Developmental changes in character-complexity and word-length effects when reading Chinese script

Yi-Fen Su · S. Jay Samuels

Abstract The present study was designed to examine the developmental changes in the character-complexity and word-length effects when reading Chinese script. Character complexity was defined in two ways: (1) the number of constituent strokes for characters (Experiment 1), and (2) the number of constituent radicals for characters (Experiment 2). The word length was defined as the number of characters in words (Experiment 3). The three experiments involved a lexical recognition task, and the participants consisted of 25 second graders, 24 fourth graders, 24 sixth graders, and 25 university students. In Experiment 1, it was found that the response latencies increased with the number of strokes in characters for second graders, whereas no effects were evident for fourth graders, sixth graders, or university students. In contrast, in Experiment 2 no character-complexity effect was found for the number of radicals. In Experiment 3, only a partial word-length effect was found for the number of characters for second graders. These results suggest that beginning Chinese readers process characters in an analytical way, but that the decoding process changes gradually from analytic to holistic as their reading skills develop. The educational implication of this result is discussed.

Keywords Character-complexity effect · Chinese script · Developmental changes · Word-length effect

Y.-F. Su (✉)
Department of Educational Psychology and Counseling, National Taiwan Normal University, 162, Ho-Ping E. Rd., Section 1, Taipei 10610, Taiwan
e-mail: yifensu@ntnu.edu.tw

S. J. Samuels
Department of Educational Psychology, University of Minnesota, 250 Education Science Building, 56 East River Road, Minneapolis, MN 55455-0364, USA

Introduction

Word-length effect refers to response latency increases with the length of words. This effect is often considered indicative of word recognition involving a sublexical or analytic reading process. The differential effect of length on high-frequency and low-frequency words in skilled readers, beginning readers, and dyslexic readers suggests that the nature of word recognition changes with the development of reading ability (Martens & de Jong, 2006). However, does the word-length effect found in alphabetic written languages also occur in logographic script, such as Chinese? Can the theoretical explanation of word-length effect be generalized to a language that is not spelled, such as Chinese? Is the nature of developmental changes in Chinese word recognition similar to that in alphabetic written languages? These are questions which the present study attempts to answer. Investigating these questions not only provides cross-language evidences for theoretical implications of word-length effects, but also helps to understand how Chinese word recognition skills develop.

In past research, various techniques have been used to investigate word-length effect in alphabetic written language. The results from adult studies have shown a mixture of inhibitory effects (longer words are harder) and null effects (reviewed in New, Ferrand, Pallier, & Brysbaert, 2006). A common finding is that the speed of reading high-frequency words is not affected by word length for adult skilled readers, but the latencies for low-frequency words and pseudowords are affected by length (Ferrand, 2000; Jared & Seidenberg, 1990; Juphard, Carbonnel, & Valdois, 2004; Weekes, 1997).

The results of research involving young readers and dyslexics are more consistent than those of the adult studies. In young readers, the word-length effect has been found to exhibit developmental changes that gradually diminish with age (Aghababian & Nazir, 2000; Bijeljac-Babic, Millogo, Farioli, & Grainger, 2004; Samuels, LaBerge, & Bremer, 1978; Spinelli, De Luca, Di Filippo, Mancini, Martelli, & Zoccolotti, 2005; Su, 1997; Zoccolotti, De Luca, Di Pace, Gasperini, Judica, & Spinelli, 2005). Samuels et al. (1978) conducted a study using a word-categorization task in which the subjects had to answer (by pressing a button) whether or not a word presented on a screen related to an animal. They found that the response latencies of second graders increased with word length, whereas the word-length effects gradually decreased for fourth and sixth graders, and was nonexistent for university students. Su (1997) replicated the procedure of Samuels et al. (1978) and obtained the same results.

Aghababian and Nazir (2000) also found similar developmental changes in French words when using a perceptual identification task in which word stimuli were presented briefly with forward and backward masking. They observed an interaction between grade and word length, with the magnitude of word-length effect decreasing from first to fifth grade. The same developmental changes were also found by Bijeljac-Babic et al. (2004). In Italian-speaking readers, Spinelli et al. (2005) and Zoccolotti et al. (2005) obtained the same results for a naming task.

Furthermore, the influence of length on reading speed is even more evident in dyslexic readers (e.g., Juphard, et al., 2004; Martens & de Jong, 2006;

Spinelli, et al., 2005; van der Leij & van Daal, 1999; Ziegler, Perry, Jacobs, Ma-Wyatt, Ladner, & Schulte-Korne, 2003; Zoccolotti, et al., 2005). Even native dyslexic speakers of a language characterized by regular orthography, such as Italian, showed a clear word-length effect for high-frequency words, low-frequency words, and pseudowords (De Luca, Di Pace, Judica, Spinelli, & Zoccolotti, 1999; De Luca, Borrelli, Judica, Spinelli, & Zoccolotti, 2002).

The differential effect of length on high-frequency, low-frequency words, and pseudowords in skilled readers, beginning readers, and dyslexic readers suggests that the nature of word decoding changes with the development of reading ability (Martens & de Jong, 2006). These results also support the hypothesis of Samuels et al. (1978) that beginning readers and non-skilled readers process a word on a component basis, whereas more-skilled readers process a word in a more holistic way depending on their decoding skill. The extent of parallel word processing depends on the number of times a word has been identified after a single fixation on it (Nazir, Ben-Boutayab, Decoppet, Deutsch, & Frost, 2004). That is, the establishment of automaticity of word recognition in normal readers requires an abundance of practice or exposure to reading words that recur in different texts (Samuels, 2002).

However, do the developmental changes of word-length effects and its theoretical explanations be generalized to a language that is not spelled, such as Chinese? Before reviewing relevant research in Chinese, we first describe the main characteristics of Chinese script.

Written Chinese is a logographic orthography that differs greatly from alphabetic writing systems. Firstly, these two types of script exhibit different orthographic structures. The distinct orthographic units in alphabetic writing systems are words, which comprise letters arranged horizontally from left to right. However, the distinct orthographic units in Chinese are *characters*, which are more like single-syllable morphemes than phonemes or words in English, and are arranged in squares of similar size. Most Chinese words are represented by two or more characters, and all characters are constructed from two types of element called *strokes* and *radicals.* The stroke (e.g., a line, "一" and "丨"; or a curve, "乚" and "丿") is the minimal compositional unit, which is ended when the writing pen leaves the paper.

The literature contains inconsistent definitions of the radical. The first definition refers to the smallest stroke pattern that comprises one or more strokes (e.g., "亻" and "木") and recurs independently in different characters (Chen, Allport, & Marshall, 1996; Honorof & Feldman, 2006; Taft & Zhu, 1997). According to this definition, the character "響" (meaning "loud") has five radicals: three radicals in the top part (i.e., 鄉, left-to-right) and two radicals in the bottom part (i.e., 音, top-to-bottom). Another character, "排" (meaning "row"), has two radicals, although it appears to contain three components. Because either the left or the right part of "非" (meaning "not" in isolation) cannot recur independently in other characters, "非" can only be counted as one radical according to the first definition. The second definition refers to *Bu Shou* (sometimes called a "semantic radical"), which means "semantic root" or "category headers" of a character (Chen et al., 1996; Feldman & Siok, 1999). Based on this definition, the semantic radical of the character "響" is the bottom part "音" (meaning "sound"). The third definition refers to "semantic

radicals" and "phonetic radicals" in semantic-phonetic compound characters (Chen et al., 1996; Ho, Ng, & Ng, 2003). By this definition, the semantic and phonetic radicals of the character "響" are the bottom part "音" and the top part "鄉", respectively. The phonetic radical "鄉" shares the same pronunciation as the character "響", except for the tone. Irrespective of which definition is adopted, a radical itself in some cases is a character (e.g., "木"), but in most cases a character consists of more than one radical (e.g., "休") (Perfetti & Tan, 1999; Sun & Feng, 1999; Taft & Zhu, 1997).

