

ORIGINAL ARTICLE

Analytic visual word recognition in L2 learners: evidence from the length effect among ESL speakers

Nan Jiang¹ , Zhiyi Wu¹  and Ying Zhou² 

¹The University of Maryland, College Park, USA and ²Xinjiang University, Ürümqi, China

Corresponding author: Nan Jiang; Email: njiang@umd.edu

(Received 18 October 2024; revised 29 June 2025; accepted 13 August 2025)

Abstract

Research on first language acquisition has shown that children initially approach word recognition analytically and gradually shift to holistic processing as their reading experience increases, as evidenced by a reduction and eventual disappearance of the word length effect in word recognition. The present study aimed to investigate visual word recognition strategies among non-native speakers of English, specifically examining whether these speakers recognize words analytically or holistically and whether their first language influences their recognition strategies. The study tested native and non-native speakers of English with Chinese or Latin-script language backgrounds on 160 English words and 160 nonwords, ranging from 4 to 8 letters in length. The results indicated that Chinese ESL speakers exhibited a positive length effect, showing slower response times to longer words, in contrast to native English speakers, who demonstrated a reversed length effect, and to the Latin-script group, who exhibited no significant length effect. These findings suggest that non-native speakers are more likely to adopt an analytic word recognition strategy when the writing systems of their first and second languages differ. Conversely, same-script second language learners appear able to transfer holistic word recognition strategies from their first language to English.

Keywords: analytic strategy; English as a second language (ESL); holistic word recognition; length effect; lexical decision; visual word recognition

Introduction

Visual word recognition is a basic process underlying reading. A distinction is made in visual word recognition research between analytic (or serial) and holistic (or parallel) processing (e.g., Bijeljac-Babic et al., 2004; Marinus et al., 2015). The former is characterized by the serial processing of sublexical units that precedes the recognition of a word, while the latter is believed to involve the parallel analysis of the entire word as a single processing unit (e.g., Rau et al., 2014). In the latter case,

decomposition may occur after or in addition to whole-word recognition. This distinction can be best viewed as a continuum rather than a dichotomy, as an individual may adopt both processing strategies and apply them in the recognition of different types of words, e.g., adopting holistic processing for high-frequency words and analytic processing for low-frequency words (e.g., Jiang & Feng, 2022). Thus, an analytic or holistic reader is one who relies mainly rather than exclusively on analytic or holistic processing.

In the empirical research on visual word recognition, the adoption of these strategies is often assessed in terms of a length effect in a word recognition task such as a lexical decision task (LDT) for alphabetic languages and in terms of a stroke number effect where logographic languages are involved. The length effect refers to a longer response time or a lower accuracy rate associated with longer words (e.g., *library* versus *desk*) when other lexical properties are controlled (e.g., Bijeljac-Babic *et al.*, 2004; Rau, Moeller, & Landerl, 2014; Tiffin-Richards & Schroeder, 2015). Similarly, the stroke number effect refers to a longer response latency or a higher error rate for words of more strokes than for words of fewer strokes (e.g., Jiang & Feng, 2022; Su & Samuels, 2010). In this research, the size of these effects is interpreted as reflecting the extent to which an analytic processing strategy is employed, with a stronger length or stroke number effect reflecting more reliance on analytic processing. In contrast, a reduced or an absence of this effect is considered as reflecting holistic processing whereby words of different lengths or stroke numbers are treated as single and holistic processing units. The rationale underlying this interpretation is that if words are represented and processed holistically, the number of letters (or strokes) within a certain range (e.g., 4 to 8 letters) should not affect processing time, as they are treated as single units or processed in parallel. If words are represented and processed sublexically or serially, longer words mean more processing units, which would require more time for processing.

Research on word recognition strategies in L1 research

A large number of studies have demonstrated a developmental trend in the adoption of such word recognition strategies in first-language (L1) speakers. Early learners often rely more on analytic processing. As their reading experience increases, they become more of a holistic word recognizer. For example, in comparing the performance of third graders, fifth graders, and adults in identifying and naming words of varying length, Bijeljac-Babic *et al.* (2004) found a decreasing trend in the size of the length effect as assessed in terms of response latencies across the three age groups. The third graders showed a robust length effect, but the effect was much weaker for fifth graders and was absent for adults. Similar findings have also been reported by Rau, Moeller, and Landerl (2014) and Tiffin-Richards and Schroeder (2015). It was also found that when school children of the same ages were divided into slower and faster readers based on their reading performance, only slower readers showed a length effect, but faster readers did not (Marinus, Nation, & de Jong, 2015). Additionally, dyslexic readers were found to produce a stronger length effect than age-matched individuals; the length effect of dyslexic readers was often the same in size as that of nondyslexic readers two to three years younger, and they

demonstrated different eye fixation patterns in sentence reading with more fixations and fewer skipped words than unimpaired readers (e.g., Hawelka, Gagl, & Wimmer, 2010; Hyönä & Olson, 1995; Martens & de Jong, 2006; Ziegler et al., 2003; Zoccolotti et al., 2005). This developmental trend has also been found among school children of a logographic language such as Chinese. For example, Su and Samuels (2010) tested Chinese-speaking second, fourth, and sixth graders and adults in a word recognition task. Only the second graders showed a stroke number effect; the fourth and sixth graders and adults showed no such effects. In short, developing, slower, and dyslexic readers are more likely than skilled readers to adopt an analytic processing strategy, as assessed by the length or stroke number effect. Holistic processing occurs among more experienced and competent readers.

