

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Fu, Yang; Bermúdez-Margaretto, Beatriz; Wang, Huili; Tang, Dong; Cuetos, Fernando; Alberto, Domínguez

Title: Struggling with L2 alphabet : the role of proficiency in orthographic learning

Year: 2023

Version: Accepted version (Final draft)

Copyright: © Authors 2023

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Fu, Y., Bermúdez-Margaretto, B., Wang, H., Tang, D., Cuetos, F., & Alberto, D. (2023). Struggling with L2 alphabet : the role of proficiency in orthographic learning. *Quarterly journal of experimental psychology*, 76(12). <https://doi.org/10.1177/17470218231154910>

Struggling with L2 alphabet: the role of proficiency in orthographic learning

Journal:	<i>Quarterly Journal of Experimental Psychology</i>
Manuscript ID	QJE-STD-22-016.R4
Manuscript Type:	Standard Article
Date Submitted by the Author:	10-Jan-2023
Complete List of Authors:	Fu, Yang; Universidad de La Laguna - Campus de Guajara, Facultad Psicología y Logopedia Bermúdez-Margaretto, Beatriz; Universidad de Salamanca, Faculty of Psychology Wang, Huili; Hangzhou City University, School of Foreign Languages Tang, Dong; University of Jyväskylä, Faculty of Information Technology Cuetos, Fernando; University of Oviedo, Psychology Alberto, Domínguez; Universidad de La Laguna - Campus de Guajara, Facultad Psicología y Logopedia
Keywords:	Orthographic learning, phonological decoding, biliteracy, self-teaching, GAMM

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10 **Struggling with L2 alphabet: the role of proficiency in orthographic learning**
11
12

13
14 Fu Yang¹, Bermúdez-Margaretto Beatriz², Wang Huili³, Tang Dong⁴, Cuetos Fernando⁵,
15 Alberto Domínguez¹
16

17
18
19 ¹Universidad de La Laguna, Spain

20 ²Universidad de Salamanca, Spain

21 ³Hangzhou City University, China

22 ⁴University of Jyväskylä, Finland

23 ⁵University of Oviedo, Spain
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

44 ***Corresponding author:***

45 Bermúdez-Margaretto, Beatriz

46 Universidad de Salamanca - Faculty of Psychology

47 Salamanca, Castilla y León 37008

48 Spain

49 bermudezmargaretto@usal.es
50
51
52
53
54
55
56
57
58
59
60

Struggling with L2 alphabet: the role of proficiency in orthographic learning**Abstract**

The present study examined the process of L2 orthographic learning in bilinguals with distant L1-L2 orthographies. Chinese-English bilinguals with various English proficiency levels were trained with novel L2 words during a reading task. In contrast to higher proficient learners, those with lower L2 proficiency exhibited increased effects of length, frequency and lexicality across exposures and at-chance recognition of trained words. Importantly, an additional post-training task assessing the lexical integration of trained words evidenced the engagement in different L1-L2 reading strategies across different levels of L2 proficiency, hence suggesting the L1 holistic processing at the base of the effortful establishment of L2 orthographic representations shown by lower-proficient learners. Overall, these findings indicate the role of L2 proficiency in the influence that cross-linguistic variation exerts on L2 orthographic learning and highlight the need for English education programs to tackle specific grapheme-to-phoneme skills in non-alphabetic target communities.