Secondly, the orthography–phonology relationships are different between alphabetic scripts and written Chinese. In alphabetic written languages, the letters approximate the phonemes, although languages which use the alphabetic system differ in how close the correspondence between letters and sounds is. Some languages such as Finnish and Spanish have a close correspondence (shallow orthography), but other languages (e.g. English) often lack a correspondence between letters and sounds (deep orthography) (Gazzaniga, Ivry, & Mangun, 2002). However, written Chinese is not spelled. The orthography–phonology relationship in Chinese script is quite opaque (Cheung & Ng, 2003). Although in some characters (i.e., semantic–phonetic compound characters), phonological information is loosely related to the phonetic components (i.e., phonetic radicals), the components are not sufficiently reliable to support a systematic approach to pronunciation (Honorof & Feldman, 2006; Perfetti & Liu, 2006). The semantic–phonetic compounds refer to the characters that are composed of at least two components: one or more components provide semantic categorical information, and one component provides pronunciation cues (Cheung & Ng, 2003; Honorof & Feldman, 2006; Perfetti & Tan, 1999). About 85% of modern Chinese characters are referred to as semantic–phonetic compounds (Perfetti & Tan, 1999). However, only 26.3% of semantic–phonetic compounds have an identical pronunciation (including the same tone) to their phonetic radicals. Most of the phonetic radicals cue a variety of initials, finals, or tones in the various characters in which they appear (Fan, Gao, & Ao, 1984).

These large differences between Chinese script and alphabetic writing systems lead to questions such as, does the word-length effect found in alphabetic scripts also exist in Chinese writing system, and to what extent does the explanation of the word-length effect (such as sublexical and lexical processing or analytic and holistic processing) be generalized to Chinese script? Some previous adult studies (Chen, et al., 1996; Chen & Liu, 2000; Cheng, 1981; Fang, 1994; Just & Carpenter, 1987; Leong, Cheng, & Mulcahy, 1987; Tan & Peng, 1990), one children study (Chiang, 2003), and one dyslexic study (Yang, 1998) have explored the analogous word-length effects (i.e., the Chinese character-complexity and word-length effects), using strokes, radicals, or characters as the unit of analysis. In studies examining how the number of strokes influences character-recognition latency (i.e., the character-complexity effect), Just, Carpenter, and Wu (cited in Just & Carpenter, 1987) used an eye-tracking paradigm and found that the gaze duration on Chinese characters increased with the number of strokes comprising a character. The same result was obtained by Tan and Peng (1990) in a lexical decision task. Furthermore, in a tachistoscopic identification task, Cheng (1981), Experiment 2 found that the response latencies of adults were longer for high-stroke than for low-stroke

low-frequency characters, but that they did not differ for high-frequency characters. These findings were replicated by Leong et al. (1987) using a naming task.

In addition, Yang (1998) recruited 18 normal and 18 dyslexic fifth-grade readers and found that the latencies of normal readers were not influenced by the number of strokes in a lexical decision task, whereas the reaction times of dyslexic readers were affected by character complexity. Although both high- and low-frequency characters were selected, the interaction effect between the number of strokes and frequency of characters was not analyzed in that study.

Few studies have investigated how the number of radicals influences character-recognition latency (i.e., another type of character-complexity effect using the radical as the unit of analysis). Fang (1994, Experiment 3) adopted a mixed task of lexical decision and stimulus identification, and found that the response latencies of Chinese adult readers increased with the number of radicals in a character. However, in this experiment the radical factor was confounded by the stroke factor. That is, the mean number of strokes was higher in the four-radical characters (18.6) than in the two-radical characters (9.7). Thus, it is unclear whether the number of radicals or the number of strokes was responsible for the different response latencies.

Chen et al. (1996) applied a simultaneous same-different comparison task that required skilled adult readers to make a judgment about the physical identity of two stimuli presented simultaneously on a computer screen. They found that when the number of strokes was between 8 and 11, the response latency was longer for three-radical stimuli than for two-radical stimuli comprising real characters in the "same" pairs (i.e., two stimuli were identical). But the effect of the number of radicals was not significant for pseudocharacters and noncharacters. In the "different" pairs (i.e., two stimuli were different), for the two-radical stimuli, the main effect of the proportion of radicals differing (i.e., one radical or two radicals differing) was significant for real characters, pseudocharacters, and noncharacters. For the three-radical stimuli, the main effect of the proportion of radicals differing (i.e., one, two, or three radicals differing) was only significant for real characters, but not for pseudocharacters and noncharacters. For the completely different pairs (i.e., two radicals differing in two-radical stimuli, and three radicals differing in three-radical stimuli), comparisons were significantly faster when comparing different two-radical pairs than when comparing different three-radical pairs for real characters, pseudocharacters, and noncharacters. Although an effect of the number of radicals was shown in the study of Chen et al. (1996), this might be partly due to the nature of a simultaneous same-different comparison task in which participants have to judge the physical identity of the two stimuli, and also partly due to the "different" pairs being manipulated by the proportion of mismatching radicals in each character. The simultaneous same-different comparison task and the manipulation of stimuli in the "different" pairs might lead participants to use a strategy comparing pairs of stimuli radical by radical.

Chen and Liu (2000) attempted to replicate Chen et al.'s (1996) study in their Experiments 1 and 2. They selected stimuli in a more controlled manner than those used in Chen et al.'s (1996) study in several respects: (1) not only high frequency characters but also low frequency characters were included; (2) the number of strokes and the number of radicals were organized in a completely orthogonal

design; (3) the mismatching portions between the two stimuli in the "different" pairs were made comparable across the manipulation of the number of radicals. The results revealed that the effect of radical number was only obtained for the "same" pairs of high-frequency characters, but not for low-frequency characters of the "same" or of the "different" pairs. Moreover, in the "different" judgment and in the character decision task, the effect of stroke number was found.

Chiang (2003) adopted the design and stimuli used in Chen and Liu's (2000) Experiment 1 to investigate the functional orthographic units in children's recognition of Chinese characters. Fifth-grade elementary school students were recruited. In the same-different comparison task, Chiang observed the effect of radicals only for high-frequency characters, but not for low-frequency characters, which replicated Chen and Liu's (2000) finding. These results suggested that the effect of radicals observed in the "same" judgments might result from specific task-demands only for familiar characters.

The four studies mentioned above all adopted the first definition of radicals and investigated if the number of radicals affected response latency. However, some other studies adopted the second definition of radicals to explore the role of semantic radical in character recognition (e.g., Feldman & Siok, 1999) or the third definition of radicals to explore the role of semantic and phonetic radicals in character recognition (e.g. Leck, Weekes, & Chen, 1995) or the development of metalinguistic awareness (e.g., Ho et al., 2003). The first definition was used in the present study because its main purpose was to elucidate the analogous word-length effect, which is also called the character-complexity effect when the radical is the unit of analysis.

Some studies have examined how the number of characters influences the word-recognition latency, Just et al. (cited in Just & Carpenter, 1987) found that the gaze duration on Chinese words increased with the number of characters in a word. In addition, Fang identified length effects for low-frequency words that were two to five characters long in a mixed task of lexical decision and stimulus identification (Fang, 1994, Experiment 1), and for foreign geographical and biographical names that were two to four characters long in a categorization task and a target detection task (Fang, 2003, Experiments 1 and 2). However, length effects were not found for high-frequency words that were two and four characters long in a mixed task of lexical decision and stimulus identification (Fang, 1994, Experiment 2).

The studies reviewed above on the character-complexity and word-length effects in Chinese script suggest that Chinese adult readers take more time to process characters comprising more strokes, especially for low-frequency characters, and show longer response latencies for low-frequency words comprising more characters. In addition, the reaction time of Chinese dyslexic children was also affected by character complexity for the number of strokes. However, it remains unclear whether there is a character-complexity effect for the number of radicals. In addition, the developmental changes in the analogous length effects from beginning to mature Chinese readers remain unknown. Thus, the present study was designed to systematically investigate developmental changes in character-complexity and word-length effects when decoding Chinese script. Three experiments were conducted in this study: Experiments 1, 2, and 3 used strokes, radicals, and characters as the unit of analysis, respectively, to explore whether there are developmental

changes in character-complexity and word-length effects in Chinese written language. The results of this study are useful for elucidating how the nature of Chinese-character recognition and Chinese-word recognition changes with the development of reading ability, and also the extent to which the explanation of word-length effect applied to alphabetic written language can be generalized to Chinese script.