Also relevant to the present study was an interaction between length and lexical status. Many previous studies have shown a robust length effect for nonwords, while words either showed no length effect or a reduced length effect (e.g., Di Filippo et al., 2006; Yap et al., 2015; Ziegler, Jacobs, & Klüppel, 2001). One may argue that the recognition of words can become holistic with increasing exposure. In contrast, nonwords do not have a lexical entry and thus cannot be processed holistically. Thus, this interaction provided further evidence for the view that the analytic or serial processing is underlying the length effect.

All these findings involving alphabetic languages are consistent with word recognition models such as the Dual Route Cascaded model (DRC, Coltheart et al., 2001), where word recognition can be achieved through both a sublexical route involving serial mapping of orthography to phonology and a lexical route that involves the parallel activation of all letters or the orthography of the whole word, and with the idea of a transition from sublexical to lexical reading as an individual becomes a more skilled reader (Jackson & Colheart, 2001). Whether the same mechanisms apply to a logographic language such as Chinese is yet to be determined.

Analytic word recognition in L2: an emerging topic

These findings from L1 research raise several questions about second language (L2) readers. The most basic question is whether L2 learners tend to be analytic readers in their early stages of L2 learning, just like L1 readers. Two scenarios may be conceived. One is that individuals who have become competent readers in their L1 may transfer their holistic reading strategy to the new language, thus bypassing the early analytic stage. Thus, they read holistically in the new language. Another possibility is that the development of a holistic word recognition strategy is always language-specific and only occurs along with increasing reading experiences in the new language. In this scenario, L2 learners would also begin with a heavy reliance on analytic processing and only become more holistic readers later. A potential moderating factor in this context may be the script similarity between a learner's L1 and L2. Research has shown that higher orthographic overlap creates more cross-linguistic activation (e.g., Coderre & van Heuven, 2014), potentially increasing the likelihood of an L2 learner transferring their holistic strategies from L1 word recognition. Thus, an L2 learner may be more likely to transfer their holistic L1 reading strategy to an L2 if the two languages share the same script, e.g., Spanish and

English, compared to languages that do not overlap in the script, e.g., Chinese and English. Thus, the role of script similarity in this development is a second question. Finally, assuming that L2 learners also begin as analytic readers, particularly those different-script L2 learners, the question is whether they would also demonstrate a transition from analytic to holistic processing and be able to eventually develop native-like holistic word recognition. Lastly, one may also explore what factors may affect the trajectory of this development.

These issues have received very limited attention in L2 processing research. Even though word length has been included in some comparisons of visual word recognition in L1 and L2, it was not linked to analytic word recognition, and the results were not consistent. For example, De Groot, Borgwaldt, Bos, & Van Den Eijnden (2002) included 18 predictor variables in their comparison of word recognition in L1 and L2, including length (operationalized as the number of letters, phonemes, syllables, and morphemes). The participants produced a significant length effect in both naming and lexical decision in L2 English, but they showed no such effects in L1 Dutch, thus suggesting a difference between L1 and L2 processing in this area. A subsequent study reported by Lemhöfer *et al.* (2008) also compared visual word recognition in L1 and L2 speakers. Fifteen predictor variables were examined, including the length variable (as assessed in the number of letters, range: 3 to 5 letters). In contrast to what was reported in De Groot *et al.* (2002), this megastudy involving 1025 words showed no difference in how length affected word recognition in L1 and L2 speakers. Both groups produced a comparable length effect. Berger, Crossley, and Skalicky (2019) also included the length effect in their study of L2 lexical processing, but it served more as a moderating variable without a comparison of the effect in L1 and L2. None of these studies examined the length effect with a focus on analytic and holistic word recognition strategies.

One of the earliest attempts to explore analytic word recognition in L2 was a study of the missing-letter effect among L2 learners reported by Liang *et al.* (1997). The missing-letter effect refers to an individual's failure to detect a target letter while reading a text. It is considered a result of what they referred to as unitization, the processing of words with an increasingly larger unit, i.e., processing words at the syllable or word level rather than at the letter level. Individuals are more likely to miss target letters for higher-frequency words when they become faster and more efficient readers as a result of unitization in reading. Thus, the proportion of missed letters, or the size of the missed-letter effect, may help reveal the extent to which a reader processes words analytically or holistically. In their study, Liang *et al.* (1997) asked Chinese learners of English as a second language (ESL) and English L1 speakers to read two texts and circle the letter *t* in *the* and *f* in *of*. Two findings are relevant to the current study. First, the proportion of missed letters increased as the participants' L2 proficiency increased (e.g., 0.049, 0.097, and 0.319 for the three proficiency groups for the letter *f*), and second, even the participants of the highest L2 proficiency had a lower proportion of missed letters compared to English L1 speakers (0.528). These results may be the first piece of evidence in support of the adoption of a more analytic strategy in visual word recognition among L2 speakers than L1 speakers. We are not aware of any follow-up research related to this study and to unitization in L2 reading.

Three recent studies represented an emerging interest in this topic. In an effort to examine how lexical characteristics affected visual word recognition among children L2 speakers, Schröter and Schroeder (2018) included word length as a variable, along with frequency and neighborhood size. The length variable was included in order to examine if there was a developmental trend from sublexical or analytic processing to lexical or holistic processing. They tested school children who were L1 or L2 speakers of German in an LDT. The test materials included German words that were 3 to 12 letters in length. They reported a significantly larger length effect among L2 than L1 children. This may be the first study that compared L1 and L2 word recognition using the length effect as an assessment of analytic word recognition strategies. Two recent studies reported by Jiang and colleagues examined analytic Chinese word recognition among L2 speakers as assessed through the stroke number effect. In this research, the stroke number effect refers to longer response latencies for words with more strokes. Similar to the length effect for alphabetic languages, the presence or the size of the stroke number effect was also considered as an indication of analytic word recognition. They tested speakers of Chinese as a second language (CSL) and Chinese L1 speakers on disyllabic Chinese words that varied between 5 and 27 strokes (Jiang, Hou, & Jiang, 2020) or monosyllabic words of 2 to 13 strokes (Jiang & Feng, 2022) in a lexical decision task. In both studies, CSL speakers produced a reliable stroke number effect in reaction time (RT), but L1 speakers tested with the same set of materials showed no such effect. These results provided compelling evidence in support of the idea that L2 speakers relied more on analytic processing in visual word recognition than L1 speakers did.

