Studies in Second Language Acquisition (2024), **46**, 119–140
doi:10.1017/S0272263123000426

RESEARCH ARTICLE

Language proficiency modulates L2 orthographic learning mechanism: Evidence from event-related brain potentials in overt naming

Yang Fu[1,2], Beatriz Bermúdez-Margaretto[3,4], David Beltrán[5], Wang Huili[1] and Alberto Dominguez[2]

[1]School of Foreign Languages, Hangzhou City University, China; [2]Instituto Universitario de Neurociencia, Universidad de La Laguna, Spain; [3]Departamento de Psicología Básica, Psicobiología y Metodología de las Ciencias del Comportamiento, Facultad de Psicología, Universidad de Salamanca, Salamanca, Spain; [4]Instituto de Integración en la Comunidad - INICO, Universidad de Salamanca, Salamanca, Spain; [5]Psychology Department, Universidad Nacional de Educación a Distancia, Spain
Corresponding author: Wang Huili; Email: wanghl@hzcu.edu.cn

(Received 17 July 2022; Revised 24 July 2023; Accepted 04 August 2023)

Abstract
The present study investigates bilinguals' capacity to rapidly establish memory traces for novel word forms in a second language (L2), as a function of L2 linguistic proficiency. A group of Chinese-English bilinguals with various English proficiency levels were presented with a reading-aloud task, consisting of 16 pseudowords and 16 English words repeatedly presented across six training exposures. Behavioral and neurophysiological data were collected, and modulations in the word-length effect across repetitions were measured as an index of transition from sublexical to lexical involvement. Results revealed that higher L2 proficiency was associated with decreased word-length effect on novel words, reflected in both naming latencies and early N1 and P200 brain responses. In contrast, lower proficiency learners appeared to engage in effortful letter-to-sound decoding processes, with higher attentional allocation to the letter sequence and greater use of sublexical processing across exposures. Our findings highlight the need to tackle specific grapheme-to-phoneme skills for efficient learning of L2, particularly in populations where the L1 is nonalphabetic.

Introduction
Visual word learning requires the formation of mental representations for orthographic codes, a process that occurs in response to the specific properties of a given language. Despite the variation in orthographic complexity across languages, the acquisition of word-specific orthographic information and the ability to "convert the visual print into its corresponding spoken forms" (i.e., *phonological decoding*, cf. Nassaji, 2014, p. 9) are universal processes that underlie word identification and are essential for the development of reading fluency (Share, 2008). According to the *self-teaching* hypothesis (Share, 1995, 1999), every successful activation of phonological decoding affords the chance to acquire orthographic representations of particular word forms. In response, over the

course of encounters with a written word, its orthographic representation can be stored and hence be rapidly and automatically accessed without the need for serial decoding, a process referred to as *orthographic learning* (Nation et al., 2007). To date, insight into the behavioral (Nation et al., 2007; Tucker et al., 2016) and neural (e.g., Bermúdez-Margaretto, Shtyrov, et al., 2020; Partanen et al., 2018) mechanisms underlying the acquisition of novel orthographic representations has largely been derived from studies conducted in the native language (L1). Therefore, evidence in support of L1 orthographic learning, including in English (Kwok & Ellis, 2015; Maloney, 2008), Spanish (Álvarez-Cañizo et al., 2018; Suárez-Coalla et al., 2016), French (e.g., Bosse et al., 2015; Ginestet et al., 2020), Dutch (de Jong & Share, 2007; van Viersen et al., 2022), Hebrew (Share, 1999), and Chinese (Li et al., 2020), has become increasingly grounded in the conception that the typical orthographic formation of novel word form per se primarily occurs through phonological decoding and can be achieved in a relatively automatic and rapid manner over a brief period of exposure.

Importantly, it is still unclear how orthographic learning occurs in bilinguals, and whether L2 linguistic proficiency affects the brain's capacity to learn. The current study intends to fill this gap by exploring the behavioral and neurophysiological indices of the rapid formation of memory representations for L2 novel word forms and the possible effects of L2 proficiency on this process.

Tracking the early stages of orthographic acquisition

In general, the acquisition of orthographic representations in a mature recognition system consists of a succession of phases that converts the visual input into its matching sound and subsequently applies general print-sound mappings to activate whole-word phonology. Influential models of this process include computational dual-route models of visual word recognition, such as the dual-route cascaded model (DRC, Coltheart et al., 2001) and connectionist dual-process models (CDP (+)+, Perry et al., 2007). The underlying idea of the dual-route architecture is that the formation of novel orthographic representations involves a transition from a serial, sublexical route that operates through the application of letter-to-sound conversion, to parallel retrieval and activation in a lexical route. Operationally, the strength of this change from sublexical to lexical reading can be quantified by the effect of letter length (i.e., the difference in naming latencies between short and long words, Weekes, 1997).

The simulation of the differences in length-sensitivity between novel and familiar words in dual-route models proceeds as follows. Naming novel word forms for the first time starts with sublexical letter-to-sound conversion in a serial, left-to-right fashion, with the consequence that naming latencies increase with the number of letters in the word. As unknown words become familiar through repeated exposure, both visual and speech-based representations of these words are established in the orthographic input lexicon and the phonological output lexicon, respectively. Those representations enable words to be accessed holistically and therefore independently of their letter length. Evidence in support of this proposal comes from studies that report a progressive reduction across exposures in the length effects exhibited by novel words (Kwok et al., 2017; Kwok & Ellis, 2015; Maloney, 2008). These studies have consistently argued that when a novel word is encountered sufficiently, the acquisition of orthographic knowledge permits a direct visual recognition and a consequent switch from sublexical, length-sensitive decoding to a lexical mode of processing, as evidenced by the reduction or absence of length effects through repeated exposure. Notably, a relatively short training with novel words, involving four or five exposures, is enough for native English readers to create durable lexical representation (Li et al., 2020, in Chinese; Nation et al., 2007, in English; Suárez-Coalla et al., 2016, in Spanish).

Tracking the early time course of orthographic learning

Beyond behavioral measurements, time-sensitive neurophysiological approaches, such as continuous electro- or magnetoencephalogram (MEG or EEG) registration, are well suited for assessing the temporal course of reading processes and the subtle changes that occur through sublexical to lexical stages. Two major event-related potentials (ERPs), the N1/P1 complex and the P200, have been documented as associated with varied processes in the early stages of visual word recognition. The early N1/P1 (~100 ms) is known as a marker of sublexical processes in the visual domain (Comesaña et al., 2012; Vergara-Martínez et al., 2020), which is sensitive to physical variations of visual features (e.g., letter case, size, or font; Chauncey et al., 2008), sequences of letters (Dufau et al., 2008) or word length (Hauk et al., 2006). It is therefore considered to reflect the extraction of low-level perceptual features involved in letter identity—that is, mapping visual information onto the location of each letter within the alphabetic string (i.e., location-specific orthographic codes, Grainger & Holcomb, 2009). The later component, P200 (~200 ms), is commonly assumed to be connected to the extraction of orthographic and phonological representations of words during the early stages of word-specific processing, reflecting the mapping of the word form onto its phonological properties (e.g., Kong et al., 2010; Wu et al., 2012).

Crucially, both brain responses have been found to alter their patterns throughout the process of orthographic acquisition. In particular, a decrease in N1 activity (Bermúdez-Margaretto, Kopytin, et al., 2021; Kimppa et al., 2016) and an increase in P200 amplitude (Bermúdez-Margaretto, Shtyrov, et al., 2020; Partanen et al., 2018) have been reported as a consequence of repeated encounters with novel word forms. For instance, Bermúdez-Margaretto, Shtyrov, et al. (2020) reported a strikingly fast (i.e., from first to second exposure) and stable enhancement (across four subsequent repetitions) of the early positivity (i.e., P200) elicited by novel word forms, reflecting neural plasticity in rapid novel word learning. Thus, in line with the dual-route framework, the cortical mechanisms for rapid formation of neural word memory circuits may reflect a switch in the early phases of orthographic learning, from sublexical, serial letter decoding to a more holistic, lexically oriented access to novel word forms over the course of exposures.