Experiment 1

The purpose of Experiment 1 was to examine whether developmental changes in the character-complexity effect for the number of strokes occur for the Chinese written language. This experiment adopted a lexical recognition task that was modified from the lexical decision task by Hue and Tzeng (2000). These two tasks differ only with regard to the type of instruction: in a lexical decision task, participants are usually asked to decide as rapidly as possible whether strings of letters presented on a screen are real words, whereas in a lexical recognition task, participants are instructed to indicate whether they know the stimulus on the screen by pressing a key. It is more difficult for children to understand the instruction of a lexical decision task, so a lexical recognition task was used in the present study.

Method

Participants

The participants were 25 second graders (14 boys, 11 girls; mean age 8.1 years), 24 fourth graders (13 boys, 11 girls; mean age 10.0 years), 24 sixth graders (13 boys, 11 girls; mean age 12.0 years), and 25 university students (4 males, 21 females; mean age 19.6 years). The participants in the first three groups were selected from a primary school in Taipei County, Taiwan. They all exhibited average decoding abilities for their grade level: their T scores on the Graded Chinese Character Recognition Test (Huang, 2001) were all in the range 49–51. The university students were recruited from enrollees in an introductory psychology class at the National Taiwan Normal University, Taipei, Taiwan.

Design and stimuli

The stimuli used were 45 real characters and 45 pseudocharacters. The 45 real characters comprised 15 low-complexity, 15 moderate-complexity, and 15 high-complexity characters, for which the mean numbers of strokes were 5.53, 13.40, and 21.33, respectively (see Appendix Table 4). All characters were level-A (i.e., high-frequency) characters, as defined by a frequency count of primary-school reading material (National Institute for Compilation and Translation, 1999)–they exhibited frequencies higher than 196 in a corpus of 1,419,219 Chinese characters of primary-school reading material. The mean frequencies of the low-, moderate-, and high-complexity characters were 556.07, 557.60, and 557.13, respectively.

The manipulation was checked by performing two analyses of variance (ANOVAs), which showed that the number of strokes differed between the three groups of characters, $F(2, 42) = 589.78$, $MSE = 1.59$, $p < .001$. The frequencies of the three groups of characters did not differ ($F < 1$).

The 45 pseudocharacters comprised 15 low-complexity, 15 moderate-complexity, and 15 high-complexity characters, for which the mean numbers of strokes were 5.53, 13.33, and 21.47, respectively. The pseudocharacters contain real components of Chinese characters, and the components are in the correct position. In other words, the structure of the pseudocharacters follows Chinese orthographic rules. However, these are nonexistent Chinese characters. The pseudocharacters were included to avoid participants pressing "Yes" without making a judgment.

The characters were set up in standard Kai Shu (標楷) font. Each character was presented on the 13-inch screen of an IBM 390E notebook computer at a size of approximately 6 cm high and 6 cm wide, and appeared white on a black background to minimize flicker.

A 4×3 (grade level $\times$ character complexity) mixed-factor design was adopted. The between-subjects factor was grade level (grades two, four, six, and university), and the within-subjects factor was character complexity for the number of strokes (low, moderate, and high complexity). All participants were tested with the complete set of 90 characters, and the order of stimuli was randomized for each participant.

Procedure

All participants were tested individually. Prior to data collection, the primary-school students were shown the 45 real characters and 45 pseudocharacters on index cards. They were requested to answer whether or not they knew the characters on the index cards. If they stated that they knew a character, this was confirmed by asking them to read it aloud. If a student had difficulty with this, the student was helped in pronouncing the character, and the character was subsequently reintroduced until the student could recognize all of the characters successfully. This procedure was carried out to ensure that the primary-school students could read the characters to be tested (Samuels et al., 1978).

The experimental procedure was a lexical recognition task. Each trial began with a beep sound signifying the ready signal, and a fixation cross (+) was displayed at the center of the screen for 1000 ms. The fixation cross then disappeared for 1100 ms and the stimulus was displayed for a maximum of 8 seconds. What the participants had to do was to indicate as rapidly as possible whether they knew the character by pressing the "Yes" key.

Ten practice trials were presented at the beginning of the task. Five high-frequency real characters and five pseudocharacters were used in the practice trials. The practice stimuli were different from the 45 characters and 45 pseudocharacters to be tested. The order of the practice stimuli was randomized. Each participant had to exhibit an accuracy rate of at least 80% in order to pass the practice trials; otherwise they had to practice again. The entire experimental procedure was controlled by a program designed by Chen and Cho (2000).

Results

The analyses were based on the data of the 45 real characters. All incorrect responses were excluded from the analysis of latency. In addition, outliers (defined as a latency of less than 200 ms or more than three standard deviations away from a subject's condition mean) were also eliminated (Bush, Hess, & Wolford, 1993). This procedure eliminated 3.99% of the data. The mean and standard deviation of the latency and the percentage of correct responses to real characters as functions of character complexity across grade level are listed in Table 1. The data in Table 1 suggest that there was no trade-off between speed and accuracy in this lexical recognition task.

The latency data were analyzed using two-way ANOVAs of mixed design. The two factors were the grade level (grades two, four, six, and university) and the complexity of characters for the number of strokes (low, moderate, and high complexity). The first factor was manipulated between subjects and the second factor was manipulated within subjects. The results revealed main effects for the grade level [$F_1(3, 94) = 63.29$, $MSE = 55,273.0$, $\eta^2 = .67$, $p < .001$; $F_2(3, 126) = 357.58$, $MSE = 6,128.96$, $\eta^2 = .90$, $p < .001$ (where F_1 is the result of analysis across subject means and F_2 is the result of analysis across item means)], and the complexity of characters [$F_1(2, 188) = 18.14$, $MSE = 5,193.07$, $\eta^2 = .16$, $p < .001$; $F_2(2, 42) = 3.80$, $MSE = 14,086.66$, $\eta^2 = .15$, $p < .05$]. The interaction effect was also significant [$F_1(6, 188) = 7.28$, $MSE = 5,193.07$, $\eta^2 = .19$, $p < .001$; $F_2(6, 126) = 3.54$, $MSE = 6,128.96$, $\eta^2 = .14$, $p < .01$].

An adjusted comparison-wise error rate ($p = .05/7 = .007$) was used to analyze the simple effect of character complexity, in order to control for type I error inflation. The analysis showed that the simple effect of character complexity

Table 1 Mean and standard deviation (SD) of latencies (reaction time, RT) in milliseconds and the percentage of correct responses to real characters as functions of character complexity and grade level (Experiment 1)

		Low complexity	Moderate complexity	High complexity
Grade 2	Mean RT	987.58	1053.66	1158.62
	SD	181.09	180.22	243.29
	% Correct	95.68	96.44	92.96
Grade 4	Mean RT	875.16	906.85	931.18
	SD	162.95	175.64	159.11
	% Correct	98.00	99.42	97.46
Grade 6	Mean RT	712.88	695.54	719.96
	SD	135.76	69.48	99.46
	% Correct	98.00	96.87	95.75
University	Mean RT	568.10	574.27	578.57
	SD	69.39	82.38	100.29
	% Correct	98.32	99.72	99.20
Total	Mean RT	785.77	807.71	847.52
	SD	214.41	230.98	273.55
	% Correct	97.49	98.11	96.34

was significant in second graders [$F_1(2, 188) = 35.82$, $MSE = 5{,}193.07$, $p < .007$; $F_2(2, 168) = 13.10$, $MSE = 8{,}118.38$, $p < .007$], but not in fourth graders [$F_1(2, 188) = 3.65$, $MSE = 5{,}193.07$, $p = .028$; $F_2(2, 168) = 1.17$, $MSE = 8{,}118.38$, $p = .31$], sixth graders, or university students (all $F_s < 1$). Post-hoc comparisons using the Bonferroni correction procedure ($p = .007/3 = .0023$) revealed that the second graders took longer to recognize high-complexity characters (1159 ms) than low- and moderate-complexity characters (988 ms and 1054 ms, respectively).