The present study

The purpose of the present study was to examine the topic among adult ESL speakers. A set of English words and nonwords was selected based on the data from the English Lexicon Project (Balota et al., 2007). Both words and nonwords varied in length from four to eight letters. The data collected from English L1 speakers in the English Lexicon Project showed no length effect for words but a significant length effect for nonwords. We tested two groups of ESL speakers who differed in their L1. One group consisted of Chinese ESL speakers whose L1 differed from English in script, and the other group included ESL speakers with an L1 that shared the same script with English. As the latter group used the Latin alphabet in their L1s, they are referred to as the Latin-script ESL group. This setup allowed us to explore the following two related questions.

1. First, given a set of English words that varied in length but showed no length effect among L1 English speakers, would these ESL speakers demonstrate a length effect, i.e., take longer to respond to longer words?
2. Second, would the ESL speakers' L1 background affect the presence or size of any observed length effect?

If L2 speakers indeed tended to rely more on analytic word recognition regardless of their L1, we would expect a length effect similar in size in these two groups of ESL

Table 1. Background information for English-learning participants

	Chinese ESL group <i>N</i> = 32	Latin-script ESL group <i>N</i> = 34
Age at testing (range)	21.8 (19–24)	27.9 (21–45)
Length of English Learning (range)	12.7 (9–17)	9.4 (4–14)
Self-rated English proficiency out of 10 (range)	5.9 (2–8)	8.0 (4–10)

speakers. However, if L2 learners were able to transfer the holistic strategy developed in L1 reading to L2 word recognition, the Latin-script ESL group would be more likely to benefit from this transfer and adopt a holistic strategy, thus showing a smaller or even no length effect in comparison to Chinese ESL speakers. Based on the findings of Jiang and colleagues reviewed earlier, we would predict a length effect among the Chinese ESL group. However, we kept the prediction open for Latin-script ESL speakers, as this study represented the first study that examined adult L2 speakers whose L1 and L2 shared the same script in this context. Finally, based on previous research in L1, we would expect all groups to show a length effect for nonwords.

Method

a. *Participants.* Three groups of participants took part in the study. The first group consisted of English L1 speakers who were recruited from an American university. These were undergraduate students enrolled in a psychology course and received course credit for their participation. Additionally, two groups of ESL speakers participated in the study. One group included 32 Chinese ESL speakers, all college students living in mainland China. The other group, Latin-script ESL speakers, comprised 34 participants recruited online through Prolific (www.prolific.co). This group consisted of 25 Spanish, five Portuguese, three Italian, and one German L1 speakers. Both groups of participants reported their self-rated proficiency level in English from one to ten. Table 1 provides the demographic information and self-reported proficiency ratings of the two groups.

b. *Materials.* All test items were selected from the English Lexicon Project (ELP, Balota et al., 2007), which provided RTs for a large number of English words obtained in an LDT from English L1 speakers. The development of test materials began with a selection of 35–37 words for each word length of from four to eight letters, resulting in the initial list of 178 words. These words were selected on the basis of several considerations. First, they were all content words (nouns, verbs, and adjectives) without inflections. Second, they were high-frequency words considered to be highly familiar to the ESL speakers to be tested. Finally and most importantly, these words of varying length showed no length effect among English L1 speakers based on the data from the ELP. To ensure familiarity among ESL speakers, these words were given to 30 ESL speakers (undergraduates in a Chinese university from the same population as those who participated in the main study), who completed a familiarity rating task using a 1–10 scale. A final list of 160 words was selected. All chosen words received a mean familiarity rating score higher than 8, and they were

Table 2. Characteristics of words and nonwords items

Length (# of letters)	Words <i>n</i> = 160				Nonwords <i>n</i> = 160	
	# of items	Familiarity scores	Frequency* (per million)	Mean RT in LDT*	# of items	Mean RT in LDT*
4	33	9.35	28.2	621	33	706
5	34	9.41	30.3	620	34	760
6	31	9.38	26.4	628	31	803
7	31	9.43	29.8	623	31	821
8	31	9.53	25.3	623	31	844

* Frequency based on ClearPond and RT from L1 speakers from the English Lexicon Project.

matched across the five lengths of four to eight letters for mean familiarity rating score and RT (based on the ELP). Nonwords were also selected from the ELP. They also varied from four to eight letters in length. Participants were presented with a randomly ordered list of materials constructed with 160 words and 160 nonwords. Table 2 shows the characteristics of these test items, including the number of items for each length, mean familiarity rating scores based on the norming study, mean frequency obtained from CLEARPOND (Marian et al., 2012), and mean RT in the LDT from the ELP.

c. Task and procedure. The participants were tested individually in an LDT, with slight variations in testing conditions between groups due to logistical constraints. Latin-script ESL speakers were tested on the online experimental platform Gorilla (www.gorilla.sc; Anwyl-Irvine et al., 2020), while the other two groups were tested on a personal computer using the presentation software DMDX (Forster & Forster, 2003) in a quiet space. In the LDT, each trial began with a fixation cross displayed for 1,000 ms, followed by a visual stimulus. Participants were instructed to decide if the stimulus was an English word or not by pressing two keys on a keyboard, one for Yes and the other for No. They were asked to perform the task as quickly and accurately as possible. RT was measured from stimulus onset, with a 3,000-ms response window. Ten practice trials, including five words and five nonwords, preceded the test trials. Feedback was provided on practice items but not on test items. All test items were presented as a single list randomized differently for each participant. All participants were asked to read and sign the consent form before the test.

d. Data analysis. Before the data were analyzed, data points for incorrectly responded items and outliers defined as 2.5 standard deviations of the mean of the same participant were removed, which counted for 6.06%, 6.93%, and 9.06% of the data for English L1 speakers, Roman ESL, and Chinese ESL groups, respectively.