L2 orthographic learning: The influence of bilinguals' L2 proficiency

Orthographic acquisition has also been explored in bilingualism and biliteracy contexts, but as argued below, knowledge of this area remains incomplete due to its focus on samples of children, and the results have been mixed. In general, it is a steady and robust finding that L2 learners with either alphabetic (Chung et al., 2019; Schwartz et al., 2014) van Daal & Wass, 2017) or nonalphabetic (Li et al., 2021) L1 backgrounds can develop word-specific L2 orthographic representations during independent reading. The achievement of orthographic acquisition after four/five decoding attempts confirms that phonological decoding has the potential to enhance the cognitive skills responsible for reading, recall, and retention of L2 orthographic representations. Yet this body of empirical evidence may not fully address controversial aspects related to

https://doi.org/10.1017/S0272263123000426 Published online by Cambridge University Press

orthographic learning. For instance, a seemingly contradictory conclusion arises from the findings of Schwartz et al. (2014) and van Daal and Wass (2017), who reported that among children with limited L2 experience, successful orthographic learning was only evident for bilinguals with a close L1-L2 orthographic distance (e.g., L1 Russian/Swedish-L2 English) rather than those with more distinct orthographies (i.e., L1 Hebrew/Danish-L2 English). In contrast, studies conducted by Chung et al. (2019) and Li et al. (2021) involving highly proficient L2 learners demonstrated an opposing trend: the acquisition of word-specific orthographic representations occurred through repeated exposure, irrespective of the learners' L1 orthographic backgrounds.

As posited by Chung et al. (2019), these inconclusive findings might stem from variations in individual L2 proficiency levels. That is to say, even though there is an expected learning trajectory at the group level, individuals' abilities to learn L2 orthographic representations can vary. Those with higher proficiency might be more adept at identifying the most effective strategy for orthographic acquisition than their less proficient counterparts. Indeed, it stands to reason that increasing L2 proficiency is likely linked to optimized cortical signals of learning efficiency. For example, in the field of visual word recognition, evidence from neurophysiological studies (e.g., Abutalebi, 2008; Stein et al., 2009) has proposed that more automatic and effortless native-like L2 word processing is engaged with increased L2 proficiency, indicated by decreased amplitude in N400 and late positive components (see also, Stein et al., 2006). With higher proficiency, L2 lexical representations become more independent and can be developed by recruiting fewer cognitive resources. These L2 learning findings have linked linguistic proficiency to the activation of semantic information and the retrieval of a lexical entry corresponding to the target word, with an emphasis on the late central or posterior modulations of lexical access (at ~ 350–600 ms poststimulus onset) for single-word processing. Little is known, however, about the interplay between novel word-form acquisition and language proficiency during the early phases of visual word processing.

To the best of our knowledge, no behavioral or electrophysiological data are yet available on the issue of whether the functional role of L2 proficiency extends to L2 learners' ability to acquire novel orthographic representations. One exception, if anything, is the perceptual learning study by Kimppa et al. (2016) that addressed the influence of individual L2 experience on the neural dynamics related to the rapid acquisition of novel spoken word forms. The data pointed to a positive association between the age of L2 acquisition (and the number of learned languages) and the development of new memory circuits for novel spoken words, as indexed by increased frontal responses at ~ 50 ms after word onset over the course of exposures. However, it is important to note that the study failed to observe a correlation between average proficiency and neural memory-trace formation for novel words. Kimppa et al. (2016) attributed the absence of proficiency-related modulation to the proficiency level being self-reported (a language history questionnaire) and the small sample size (21 participants), which could limit differentiation between groups of higher and lower proficiency levels across various nonnative languages. Therefore, how the neurophysiological correlates of L2 orthographic acquisition alter with increasing proficiency levels remains to be determined.

The present study

In the current work, we sought to elucidate the effects of linguistic proficiency on the build-up of early memory traces for L2 novel words. As reviewed earlier, word learning in

English can be characterized by a transition from sublexical decoding to lexical processing. Conversely, for Chinese readers, visual-spatial encoding serves as a significant source of facilitation for the short-term retention of visually presented linguistic input, both in their Chinese L1 (Mou & Anderson, 1981) and in their English L2 (Hamada & Koda, 2010). Thus, our interest lay in exploring the extent to which the modulation of dual-route processes varies as a function of L2 proficiency for bilinguals with distant L1-L2 orthographic systems. Behavioral and EEG data were collected from a sample of Chinese-English bilinguals with different levels of L2 proficiency. The learning outcomes were assessed using an online training measure (word-reading-aloud task[1]) and an offline free-recall memory task. The training session focused on the relative effects of word length on naming latencies and the early neurophysiological responses to two classes of L2 words (novel vs. familiar words). Specifically, our hypotheses are as follows.

1. At the behavioral level, based on previous literature and computational models, we predict that the formation of orthographic representations of nonnative novel words could be characterized by a rapid transition from sublexical, length-sensitive decoding to a lexical mode of processing. This will be reflected in reduced naming latencies and decreased length effect for novel words, as a result of repeated exposure. Additionally, we expect increasing L2 proficiency to be associated with faster convergence of naming latencies for short and long novel word forms.
2. At the neural level, focusing on the previously described neural markers underlying orthographic processing, we expect a reduction of the early event-related responses N1/P1 and enhancement of P200 for novel word forms over the course of exposures. More importantly, we predict the beneficial effects of language proficiency on the establishment of new L2 linguistic memory traces, reflected in a more significant reduction in length effects for early N1/P1 and P200 effects associated with increased L2 proficiency during training. This, in turn, would lead to a more automatized switch from serial, sublexical processing of novel word forms to faster, more parallel activation that occurs as the result of orthographic representations being established.
3. Correspondingly, assuming that L2 proficiency levels indeed modulate the rapid build-up of behavioral and neural memory traces for L2 novel word forms during online training, a potential influence of language proficiency on orthographic learning would also be expected, as measured by a free recall task. We hypothesize that higher linguistic proficiency would benefit performance on orthographic memorization for novel word forms in spelling.

[1]Prior L2 orthographic learning studies shared one common set of choices in the selection of stimulus characteristics and the tasks involved. That is, when learning novel word forms, both the sources of the word per se and its meaningful context were tapped into. It is still a matter of debate whether semantic and syntactic contexts are relevant to strengthening the memorization of novel orthographic patterns. While this notion converges with prior research on the acquisition of emotional, concrete, and abstract words (e.g., Borghi et al., 2017; Snefjella et al., 2020), other studies have found null (Duff & Hulme, 2012; Nation et al., 2007) or even negative (Landi et al., 2006) effects. Rather, it appears that a meaningful context may induce attentive analysis of detailed orthographic representations, potentially undermining the pivotal role of phonological decoding (Share, 2004). Such contradictory evidence therefore raises a question: since the orthographic learning mechanism has phonological decoding at its core, at what level, if at all, do decoding processes lead to changes in the nature and efficiency of novel word learning? From our viewpoint, probably the most appropriate approach is to explore the processes of orthographic acquisition under the conditions of how it interacts with phonological decoding without interference from the semantic factor. Thus, in the experiment reported here, we presented pseudoword stimuli in isolation to bilinguals six times in different random orders with instructions to read these stimuli aloud as quickly as possible as they appeared on the screen.

Methods

Participants

The present study included 48 Chinese-English bilinguals (M_{age} = 23, SD_{age} = 1.48, N = 19 males) from Dalian University of Technology. All participants were right-handed (according to the Edinburgh Handedness Inventory) and native Chinese speakers with normal or corrected-to-normal vision. None of them reported neurological, developmental, psychological, or learning disorders. Their average English exposure was 16 years (range = 14-19, SD = 1.72) and took place in a formal classroom environment from when they were approximately seven years old (the common age to start English learning in China). None of them were known to be early bilinguals in any other language or to have studied abroad. Five individuals were excluded due to inadequate reading or phonological short-term memory scores (see below), leaving a sample of 43 participants. All of them provided informed consent before taking part in the study and were monetarily compensated for their participation. Ethical approval for the study was provided by the Research Ethics Committee of the University of La Laguna (Comité de Ética de la Investigación y Bienestar Animal, University of La Laguna, Registration number: CEIBA2021-3104).

We conducted a careful two-stage procedure to recruit participants. In the first stage, we ensured that all recruited targets were college-level students who had attended weekly English courses, either passing the Test for English Majors-Band 8 (the highest-level standardized test of English proficiency, the minimum requirement of vocabulary size is 13,000) or with a TOEFL or IELTS (Test of English as a Foreign Language) score of more than 105/120 or 7.5/9.0 (corresponding to common European Framework score C1), respectively. Others showed TOEFL or IELTS scores of at least 80/120 or 6.0/9.0 (corresponding to common European Framework score B2), respectively, or completed the College English Exam Band 4 (the basic standardized test of English ability, the minimum requirement of vocabulary size is 5,000). Therefore, at the time of testing, participants had a high-intermediate or low-advanced level of English proficiency.

In the second stage, in accordance with the self-teaching hypothesis wherein phonological decoding is a precondition for orthographic acquisition, we controlled that all participants recruited in the current study did not differ in their decoding abilities and working memory capacity (see Hamada & Koda, 2010). Thus, a battery of tests was applied for a standardized assessment of L2 proficiency (computer-based DIALANG language diagnostic system and the LexTALE test), word/pseudoword reading (English word and pseudoword-reading tasks), and short-term phonological memory (digit span test). See the on-line supplementary material (*session: Evaluations of the English Proficiency, Word/Pseudoword Reading Ability, and Phonological Short-term Memory*) for detailed procedures.