The results of an analysis of accuracy data showed a main effect for the grade level [$F_1(3, 94) = 9.12$, $MSE = 25.91$, $\eta^2 = .23$, $p < .001$; $F_2(3, 126) = 9.93$, $MSE = 14.02$, $\eta^2 = .19$, $p < .001$]; however, the main effect of the number of strokes was not reliable found, which was significant across subject means [$F_1(2, 188) = 3.75$, $MSE = 21.13$, $\eta^2 = .04$, $p < .05$] but not significant across item means [$F_2(2, 42) = 1.46$, $MSE = 32.56$, $p = .24$]. Furthermore, the interaction effect was not significant either across subject means [$F_1(6, 188) = 1.13$, $MSE = 21.13$, $p = .35$] or across item means [$F_2(6, 126) = 1.00$, $MSE = 14.02$, $p = .43$].

Discussion

As shown in Fig. 1, the response latencies of second graders increased with the character complexity, whereas those of fourth graders, sixth graders, and university students did not. This developmental change in character recognition is similar to previous findings for alphabetic writing systems (Aghababian & Nazir, 2000; Bijeljac-Babic et al., 2004; Samuels et al., 1978; Su, 1997), and was also replicated by a follow-up study conducted by Chen and Su (2009). Chen and Su recruited poor, average, and good third-grade readers, and adopted the same task and stimuli as in our Experiment 1. They found that both reading-ability and character-complexity effects were significant, but that the interaction effect was not significant. The differences in latencies between low- and high-complexity characters were largest for poor readers, followed by average and then good readers.
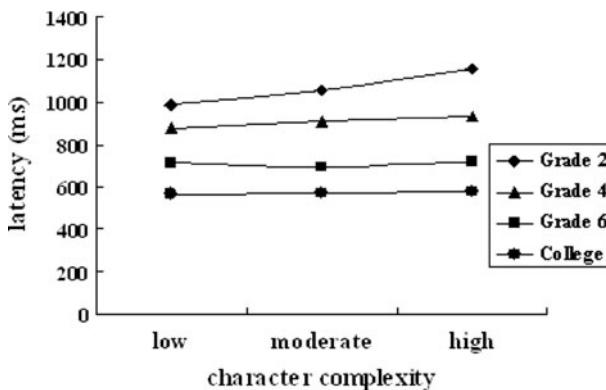

Fig. 1 Latency of lexical recognition versus character complexity for grades 2, 4, and 6, and university students (Experiment 1)

Nevertheless, the finding that university students did not exhibit character-complexity effects differs from those of some other studies (Just & Carpenter, 1987; Leong et al., 1987; Tan & Peng, 1990). This discrepancy might be due to the printed frequencies of the characters selected as stimuli. The frequencies of characters were not reported by either Just et al. (cited in Just & Carpenter, 1987) or Tan and Peng (1990), and hence it is unclear whether the stimuli were high- or low-frequency characters. In the study of Leong et al. (1987), the characters had a difficulty level at or below grade 6, and they were selected according to the printed frequency (high or low). Although a character-complexity effect was observed for low-frequency stimuli, this was not present for high-frequency characters. Because the stimuli selected in our Experiment 1 were high-frequency characters, the university students tended to process these characters as a whole rather than performing analytical processing.

The results of our Experiment 1 and Chen and Su (2009, Experiment 1) suggest that beginning Chinese readers process characters analytically, as did second graders in our Experiment 1 and third graders in their study. The word-length effect in alphabetic script suggests the use of letter-by-letter processing, but the character-complexity effect for the number of strokes implies feature processing, which means the visual analysis of primitive features, such as horizontal lines, vertical lines, intersections, dots, open curves, etc. Furthermore, more-skilled Chinese readers appear to process characters holistically, as did sixth graders and university students in our experiment. That is, the way in which Chinese readers process characters varies with reading ability. The effect of frequency found in adult studies reflects the effect of learning in a sense. The extent of parallel word processing depends on the number of times a word has been identified after a single fixation on it (Nazir et al., 2004). High-frequency characters are more likely to be read, and hence the degree of parallel processing is higher for high-frequency characters than for low-frequency characters.

One limitation of our Experiment 1 is that the number of radicals in the stimuli was inevitable confounded by the number of strokes. The mean number of radicals was significantly higher in high-complexity characters (3.93) than in moderate-complexity (2.73) and low-complexity (2.00) characters. If the results found in this experiment were attributable to the number of radicals, then controlling the stroke number and the frequency of characters should reveal the effect of the number of radicals. This was examined in Experiment 2.

Experiment 2

The purpose of Experiment 2 was to determine whether there are developmental changes in the character-complexity effect for the number of radicals on Chinese-character recognition. The definition of radical adopted in this experiment was the smallest stroke pattern comprising one or more strokes and that recurred independently in different characters. The lexical recognition task as described for Experiment 1 was adopted.

Method

Participants

The participants were the same as for Experiment 1.

Design and stimuli

The stimuli used were 39 real characters and 39 pseudocharacters. The 39 real characters comprised 13 two-radical, 13 three-radical, and 13 four-radical characters (see Appendix Table 5). All characters were level-A (i.e., high-frequency) characters, as defined by a frequency count of primary-school reading material, as mentioned in Experiment 1 (National Institute for Compilation and Translation, 1999). The mean frequencies of the two-radical, three-radical, and four-radical characters were 465.00, 469.54, and 465.92, respectively, and the mean numbers of strokes in these characters were 13.62, 14.85, and 15.69, respectively. The manipulation was checked by performing two ANOVAs, which showed that neither frequencies nor strokes differed between the three groups of characters [$F < 1$ and $F(2, 36) = 2.24$, $MSE = 6.32$, $p = .121$, respectively].

The 39 pseudocharacters comprised 13 two-radical, 13 three-radical, and 13 four-radical pseudocharacters, for which the mean numbers of strokes were 13.23, 14.31, and 15.54, respectively. The pseudocharacters contain real components of Chinese characters, and the components are in the correct position. In other words, the structure of the pseudocharacters follows Chinese orthographic rules. However, these are nonexistent Chinese characters. The pseudocharacters were included to avoid participants pressing "Yes" without making a judgment.

The font, size, and color of the stimuli were all the same as in Experiment 1. All of the stimuli were presented on an IBM 390E notebook computer.

A 4×3 (grade level $\times$ number of radicals) mixed-factor design was adopted. The between-subjects factor was grade level (grades two, four, six, and university), and the within-subjects factor was the number of radicals in characters (two, three, and four radicals). All participants were tested with the complete set of 78 characters, and the order of stimuli was randomized for each participant.

Procedure

The procedure was the same as in Experiment 1.

Results

The analyses were based on the data of the 39 real characters. All incorrect responses were excluded from the analysis of latency, as were outliers (defined as a latency of less than 200 ms or more than three standard deviations away from a subject's condition mean) (Bush et al., 1993). This procedure eliminated 4.45% of

Table 2 Mean and standard deviation (SD) of latencies (reaction time, RT) in milliseconds and the percentage of correct responses to real characters as functions of the number of radicals and grade level (Experiment 2)

		Two radicals	Three radicals	Four radicals
Grade 2	Mean RT	1188.17	1197.42	1133.05
	SD	300.86	267.19	251.83
	% Correct	96.62	94.46	93.54
Grade 4	Mean RT	984.18	982.87	978.12
	SD	220.60	179.20	226.47
	% Correct	95.19	98.08	96.47
Grade 6	Mean RT	744.54	772.29	737.34
	SD	129.70	126.62	130.24
	% Correct	96.47	96.79	98.40
University	Mean RT	605.80	603.15	603.81
	SD	89.92	83.80	90.14
	% Correct	97.23	98.77	99.08
Total	Mean RT	881.01	889.17	863.19
	SD	301.66	286.26	277.81
	% Correct	96.39	97.02	96.86

the data. The mean and standard deviation of the latency and the percentage of correct responses to real characters as functions of the number of radicals across grade level are listed in Table 2. The data in Table 2 suggest that there was no trade-off between speed and accuracy in this lexical recognition task.