For data analyses, RT data and response accuracy (RA, coded as “0 = incorrect” and “1 = correct” for each trial) were analyzed using Linear Mixed-effects Models (LMM) and Generalized Linear Mixed-effects Models (GLMM), respectively, utilizing the *lme4* (v1.1.35.5; Bates et al., 2015) and *lmerTest* (v3.1.3; Kuznetsova et al., 2017) packages in R (v4.4.2; R Core Team, 2021). We employed a maximal random effects structure with the help of the *bobyqa* optimizer in the *nloptr* package (v2.1.1; Powell, 2009) to ensure convergence and avoid overfitting, though random

Table 3. Reaction time (RT) and response accuracy (RA) for words and nonwords of different lengths for English L1 speakers, Chinese ESL speakers, and Latin-script ESL speakers

Group	Letter count	Words		Nonwords	
		RT (MS)	RA (%)	RT (MS)	RA (%)
English L1	4	648	95%	658	98%
	5	630	96%	693	98%
	6	627	98%	720	94%
	7	624	98%	742	94%
	8	610	99%	744	96%
Chinese ESL	4	678	91%	667	97%
	5	687	94%	756	92%
	6	700	95%	796	88%
	7	702	96%	818	92%
	8	709	96%	829	91%
Roman ESL	4	646	95%	699	98%
	5	637	98%	768	94%
	6	639	96%	798	89%
	7	616	98%	802	93%
	8	653	99%	864	94%

slopes were excluded due to convergence issues. Both LMM and GLMM analyses included fixed effects of Group (English vs. Chinese vs. Latin-script Languages), Lexical Status (Word vs. Nonword), mean-centered Length, and their interactions, along with random effects of Subject and Item: $RT \sim \text{Group} * \text{Lexical_Status} * \text{Length} + (1|\text{Item}) + (1|\text{Subj})$). The analyses proceeded from an omnibus analysis to separate analyses of the data for the three participant groups.

As noted by a reviewer, a subset of our stimulus words constituted cognates for Latin-script ESL speakers, potentially introducing a confounding variable. To address this issue systematically, we identified and coded all stimulus words as either cognate or non-cognate for Latin-script ESL speakers, determining that 59 out of 160 critical words (36.9%) qualified as cognates. We then conducted two complementary analyses to control for potential cognate effects. First, we analyzed the complete dataset with binary cognate status included as a covariate in our statistical model $(RT \sim \text{Group} * \text{Length} * \text{Cognate_Status} + (1|\text{Item}) + (1|\text{Subj}))$. Second, we performed a more conservative analysis with all cognate trials removed entirely $(RT \sim \text{Group} * \text{Length} + (1|\text{Item}) + (1|\text{Subj}))$.

Results

The descriptive statistics for the remaining data are shown in Table 3 and graphically in Figure 1.

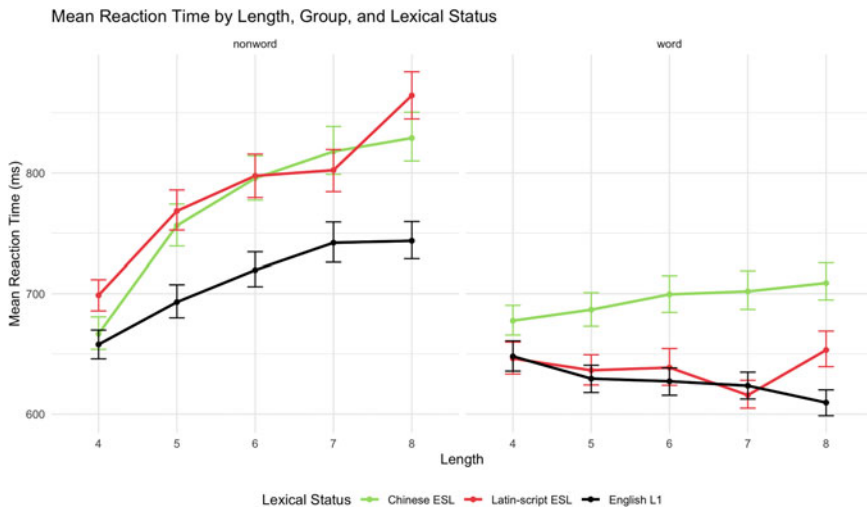


Figure 1. Mean reaction times by length, group, and lexical status.

Table 4. LMM modeling analysis results. $\text{lmer}(\text{RT} \sim \text{Group} * \text{Lexical_Status} * \text{Length} + (1|\text{Item}) + (1|\text{Subj}), \text{data} = \text{all})$

	β	Standard error	df	t value	p value
(Intercept)	675.62	16.62	108	40.66	<.001
Group-Chinese_ESL	71.49	25.28	104	2.83	0.006
Group-LatinScript_ESL	74.75	24.40	104	3.06	0.003
Lexical Status_Word	-61.80	4.59	538	-13.46	<.001
Length	16.19	2.29	541	7.06	<.001
Group-Chinese_ESL * Word	-5.99	4.55	30684	-1.32	0.19
Group-LatinScript_ESL * Word	-67.60	4.37	30682	-15.47	<.001
Group-Chinese_ESL * Length	17.30	2.27	30675	7.61	<.001
Group-LatinScript_ESL * Length	10.88	2.19	30672	4.96	<.001
Lexical Status_Word * Length	-23.03	3.25	535	-7.10	<.001
Group-Chinese_ESL * Word * Length	-2.99	3.20	30680	-0.93	0.35
Group-LatinScript_ESL * Word * Length	-6.63	3.08	30677	-2.16	0.03

