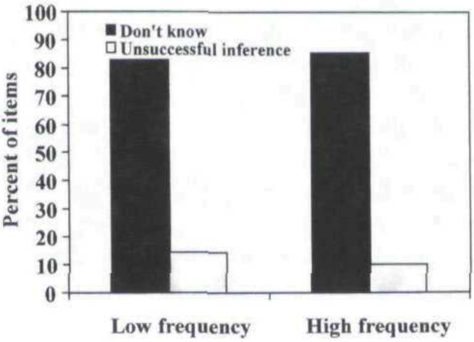

FIGURE 3 Meaning identification without probe.

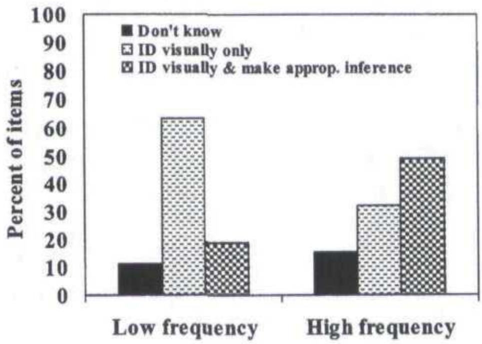

FIGURE 4 Meaning identification after probe. ID visually only = identify visually only; ID visually and make approp. inference = identify visually and make appropriate inference.

ones, $t(14) = 4.45$, $p < .01$, but were able to make more visual identifications on low-frequency radicals than on high-frequency ones, $t(14) = 5.80$, $p < .001$.

We found that besides the target semantic radicals, the participants also visually identified other known radicals from the curriculum. There were 18 such component radicals in total. In addition, the participants visually identified certain stroke combinations in the unknown characters, which would not be salient to skilled native Chinese readers. For example, 文 in 恼, and 下 in 诉. There were 10 in total. These stroke combinations are only parts of legal radicals in the whole characters. These participants chose these stroke combinations without concern for the legality of the entire radical.

After explicit instruction. The percentage of items that fell into the three coding categories after the experimenter's explicit instruction is shown in Figure 5. We were interested in comparing meaning identification performance on the second trials of the unknown characters before and after explicit instruction. A repeated measures ANOVA was performed on the accuracies of visually identifying the target radical and making appropriate inferences before and after explicit instruction. Results showed that the learning effect was significant, $F(1, 14) = 6.81, p < .05$. The learning effect represented the difference between Figure 4 and 5. The students demonstrated more use of the semantic radical information to infer the meaning of the unknown characters after explicit instruction on the meaning of the target radical, compared with their performance prior to such explicit instruction. Post hoc paired sample t tests revealed that the learning effect for low-frequency radicals was statistically significant, $t(14) = 3.06, p < .01$, but this was not the case for high-frequency radicals, $t(14) = 1.57, p = .1$. Also of interest, the effect of the frequency of the radical was significant, $F(1, 14) = 7.09, p < .05$. The students made more appropriate meaning inferences on high-frequency radicals than low-frequency ones. The interaction between frequency and learning was not significant, $F(1, 14) = 2.90, p > .05$.

These results suggest that prior to probing, these beginning learners did not demonstrate any implicit or explicit knowledge of the function of semantic radicals. Most responses were "Don't know." The students made attempts to infer meaning from certain parts of the characters but appeared to be unsuccessful. After probing, the students obviously performed better; they apparently started to pay attention to and visually identify the familiar radicals. Furthermore, they showed some successful trials of meaning inference. It is interesting that the students extracted meaning information more from the high-frequency radicals than from the

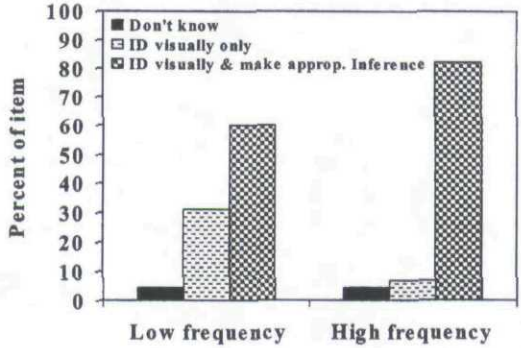

FIGURE 5 Meaning identification after explicit instruction. ID visually only = identify visually only; ID visually and make approp. inference = identify visually and make appropriate inference.

low-frequency radicals. This result suggests that print exposure plays an important role in learning orthographic regularities. After explicit instruction, the students performed even better. Especially for low-frequency radicals, the students extracted significantly more meaning information from known semantic radicals compared with their performance before the explicit instruction.