The latency data were analyzed using two-way ANOVAs of mixed-factor design. The two factors were the grade level (grades two, four, six, and university) and the number of radicals in characters (two, three, and four radicals). The first factor was manipulated between subjects and the second factor was manipulated within subjects. The results showed a main effect for the grade level $[F_1(3, 94) = 48.03, MSE = 97,654.59, \eta^2 = .61, p < .001; F_2(3, 108) = 423.51, MSE = 5,952.11, \eta^2 = .92, p < .001]$; however, the main effect of the number of radicals was not reliably identified, which was significant across subject means $[F_1(2, 188) = 3.31, MSE = 5,155.23, \eta^2 = .03, p < .05]$, but not across item means $(F_2 < 1)$. The mean response latencies of two-, three-, and four-radical characters were 881 ms, 889 ms, and 863 ms respectively, which did not reveal a length effect. Furthermore, the interaction effect was not significant across either subject means $[F_1(6, 188) = 1.39, MSE = 5,155.23, p = .222]$ or item means $(F_2 < 1)$.

The results of an analysis of accuracy data showed a main effect for the grade level $[F_1(3, 94) = 4.89, MSE = 32.51, \eta^2 = .14, p < .01; F_2(3, 108) = 4.06, MSE = 20.55, \eta^2 = .10, p < .01]$; however, the main effect of the number of radicals was not significant across either subject means or item means $(F_s < 1)$. Furthermore, the interaction effect was not reliably identified, which was significant across subject means $[F_1(6, 188) = 2.22, MSE = 22.82, \eta^2 = .07, p < .05]$ but not across item means $[F_2(6, 108) = 1.31, MSE = 20.55, p = .26]$.

Discussion

In Experiment 2 the number of radicals in a character had no reliable effect on the latencies of character recognition. Even though the effect of the number of radicals was significant across subject means, the mean response latencies of two-, three-, and four-radical characters were 881 ms, 889 ms, and 863 ms respectively, which did not reveal a length effect. This result suggests that the findings of Experiment 1 are not attributable to the number of radicals. These results were replicated by a follow-up study (Chen & Su, 2009, Experiment 2) for poor, average, and good third-grade readers, but differ from the findings of Fang (1994) and Chen et al. (1996). As mentioned in the Introduction, the radical factor was confounded by the stroke factor in Experiment 3 of Fang (1994), whereas in our Experiment 2– which controlled for the number of strokes – the mean number of strokes did not differ significantly between the three levels of the radical factor. Hence, the differing response latencies in Experiment 3 of Fang (1994) might be at least partially attributable to the number of strokes rather than the number of radicals.

In addition, the results obtained by Chen et al. (1996) might be partly due to the nature of a simultaneous same-different comparison task in which participants have to judge the physical identity of the two stimuli, and also partly due to the "different" pairs that were manipulated by the proportion of mismatching radicals in each character. In contrast, our Experiment 2 adopted the lexical recognition task modified from the lexical decision task that is commonly used in word-recognition research.

Nevertheless, caution is still required when drawing conclusions from our experiment, since some other studies have shown that radical activation is involved in character recognition (e.g., Taft & Zhu, 1997), though they were not designed to investigate the character-complexity effect for the number of radicals. Although the results of our experiment appear to indicate that Chinese readers do not process characters on a radical-by-radical basis, at least for high-frequency moderate-complexity characters, this does not mean that radical activation is not involved in the process of character recognition. However, it remains to be determined whether this finding only applies to high-frequency moderate-complexity characters (mean number of strokes is around 14), or also to high-frequency high-complexity characters and low-frequency characters.

In addition, one follow-up experiment is also suggested to confirm the results obtained in Experiment 1 do result from the effect of stroke number rather than the effect of radical number. This experiment should test if the manipulation of the number of strokes while the number of radicals is controlled still reveals a significant effect independently.

Experiment 3

The purpose of Experiment 3 was to determine whether there are developmental changes in the word-length effect for the number of characters on Chinese-word recognition. The lexical recognition task as described for Experiment 1 was adopted.

Method

Participants

The participants were the same as in Experiment 1.

Design and stimuli

The stimuli were 45 real words and 45 nonwords. The 45 real words comprised 15 two-character, 15 three-character, and 15 four-character words (see Appendix Table 6). Because no data were available on the word frequency count based on primary-school reading material, the frequencies of all words selected in this experiment were estimated from the frequency count in the Corpus-Based Frequency Count of Words in Journal Chinese surveyed by Academia Sinica in Taiwan (Chinese Knowledge Information Processing Group, 1993). This frequency count of words was based on adult reading material in general. In addition, since it was difficult to find 15 high-frequency four-character words that would be expected to be known by primary-school students, moderate- to high-frequency words were used in this experiment. The mean frequencies of two-character, three-character, and four-character words were 304.73, 304.73, and 304.80, respectively, in a corpus of 9,529,233 Chinese words in journals. The frequencies did not differ among the three groups of words ($F < 1$). The stimuli were presented on an IBM 390E notebook computer as in Experiment 1.

The 45 nonwords comprised 15 two-character, 15 three-character, and 15 four-character nonwords. The characters of the nonwords were randomly selected from the 45 real words, but the combinations of the characters are nonexistent Chinese words. The nonwords were included to avoid participants pressing "Yes" without making a judgment.

A 4×3 (grade level $\times$ word length) mixed-factor design was adopted. The between-subjects factor was grade level (grades two, four, six, and university), and the within-subjects factor was word length (two, three, and four characters). All participants were tested with the complete set of 90 stimuli, and the order of stimuli was randomized for each participant.

Procedure

The procedure was the same as in Experiment 1.

Results

The analyses were based on the data of the 45 real words. All incorrect responses were excluded from the analysis of latency, as were outliers (defined as a latency of less than 200 ms or more than three standard deviations away from a subject's condition mean) (Bush et al., 1993). This procedure eliminated 2.70% of the data. The mean and standard deviation of the latency and the percentage of correct

Table 3 Mean and standard deviation (SD) of latencies (reaction time, RT) in milliseconds and the percentage of correct responses to real words as functions of the number of characters and grade level (Experiment 3)

		Two characters	Three characters	Four characters
Grade 2	Mean RT	1189.01	1149.45	1233.17
	SD	248.66	238.13	283.75
	% Correct	95.47	98.40	97.87
Grade 4	Mean RT	1007.39	982.46	987.21
	SD	174.30	191.50	154.25
	% Correct	96.67	99.17	99.72
Grade 6	Mean RT	800.09	771.47	821.20
	SD	112.37	75.49	113.45
	% Correct	97.50	99.17	99.44
University	Mean RT	617.73	626.01	654.60
	SD	71.24	76.39	91.32
	% Correct	99.20	99.73	98.40
Total	Mean RT	903.55	882.46	924.45
	SD	272.26	257.62	278.61
	% Correct	97.21	99.12	98.84

responses to real words as functions of word length across grade level are listed in Table 3. The data in Table 3 suggest that there was no trade-off between speed and accuracy in this lexical recognition task.