The omnibus analysis revealed significant main effects for all three independent variables (Participant Group: $\chi^2 = 7.67$, $p = .02$; Length: $\chi^2 = 67.95$, $p < .01$; Lexical Status: $\chi^2 = 459.89$, $p < .01$), as well as significant interactions among them ($ps < .01$). We then examined the effect of length for each group separately by rotating the reference levels. See Table 4 for the LMM modeling results. All groups exhibited a significant interaction effect between length and lexical status (English L1 speakers: $t = -7.10$, $p < .001$; Chinese ESL: $t = 7.44$, $p < .001$; Latin-script ESL:

$t = 8.78, p < .001$), suggesting that words and nonwords were processed differently with respect to length. For nonwords, all three groups showed a positive length effect, with longer RTs for longer words ($ps < .001$). For real words, however, the results varied by group: English L1 speakers showed a significant negative length effect ($t = -2.98, p = .003$), responding faster to longer words; Chinese ESLs exhibited a significant positive length effect ($t = 3.02, p = .003$); and Latin-script ESLs showed no significant effect ($t = -1.09, p = .28$).

The two additional analyses examining the influence of cognate status provided robust confirmation of our main findings. When we included binary cognate status as a covariate in our statistical model, the critical group-by-length interactions remained significant, with the pattern of length effect unchanged: English L1 speakers showed a significant negative length effect ($t = -3.36, p = .0009$), Chinese ESL speakers demonstrated a significant positive length effect ($t = 2.30, p = .02$), and Latin-script ESL speakers showed no significant length effect ($t = -.57, p = .57$). Notably, cognate status itself did not significantly affect performance in any group (English: $t = -.74, p = .46$; Chinese: $t = -.47, p = .64$; Latin-script: $t = -1.04, p = .30$). Our more conservative analysis with all cognate trials entirely removed yielded identical patterns across groups (English: $t = -3.30, p = .001$; Chinese: $t = 2.25, p = .03$; Latin-script: $t = -.56, p = .58$). The consistency across both analytical approaches provides evidence that the differential length effects between participant groups reflect genuine differences in word recognition strategies rather than artifacts of cognate facilitation.

The GLMM analysis of RA included the same fixed and random effects as the LMM analysis. In the omnibus analysis, all three variables showed a main effect (Participant Group: $\chi^2 = 19.45, p < .001$; Length: $\chi^2 = 8.67, p = .003$; Lexical Status: $\chi^2 = 11.80, p = .001$). Furthermore, there were significant two-way interactions between Participant Group and Lexical Status ($\chi^2 = 8.11, p = .017$) and between Lexical Status and Length ($\chi^2 = 40.60, p < .01$). However, the interaction between participant group and length was not significant ($\chi^2 = .28, p = .87$). The three-way interaction was significant ($\chi^2 = 6.43, p = .04$). By rotating the levels, further examination of within-group length effects revealed significant interactions between length and lexical status for all groups (English L1: $z = 6.37, p < .001$; Chinese ESL: $z = 5.01, p < .001$; Latin-script ESL: $z = 4.04, p < .001$). See Table 5 for the GLMM modeling results. Length negatively affected the accuracy of nonword items for all three groups ($ps < .01$), indicating that longer nonwords were more challenging for all participants to reject correctly. In contrast, length positively affected accuracy for word items for all three groups ($ps < .002$), with participants responding more correctly to longer words.

Discussion

Several findings emerged from the study. First, all three groups produced a positive length effect for nonwords, showing a longer RT for longer nonwords. This was consistent with previous findings that nonwords were more likely to show a positive length effect (e.g., Di Filippo *et al.*, 2006; Yap *et al.*, 2015; Ziegler, Jacobs, & Klüppel, 2001), indicating the adoption of an analytical word recognition strategy while

Table 5. GLMM modeling analysis results. *glmer*(Accuracy~ Group * Lexical_Status * Length + (1|Item) + (1|Subj), data = all, family = binomial, control = glmerControl(optimizer = "bobyqa"))

	β	Standard error	z value	p value
(Intercept)	3.72	0.15	25.63	<.001
Group-Chinese_ESL	-0.83	0.19	-4.39	<.001
Group-LatinScript_ESL	-0.45	0.19	-2.40	0.02
Lexical_Status_Word	0.49	0.14	3.44	<.001
Length	-0.20	0.07	-2.94	0.003
Group-Chinese_ESL * Word	-0.08	0.14	-0.60	0.55
Group-LatinScript_ESL * Word	0.29	0.15	1.93	0.05
Group-Chinese_ESL * Length	0.02	0.07	0.29	0.77
Group-LatinScript_ESL * Length	0.04	0.07	0.52	0.60
Lexical_Status_Word * Length	0.66	0.10	6.37	<.001
Group-Chinese_ESL * Word * Length	-0.20	0.10	-2.02	0.04
Group-LatinScript_ESL * Word * Length	-0.26	0.11	-2.41	0.02

processing nonwords, as nonwords are inherently unfamiliar and thus more likely to be processed analytically than holistically. However, the more important findings came from the processing of word items for which the three groups showed different patterns. The English L1 speaker group unexpectedly produced a negative length effect, responding to longer words faster than shorter words. It is not immediately clear why the English L1 speakers showed this pattern when the same set of words showed no length effect in the English Lexicon Project, which was the primary reason for their selection in our study. Interestingly, New et al. (2006) also reported a negative length effect among English native speakers for words of three to five letters in a large-scale analysis of the English Lexicon Project data, finding that RTs initially decreased with word length (for words of 3–5 letters) before increasing again for longer words. Therefore, it is not the negative length effect itself that was unexpected, but rather the inconsistency between what we found and the subset of the English Lexicon Project that informed our original item selection. This discrepancy represents a limitation of our study and highlights the complexity of the length effect even among native speakers. The two ESL groups showed patterns that differed both from the L1 speaker group and from each other. Instead of showing no length effect (as seen in the data from the English Lexicon Project) or a negative length effect (as shown among the L1 speakers tested in the present study), the Chinese ESL group showed a positive length effect. They responded to longer words more slowly. In contrast, the Latin-script ESL showed no length effect.