GENERAL DISCUSSION

The students who were learning Chinese as a new writing system in our study did not receive explicit instruction in their curriculum concerning the visual-orthographic structure of the characters. Nevertheless they acquired implicitly these structural features through repeated exposure. They demonstrated a quick learning of the internal character structure. They were able to detect the legality of the radical form and the legality of the radical position. This powerful implicit learning could be attributed to two important characteristics of Chinese character learning by alphabetic readers. First, because each Chinese character is a salient perceptual unit, visual-orthographic skills are critical in learning to identify characters. Beginning learners of Chinese will naturally attend to orthographic details in acquiring character recognition skills. Another important feature of these learners is that they were forced to learn to speak and read simultaneously. Because the students lacked strong support from foundational oral language skills for reading acquisition, the graphic form of characters became an indispensable source of information in character learning.

One may then question whether this type of implicit visual-orthographic learning is unique to Chinese learning. Implicit learning of visual-orthographic features has also been found in spelling acquisition of young alphabetic learners. For example, Treiman (1993) analyzed Grade 1 English-speaking children's errors on invented spelling. She found that children's errors reflected their implicit understanding of some basic English orthographic regularities. For example, higher frequency consonant doublets (e.g., *ee*, *bb*) were utilized in spelling attempts more often than lower frequency ones (e.g., *hh*, *kk*) by the children in her study. The children also produced more consonant doublets at a legal position in the word than at an illegal position. Cassar and Treiman (1997) carried out a series of experiments to investigate further implicit knowledge of English double letters by young children. They found that late kindergarten English-speaking children have acquired some knowledge of the legal form and position of consonant doublets. The young children preferred spellings with final doublets (e.g., *baff*) to those with beginning doublets (e.g., *bbaf*). They preferred the legal doublets (e.g., *yill*) to those with illegal ones (e.g., *yihh*). The authors argued that this type of learning occurs via children's experience with print but not from explicit teaching from parents or teachers. Thompson et al. (1996) also observed young children's learning of sublexical orthographic–phonological relations via "spontaneous induction" from print word exposure at a very early stage: during the first few months of school instruction in New Zealand. It is noteworthy that this type of implicit learning occurred in a learning context without explicit phonics instruction. These findings from both Chinese and English suggest that implicit learning might be sufficient for acquiring basic visual-orthographic structures of a writing system.

However, it is important to emphasize the difference between Chinese visual-orthographic learning and alphabetic orthographic learning. The orthographic structure of Chinese characters contrasts sharply with that of English words. The radical-based composition and positional constraints of the Chinese character work in a very different way from English letter forms and positional patterns.

The significant effect of curriculum frequency in lexical decisions is also a good indication of implicit learning. The fact that the students made faster and more accurate lexical decisions on high-frequency characters indicates that the more times the character is encountered, the better the lexical knowledge of the character (Experiment 1). Print exposure seems to be a critical source of implicit learning. The effect of frequency of exposure on word identification reflects a very general learning principle. Connectionist learning models (e.g., Seidenberg & McClelland, 1989) offer a very strong explanation for this frequency effect. According to these models, the weighted connections between lexical units (phonological, orthographic, and semantic units) are stronger for words experienced more often. Strong connections between the lexical units result in rapid and accurate word identification. In Experiment 2, we see the effects of frequency also in the significant main effect of frequency of the target semantic radical for performance on meaning inferences at the probing stage. The students made more appropriate meaning inferences on high-frequency radicals than on low-frequency radicals. This suggests that the more students are exposed to the orthographic components signifying certain functions, the higher the probability that they will extract the functional regularities of the components. However, it is important to stress that the beginning learners in our study did not clearly show implicit knowledge of the functional cues of the orthographic components. When they were required to identify the unknown characters without any probing, they demonstrated great difficulty. At the probing step, the students were able to identify visually the majority of the target semantic radicals. However, the percentage of items that can be identified not only visually but also with an attempted meaning inference was still lower than 50% for both high- and low-frequency target radicals. This result implies that the learners have only limited implicit knowledge of the function of the radical component in their 1st year of learning Chinese characters. This might be because the students' limited vocabulary would not allow for extraction of the radical regularities for inference. It is not surprising that beginning learners of Chinese in a foreign language program in America had such difficulties in their 1st year. The first-grade Chinese children in Shu and Anderson (1997) demonstrated similar kinds of difficulties in processing the morphemic radical in compound Chinese two-character words despite the fact that two-character words had high print exposure within and outside of their curriculum. There is also evidence in English that children in early grades have not spontaneously developed knowledge of the morphological function of the orthographic component (e.g., Nagy, Diakidoy, & Anderson, 1993; Tyler & Nagy, 1989). Another reason for such a difficulty might be related to the lack of explicit instruction facilitating this type of learning.