As shown in Figure 1, the overall English proficiency ranged widely from 7 to 29 in the DIALANG self-assessment ($M_{DIALANG}$ = 16, $SD_{DIALANG}$ = 5.72, Max = 30) and from 48 to 86 in the LexTALE test ($M_{LexTALE}$ = 61, $SD_{LexTALE}$ = 4.33, Max = 100), with both measures highly correlated, rho = .91, HDI [0.85, 0.96], suggesting that different L2 proficiency levels were likely represented in the current sample. Furthermore, participants scored relatively high in both word ($Range_{word}$ = 191–200, SD_{word} = 2.11, Max = 200), pseudoword naming tests ($Range_{pseudoword}$ = 58–61, $SD_{pseudoword}$ = 0.77, Max = 61) as well as in the digit memory test ($Range_{digitspan}$ = 93–105, $SD_{digitspan}$ = 2.52, Max = 132). Additionally, no correlation between L2 proficiency and the other three preliminary tests was found (*rho*s < .19). Results obtained in the preliminary tests can be summarized by saying that all participants had word-reading skills and basic decoding

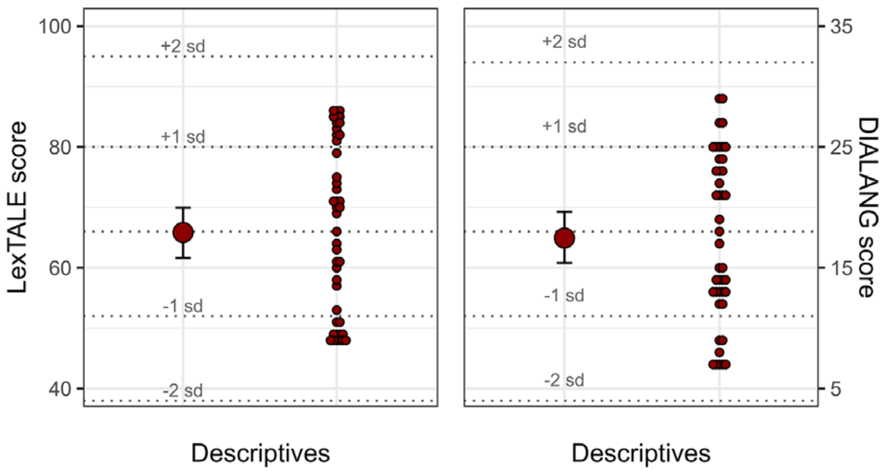

Figure 1. LexTale (left panel) and DIALANG (right panel) data. Plots include the mean ± 95% highest posterior-density interval (HPDI) and the distribution spread (± 1 and 2 standard deviations from the mean).

ability to translate printed input into corresponding spoken forms, as well as similar phonological short-term memory capacity, regardless of their proficiency levels.

Stimuli

A list of 16 pseudowords (eight monosyllabic, four- to five-letter short items and eight bisyllabic, seven- to eight-letter-long items) was used as novel words to test orthographic learning. The stimuli were drawn from previous studies focused on orthographic learning using the same or similar methodology as used here (Kwok et al., 2017; Kwok & Ellis, 2015; Maloney, 2008), and only pseudowords with either no or few neighbors were used (maximum number of neighbors = 2). None of them were homophones of English words or began with voiceless fricative phonemes (“f,” “s,” “sh,” or “th”) to optimize voice-key activation. Each pseudoword was phonotactically legal in the English language and had one obvious, canonical pronunciation based on the standard grapheme-to-phoneme conversion rules of English.

As control words, 16 high-medium frequency words (expressed as SUBTLEX-UK log frequency Zipf values, Van Heuven et al., 2014) with matching length and syllable structure to the pseudowords were used as familiar words. Type and token bigram frequency was assessed using British Lexicon Project (Balota et al., 2007). Coltheart’s N (Coltheart et al., 1977) was calculated by using the *coltheart.N* function from the vwr package. All stimuli were matched based on their initial letters, neighborhood density, and mean log bigram frequency (details for the matching of psycholinguistic variables are provided in supplementary materials Table S1).

Procedure

During the training session, stimuli were displayed on a gray background in lowercase Courier New 18-pt font at the center of a computer screen by means of E-Prime software (Version 2.0, Psychology Software Tools, Inc., Pittsburgh, PA). Each trial began with a fixation cross presented at the center of the screen for 1,000 ms, followed

https://doi.org/10.1017/S0272263123000426 Published online by Cambridge University Press

by a stimulus (word or pseudoword) for 2,000 ms, a blank screen for 500 ms, and then a blink command for 1,000 ms before the next trial began. Words and pseudowords were interleaved and repeatedly presented across six different blocks; each stimulus was encountered only once per block (i.e., repeated six times) and in a different random order across blocks. Participants were asked to read each item aloud as quickly and as accurately as possible; no feedback was given, and the stimulus remained on the screen for the whole 2,000 ms interval regardless of whether it was pronounced or not. Participants were informed that their learning performance would be evaluated in a task afterward. Both behavioral (naming latencies) and EEG data were recorded during the task. A voice key integrated into a serial response box (SR-Box, Psychology Software) was employed to record participants' utterances during the naming task. After each block, participants were allowed to take a break and then to press the space bar when they were ready to continue. Following the learning phase, participants completed a free recall task in which they were told to write down the experimental stimuli that they had previously seen. There was no time restriction or feedback given to the participants. The whole experimental session lasted around 40 min.

EEG data recording, preprocessing and articulation artifacts trimming

EEG signals were registered during the training session by means of 64Ag/Cl active electrodes, amplified, and digitized with an ActiChAmp amplifier (Brain Products GmbH, Gilching, Germany) at 1,000 Hz sampling rate. Electrode impedances were kept below 5 kΩ. Two electrodes were placed below and at the canthus of the left eye to monitor horizontal movements and eye blinks. An additional analog channel was used as an input signal recorded by a sound sensor for the detection of utterances' onset time. The continuous EEG recording was referenced to the vertex electrode (Cz) and filtered (low cutoff: 0.1 Hz; high cut off: 100 Hz). A notch filter was set at 50 Hz.

EEG data were preprocessed and analyzed using MNE-Python (Gramfort et al., 2013) analysis package. The automated EEG data processing workflow was operated as follows: Raw data were re-referenced offline to the averaged mastoid signal and were band-pass filtered (using a Butterworth 6th-order high-pass filter at 0.1 Hz and a Butterworth 8th-order low-pass filter at 40 Hz). Responses time-locked to the onset of the stimuli were extracted between −200 and 1,000 ms and down-sampled to 250 Hz. Baseline correction was applied to each epoch using the 200 ms prestimulus interval. Blinks and saccade artifacts were detected via automated MNE functions and used in an independent component analysis (ICA) to help identify the electrooculographic- and electrocardiographic-related signals and sources with sub-Gaussian distribution (e.g., length noise and slow activity). The automated identification of ICA components was implemented by python function run_icausing the fastica algorithm (Hyvärinen, 1999, see Haumann et al., 2016 for details) and then verified manually to avoid overremoval of components; the data were visually inspected again after epoching. In addition, all data points for naming errors, no responses, and outliers of naming latencies that deviated from a range of ± 2.5 standardized residual errors were discarded from the analysis (i.e., behavioral trial trimming, see below).

Several ERP studies examining overt naming production in monolinguals (Riès et al., 2013) and bilinguals (Misra et al., 2012) have suggested that articulation movements tend to be initiated within the early segment of the recording epoch. Therefore, an extended strategy using Residue Iteration Decomposition (RIDE) was applied to avoid contamination of the EEG signal by articulation artifacts (Ouyang

https://doi.org/10.1017/S0272263123000426 Published online by Cambridge University Press

et al., 2016). Based on the time locking to the stimulus onset, articulation times, and an iterative scheme (i.e., step for decomposition and time window updating were iterated until convergence, Wang et al., 2015), RIDE was performed to differentiate stimulus-associated linguistic processes and speech-related EEG activity. The decomposing procedure is detailed in the online supplementary material (*session: Residue Iteration Decomposition*). Overall, a minimum of 26 trials were included per training exposure (out of 32 trials per exposure, thus > 80% of trials). One participant was excluded from further analyses as the number of valid segments in two conditions was lower than 60%.