Based on the interpretation of the positive length effect as reflecting the adoption of an analytic word recognition strategy, these findings suggest that when L2 learners' L1 and L2 differ in script, as with the Chinese ESL group, they undergo a prolonged period when their word recognition is more analytic. This extended phase contrasts with children learning to read in their L1. For example, Marinus

et al. (2015) found that English-speaking and Dutch-speaking children no longer showed a length effect by third grade, while Su and Samuels (2010) observed that the stroke number effect that was present among Chinese second graders disappeared among the fourth graders, indicating a successful transition from analytic to holistic processing. Notably, the Chinese ESL group tested in the present study, despite having been exposed to English for more than 10 years on average, albeit less intensively, still showed a reliable length effect.

We may interpret the lack of a length effect among the Latin-script ESL group in two different ways. One possibility is related to the L1 transfer effect. Studies have shown that transfer effects from the native language were more pronounced for a within-script L2 than a cross-script L2 (e.g., Spinelli *et al.*, 2021). In this study, it is likely that due to the involvement of the same script in their L1 and L2, these ESL speakers can adopt a holistic word recognition strategy early on as a result of L1 transfer. Thus, they started as holistic readers while learning a new L2. Alternatively, they may begin with analytic word recognition, but their L1 experiences with the same script allow them to make a faster transition from analytic to holistic word recognition. As the participants tested in this study were already quite experienced and proficient in their L2 English, our results cannot definitively support either interpretation. Less proficient ESL speakers with similar language backgrounds would help differentiate these two possibilities.

The findings from the Chinese ESL group provide further evidence for the adoption of an analytic word recognition strategy among intermediate adult L2 speakers and the first such evidence involving an alphabetic L2 in the form of a length effect. This is built upon previous research by Jiang, Hou, and Jiang (2020) and Jiang and Feng (2022), who already demonstrated this tendency among L2 speakers of a logographic language in the form of a stroke number effect. Similarly, Liang *et al.* (1997) also demonstrated this tendency in the form of the missing-letter effect in an alphabetic language among Chinese ESL speakers. The sharp contrast in the direction of the length effect between the Chinese ESL speakers (positive) and English L1 speakers (negative) highlights the difference in visual word recognition between these two populations. The findings from the Latin-script ESL group provide novel evidence for the influence of a learner's L1 in this context, an issue not addressed in previous research related to this topic. They are also consistent with previous research showing the transfer of processing strategies from L1 to L2. Such transfer has been observed in the differential reliance on orthographic and phonological information in visual word recognition among L2 learners with a logographic or alphabetic L1 background (e.g., Akamatsu, 2003; Wang, Koda, & Perfetti, 2003) and the differential involvement of decomposition in processing L2 complex words (e.g., Portin & Laine, 2001; Vainio, Pajunen, & Hyönä, 2014).

To consider the findings in a broader context of L2 word recognition research, the observation of a positive length effect among L2 speakers, contrasting with the absence of such an effect among L1 speakers, is consistent with emerging evidence for the enhanced role of sublexical processing in L2 word recognition. One line of evidence comes from morphological priming studies in L2. L1 speakers typically exhibit a reliable morphological priming effect in the masked priming paradigm involving morphologically related prime-target pairs such as *disagree-AGREE*). To ensure that such priming effects were morphological in nature rather than due to

letter overlap, a control comparison was often built in such studies that involved prime-target pairs that were orthographically similar but morphologically unrelated, such as *freeze-FREE*. L1 speakers usually show no priming effect on these items, thus helping to validate the morphological nature of the morphological priming effect. Interestingly, when the same set of test materials was used to test L2 speakers, they showed a robust and consistent orthographic priming effect in multiple studies (e.g., Diependaele et al., 2011; Heyer & Clahsen, 2015; J. Li, Taft, & Xu, 2017; M. Li, Jiang, & Gor, 2017; J. Li & Taft, 2020; Jiang & Wu, 2022).

A similar L1–L2 difference was present in the prime lexicality effect. In L1 processing research, it has been demonstrated that a nonword prime would produce a positive priming effect if it is an orthographic neighbor of the target, e.g., *feason-SEASON*, but a word prime would show no priming or a negative priming effect, e.g., *reason-SEASON* (e.g., Andrews & Hersch, 2010; Davis & Lupker, 2006; Forster & Veres, 1998). In contrast, neighbor primes always produced a priming effect in L2 speakers, regardless of whether they were words or nonwords (e.g., Jiang, 2021; Nakayama & Lupker, 2018; Qiao & Forster, 2017).