Keywords

Orthographic learning; phonological decoding; self-teaching; GAMM; biliteracy

Introduction

Orthographic learning refers to the dynamic ability to store word-specific orthographic representations and acquire generic spelling patterns, which is fundamental for the development of reading fluency (Deacon et al., 2019). According to the self-teaching hypothesis (Share, 1995, 1999), phonological decoding contributes to the formation of orthographic memory traces through repeated exposure to novel words, either embedded in meaningful texts (Share, 1999, Exp. 1) or presented in isolation (Share, 1999, Exp. 2 & 3). This hypothesis highlights the central mechanism of phonological decoding in developing orthographic knowledge. In this sense, readers have more chances to store the orthographic form of a novel item when they can decode it successfully (Pritchard et al., 2018; Ricketts, 2011; Ziegler et al., 2014) and hence to rely on the newly acquired word-specific orthographic representation for its direct word recognition (Share, 2008). The self-teaching hypothesis has been successfully tested in studies conducted with adults (e.g., Kwok & Ellis, 2015; Maloney et al., 2009) and children (e.g., Bowers et al., 2005; Suárez-Coalla et al., 2016) and across different languages including Spanish (Álvarez-Cañizo et al., 2018; Suárez-Coalla et al., 2016), Dutch (de Jong & Share, 2007; Martens & de Jong, 2006), French (Bosse et al., 2015; Ginestet et al., 2020), English (Cunningham, 2006; Cunningham et al., 2002; Wang et al., 2011a) and Chinese (Li, Marinus et al., 2020; Li, Li et al., 2020). Findings in these studies consistently prove that a relatively small number of decoding attempts, involving from four to six exposures, is sufficient for native skilled readers to create durable lexical representations, which can indeed remain largely intact over variable periods of time tested, from a short interval (Nation et al., 2007) to some days (Bowey & Muller, 2005; Kwok & Ellis, 2015; Wang et al., 2011), or even a month later (Kwok et al., 2017; Share, 2004; Suárez-Coalla et al., 2016).

The aforementioned findings provide robust evidence regarding the acquisition of reading fluency in the native language. Nonetheless, considering the current globalized community in which proficient reading and written communication skills are also essential in a second language (L2), there is an impelling need to understand the nature of orthographic learning

1
2
3
4 processes in L2. Whereas there are some recent examples of self-teaching research conducted
5
6 in L2 (Chung et al., 2019b; Li, Wang et al., 2021; Schwartz et al., 2014), this literature is still
7
8 very scarce and important questions remain unsolved, particularly regarding the influence of
9
10 orthographic variations across both L1 and L2. The aim of the present study is precisely to
11
12 investigate the impact of distant L1-L2 reading strategies on the process of L2 orthographic
13
14 learning.

15
16
17 In this line, L1-L2 orthographic variation might compromise the efficiency of visual word
18
19 recognition processes in L2 and ultimately, the development of L2 reading fluency. L1 features
20
21 might influence the preferred reading strategy for L2, thus interfering with L2 processing,
22
23 primarily when both orthographic systems rely on different processes. The case of Chinese-
24
25 English, as one of the most distant L1-L2 pairs, is particularly relevant, considering that English
26
27 is the most prevalent L2 spoken worldwide (Eberhard et al., 2021), and Chinese is the language
28
29 spoken by the greatest number of people in the world (Eberhard et al., 2020). The differential
30
31 mechanisms underlying orthographic processing in these two languages have been manifested
32
33 in simulation data of English (Harm & Seidenberg, 2004) and Chinese reading networks (Yang
34
35 et al., 2013; Yang et al., 2006), pointing to more systematic letter-to-sound mappings and finer-
36
37 grained phonological processing in English than in Chinese (see Smith et al., 2021 for a recent
38
39 connectionist triangle model of reading). Chinese reading, in contrast, is primarily characterized
40
41 by a holistic orthographic analysis, with direct mapping between visual form and sound (Hu,
42
43 2013; Perfetti et al., 2012). Accordingly, experimental data have demonstrated that Chinese-
44
45 English bilinguals generally adopt a strategy biased towards L1 when reading English as a L2,
46
47 exhibiting a visual orthographic analysis (Ben-Yehudah et al., 2019; Hamada & Koda, 2008,
48
49 2010a; Pae et al., 2017; Wang & Koda, 2005), that ultimately affects the efficient decoding and
50
51 visual word recognition of L2 English words (Hamada & Koda, 2008). Such a disadvantageous
52
53 position for Chinese-English learners is in fact common to many other bilinguals, considering
54
55 that English exhibits an outlier orthography compared to most languages (Share, 2008). In
56
57 addition to these findings, a good number of studies have reported that the likelihood of L1
58
59 activation (and hence interference) during L2 word processing decreases as a function of L2
60