EEG data analyses

The resulting epoched EEG data were submitted to a permutation test using threshold-free cluster enhancement (TFCE, Smith & Nichols, 2009) to tentatively explore the potential time windows and the regions of interest relative to significant differences between conditions. The principal advantage of the TFCE method lies in its ability to evaluate effects at all channels, without the need for averaging data within a specific time window or a designated region of interest pertaining to significant differences between conditions (Maris & Oostenveld, 2007). Consequently, this approach enables an objective determination of the precise timing of ERP effects that reflect neural modulations of the length effect throughout the exposures. We refer readers to the on-line supplementary material (*session: Exploratory Analysis*) for the statistical analysis and results. Data averaged across channels and time windows identified in exploratory analysis were further analyzed by linear mixed-effect models (LMMs). Using mixed-effects modeling can easily accommodate both quantitative and qualitative predictors, allowing us to integrate individual proficiency levels and all training exposures without relying on dichotomization and associated loss of power (see Alday et al., 2017).

Statistical modeling

Statistical analyses for both naming latencies (on ms, log transformed for a better correction of the right skewness) and EEG data (μV, default Gaussian family) collected during the training were carried out using LMMs with the lme4 package 1.1.26 (Bates, Mächler, et al., 2015) in R (R Core Team, 2018). For behavioral data, only naming latencies were analyzed, as accuracy was at ceiling across all trials (>95%) with few occasional reading errors. We used LMMs where behavioral and brain responses were predicted by fixed effects of length (short vs. long items), lexicality (familiar vs. novel words), proficiency (individual scores), exposures (1 to 6) and their interactions. We called the hypr package 0.1.7 (Rabe et al., 2020) to design sequential difference contrasts for categorical variables (two-level predictors length and lexicality: 1/2, -1/2). Exposures was encoded with an adapted treatment contrast, with the intercept as the grand mean, and therefore the coefficients reflect comparisons of each condition (exposure 2–6) to the baseline condition (exposure 1). Scores from the DIALANG language diagnostic system and the LexTALE were standardized and summed into the proficiency factor and centered on the mean in the model. Thus, the intercept was estimated as the grand average across factors over successive exposures and the resulting fixed-effect estimates can be interpreted as simple main effects. Scores of preliminary tests and average years of English exposure were included as covariates. Moreover, interleaving familiar and novel words in mixed blocks may result in block/

list context effects—namely, that the speed of processing might be adjusted according to the ease or difficulty of the preceding trial (Taylor & Lupker, 2001) or not (Reynolds et al., 2012). We thus included the order of trials as a covariate as well. All covariates were standardized.

For the random structure, we chose a compromise between the maximal building (Barr et al., 2013) and the parsimonious (Bates, Kliegl, et al., 2015) approaches, by maintaining the simplest random structure when none of the adopted models with maximal ones could reach algorithm convergence. The model-trimming procedure included the comparison of nested models using likelihood-ratio tests (Seedorff et al., 2019). The final model for behavioral data included random-effect terms for the intercepts of individual participants and items, and the model for ERP data included random intercepts for participants (here we analyzed an aggregated voltage amplitude across trials, see Alday et al., 2017, for the implementation of LMMs in the psycholinguistic field). The resulting models and their corresponding more complex models did not diverge in their results (see table S4 for a summary of estimates and standard errors across all models reported). All models were fit based on restricted maximum-likelihood estimation (REML). For the model summaries, we considered the estimate of the contrast coefficient with absolute t values larger than 1.96 as being indicative of a precise estimate (Baayen et al., 2008). The data, code, and experimental materials necessary to reproduce the present study are freely available at https://osf.io/dpcvs/.

Results

Behavioral data: Naming latencies in the word-reading task

Of the overall 8,256 data points, inaccurate reading/no responses ($n = 273$, 3.3%) and data points that deviated from a range of ± 2.5 standardized residual errors ($n = 63$, 0.8%; model criticism, see Baayen et al., 2008, for the application of model criticism in identifying overly influential outliers) were discarded, leaving a total of 7,920 data points for further LMMs analyses. A summary of model fit for the general analysis of naming latencies data can be seen in Table S2.1 in the supplementary materials. The grand mean naming latencies had an estimate of 6.6 as represented by the intercept. The statistical results were placed alongside the higher order interaction for ease of interpretation. A four-way interaction effect between lexicality, length, proficiency, and exposures became marginally significant in the fourth exposure (-0.0028, $t = -1.8$) and significant in the fifth (-0.0035, $t = -2.3$) and the last (-0.0044, $t = -2.7$) encounters, reflecting that after the fourth exposure, the patterns of length were gradually changed according to the type of stimulus and the L2 proficiency levels of participants. As shown in the diptych plot in Figure 2, the length effect for familiar words decreased over exposures, independently of L2 proficiency (right panel). Conversely, from the fourth exposure onwards, the length effect for novel word forms gradually decreased with increasing proficiency levels in a linear fashion, with the magnitude of the reduced length effect associated with higher proficiency levels (left panel).

Separate analyses of the data were conducted to better understand the direction of the four-way interaction. We explored the effect of L2 proficiency level and converted participants' scores into z scores—(proficiency score – average score) / standard deviation. Two sets of analyses were computed by treating proficiency as a discrete variable and then defining two groups of participants with higher (hereinafter HPG, > z_{mean}, 19.91% probability of language proficiency that follows a gamma distribution with parameters 2 and 1.5) and lower (hereinafter LPG, < z_{mean}, 19.44% probability)

https://doi.org/10.1017/S0272263123000426 Published online by Cambridge University Press

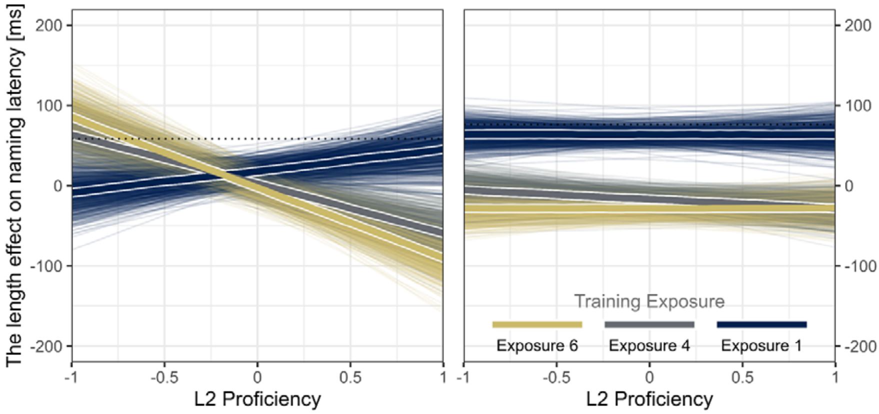

Figure 2. Changes in the effect of length on naming latencies (differences between short and long items [long minus short]) across exposures as a function of L2 proficiency level, separately for novel words (left panel) and for familiar words (right panel). Thick lines represent the reported models' estimates for the first, fourth, and sixth exposures, with shadow areas indicating the standard errors. Thin lines represent the raw data from a participant-based level. The dotted lines indicate the model intercept.

proficiency levels (see table S2.2 and S2.3 for a summary of model fit and, table S3 for descriptive results of the HPG and the LPG, respectively). Figure 3 shows the estimated marginal means (EMMs) and confidence intervals from the models for the HPG and the LPG. The difference between both groups was mainly identified from the fourth repetition onward. In the analysis for the HPG, contrasts involving two-way interactions between lexicality/length and exposures exceeded the threshold of absolute t values larger than 1.96, but no three-way interaction contrast was found. Thus, a similar reduction of the length effect was observed in familiar and novel word forms across exposures, with the difference in naming latencies between long and short novel words reducing from 116 ms in the first exposure to 7 ms in the last exposure (see Figure 3, left panels). In the LPG, contrary to the two-way interaction effects observed for the HPG, all contrasts remained below the threshold. Nevertheless, the effect of the three-way interaction was marginally significant in the fourth exposure (0.096, $t = 2$), the fifth exposure (0.12, $t = 2.2$), and the last exposure (0.17, $t = 3.1$). This indicated that from the fourth repetition onward, the reduction in naming latencies was greater for short than for long novel words. Consequently, the length effect was larger in the last (123 ms) than the first (91 ms) exposure, whereas naming latencies for long and short familiar words converged at the end of training (15 ms), as shown in Figure 3 (right panels).