The findings in Experiment 2 provide evidence for the role of explicit instruction in learning orthographic regularities. Acquisition of the functional use of orthographic components of characters appeared to be speeded by explicit instruction. The students in our study received only a short period of instruction, and they demonstrated rapid learning. The explicit instruction helped the students decompose a newly encountered character into smaller known components and then make an inference about the functional component to acquire the meaning information of the whole. It is also worth noting that the frequency of the semantic target radical played a role in helping students make an inference about the meaning of the functional component during the explicit instruction. The students made more appropriate meaning inferences on high-frequency radicals than low-frequency ones. In the meantime, the students made more improvement in learning the function of low-frequency radicals than high-frequency ones. This result suggests that the low-frequency radicals are particularly sensitive to the explicit instruction, and the explicit instruction can facilitate the students' learning of low-frequency semantic radicals more than high-frequency ones. Previous studies on training Chinese children using semantic radical information in learning to read have also revealed strong training effects. For example, both Ho, Wong, and Chan (1999) and Nagy et al. (2002) demonstrated that after explicit teaching of the function of semantic radicals, Chinese children as early as Grade 1 were able to improve their character reading skills. However, their training was dramatically longer than the training in the present study, for example, it was a yearlong classroom-based intervention in Nagy et al.'s study. These authors thus highly recommended that Chinese primary-grade teachers consider employing explicit instruction on the structural and functional features of semantic radicals for young Chinese children. In another related study, Taft and Chung (1999) reported a successful case of teaching semantic radical knowledge to naive Chinese learners. This explicit teaching also facilitated individuals' character learning, especially when the radical instruction was provided at the initial stage of learning. These findings on learning Chinese characters are also in line with a claim among second-language researchers about the effectiveness of explicit teaching in general. These researchers argued that explicit second-language instruction can result in significant target-oriented gains and that the effectiveness of second-language instruction is long lasting (DeKeyser, 1997; MacWhinney, 1997; see Ellis, 2002, for a detailed review).

An interesting finding from the probing step in identifying the unknown character is that the students identified certain stroke combinations that skilled Chinese readers would find peculiar. This finding was consistent with a study by Yeh, Li, Takeuchi, Sun, and Liu (1999). They demonstrated in a character sorting and recall task that, compared to Taiwanese and Japanese students who had extensive character-learning experience, American college students who had no experience of learning Chinese used some salient strokes or stroke patterns such as L-shaped, P-shaped, or enclosed strokes as cues to represent the visual-orthographic structure of the characters. These stroke patterns were embedded in a larger component according to skilled Chinese readers and hence would not be attended to by these readers. However, these features became attractive to the naive Chinese readers in Yeh et al. and the beginning Chinese readers in our study. These results suggest that less skilled readers rely on lower level perceptual cues and less structured strategies in processing written symbols than skilled readers. High-level knowledge and learning experience in a particular writing system can change and direct visual perceptions of the written symbols. Another possible interpretation for the American students' tendency to identify L- and P-shaped or enclosed strokes is their familiarity with the English alphabet.

Two future research directions are suggested. First, Experiment 1 did not inform us as to whether the learners acquired explicit understanding of the internal structure of the character. Future research could follow the lexical decision task with a series of, postexperiment questions to investigate this, such as "How did you know that this stimulus is not a real character?" and "Why did you decide that this item is not a real character?" Answers to these questions would allow us to gather important information about learners' explicit knowledge of characters' orthographic structural constraints. Second, in Experiment 2 it would be important to examine whether learners' explicit learning of the radical function of the character can generalize to characters containing radicals that the students encountered in their curriculum but were not trained on in the experiment.