1
2
3
4 proficiency increase, resulting in less effortful and more L1-independent processing of the L2
5 (e.g., Berghoff et al., 2021; Cuppini et al., 2013; Stein et al., 2009). Therefore, the level of L2
6
7 proficiency may indeed dictate the employment of phonological decoding or more holistic
8
9 identification of L2 written words in distant L1-L2 bilinguals.
10

11
12
13 Despite the fact that the impact of such L1-L2 orthographic variation in L2 visual word
14
15 recognition has been explored together with the modulatory role of L2 proficiency, the putative
16
17 influence of L1 strategies on the functioning of the self-teaching mechanism remains unclear.
18
19 Indeed, the few studies that focused on L2 orthographic learning have yielded rather
20
21 inconsistent findings. Some of them found poor orthographic learning outcomes and suggested
22
23 that orthographic L1-L2 distance may impair orthographic learning (van Daal & Wass, 2017;
24
25 Cheng, 2017; Chung et al., 2019b; Schwartz et al., 2014). These studies inform of weaker
26
27 decoding skills in L2 learners (e.g., Chung et al., 2019b), particularly evident in distant rather
28
29 than close L1-L2 pairs (e.g., Schwartz et al., 2014). Therefore, such underperformance in L2
30
31 learners might indeed be particularly detrimental to self-teaching mechanisms through
32
33 phonological decoding. Moreover, as observed in L2 visual word recognition studies, L2
34
35 proficiency seems to play a crucial role in the influence of L1 orthographic features on L2
36
37 orthographic learning, since this influence was found attenuated at relatively high L2
38
39 proficiency levels (Chung et al., 2019b). In this regard, the proposed *Interactive Transfer*
40
41 *Framework* (Chung et al., 2019a) points to L2 proficiency as one of the factors that determine
42
43 the impact of L1 orthographic features on L2 orthographic acquisition. On the contrary, other
44
45 L2 self-teaching studies have provided results indicative of successful orthographic learning in
46
47 L2 (Li, Wang, et al., 2021, 2022; Webb, 2005, 2007), even in the case of highly
48
49 orthographically distant L1-L2 pairs (e.g., Chinese-English, Li, Wang, et al., 2021). In the study
50
51 of Li, Wang, et al. (2021), post-training accuracy rates obtained in spelling and visual
52
53 recognition tasks revealed a fast L2 orthographic learning in Chinese-English learners, after
54
55 just a few exposures to novel words, hence in line with self-teaching L1 findings. In that study,
56
57 novel L2 words were embedded in texts which, importantly, could lead to the activation of top-
58
59 down mechanisms that might ameliorate the putative impact of L1-L2 orthographic variation.
60

1
2
3
4 Nonetheless, Li, Wang, et al. (2021) reported no significant correlation between the decoding
5 accuracy of pseudowords and orthographic learning outcomes. Although the authors
6 hypothesized that this pattern might be a consequence of their participants utilizing an L1
7 reading approach during L2 orthographic learning, hence relying more on lexical than
8 phonological information, this alternative explanation needs specific verification.
9
10
11
12

13
14
15 Considering this body of evidence, the question regarding the putative influence of L1-L2
16 variation in the process of L2 orthographic acquisition remains unclear. The present study
17 sought to extend this research and further contribute to the understanding of L2 orthographic
18 learning, making use of more informative behavioral indices collected both *during* the training
19 and in the post-training phase. In particular, L2 self-teaching was examined in Chinese-English
20 readers, focusing on the role of L2 proficiency over the course of exposures to novel L2 word-
21 forms, while controlling the access to semantic information. This control would prevent
22 compensatory effects led by the context, allowing us to scrutinize the modulation of purely
23 orthographic processes over repeated exposure to novel word-forms.
24
25
26
27
28
29
30
31
32