Electrophysiological data in the word-reading task

The mass univariate analyses (see online supplementary material. *session: Exploratory Analysis*) revealed the modulation of two brain potentials typically linked to the early stages of orthographic processing. In particular, repeated exposure to novel words exhibited a reduction of early frontal negativity between 90 and 120 ms poststimulus onset (i.e., N1 component), followed by a positive enhancement extending approximately from 150 to 210 ms over fronto-central scalp sites (i.e., P200). ERP data across each condition and participant were averaged over these time windows (90–120 and

https://doi.org/10.1017/S0272263123000426 Published online by Cambridge University Press

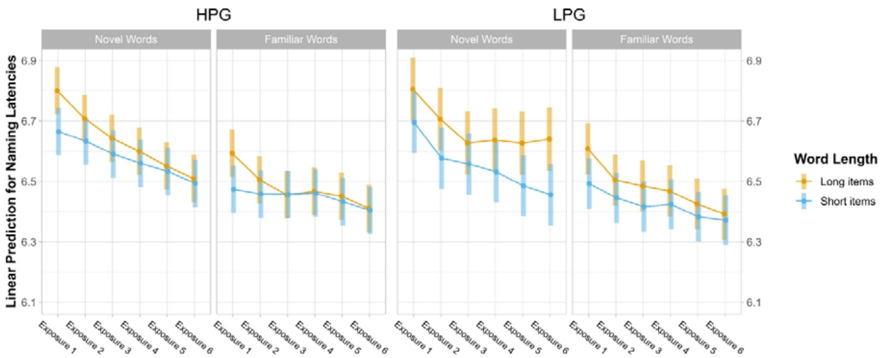

Figure 3. Visualization of fixed-effect estimates for model fit per group. The models included random-effects terms for the intercept of individual participants and items and the fixed effects for lexicality (separate panels), length (in color), and exposures (*x*-axis). Linear prediction for naming latencies (i.e., estimated coefficients, *y*-axis) from two models showed substantial effects for the two-way interaction contrasts between lexicality and exposures and between length and exposures (HPG) and for the three-way interaction contrast among lexicality, length, and exposures (LPG). Dots represent the EMMs, and bars correspond to 95% confidence intervals. Lines indicate changes in naming latencies across repetitions for each condition.

150–210 ms) across the following electrodes: Fz/F1/F1/FCz/FC1/FC2 for the mean amplitudes of the N1 time window and FCz/FC1/FC2/Cz/C1/C2 for the P200 time window. Data averaged across preselected channels and time windows were further analyzed by LMMs to obtain a general picture of the changes in the early stage of L2 orthographic learning as a function of L2 proficiency.

Figure 4 displays averaged ERP waveforms and scalp distribution of the activity elicited by each condition at N1 and P200 time windows. LMM analysis conducted for the early N1 component revealed that the reduction of the length effect in the N1 component was associated with proficiency levels from the third exposure onward (i.e., three-way interaction: Length × Proficiency × Exposures, E3: -0.04, $t = -2.1$ to E6: -0.046, $t = -2.4$), indicating a larger magnitude of decreased length effect with higher L2 proficiency levels (see Figure 5). No other contrast exceeded the threshold of absolute t values larger than 1.96. A summary of model fit for the electrophysiological data of the N1 component can be seen in table S2.4 in the supplementary materials.

Regarding the P200 time window, the model (see table S2.5 for a summary of model fit) illustrated a significant four-way interaction effect for lexicality, length, proficiency, and exposures in the fifth (0.14, $t = 3.5$) and the last (0.14, $t = 3.7$) exposures. Separate analyses were then performed on the P200 activity exhibited by familiar and novel words in the first and last exposures, with length, exposures and proficiency remaining as fixed effects (see Table S2.6, Table S2.7 for a summary of model fit for pseudowords and words respectively). In the later repetitions (E3 to E6), contrasts involving three-way interactions of length, proficiency, and exposures exceeded the threshold of absolute t values larger than 1.96, indicating that the length effect in P200 for novel words decreased as L2 proficiency increased (see Figure 6, left panel). Conversely, the model fit for familiar words did not show any significant effect in relation to Proficiency, and neither did the interaction effect for Length × Exposures (all t values below the threshold of 1.96), indicating no modulation of proficiency or reduction in the length effect for familiar words (right panel).

https://doi.org/10.1017/S0272263123000426 Published online by Cambridge University Press

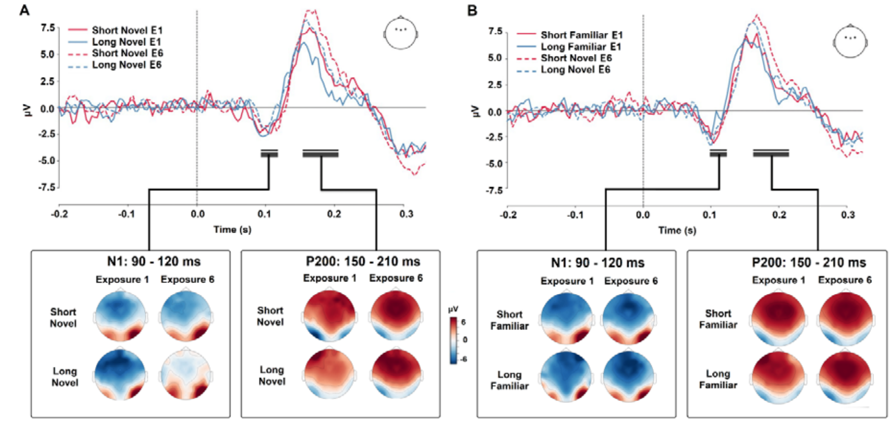

Figure 4. Grand average ERP waveforms at fronto-central channels (FC1, FCz, FC2) for (A) novel words and for (B) familiar words. Black rectangles indicate early (90–120 ms) and late (150–210 ms) time windows, showing the effects of exposures and length for novel and familiar words. Topographic maps depict the distribution of ERP activity elicited in each condition and time window.

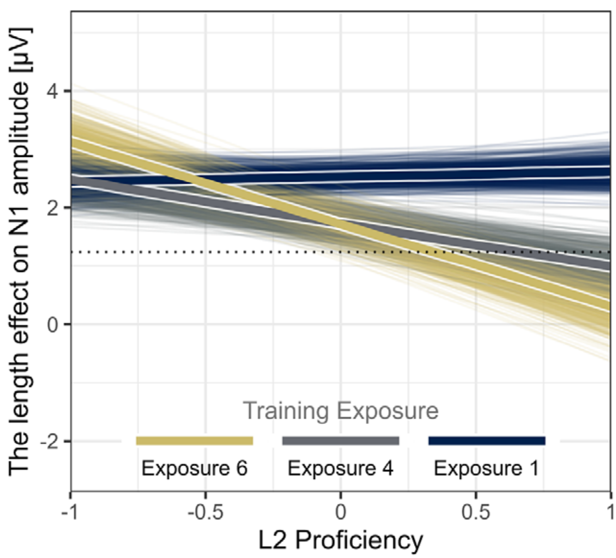

Figure 5. Changes in the effect of length on N1 component (differences between short and long items [short minus long]) across training exposures as a function of L2 proficiency levels. Thick lines represent the reported models' estimates for the first, fourth, and sixth exposures, with shadow areas highlighting the standard errors. Thin lines indicate the electrophysiological data from a participant-based level. The dotted lines indicate the model intercept.

Posttraining behavioral data

The analyses of the free-recall data were focused on novel words, as our observation showed that spelling accuracy to familiar words was at ceiling for the given sample, with only occasional spelling errors (> 88% accuracy rate trials per subject). Two sets of scores were computed and submitted to separate analyses: (1) a whole-word spelling

https://doi.org/10.1017/S0272263123000426 Published online by Cambridge University Press

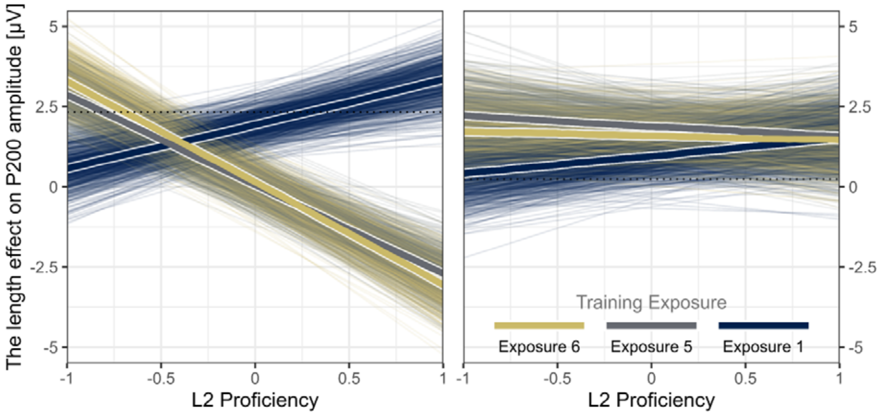

Figure 6. Changes in the effect of length on P200 component (differences between short and long items [short minus long]) across training exposures as a function of L2 proficiency levels, separately for novel words (left panel) and for familiar words (right panel). Thick lines represent the reported models' estimates by exposures, with shadow areas indicating the standard errors. Thin lines represent the electrophysiological data from a participant-based level. The dotted lines indicate the model intercept.

score that corresponded to the correct spelling that was strictly identical to the target orthography and (2) a by-grapheme spelling score that corresponded to the percentage of target graphemes accurately spelled (adapted from Ginestet et al., 2020).We ran Bayesian multilevel logistic regression models fitted in stan (Stan Development Team, 2018) via the library brms (Bürkner, 2017) to examine the learning outcomes, given that all frequentist models fitted for the accuracy data that included relevant random structures failed to reliably converge. Having convergence warnings in the context of mixed-effects modeling of accuracy is hardly surprising (Matuschek et al., 2017); our solution to these difficulties was opting to assess learning performance using Bayesian approaches. The fixed effects were associated with length (1/2, −1/2), proficiency levels (centered on the mean), and their interactions. The random structure included by-participants and by-item intercepts with by-participant random slopes for length and by-item random slopes for proficiency. The same model formula fitted the whole-form-based and by-grapheme-based data. Markov chain Monte Carlo sampling (Hoffman & Gelman, 2014) was implemented with four chains distributed between 16 processing cores to draw samples from the posterior probability distribution. The models included diffuse conjugated priors with a Cauchy distribution for the parameters of the logistic regressions (Gelman et al., 2017), and the 95% highest posterior-density interval (HPDI) was reported. The models were considered to converge successfully based on the potential scale reduction factor (R-hat) which were all reported below 1.01 (Brooks & Gelman, 1998).