Within an interactive activation model such as the one proposed by McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982), the absence of priming from form-related primes (e.g., *freeze-FREE*) and from word neighbors (e.g., *reason-SEASON*) can be explained in terms of lateral inhibition. In both cases, priming may occur due to letter overlap between the primes and the targets (sublexical processing). However, at the lexical level, the recognition of the prime word would automatically inhibit the activation of all other words, including the target. In the case of L1 speakers, lateral inhibition cancels or overpowers sublexical facilitation, thus leading to no priming or negative priming for orthographically related word primes. However, in L2 processing, researchers have suggested that there is less lexical competition; instead, sublexical overlap becomes the primary driving force in visual word recognition (e.g., Jiang & Wu, 2022; Li & Taft, 2020; Nakayama & Lupker, 2018). This explains why L2 speakers produced orthographic priming effects where L1 speakers did not. Thus, the orthographic priming effect corroborates with the observation of the length effect in L2 speakers in the present study in showing that sublexical or analytic processing plays a particularly important role in L2 word recognition.

However, we are just beginning to understand this topic, and many issues remain to be explored. A question related to the interpretation of the results of the present study is whether the difference between the two ESL groups reflected L1 influence or L2 proficiency. Note that the Latin-script group showed a higher mean self-rating English proficiency score of 8 as compared to that of 5.9 for the Chinese ESL group. This was corroborated by faster response times for the Latin-script group (737 ms) compared to the Chinese group (787 ms). These observations raise the possibility of the absence of a length effect, and thus the inferred adoption of a more holistic strategy in the Latin-script group might be attributed to their higher proficiency in L2 English. However, it's worth noting that L1–L2 script differences could also have affected response latencies, either instead of or in addition to L2 proficiency.

A more theoretically important issue is whether and how quickly L2 speakers are able to develop native-like holistic word recognition strategies and what factors

(other than L1) would play a role in this process. One possibility is that L2 speakers may also show a developmental analytic-to-holistic transition, eventually achieving native-like word recognition strategies, even though this process may take longer. An alternative is that a native-like word recognition strategy could be difficult to develop among different-script L2 speakers, possibly due to their limited quantity or intensity of L2 reading and exposure.

What it takes for L2 speakers to develop a native-like word recognition strategy, as assessed in terms of the length effect, or for L2 speakers to develop native-like interactive lexical knowledge, as assessed in terms of orthographic priming or the prime lexicality effect, is certainly an intriguing issue to explore in future research.

Limitations and future research

As the first study specifically designed to explore analytic word recognition strategies in adult L2 speakers using the length effect, our focus was on the length effect and the role of native languages with familiarity as a controlled lexical variable. This study, however, has important limitations to acknowledge, and these have to be considered in future research. First, we relied on self-reported rather than standardized measures of L2 proficiency, making precise comparisons between our participant groups challenging. Second, the unexpected negative length effect found in our native speaker group, contrary to the English Lexicon Project data that guided our stimulus selection, suggests possible influences from methodological differences in testing conditions or participant sampling. Furthermore, lexical variables such as cognate status and neighborhood size should be considered in developing test materials, as they may affect participants' performance. A potential speed-accuracy tradeoff in the results needs further exploration, too. These limitations, while not undermining our main findings regarding differential word recognition strategies among L2 speakers with different script backgrounds, indicate the need for cautious interpretation and provide direction for methodology refinements in future research.

Future research could consider examining the word recognition processing by the novice ESL speakers with an alphabetic script background to provide a more comprehensive and definitive answer to whether these learners show a lack of length effect from the beginning of their learning phase. Moreover, to get a better understanding of the extent to which L2 proficiency contributed to the observed group difference, future research could use one of the following three approaches. First, researchers could compare two groups that were matched for L2 proficiency but with different L1 backgrounds. If L1 influence is the main factor for the group difference observed in this study, the group difference should remain in the two proficiency-matched groups. Otherwise, if L2 proficiency is the key factor, the group difference should disappear. An alternative approach would be to test groups of Chinese ESL speakers that differed in English proficiency (to mimic the proficiency difference between the two ESL groups of the present study). If L2 proficiency is the primary factor, this group difference should remain even among L2 speakers who share the same L1 but differ in L2 proficiency. Finally, a longitudinal approach to

tracking the length effect within the same group of L2 speakers would offer a more precise assessment of the trajectory of changes in their word recognition strategies.

Conclusion

There has been a surge of research on visual word recognition among L2 speakers in recent years. This research has offered new insights into how visual word recognition processes and the representation of lexical information differ in L1 and L2 speakers. The findings from this study (a) provide further evidence for an intriguing difference in visual word recognition between L1 and L2 speakers, i.e., word recognition in L2 being more analytic and relying more on sublexical processes, (b) demonstrate the influence of an L2 speaker's L1 in this process, and (c) raise a developmental question regarding the extent to which L2 speakers are able to develop native-like word recognition strategies or lexical competence.

In conclusion, we emphasize the last point as particularly crucial for future research. From a cognitive perspective, lexical competence may refer to an L2 speaker's ability to represent and access lexical knowledge in a native-like manner, rather than or beyond the number of words an L2 speaker knows. The length effect investigated in the present study, along with other phenomena such as the orthographic priming effect and the lack of the prime lexicality effect among L2 speakers, may serve as important tools for studying such lexical development. If native-like lexical access is characterized by holistic and interactive access of lexical knowledge, the presence, absence, or size of the length effect, the orthographic priming effect, and the prime lexicality effect may help us determine where an L2 speaker is in this development.

Competing interests. The authors declare none.

References

- Akamatsu, N. (2003). The effects of first language orthographic features on second language reading in text. *Language Learning*, 53, 207–231.
- Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, 139(2), 299–318.
- Anwyl-Irvine, A.L., Massonnié, J., Flitton, A. et al. (2020). Gorilla in our midst: an online behavioral experiment builder. *Behavior Research* 52, 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . & Treiman, R. (2007). The English lexicon project. *Behavior research methods*, 39, 445–459.
- Berger, C., Crossley, S., & Skalicky, S. (2019). Using lexical features to investigate second language lexical decision performance. *Studies in Second Language Acquisition*, 41(5), 911–935.
- 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*, 17, 411–431.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204–256.
- Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 668–687.
- De Groot, A. M. B., Borgwaldt, S., Bos, M., & Van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language*, 47, 91–124.