33
34 Therefore, during the *training phase*, Chinese-English adult participants with various
35 levels of L2 English proficiency were presented with short and long novel and high- and low-
36 frequency words throughout ten exposures; modulations in the length effect (i.e., the difference
37 between long and short items, e.g., Weekes, 1997), the lexicality effect (i.e., the difference
38 between very familiar and novel words, e.g., Forster & Chambers, 1973) and the frequency
39 effect (i.e., the difference between high- and low-frequency words, e.g., Monsell et al., 1989)
40 were evaluated across repetitions, both in terms of reading latencies and accuracy rates, as a
41 function of L2 proficiency. Reductions of these effects through exposures with novel and low-
42 frequency words have been proven as robust indices of orthographic representation and lexical
43 route usage in L1 learning studies. Thus, in contrast to a highly familiar word, when novel
44 words (or very unfamiliar words) are encountered, their reading can only be achieved via sub-
45 lexical processing by applying serial grapheme-to-phoneme decoding, with reading latencies
46 increasing with the number of letters (Coltheart et al., 2001; Pritchard et al., 2018; see also Fu
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 et al., 2020 for a recent review). As novel words become familiar through repeated exposures,
5
6 the effects of length (Maloney et al., 2009) and lexicality (Bermúdez-Margaretto et al., 2022)
7
8 become reduced or even absent, indicating the formation of new representations that allow a
9
10 direct, parallel visual recognition (see also Álvarez-Cañizo et al., 2018; Kwok & Ellis, 2014;
11
12 2015; Kwok et al., 2017; Maloney et al., 2009; Suárez-Coalla & Cuetos, 2016). As similarly
13
14 occurs for lexicality, the frequency effect tends to be stronger for long than for short items and
15
16 it is also found reduced with repeated exposures (e.g., Kwok et al., 2017), a result considered
17
18 to reflect the strengthening of existing representations through repetition. Therefore, a detailed
19
20 analysis of the modulation of these variables through an exposure-by-exposure scrutinization
21
22 of the L2 self-teaching mechanism and, importantly, in the absence of semantic confounding
23
24 factors, was expected to provide specific information regarding the putative cost of L1-L2
25
26 variation for the acquisition of new L2 orthographic representations.
27
28

29
30 The *post-training phase* included a word recognition task of previously trained words as a
31
32 classical assessment of orthographic learning. Additionally, we decided to employ a more
33
34 specific manipulation in this phase that could help to reveal the transfer of L1 strategies in the
35
36 process of L2 orthographic learning. In particular, an interference paradigm was used during a
37
38 lexical decision task. Participants were presented with non-trained pseudowords
39
40 orthographically related to those stimuli previously trained (differing in one letter), together
41
42 with unrelated pseudowords and filler familiar words. This paradigm has been proven to inform
43
44 about the lexical integration of trained stimuli in previous L1 learning studies (Bermúdez-
45
46 Margareto et al., 2022; Bowers et al., 2005; Leminen et al., 2016; Merckx et al., 2011), as a
47
48 function of the interference caused by newly established representations in the categorization
49
50 of orthographically related pseudowords as non-lexical items. Importantly, this paradigm was
51
52 also expected to provide valuable information regarding the putative L1 transfer during L2
53
54 visual word recognition processes.
55
56

57
58 Regarding our hypotheses, since previous works have reported successful L2 self-teaching
59
60 even in distant Chinese-English orthographies (e.g., Li, Wang et al., 2021), one might expect