Table S5 of supplementary materials provides a summary of posterior predictive distributions with 95% HPDI. Higher L2 proficiency levels were associated with increasing probabilities of spelling correctly the whole-form-based, 0.74, 95% HPDI [0.53, 0.95], or the by-grapheme-based, 0.42, 95% HPDI [0.34, 0.49], novel word forms. Thus, the results suggest a positive link between effective orthographic learning and L2 linguistic proficiency. Additionally, a significant interaction effect between length and proficiency on by-grapheme-based scores was observed, 0.26, 95% HPDI [0.11, 0.40], showing within-group variability in participants with different proficiency levels. An analysis of individual differences (see supplementary material Figure S3 of posterior

https://doi.org/10.1017/S0272263123000426 Published online by Cambridge University Press

intercept and length estimates for individual posterior medians) suggested that individuals with lower proficiency levels could spell short novel words more accurately than long items. This performance tends to be closer to that of individuals with higher proficiency levels. However, higher proficient learners showed better learning performance on long novel word forms, either considering the whole-form-based or by-grapheme-based scoring.

Power considerations

We called function *simulate ()* from the *lme4* package 1.1.26 (Bates, Mächler, et al., 2015) to sample from the fixed effects and the random structure of overall-effect models by using the behavioral and EEG data obtained. A total of 1,000 new data sets were created by randomly labeling trials as missing trials, which were then excluded from the data sets. Then, a model with the same effect structure was estimated for each new data set. We considered the percentage of models where three-way or four-way interactions were detected (i.e., for which $p < .05$) as the estimate for statistical power. Based on 1,000 simulated samples, the estimated statistical power of .863 for behavioral data, 1. for N1, and .967 for P200 amplitude data (i.e., in 863/967/1,000 out of 1,000 simulation runs, the model detected significant four-way interactions for behavioral data and P200 data and significant three-way interaction for N1 data), exceeding the threshold value of .80. We thus consider that our study has power enough to be informative.

Discussion

The present study examined the neural correlates of L2 orthographic learning and the effect of L2 linguistic proficiency on the automatic build-up of memory traces for L2 novel written word forms. Behavioral data (naming latencies) and ERPs were obtained from Chinese-English bilinguals with various proficiency levels to explore visual word-learning processes and brain dynamics during orthographic acquisition (within the first 200 ms poststimulus onset). Modulations in the length effect to novel words across the training (six exposures) at both behavioral and neural levels were taken as a direct index of the formation of L2 novel orthographic representations. Our results revealed that the process of L2 orthographic acquisition and its underlying neural mechanisms varied as a function of L2 proficiency. Namely, from around the fourth exposure onward, increasing L2 proficiency levels were associated with a sharper decline in the length effect for novel words, in terms of both behavioral (i.e., naming latencies) and electrophysiological responses (i.e., N1 reduction and P200 enhancement) in the early stages of orthographic processing. In parallel, a posttraining free-recall task revealed that individuals with higher proficiency were more efficient at retrieving whole or part of the orthographic knowledge of the novel word forms. In contrast, less proficient learners could stabilize the coarse spelling of short novel words but lacked the ability to establish new representations of long items in response to training. Overall, individual L2 proficiency, from a quantitative perspective, was shown to alter the degree of neuronal plasticity associated with L2 word recognition in general and the online establishment of new memory traces to novel words within a brief training in particular. In what follows, we discuss in detail the behavioral and ERP results within the framework of computational models of reading before moving to the specific findings concerning L2 proficiency revealed in the present study.

The orthographic learning observed in the naming latencies of highly proficient L2 learners during training was consistent with previous studies on skilled English readers (Kwok et al., 2017; Kwok & Ellis, 2015; Maloney, 2008). The length effect on novel words was reduced as soon as after one repetition and continued to decline with the subsequent encounters, in line with prior findings showing rapid and automatic orthographic learning occurring after a single decoding opportunity (Ginestet et al., 2020; Nation et al., 2007). The difference in naming latencies between long and short novel words was no longer noticeable after the fourth or fifth repetition. Within the framework of the DRC and CDP+ models, the result can be interpreted as indicating the formation of larger scale orthographic and phonological representations for novel word forms. Those representations allow a relatively rapid transition from a serial, length-sensitive sublexical route to a more parallel, lexical process that has direct access to stored representations in the orthographic input and phonological output lexicons.

In accordance with predictions of dual processing models (i.e., DRC and CDP +), the early brain activity relative to novel words demonstrated a rapid modulation within the first 250 ms after stimulus onset, as a result of repeated exposure. Specifically, long novel word forms exhibited a larger magnitude of reduction in the N1 effect (90–120 ms poststimuli onset) than short ones. These results are in line with Grainger & Holcomb (2009), showing that orthographic code with long visual structure recruited more cognitive resources to transform location-specific, retinotopic mapping (i.e., visual object) to location-invariant, word-centered representations (i.e., linguistic stimuli). Thus, the diminished influence of length on this neural responsiveness could indicate optimal processing of novel words across the full spectrum, from a slow, attention-demanding phonological decoding mechanism to whole-word-based recognition. Following this interpretation, the subsequent P200 enhancement (150–200 ms post-stimuli onset) likely reflects the increase in automatic access to newly established representations. Such positivity elicited by novel word forms increased over exposures, and the P200 discrepancies between long and short novel words were eliminated in later repetitions. This pattern suggests that more holistic, lexical-based processing of novel orthographic representations was acquired across repetitions. In sum, the repetitive exposure to novel words induced a decreased N1 and increased P200 amplitude across the training, leading to reduced length effects on these ERP components. These neural responses reflect modulation in a set of orthographic processes underlying the acquisition of L2 novel word forms, supporting the available behavioral evidence, and are also consistent with previous EEG/MEG studies conducted in monolingual (Partanen et al., 2018) and bilingual populations (Bermúdez-Margaretto, Kopytin, et al., 2020).

Notably, the present results pinpoint the influence exerted by L2 proficiency on early brain responses (~ 100 ms poststimulus onset) related to the acquisition of novel words. From around the fourth exposure onwards, L2 proficiency was favorably associated with the reduction of length effects on N1 and P200 responses as well as on naming latencies. Together, these findings indicate a positive relationship between L2 proficiency levels, the engagement of lexical processing, and the magnitude of learning-related brain response to the novel, nonnative orthographic input. It should be emphasized that the overall variability in the length effect for familiar words over the course of exposures seems to be unaffected by L2 proficiency level. Thus, the results suggest that access to lexicalized items occurs regardless of individual variability and support the idea that the early modulations of linguistic proficiency described here are mainly related to learning processes underlying new memory trace formation.

As outlined in the introduction, previous literature has widely reported that four or five repetitions of novel words within a single experimental session are adequate to achieve the orthographic acquisition that enables rapid visual word recognition. Correspondingly, the modulation of proficiency observed here in behavioral and brain responses related to the establishment of orthographic representations has occurred from the fourth presentation onward. As the individual knowledge of a word grows with cumulative exposure, the observed variations in written word-form identification across bilinguals with varied proficiency levels are likely to reflect the distinction between bottom-up input analysis (attentional and visual processes) and automatic, top-down predictions (the influence of lexicalized representations, Dehaene and Cohen, 2011; Twomey et al., 2011). Specifically, given that the accuracy of all participants was reasonably high in the reading task, lower proficient participants might have benefited from visual attention and low-level specialization to acquire orthographic codes for novel words when mapping letters onto their corresponding sounds (evidenced by the large length effect in N1). In turn, this pattern could result in a greater reliance on sublexical processing and less efficient learning of long-structure items due to the nonautomatized functioning of grapheme-phoneme decoding (large length effect for P200 amplitude and naming latencies). Therefore, attentional bias to form orthographic representations of learned words is constrained by the length structure of novel words, leading to higher cognitive loads caused by exclusively bottom-up processing in long novel word forms, particularly during later training. In contrast, the efficient build-up of new memory traces by more highly proficient learners may reflect the acquisition of word-specific knowledge arising from the combination of automatic, top-down predictions and bottom-up processing of the visual input. In this sense, graphemes activate phonemes, leading to particular combinations of phonemes into syllables or words and thus narrowing the distance between output and target phonology.