The latency data were analyzed using two-way ANOVAs of mixed-factor design. The two factors were the grade level (grades two, four, six, and university) and the word length (two, three, and four characters). The first factor was manipulated between subjects and the second factor was manipulated within subjects. The results showed a main effect for grade level [$F_1(3, 94) = 59.56$, $MSE = 73{,}044.38$, $\eta^2 = .66$, $p < .001$; $F_2(3, 126) = 713.03$, $MSE = 3{,}716.79$, $\eta^2 = .94$, $p < .001$]; however, the main effect of word length was not reliably identified, which was significant across subject means [$F_1(2, 188) = 6.78$, $MSE = 6{,}277.31$, $\eta^2 = .07$, $p < .01$], but not across item means [$F_2(2, 42) = 1.39$, $MSE = 19{,}844.38$, $p = .261$]. Furthermore, the interaction effect was not significant across subject means [$F_1(6, 188) = 1.55$, $MSE = 6{,}277.31$, $p = .165$] or item means [$F_2(6, 126) = 1.61$, $MSE = 3{,}716.79$, $p = .150$].

Based on the data applying to across subject means, the simple effect of word length was significant in second graders [$F_1(2, 188) = 6.99$, $MSE = 6{,}277.31$, $p < .007$] but not in fourth graders ($F_1 < 1$), sixth graders [$F_1(2, 188) = 2.38$, $MSE = 6{,}277.31$, $p = .10$], or university students [$F_1(2, 188) = 1.49$, $MSE = 6{,}277.31$, $p = .23$]. Post-hoc comparisons using the Bonferroni correction procedure ($p = .007/3 = .0023$) only revealed that the latency of three-character words (1149 ms) was marginally significantly shorter than that of four-character words (1233 ms) in second graders, with there being no other significant differences in paired comparisons.

The results of an analysis of accuracy data showed a main effect for the grade level [$F_1(3, 94) = 4.99$, $MSE = 9.73$, $\eta^2 = .14$, $p < .01$; $F_2(3, 126) = 3.17$, $MSE = 9.21$, $\eta^2 = .07$, $p < .05$] and a main effect of the number of characters

$[F_1(2, 188) = 12.11, MSE = 8.67, \eta^2 = .11, p < .001; F_2(2, 42) = 4.71, MSE = 13.67, \eta^2 = .18, p < .05]$. Furthermore, the interaction effect was not reliably identified, which was significant across subject means $[F_1(6, 188) = 2.25, MSE = 8.67, \eta^2 = .07, p < .05]$ but not across item means $[F_2(6, 126) = 1.28, MSE = 9.21, p = .27]$. According to the analysis of post-hoc comparisons, although the accuracy rate was significantly lower for two-character words (97.21%) than for three-character (99.12%) and four-character (98.86%) words, this did not show a word-length effect where longer words were harder.

Discussion

As mentioned in Method section, since no data were available on the word frequency count based on primary-school reading material, the frequencies of all words selected in Experiment 3 were estimated from an adult corpus of word frequency. However, as Table 3 shown, the accuracy rates of two-character, three-character, and four-character words for second to sixth graders ranged from 95.47 to 99.72 percent. If the selected high frequency words for adults were actually low frequency words for children, the accuracy rates would not be that high.

Based on the analysis of latency data, Experiment 3 showed that the number of characters had a partial effect on word recognition for second graders, but not for fourth graders, sixth graders, or university students. This results were not identical to those of the follow-up study conducted by Chen and Su (2009, Experiment 3). In their study, the main effects of reading ability and word length were both significant. However, the interaction effect was significant across item means but only marginally significant across subject means. For the third-grade poor readers, the latencies differed significantly between two- and three-character words, two- and four-character words, and three- and four-character words, with longer words taking more time to process than shorter words. For the average and good third-grade readers, the latencies differed significantly between two- and four-character words and between three- and four-character words; however, the reaction time did not differ significantly between two- and three-character words. The difference in findings between these two studies might be partly attributable to the characteristics of the participants recruited. The socioeconomic status of the school district was higher for the children of the present study than for those in the study of Chen and Su (2009), and the literature on reading research contains many correlational studies showing that family resources influence the reading abilities of children. It appears that beginning readers and non-skilled readers demonstrate a clear word-length effect, but this becomes less apparent as reading skills mature.

Moreover, the results of more skilled readers in the present study are similar to those of Fang (1994, Experiment 2), but differ from those of some other studies (Fang, 1994, Experiment 1; Fang, 2003, Experiments 1 and 2; Just & Carpenter, 1987). The main reason for this discrepancy could be the printed frequency of the words selected as stimuli. The frequency of words was not reported in the study of Just et al. (cited in Just & Carpenter, 1987). The stimuli selected in Experiment 1 of

Fang (1994) were low-frequency words, and the words used in Experiments 1 and 2 of Fang (2003) were foreign geographical and biographical names. However, the stimuli used in Experiment 2 of Fang (1994) were high-frequency words. Our Experiment 3 also selected moderate- to high-frequency words as stimuli. It appears that length effects are present when low-frequency words or unfamiliar words (such as foreign geographical and biographical names) are used, but are less likely to be found when high-frequency words are adopted and skilled readers are recruited. As discussed for Experiment 1, the effect of frequency reflects reading experience or the effect of learning in a broad sense. Because high-frequency words are more likely to be read, the degree of parallel processing is higher for high-frequency words than for low-frequency words.

Nevertheless, one limitation of our Experiment 3 is that the numbers of strokes and radicals were not controlled. It is difficult to select three sets of moderate-to-high-frequency Chinese words with different lengths, and also control the numbers of strokes and radicals. In the cases of fourth graders, sixth graders, and university students, since the word-length effects were not significant, this result cannot be attributed to the number of strokes or radicals. In the case of second graders, because the effect of the number of radicals was not significant in Experiment 2, the word-length effect found in Experiment 3 might not be due to the number of radicals. However, the stroke effect in Experiment 1 was significant for second graders, and hence further research is needed to clarify whether the word-length effect found here was affected by the number of strokes or the number of characters.

General discussion

The present study conducted three experiments to examine the developmental changes in character-complexity and word-length effects when reading Chinese script. The major findings are based on the analysis of latency data for two reasons. First, word-length effect refers to response latency increases with the length of words. Secondly, the accuracy rates were quite high in the present study. In the analysis of accuracy data, neither the main effects of analogous word-length nor the interaction effects between grade and analogous word-length were reliably found in all three experiments.

The results of the three experiments reported in this article can be summarized as follows. When high-frequency characters were selected as stimuli and the character complexity was defined as the number of constituent strokes of characters, the response latencies of second graders increased with the character complexity, whereas those of fourth graders, sixth graders, and university students did not. However, when character complexity was defined as the number of constituent radicals of characters, no effect of character complexity was found. Moreover, when word length was defined as the number of constituent characters of words, only a partial length effect was found for second graders. Although these three experiments were conducted in a single session, the order of the three experiments was the same

for all participants. In addition, all participants were allowed to rest if they wanted to at the end of each experiment, and even second graders were able to finish the three experiments in about 20 min. Therefore, it appears that the reported results were not contaminated by a carry-over effect.

The developmental changes found in Experiment 1 are similar to the effects found in previous research on alphabetic writing systems. The results support the hypothesis of Samuels et al. (1978) that beginning readers process a word on a component basis, whereas more-skilled readers process a word in a more holistic way depending on their decoding skill. When reading Chinese script, beginning readers recognize characters analytically, and hence their response latencies increase with the number of constituent strokes. However, the decoding process gradually changes from analytic to holistic as their reading skill develops. Therefore, the latencies of the fourth graders, sixth graders, and university students did not increase with the number of strokes. Bijeljac-Babic et al. (2004) considered the developmental transition from less-skilled to skilled word recognition to represent a shift from sequential to parallel processing. They postulated that less-skilled readers process a stimulus string in an alphabetic writing system serially from left to right, whereas skilled readers process a letter string in chunks. However, left-to-right serial processing is not directly applicable to Chinese-character decoding. As mentioned in the introduction, Chinese characters are arranged in squares of similar size irrespective of the number of constituent strokes or radicals, and hence character recognition by beginning readers is better represented by analytical processing than by serial processing.