- Di Filippo, G., De Luca, M., Judica, A., Spinelli, D., & Zoccolotti, P. (2006). Lexicality and stimulus length effects in Italian dyslexics: Role of the overadditivity effect. *Child Neuropsychology*, *12*(2), 141–149.
- Diependaele, K., Duñabeitia, J. A., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, *64*, 344–358.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods*, *35*, 116–124.
- Forster, K. I., & Veres, C. (1998). The prime lexicality effect: Formpriming as a function of prime awareness, lexical status, and discrimination difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 498–514.
- Hawelka, S., Gagl, B., & Wimmer, H. (2010). A dual-route perspective on eye movements of dyslexic readers. *Cognition*, *115*(3), 367–379.
- Heyer, V., & Clahsen, H. (2015). Late bilinguals see a scan in scanner AND in scandal: dissecting formal overlap from morphological priming in the processing of derived words. *Bilingualism: Language and Cognition*, *18*, 543–550.
- Hyönä, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1430–1440.
- Jackson, N. E., & Coltheart, M. (2001). *Routes to Reading Success and Failure: Toward an Integrated Cognitive Psychology of Atypical Reading*. Ann Arbor, MI: Sheridan Books.
- Jiang, N. (2021). Orthographic friends and lexical strangers in the L2 lexicon. *Journal of Second Language Studies*, *4*(2), 224–244.
- Jiang, N., & Feng, L. (2022). Analytic visual word recognition among Chinese L2 learners. *Foreign Language Annals*, *55*(2), 540–558.
- Jiang, N., Hou, F., & Jiang, X. (2020). Analytic versus holistic recognition of Chinese words among L2 learners. *The Modern Language Journal*, *104*(3), 567–580.
- Jiang, N. & Wu, X. (2022). Orthographic priming in a second language. *Language Learning*, *72*(3), 625–645.
- Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: a megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 12–31.
- Li, J., & Taft, M. (2020). The processing of English prefixed words by Chinese-English bilinguals. *Studies in Second Language Acquisition*, *42*, 239–249.
- Li, J., Taft, M., & Xu, J. (2017). The processing of English derived words by Chinese-English bilinguals. *Language Learning*, *67*, 858–884.
- Li, M., Jiang, N., & Gor, K. (2017). L1 and L2 processing of compound words: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition*, *20*, 384–402. <https://doi.org/10.1017/S1366728915000681>
- Liang, T., Healy, A. F., & Bourne Jr, L. E. (1997). Unitization in second-language learning: evidence from letter detection. *The American journal of psychology*, *110*(3), 385–395.
- Marian V, Bartolotti J, Chabal S, Shook A (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE* *7*(8), e43230.
- Marinus, E., Nation, K., & de Jong, P. F. (2015). Density and length in the neighborhood: Explaining crosslinguistic differences in learning to read in English and Dutch. *Journal of Experimental Child Psychology*, *139*, 127–147.
- 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.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.
- Nakayama, M., & Lupker, S. J. (2018). Is there lexical competition in the recognition of L2 words for different-script bilinguals? An examination using masked priming with Japanese-English bilinguals. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(8), 1168–1185.
- 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 & Review*, *13*, 45–52.
- Portin, M. & Laine, M. (2001). Processing cost associated with inflectional morphology in bilingual speakers. *Bilingualism: Language and Cognition*, *4*, 55–62.

- Qiao, X., & Forster, K. I. (2017). Is the L2 lexicon different from the L1 lexicon? Evidence from novel word lexicalization. *Cognition*, *158*, 147–152.
- Schröter, P., & Schroeder, S. (2018). Differences in visual word recognition between L1 and L2 speakers: the impact of length, frequency, and orthographic neighborhood size in German children. *Studies in Second Language Acquisition*, *40*, 319–339.
- Su, Y. F., & Samuels, S. J. (2010). Developmental changes in character-complexity and word-length effects when reading Chinese script. *Reading and Writing*, *23*(9), 1085–1108.
- Tiffin-Richards, S. P., & Schroeder, S. (2015). Word length and frequency effects on children's eye movements during silent reading. *Vision research*, *113*, 33–43.
- Vainio, S., Pajunen, A., & Hyönä, J. (2014). L1 and L2 word recognition in Finnish. *Studies in Second Language Acquisition*, *36*, 133–162.
- Rau, A. K., Moeller, K., & Landerl, K. (2014). The transition from sublexical to lexical processing in a consistent orthography: an eye-tracking study. *Scientific Studies of Reading*, *18*(3), 224–233.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effects and some tests and extensions of the model. *Psychological Review*, *89*, 60–94.
- Spinelli, G., Forti, L., & Jared, D. (2021). Learning to assign stress in a second language: The role of second-language vocabulary size and transfer from the native language in second-language readers of Italian. *Bilingualism: Language and Cognition*, *24*(1), 124–136.
- Wang, M., Koda, K., & Perfetti, C. A. (2003). Alphabetic and nonalphabetic L1 effects in English word identification: a comparison of Korean and Chinese English L2 learners. *Cognition*, *87*, 129–149.
- Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2015). Responding to nonwords in the lexical decision task: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 597–613.
- Ziegler, J. C., Jacobs, A. M., & Klüppel, D. (2001). Pseudohomophone effects in lexical decision: Still a challenge for current word recognition models. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 547–559.
- Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology*, *86*, 169–193.
- Zoccolotti, P., Maria De, L., Enrico Di, P., Filippo, G., Anna, J., & Donatella, S. (2005). Word length effect in early reading and in developmental dyslexia. *Brain & Language*, *93*, 369–373.