Moreover, the influence of L2 proficiency on orthographic learning was evident in the ability to retrieve newly acquired orthographic representations in the free-recall task. Compared with more highly proficient individuals, participants with lower proficiency performed better in spelling short novel words when using fine-grain grading based on trained graphemes. However, they failed to show signs of learning in long items, either considering whole-form or by-grapheme spelling rubrics. It is possible that repeated exposure helped these participants develop exact orthographic representations of the short novel items, even if they were susceptible to the degree of visual analysis. In contrast, this brief period of exposure within a single training session seems insufficient for lower proficient bilinguals to achieve and subsequently retrieve the accurate spelling of long word forms, unless they enhance their abilities to develop specialized and efficient word-recognition mechanisms optimized specifically for reading. Therefore, we propose that the general notion that "orthographic learning occurs after several exposures to printed word forms" should be evaluated with caution in future orthographic learning studies within the context of a second language.

Conclusion, limitation, and future pedagogy

To conclude, the present work demonstrates how L2 proficiency levels can predict the degree and extent of neural plasticity for L2 orthographic acquisition, showing a differential magnitude of the length effect in naming latencies and early neural responses—N1 and P200—related to orthographic processing. Additionally, a glimpse of an important but understudied aspect of orthographic learning emerges when the current findings are combined with previous evidence obtained in studies with dyslexic populations. In particular, the learning performance reported here in bilinguals with lower proficiency levels resembles the deficit in reading automatization and low degree of expertise in rapid orthographic acquisition shown by dyslexic populations, either children (Martínez-García et al., 2019) or adults (Kwok & Ellis, 2014). Importantly, if these word-level reading difficulties in both dyslexic readers and bilinguals with low-advanced levels of L2 proficiency are not adequately addressed, deficiency in the process of building up lexical entries may lead to more pervasive challenges related to native or L2 literacy development and language comprehension, which in turn may cause long-term negative consequences that handicap academic and lifelong success.

Given the evidence presented here, a testable hypothesis emerges: for bilingual individuals with lower L2 proficiency, conventional learning methods (i.e., four/five exposures within a single training session) may prove inadequate to significantly contribute to the development of L2 orthographic representations. Future research may explore modifications to the orthographic learning paradigm for improved outcomes. On a more theoretical level, lexicalization requires time for the consolidation of newly formed memory circuits through information transferred from hippocampal to cortical connections (see McClelland et al., 1995, for the complementary learning system approach; also see Davis and Gaskell, 2009, for its application to word learning), a process in which sleep is possibly crucial (Tamminen et al., 2010; but see the opposite view, Lindsay and Gaskell, 2013). Given this, we advocate a learning program comprising various, spaced sessions with very short exposure spreading across several days rather than massively exposing novel words in a single experimental session (see Namaziandost et al., 2020; Wegener et al., 2022). Another alternative is to consider integrated reading and writing instruction (Al Otaiba et al., 2022), as recent evidence indicates that reading and writing outputs draw on shared information and cognitive mechanisms. Writing practice has been documented to positively influence phonemic awareness, alphabetic awareness, and sight word identification (Ehri, 2005), which we believe may help less proficient L2 readers achieve automatization of low-level decoding abilities and improve their reading performance. Future studies might consider the number of exposures, spaced versus massive training periods, and the role of sleep consolidation to investigate the improvement of orthographic learning in readers with a deficit at the level of whole word-form specialization.

One limitation of the present study is related to the nature of English orthography. Given that English is a deep alphabetic language with the high degree of inconsistency of spelling-sound mappings, its word-learning mechanisms warrant further investigation to address how L2 learners deal with inconsistent grapheme-to-phoneme mappings during orthographic acquisition.

Acknowledgments. This research was supported by the Spanish Ministry of Science, Innovation PID2020-114246GB-100, "Plan General del Conocimiento", through the project "Orthographic learning in a second language: cross-linguistic and sensorymotor factors" awarded to University of La Laguna, and by the Consejería de Economía, Conocimiento y Empleo (CEI Canarias-ULL, SD-19/02) of the Canary Island Government.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0272263123000426.

Data availability statement. The experiment in this article earned the Open Materials badges for transparent practices. The data are available at https://osf.io/dpcvs/

Competing interest. No potential conflict of interest was reported by the author(s).

References

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica*, *128*, 466–478.

Al Otaiba, S., McMaster, K., Wanzek, J., & Zaru, M. W. (2022). What we know and need to know about literacy interventions for elementary students with reading difficulties and disabilities, including dyslexia. *Reading Research Quarterly*, *58*, 313–332.

Alday, P. M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2017). Electrophysiology reveals the neural dynamics of naturalistic auditory language processing: Event-related potentials reflect continuous model updates. *Eneuro*, *4*.

Álvarez-Cañizo, M., Suárez-Coalla, P., & Cuetos, F. (2018). The role of sublexical variables in reading fluency development among Spanish children. *Journal of Child Language*, *45*, 858–877.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, D., Nelson, D., Simpson, G., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*, 445–459.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models.* arXiv. arXiv:1506.04967.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bermúdez-Margaretto, B., Kopytin, G., Myachykov, A., & Shtyrov, Y. (2021). Behavioral and neurophysiological correlates of orthographic learning in L1 and L2 alphabets. In Boris M. Velichkovsky, Pavel M. Balaban, & Vadim L. Ushakov (Eds.), *Advances in cognitive research, artificial intelligence and neuroinformatics: Proceedings of the 9th International Conference on Cognitive Sciences, Intercognsci-2020, October 10-16, 2020*, Moscow, Russia (pp. 345–358). Springer International Publishing.

Bermúdez-Margaretto, B., Shtyrov, Y., Beltrán, D., Cuetos, F., & Domínguez, A. (2020). Rapid acquisition of novel written word-forms: ERP evidence. *Behavioral and Brain Functions*, *16*, 1–17.

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. *Psychological Bulletin*, *143*, 263–292.

Bosse, M., Chaves, N., Largy, P., & Valdois, S. (2015). Orthographic learning during reading: The role of whole-word visual processing. *Journal of Research in Reading*, *38*, 141–158.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28.

Chauncey, K., Holcomb, P. J., & Grainger, J. (2008). Effects of stimulus font and size on masked repetition priming: An event-related potentials (ERP) investigation. *Language and Cognitive Processes*, *23*, 183–200.

Chung, S. C., Chen, X., Commissaire, E., Krenca, K., & Deacon, S. H. (2019). Testing the self-teaching hypothesis in second language reading. *Writing Systems Research*, 11, 1–*11*.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In In S. Dornič (Ed.), *Attention and Performance VI* (pp. 535–555). Academic Press.

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.

Comesaña, M., Sánchez-Casas, R., Soares, A. P., Pinheiro, A. P., Rauber, A., Frade, S., & Fraga, I. (2012). The interplay of phonology and orthography in visual cognate word recognition: An ERP study. *Neuroscience Letters*, *529*, 75–79.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 3773–3800.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, *15*, 254–262.

de Jong, P. F., & Share, D. L. (2007). Orthographic learning during oral and silent reading. *Scientific Studies of Reading*, *11*, 55–71.

Dufau, S., Grainger, J., & Holcomb, P. J. (2008). An ERP investigation of location invariance in masked repetition priming. *Cognitive, Affective, & Behavioral Neuroscience*, *8*, 222–228.

Duff, F. J., & Hulme, C. (2012). The role of children's phonological and semantic knowledge in learning to read words. *Scientific Studies of Reading*, *16*, 504–525.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*, 167–188.

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*(10), 1–13.

Ginestet, E., Valdois, S., Diard, J., & Bosse, M.-L. (2020). Orthographic learning of novel words in adults: Effects of exposure and visual attention on eye movements. *Journal of Cognitive Psychology*, *32*, 785–804.

Grainger, J., & Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Language and Linguistics Compass*, *3*, 128–156.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-python. *Frontiers in Neuroscience*, *7*, Article 267.

Hamada, M., & Koda, K. (2010). The role of phonological decoding in second language word-meaning inference. *Applied Linguistics*, *31*, 513–531.

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage*, *30*, 1383–1400.