In summary, the findings from our study suggest that learning of visual-orthographic structure of a new writing system can occur very early and without explicit instruction, whereas learning of orthographic regularity related to meaning may be more demanding and can be fostered by explicit instruction.

ACKNOWLEDGMENTS

This research was funded in part by National Science Foundation Grant SBR-9616519 to Charles A. Perfetti and a postdoctoral fellowship to Min Wang by the Social Sciences and Humanities Research Council of Canada. Preparation of the article was supported by a National Academy of Education/Spencer Fellowship to Min Wang.

We thank the students and instructors in the Chinese program at the University of Pittsburgh and Carnegie Mellon University for their participation and cooperation in this study. We also thank Baozhi Qu for his assistance with coding the data in Experiment 2 and Stephen Tonks and Jennifer Cromley for proofreading the draft. Finally, we very much appreciate the valuable comments and suggestions from the editor, Marcus Taft, Richard Anderson, and an anonymous reviewer.

REFERENCES

Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology, 89,* 631–644.

Chan, L., & Nunes, T. (1998). Children's understanding of the formal and functional characteristics of written Chinese. *Applied Psycholinguistics, 19,* 115–131.

Chen, Y.-P. (1993). *Word recognition and reading in Chinese.* Unpublished doctoral dissertation, University of Oxford, London.

Chen, Y.-P., & Allport, A. (1995). Attention and lexical decomposition in Chinese word recognition: Conjunctions of form and position guide selective attention. *Visual Cognition, 2,* 235–268.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19,* 195–221.

Ehri, L. C. (1998). Research on learning to read and spell: A personal-historical perspective. *Scientific Studies of Reading, 2,* 97–114.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24,* 143–188.

Fang, S.-P., Horng, R.-Y., & Tzeng, O. J. L. (1986). Consistency effects in the Chinese character and pseudo-character naming task. In H. S. R. Kao & R. Hoosain (Eds.), *Linguistics, psychology, and the Chinese character* (pp. 11–21). Hong Kong: Center of Asian Studies, University of Hong Kong.

Feldman, L. B., & Siok, W. W. T. (1999). Semantic radicals in phonetic compounds: Implications for visual character recognition in Chinese. In J. Wang, A.W. Inhoff, & H.C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 19–35). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Fletcher-Flinn, C. M., & Thompson, G. B. (2000). Learning to read with underdeveloped phonemic awareness but lexicalized phonological recoding: A case study of a 3-year-old. *Cognition, 74,* 177–208.

Ho, C. S.-H., & Bryant, P. (1997). Phonological skills are important in learning to read Chinese. *Developmental Psychology, 33,* 946–951.

Ho, C. S.-H., Wong, W.-L., & Chan, W.-S. (1999). The use of orthographic analogies in learning to read Chinese. *Journal of Child Psychology and Psychiatry, 40,* 393–403.

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78,* 243–255.

Leong, C. K. (1997). Paradigmatic analysis of Chinese word reading: Research findings and classroom practices. In C. K. Leong & R. M. Joshi (Eds.), *Cross-language studies of learning to read and spell: Phonological and orthographic processing* (pp. 378–417). Dordrecht, The Netherlands: Kluwer Academic.

MacWhinney, B. (1997). Implicit and explicit processes. *Studies in Second Language Acquisition, 19,* 277–281.

Nagy, W. E., Diakidoy, I.-A. N., & Anderson, R. C. (1993). The acquisition of morphology: Learning the contribution of suffixes to the meanings of derivatives. *Journal of Reading Behavior, 25*, 155–170.

Nagy, W., Kuo-Kealoha, A., Wu, X., Li, W., Anderson, R. C., & Chen, X. (2002). The role of morphological awareness in learning to read Chinese. In W. Li, J. S. Gaffiney, & J. L. Packard (Eds.), *Chinese children's reading acquisition: Theoretical and pedagogical issues* (pp. 58–85). Norwell, MA: Kluwer Academic.

Peng, D.-L., Li, Y.-P., & Yang, H. (1997). Orthographic processing in the identification of Chinese characters. In H. C. Chen (Ed.), *Cognitive processing of Chinese and related Asian languages* (pp. 86–108). Hong Kong: The Chinese University Press.