In addition, the word-length effect in alphabetic scripts implies letter-by-letter processing, whereas the character-complexity effect for the number of strokes in Chinese script implies feature processing. Feature processing by beginning Chinese readers takes time when they are attempting to recognize characters. As their reading abilities develop, the character-recognition process might gradually change from feature processing to feature-cluster processing or holistic processing of the whole character. However, whether or not the feature-cluster processing is equivalent to radical processing it is still not clear from the present study. Nevertheless, the findings of Experiment 1 have implications for education. That is, if readers process characters analytically, as do beginning readers and nonskilled readers, their reading speed must be low and their reading comprehension should be affected due to the limitation of cognitive resources.

In addition, the present study found that the response latencies of second graders in character recognition increased with the number of strokes in a character but not with the number of radicals therein. These results appear to indirectly support the interactive constituency model of Chinese-character identification proposed by Perfetti and Tan (1999). The interactive constituency model assumes that two levels of processing are involved in Chinese-character recognition. The first level is stroke analysis, in which each stroke and its positional relationships with other strokes are detected. These detected features send activation to the second level, which consists of four constituent representation subsystems: (1) the character orthographic subsystem, (2) the noncharacter orthographic subsystem, (3) the phonological subsystem, and (4) the meaning subsystem. The character orthographic subsystem

represents simple and compound characters, including those phonetic and semantic radicals that are themselves legal characters. The noncharacter orthographic subsystem stores radicals that are not independent characters. According to this model, the character-recognition latencies of beginning readers should increase with the number of strokes. That is exactly what was found in the present study.

Based on the interactive constituency model, the developmental progress in decoding skills found in Experiment 1 may be mainly due to a reduction in the contribution of activation of each stroke code to the character code. Therefore, for less-skilled readers, the character code requires outputs from relatively well-activated constituent stroke codes. In contrast, for skilled readers, the position and feature information reach the character code before the processing of individual stroke codes is complete.

Caution is required when drawing conclusions from Experiment 2 finding no effect of the number of radicals, because some other adult studies found that radical activation was involved in character recognition (e.g., Taft & Zhu, 1997); however, those studies were not designed to investigate the character-complexity effect for the number of radicals. The results of this experiment suggest that Chinese readers do not process characters on a radical-by-radical basis, at least for high-frequency moderate-complexity characters, but this does not mean that radical activation is not involved in the process of character recognition. However, further research is needed to examine whether or not this finding only applies to high-frequency moderate-complexity characters, or also to high-frequency high-complexity characters and low-frequency characters.

In relation to the effects of the number of characters, the findings of both the previous and the present research appear to indicate that length effects are present for skilled readers when low-frequency or unfamiliar words (e.g., foreign geographical and biographical names) are used, but not when high-frequency words are used. However, for beginning readers and nonskilled readers, word-length effects were found even when using high-frequency words. These findings can be explained partly by the top two representation levels (i.e., word level and character level) of the multilevel interactive-activation model (Taft & Zhu, 1997). This model assumes four levels of representation: strokes, radicals, characters, and words. When words are encountered, the word code requires information from relatively well-activated constituent character codes. However, as reading abilities develop or reading experiences increase, the character information might reach the word code before the processing of individual character codes is complete. Therefore, when high-frequency words are encountered, a word-length effect is shown by beginning readers and nonskilled readers but not by skilled readers. However, even skilled readers show a word-length effect when low-frequency words are used.

In conclusion, developmental changes in the character-complexity effect were found when high-frequency characters were adopted as stimuli and character complexity was defined as the number of constituent strokes of characters. However, when character complexity was defined as the number of constituent radicals of characters, the complexity effect was not found in this study. In addition,

when word length was defined as the number of constituent characters of words, beginning readers showed a word-length effect for high-frequency words, but more-skilled readers did not. As our knowledge, this study is the first research designed to systematically investigate developmental changes in Chinese character-complexity and word-length effects, and demonstrated how the nature of Chinese character and word recognition changes with the development of reading skills. This study not only provides cross-language evidence supporting the hypothesis of Samuels et al. (1978), but also has implications for reading educators.

Appendix

See Tables 4, 5 and 6.

Table 4 The number of strokes and frequencies of low-, moderate-, and high-complexity characters used in Experiment 1

Low complexity			Moderate complexity			High complexity		
Character	Number of strokes	Frequency	Character	Number of strokes	Frequency	Character	Number of strokes	Frequency
加	5	1604	情	11	1593	體	23	1574
代	5	1153	連	11	1190	覺	20	1230
早	6	895	跑	12	890	歡	22	892
北	5	772	趕	14	749	讀	22	753
句	5	521	練	15	542	護	21	498
另	5	491	語	14	502	響	21	480
仁	4	462	需	14	486	鐵	21	477
切	4	416	隊	12	414	驗	23	397
冷	7	386	豬	15	378	顧	21	375
忍	7	339	靠	15	322	魔	21	336
似	7	325	盤	15	321	顯	23	326
孔	4	299	源	13	291	繼	20	295
志	7	273	換	12	283	鐘	20	278
均	7	206	輪	15	205	蘭	21	242
巧	5	199	碰	13	198	露	21	204
Mean	5.53	556.07	Mean	13.40	557.60	Mean	21.33	557.13
SD	1.19	394.21	SD	1.50	395.70	SD	1.05	395.05

Table 5 The number of strokes and frequencies of two-, three-, and four-radical characters used in Experiment 2

Two-radical characters			Three-radical characters			Four-radical characters		
Character	Number of strokes	Frequency	Character	Number of strokes	Frequency	Character	Number of strokes	Frequency
球	11	1055	結	12	1051	命	8	1029
期	12	742	積	16	841	照	13	781
養	15	672	整	16	635	隨	16	678
誰	15	627	類	19	615	寶	20	621
領	14	426	漸	14	447	慢	14	445
斯	12	419	貓	16	434	福	14	426
遇	13	418	葉	13	407	靜	16	405
鴨	16	398	暗	13	320	續	21	323
惡	12	307	諾	16	318	夢	14	307
確	15	290	趣	15	302	嘴	16	300
集	12	247	醒	16	281	模	15	291
雷	13	242	蓮	15	248	藍	18	246
顆	17	202	湖	12	205	藝	19	205
Mean	13.62	465.00	Mean	14.85	469.54	Mean	15.69	465.92
SD	1.85	246.01	SD	1.99	251.38	SD	3.40	243.47

Table 6 The frequency of two-, three-, and four-character words used in Experiment 3

Two-character words		Three-character words		Four-character words	
Word	Frequency	Word	Frequency	Word	Frequency
大樓	1384	新台幣	1276	中華民國	1384
房屋	855	計程車	1021	高速公路	849
金錢	396	消防隊	384	百貨公司	397
農田	359	加油站	350	紅十字會	361
茶葉	234	地下道	213	國家公園	234
說話	199	全壘打	201	警察局長	200
炸彈	176	來不及	177	環境衛生	178
聰明	173	時間表	173	交通安全	173
唱歌	137	科學家	136	野生動物	137
門窗	135	大自然	128	有線電視	135
飛行	133	夏令營	127	公共場所	133
獎品	110	紀念品	109	好不容易	111
願望	100	棒球場	98	日常生活	100
書店	93	大禮堂	92	身體健康	93
筆記	87	玩具槍	86	超級市場	87
Mean	304.73	Mean	304.73	Mean	304.80
SD	357.32	SD	356.87	SD	356.59

References

Aghababian, V., & Nazir, T. (2000). Developing normal reading skills: Aspects of the visual processes underlying word recognition. *Journal of Experimental Child Psychology, 76*, 123–150.

Bijeljac-Babic, R., Millogo, V., Farioli, F., & Grainger, J. (2004). A developmental investigation of word length effects in reading using a new on-line word identification paradigm. *Reading and Writing: An Interdisciplinary Journal, 17*, 411–431.

Bush, L. K., Hess, U., & Wolford, G. (1993). Transformations for within-subject designs: A Monte Carlo investigation. *Psychological Bulletin, 113*, 566–579.