Haumann, N. T., Parkkonen, L., Kliuchko, M., Vuust, P., & Brattico, E. (2016). Comparing the performance of popular MEG/EEG artifact correction methods in an evoked-response study. *Computational Intelligence and Neuroscience*, *2016*, Article 7489108.

Hoffman, M. D., & Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *5*, 1593−1623.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE transactions on Neural Networks*, *10*, 626–634.

Kimppa, L., Kujala, T., & Shtyrov, Y. (2016). Individual language experience modulates rapid formation of cortical memory circuits for novel words. *Scientific Reports*, *6*, 1–10.

Kong, L., Zhang, J. X., Kang, C., Du, Y., Zhang, B., & Wang, S. (2010). P200 and phonological processing in Chinese word recognition. *Neuroscience Letters*, *473*, 37–41.

Kwok, R. K. W., Cuetos, F., Avdyli, R., & Ellis, A. W. (2017). Reading and lexicalization in opaque and transparent orthographies: Word naming and word learning in English and Spanish. *Quarterly Journal of Experimental Psychology*, *70*, 2105–2129.

Kwok, R. K. W., & Ellis, A. W. (2015). Visual word learning in skilled readers of English. *Quarterly Journal of Experimental Psychology*, *68*, 326–349.

Kwok, R. K., & Ellis, A. W. (2014). Visual word learning in adults with dyslexia. *Frontiers in Human Neuroscience*, *8*, Article 264.

Landi, N., Perfetti, C. A., Bolger, D. J., Dunlap, S., & Foorman, B. R. (2006). The role of discourse context in developing word form representations: A paradoxical relation between reading and learning. *Journal of Experimental Child Psychology*, *94*, 114–133.

Li, Y., Li, H., & Wang, M. (2020). Orthographic learning via self-teaching in Chinese: The roles of phonological recoding, context, and phonetic and semantic radicals. *Journal of Experimental Child Psychology*, *199*, Article 104913.

Li, Y., Wang, M., & Espinas, D. (2021). Orthographic learning through self-teaching among learners of English as a second language. *Reading and Writing*, *34*, 1295–1320.

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 608.

Maloney, E. A. (2008). *Tracking the transition from sublexical to lexical processing in reading aloud: On the creation of orthographic and phonological lexical representations* [Master's thesis]. University of Waterloo.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190.

Martínez-García, C., Suárez-Coalla, P., & Cuetos, F. (2019). Development of orthographic representations in Spanish children with dyslexia: The influence of previous semantic and phonological knowledge. *Annals of Dyslexia*, *69*, 186–203.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). *Balancing Type I error and power in linear mixed models.* Potsdam Mind Research Repository. http://read.psych.uni-potsdam.de/index.php?option=com_content&view=article&id=144:beap&catid=11:publications&Itemid=13
McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457.
Mou, L. C., & Anderson, N. S. (1981). Graphemic and phonemic codings of Chinese characters in short-term retention. *Bulletin of the Psychonomic Society*, *17*, 255–258.
Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language*, *67*, 224–237.
Namaziandost, E., Mohammed Sawalmeh, M. H., & Izadpanah Soltanabadi, M. (2020). The effects of spaced versus massed distribution instruction on EFL learners' vocabulary recall and retention. *Cogent Education*, *7*, Article 1792261.
Nassaji, H. (2014). The role and importance of lower-level processes in second language reading. *Language Teaching*, *47*, 1–37.
Nation, K., Angell, P., & Castles, A. (2007). Orthographic learning via self-teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of Experimental Child Psychology*, *96*, 71–84.
Ouyang, G., Sommer, W., & Zhou, C. (2016). Reconstructing ERP amplitude effects after compensating for trial-to-trial latency jitter: A solution based on a novel application of residue iteration decomposition. *International Journal of Psychophysiology*, *109*, 9–20.
Partanen, E. J., Leminen, A., Cook, C., & Shtyrov, Y. (2018). Formation of neocortical memory circuits for unattended written word forms: Neuromagnetic evidence. *Scientific Reports*, *8*, 1–10.
Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*, 273.
R Core Team. (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/
Rabe, M. M., Vasishth, S., Hohenstein, S., Kliegl, R., & Schad, D. J. (2020). Hypr: An R package for hypothesis-driven contrast coding. *Journal of Open Source Software*, *5*, Article 2134.
Reynolds, M., Mulatti, C., & Besner, D. (2012). Reading nonwords aloud: Evidence for dynamic control in skilled readers. *Psychonomic Bulletin & Review*, *19*, 1135–1141.
Riès, S., Janssen, N., Burle, B., & Alario, F.-X. (2013). Response-locked brain dynamics of word production. *PloS One*, *8*, Article e58197.
Schwartz, M., Kahn-Horwitz, J., & Share, D. L. (2014). Orthographic learning and self-teaching in a bilingual and biliterate context. *Journal of Experimental Child Psychology*, *117*, 45–58.
Seedorff, M., Oleson, J., & McMurray, B. (2019). *Maybe maximal: Good enough mixed models optimize power while controlling Type I error.* PsyArXiv. https://doi.org/10.31234/osf.io/xmhfr
Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*, 151–218.
Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72*, 95–129.
Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology*, *87*, 267–298.
Share, D. L. (2008). Orthographic learning, phonological recoding, and self-teaching. In *Advances in child development and behavior* (Vol. *36*, pp. 31–82). Elsevier.
Smith, S., & Nichols, T. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, *44*, 83–98.
Snefjella, B., Lana, N., & Kuperman, V. (2020). How emotion is learned: Semantic learning of novel words in emotional contexts. *Journal of Memory and Language*, *115*, Article 104171.
Stan Development Team. (2018). Stan modeling language users guide and reference manual (Version 2.18.0). Stan Development Team. http://mc-stan.org
Stein, M., Dierks, T., Brandeis, D., Wirth, M., Strik, W., & Koenig, T. (2006). Plasticity in the adult language system: A longitudinal electrophysiological study on second language learning. *Neuroimage*, *33*, 774–783.

Stein, M., Federspiel, A., Koenig, T., Wirth, M., Lehmann, C., Wiest, R., Strik, W., Brandeis, D., & Dierks, T. (2009). Reduced frontal activation with increasing 2nd language proficiency. *Neuropsychologia*, *47*, 2712–2720.

Suárez-Coalla, P., Álvarez-Cañizo, M., & Cuetos, F. (2016). Orthographic learning in Spanish children. *Journal of Research in Reading*, *39*, 292–311.

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience*, *30*, 14356–14360.

Taylor, T. E., & Lupker, S. J. (2001). Sequential effects in naming: A time-criterion account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 117–138.

Tucker, R., Castles, A., Laroche, A., & Deacon, S. H. (2016). The nature of orthographic learning in self-teaching: Testing the extent of transfer. *Journal of Experimental Child Psychology*, *145*, 79–94.

Twomey, T., Duncan, K. J. K., Price, C. J., & Devlin, J. T. (2011). Top-down modulation of ventral occipito-temporal responses during visual word recognition. *Neuroimage*, *55*, 1242–1251.

van Daal, V. H., & Wass, M. (2017). First- and second-language learnability explained by orthographic depth and orthographic learning: A "natural" *Scandinavian experiment.* Scientific Studies of Reading, *21*, 46–59.

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*, 1176–1190.

van Viersen, S., Protopapas, A., Georgiou, G. K., Parrila, R., Ziaka, L., & de Jong, P. F. (2022). Lexicality effects on orthographic learning in beginning and advanced readers of Dutch: An eye-tracking study. *Quarterly Journal of Experimental Psychology*, *75*, 1135–1154.

Vergara-Martínez, M., Perea, M., & Leone-Fernandez, B. (2020). The time course of the lowercase advantage in visual word recognition: An ERP investigation. *Neuropsychologia*, *146*, Article 107556.

Wang, F., Ouyang, G., Zhou, C. S., & Wang, S. P. (2015) Re-examination of Chinese semantic processing and syntactic processing: Evidence from conventional ERPs and reconstructed ERPs by residue iteration decomposition (RIDE). *PLoS ONE*, *10*, Article e0117324.

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

Wegener, S., Wang, H.-C., Beyersmann, E., Nation, K., Colenbrander, D., & Castles, A. (2022). The effects of spacing and massing on children's orthographic learning. *Journal of Experimental Child Psychology*, *214*, Article 105309.

Wu, Y., Mo, D., Tsang, Y.-K., & Chen, H.-C. (2012). ERPs reveal sub-lexical processing in Chinese character recognition. *Neuroscience Letters*, *514*, 164–168.

Cite this article: Fu, Y., Bermúdez-Margaretto, B., Beltrán, D., Huili, W., & Dominguez, A. (2024). Language proficiency modulates L2 orthographic learning mechanism: Evidence from event-related brain potentials in overt naming. *Studies in Second Language Acquisition,* 46: 119–140. https://doi.org/10.1017/S0272263123000426