Perfetti, C. A., Liu, Y., & Tan, L. H. (2002). How the mind can meet the brain in reading: A comparative writing systems approach. In H. S. R. Kao, C. K. Leong, & D. G. Gao (Eds.), *Cognitive neuroscience studies of the Chinese language* (pp. 35–60). Hong Kong: Hong Kong University Press.

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

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recogniton and naming. *Psychological Review, 96*, 523–568.

Shu, H., & Anderson, R. C. (1997). Role of radical awareness in the character and word acquisition of Chinese children. *Reading Research Quarterly, 32*, 78–89.

Shu, H., & Anderson, R. C. (1999). Learning to read Chinese: The development of metalinguistic awareness. In J. Wang, A.W. Inhoff, & H. C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 1–18). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Taft, M., & Chung, K. (1999). Using radicals in teaching Chinese characters to second language learners. *Psychologia: An International Journal of Psychology in the Orient, 42*(4), 243–251.

Taft, M., Zhu, X., & Peng, D. (1999). Positional specificity of radicals in Chinese character recognition. *Journal of Memory and Language, 40*, 498–519.

Thompson, G. B., Cottrell, D. S., & Fletcher-Flinn, C. M. (1996). Sublexical orthographic- phonological relations early in the acquisition of reading: The knowledge sources account. *Journal of Experimental Child Psychology, 62*, 190–222.

Thompson, G. B., Fletcher-Flinn, C. M., & Cottrell, D. S. (1999). Learning correspondences between letters and phonemes without explicit instruction. *Applied Psycholinguistics, 20*, 21–50.

Treiman, R. (1993). *Beginning to spell.* New York: Oxford University Press.

Tyler, A., & Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of Memory and Language, 28*, 649–667.

Wang, M., Perfetti, C. A., & Liu, Y. (2003). Alphabetic readers quickly acquire orthographic structure in learning to read Chinese. *Scientific Studies in Reading, 72*, 183–207.

Yeh, S. L., Li, J. L., Takeuchi, T., Sun, V. C., & Liu, W. R. (1999, June). *The influence of learning experience and cognitive structure on the classification of the shapes of Chinese characters.* Paper presented at the Third Conference of the Asian Association of Social Psychology.

Zhu, X. (1988). *Analysis of the cuing function of the phonetic in modern Chinese.* Proceedings of the symposium on Chinese language and character, Beijing: Guang Ming Daily Press. [In Chinese]

Manuscript received March 21, 2003
Accepted October 24, 2003

APPENDIX

Materials Used in Experiment 1

Real characters				Non-characters			
HS	HC	LS	LC	LR-LP	LR-ILP	ILR	Symbol
飞	父	本	打	[illegible]	[illegible]	[illegible]	&
今	北	百	机	[illegible]	[illegible]	[illegible]	*
工	字	电	左	娘	[illegible]	[illegible]	@
不	多	长	走	[illegible]	[illegible]	[illegible]	!
母	分	车	对	[illegible]	[illegible]	[illegible]	#
两	没	来	级	[illegible]	[illegible]	[illegible]	%
年	那	事	近	[illegible]	[illegible]	[illegible]	$
里	吗	书	功	[illegible]	[illegible]	[illegible]	;
小	老	生	共	吧	[illegible]	[illegible]	<
太	六	少	兴	狌	[illegible]	[illegible]	>
文	师	主	台	姟	[illegible]	[illegible]	Ψ
五	去	业	问	[illegible]	[illegible]	[illegible]	∉
下	四	火	先	[illegible]	[illegible]	[illegible]	∇
天	他	白	用	[illegible]	[illegible]	[illegible]	⇔
我	她	半	早	決	[illegible]	[illegible]	ℑ
月	外	东	因	[illegible]	[illegible]	[illegible]	⇒
中	只	末	次	[illegible]	[illegible]	[illegible]	©
再	有	日	考	[illegible]	[illegible]	[illegible]	®
州	友	为	认	娃	[illegible]	[illegible]	}
西	在	牙	汉	[illegible]	[illegible]	[illegible]	Ω

Note. HS = high-frequency simple characters; HC = high-frequency compound characters; LS = low-frequency simple characters; LC = low-frequency compound characters; LR-LP = legal radicals in legal positions; LR-ILP = legal radicals in illegal positions; ILR = illegal radicals.