Chen, Y. P., Allport, D. A., & Marshall, J. C. (1996). What are the functional orthographic units in Chinese word recognition: The stroke or the stroke pattern? *The Quarterly Journal of Experimental Psychology, 49A*, 1024–1043.

Chen, H.-C., & Cho, S.-L. (2000). 認知心理學實驗週邊組件的設計 [Peripheral equipment design for cognitive experiments]. *應用心理學報 [Chinese Journal of Applied Psychology], 8*, 41-60.

Chen, S.-C., & Liu, I.-M. (2000). Functional orthographic units in Chinese character recognition. *Acta Psychologica Sinica, 32*(supplement), 13–20.

Chen, J.-L., & Su, Y.-F. (2009, in press). 國小不同認字能力學童辨識中文字詞之字元複雜度效果與詞長效果研究 [The effect of decoding ability on Character-Complexity and Word-length effects in Taiwanese beginning readers]. *教育心理學報 [Bulletin of Educational Psychology].*

Cheng, C. M. (1981). 漢字的認知歷程 [Perception of Chinese characters]. *中華心理學刊 [Acta Psychologia Taiwanica], 23*, 137–154.

Cheung, H., & Ng, L. K. H. (2003). Chinese reading development in some major Chinese societies: An introduction. In C. McBride-Chang & H.-C. Chen (Eds.), *Reading development in Chinese children* (pp. 3–17). Westport, CT: Praeger.

Chiang, W.-C. (2003). The basic orthographic units in children's recognition of Chinese characters. *Chinese Journal of Psychology, 45*(1), 47–74.

Chinese Knowledge Information Processing Group. (1993). *新聞語料詞頻統計表:語料庫研究系列之二 [Corpus-based frequency count of words in journal Chinese: Corpus-based research series no. 2].* Taipei, Taiwan: Institute of Information Science Academia Sinica.

De Luca, M., Borrelli, M., Judica, A., Spinelli, D., & Zoccolotti, P. (2002). Reading words and pseudo-words: An eye movement study of developmental dyslexia. *Brain and Language, 80*, 617–626.

De Luca, M., Di Pace, E., Judica, A., Spinelli, D., & Zoccolotti, P. (1999). Eye movement patterns in linguistic and non-linguistic tasks in developmental surface dyslexia. *Neuropsychologia, 37*, 1407–1420.

Fan, K. Y., Gao, J. Y., & Ao, X. P. (1984). Pronunciation principles of the Chinese character and alphabetic writing scripts. *Chinese character reform*, 3 (pp. 19–22). Beijing: National Commission of Chinese Character Reform.

Fang, S.-P. (1994). English word length effects and the Chinese character-word difference: Truth or myth? *Chinese Journal of Psychology, 36*(1), 59–80.

Fang, S.-P. (2003). Are there differential word length effects in the two visual fields? *Brain and Language, 85*, 467–485.

Feldman, L. B., & Siok, W. W. T. (1999). Semantic radicals contribute to the visual identification of Chinese characters. *Journal of Memory and Language, 40*, 559–576.

Ferrand, L. (2000). Reading aloud polysyllabic words and nonwords: The syllabic length effect reexamined. *Psychonomic Bulletin and Review, 7*, 142–148.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2002). *Cognitive neuroscience: The biology of the mind.* New York, NY: Norton & Company.

Ho, C. S.-H., Ng, T.-T., & Ng, W.-K. (2003). A "radical" approach to reading development in Chinese: The role of semantic radicals and phonetic radicals. *Journal of Literacy Research, 35*, 849–878.

Honorof, D. N., & Feldman, L. (2006). The Chinese character in psycholinguistic research: Form, structure, and the reader. In P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng (Eds.), *The handbook of east Asian psycholinguistics: Vol. 1. Chinese* (pp. 195–208). New York, NY: Cambridge University Press.

Huang, H.-S. (2001). *中文年級認字量表 [Graded Chinese character recognition test].* Taipei, Taiwan: Psychological Publishing Company.

Hue, C. W., & Tzeng, A. K. (2000). Lexical recognition task: A new method for the study of Chinese character recognition. *Acta Psychologia Sinica, 32*, 60–65.

Jared, D., & Seidenberg, M. S. (1990). Naming multisyllabic words. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 92–105.

Juphard, A., Carbonnel, S., & Valdois, S. (2004). Length effect in reading and lexical decision: Evidence from skilled readers and a developmental dyslexic participant. *Brain and Cognition, 55*, 332–340.

Just, M. A., & Carpenter, P. A. (1987). Orthography: Its structure and effects on reading. In M. A. Just & P. A. Carpenter (Eds.), *The psychology of reading and language processing* (pp. 287–325). Newton, MA: Allyn and Bacon.

Leck, K. J., Weekes, B. S., & Chen, M. J. (1995). Visual and phonological pathways to the lexicon: Evidence from Chinese readers. *Memory and Cognition, 23*, 468–476.

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

Martens, V. E. G., & de Jong, P. F. (2006). The effect of word length on lexical decision in dyslexic and normal reading children. *Brain and Language, 98*, 140–149.

National Institute for Compilation and Translation. (1999). 國民小學常用字彙研究字頻總表 *[A frequency count of primary school reading material]*. Taipei, Taiwan: National Institute for Compilation and Translation.

Nazir, T., Ben-Boutayab, N., Decoppet, N., Deutsch, A., & Frost, R. (2004). Reading habits, perceptual learning, and the recognition of printed words. *Brain and Language, 88*, 294–311.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English lexicon project. *Psychonomic Bulletin and Review, 13*, 45–52.

Perfetti, C. A., & Liu, Y. (2006). Reading Chinese characters: Orthography, phonology, meaning, and the lexical constituency model. In P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng (Eds.), *The handbook of east Asian psycholinguistics: Vol. 1. Chinese* (pp. 225–236). NY: Cambridge University Press.

Perfetti, C. A., & Tan, L. H. (1999). The constituency model of Chinese word identification. In J. Wang, A. W. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 115–134). Mahwah, NJ: Lawrence Erlbaum Associates.

Samuels, S. J. (2002). Reading fluency: Its development and assessment. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 166–183). Newark, DE: International Reading Association.

Samuels, S. J., LaBerge, D., & Bremer, C. D. (1978). Units of word recognition: Evidence for developmental changes. *Journal of Verbal Learning and Verbal Behavior, 17*, 715–720.

Spinelli, D., De Luca, M., Di Filippo, G., Mancini, M., Martelli, M., & Zoccolotti, P. (2005). Length effect in word naming in reading: Role of reading experience and reading deficit in Italian readers. *Developmental Neuropsychology, 27*, 217–235.

Su, Y.-F. (1997). *Indicators of automaticity in word recognition.* Unpublished doctoral dissertation, University of Minnesota, Twin Cities.

Sun, F., & Feng, D. (1999). Eye movements in reading Chinese and English text. In J. Wang, A. W. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 189–205). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Tan, L. H., & Peng, D.-L. (1990). Yujing dui hanyu danzici tezheng fenxi de yingxiang [The effects of semantic context on the feature analyses of single Chinese characters]. *Xin li xue dong tai [Journal of Psychology]*, *4*, 5–10.

van der Leij, A., & van Daal, V. H. P. (1999). Automatization aspects of dyslexia: Speed limitations in word identification, sensitivity to increasing task demands, and orthographic compensation. *Journal of Learning Disabilities, 32*, 417–428.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology, 50A*, 439–456.

Yang, H.-M. (1998). 閱讀障礙學生文字辨識自動化處理之分析研究 [An analysis of automatic processing of word recognition in dyslexics]. *特殊教育與復健學報 [Bulletin of Special Education and Rehabilitation]*, *6*, 15–37.

Ziegler, J. C., Perry, C., Jacobs, A. M., Ma-Wyatt, A., Ladner, D., & Schulte-Korne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology, 86*, 169–193.

Zoccolotti, P., De Luca, M., Di Pace, E., Gasperini, F., Judica, A., & Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and Language, 93*, 369–373.