1
2
3
4 no detrimental impact of L1-L2 orthographic variation over the course of training exposures. If
5 that is the case, the efficiency in the use of the self-teaching mechanism should be observed in
6 the progressive reduction of length, frequency and lexicality effects during the training, both in
7 terms of accuracy and speed at which reading transitions from a sub-lexical, length-sensitive
8 processing to a lexical processing as a result of storing and accessing orthographic
9 representations. Therefore, a successful phonological decoding through self-teaching would
10 help Chinese bilinguals to store new L2 orthographic representations for novel words across
11 exposures. That would also be reflected in post-training outcomes, with these participants
12 showing recognition of previously trained stimuli significantly above chance, as well as a
13 significant interference during the categorization (as non-real words) of pseudowords
14 orthographically related to those presented in training. Notably, the efficient categorization of
15 these neighbor pseudowords as non-lexical stimuli (“no” responses) is only possible using a
16 phonological, L2 decoding strategy which in turn would increase categorization latencies.
17 Therefore, it was expected that these participants would show an interference effect in terms of
18 latency. Hence, the specific orthographic neighborhood manipulation carried out in this task
19 was expected to inform about the engagement of Chinese-English bilinguals in L2 phonological
20 decoding despite their holistic-based L1 background, thus indicating that L1-L2 variations do
21 not hinder L2 orthographic learning.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 Alternatively, we hypothesized that L1-L2 variation would negatively impact L2
42 orthographic learning. If that is the case, participants’ L2 proficiency level was expected to
43 modulate the cost of such L1-L2 variation, since this variable has been shown as the main
44 modulator of L1 transfer in previous L2 word processing studies. Thus, participants with lower
45 proficiency levels would engage in a more effortful phonological decoding than their high-
46 proficient peers, hence showing a diminished capability to develop new L2 orthographic
47 representations through self-teaching. That would be observable in poor or non-significant
48 reduction of length, frequency and lexicality effects along the training, showing a flat-trend in
49 this modulation in comparison to a sharper learning trajectory expected by high L2 proficiency
50 participants. Post-training learning outcomes might also reflect differences as a function of L2
51
52
53
54
55
56
57
58
59
60

1
2
3
4 proficiency, with low proficient bilinguals showing a reduced capability to correctly recognize
5
6 previously trained stimuli, given their lack (or weakness) representation in the lexicon in
7
8 comparison to those participants with higher proficiency. Similarly, participants with lower
9
10 proficiency might show poor or no interference effect in the lexical decision task, especially in
11
12 terms of latency, during the categorization of orthographically related pseudowords in
13
14 comparison to their high-proficient peers. Critically, the engagement of these participants in the
15
16 L1-based reading strategy would lead to a superficial, holistic analysis during the categorization
17
18 of neighbor pseudowords, instead of following a letter-by-letter, phonological decoding
19
20 approach. That would result in the commission of more errors (“yes” responses for these non-
21
22 lexical stimuli), hence causing a higher interference in accuracy rather than in latency.
23
24

25 **Method**

26 27 28 *Participants*

29
30
31 Sixty-one Chinese students from the Dalian University of Technology were recruited for
32
33 the study. However, the participation of thirteen of them was discarded due to poor scores in
34
35 preliminary tests (see details of the selection procedure in supplementary materials A in
36
37 <https://osf.io/exnd5/>). The final sample comprised forty-eight native speakers of Chinese with
38
39 normal or corrected-to-normal vision. None of them was reported as an early bilingual of any
40
41 other language or with the history of reading or language disabilities. They received a monetary
42
43 compensation (8 euros) for their participation in the study. Ethical approval was obtained from
44
45 the Ethics Committee of Dalian University of Technology (Registration number: 2021-101)
46
47 and University of La Laguna (Registration number: CEIBA2021-3104).
48
49

50
51 Participants recruited for the study had English as their second language. All of them were
52
53 college-level learners who had attended weekly English courses and reached at least a low-
54
55 advanced level of English proficiency as a prerequisite for academic studies. Thus, some
56
57 achieved the College English Test Band 4 (an elementary standardized test of English
58
59 proficiency) or scored at least 80 or 6.0 in TOEFL or IELTS (Tests of English as a Foreign
60

1
2
3
4 Language), respectively, while some others either passed the Test for English Majors-Band 8
5 (the highest-level standardized test of English proficiency) or scored above 105 or 7.5 in
6 TOEFL or IELTS, respectively. Nonetheless, in order to better determine their level of L2
7 proficiency, a battery of subjective and objective preliminary tests was carried out, including
8 the Language History Questionnaire 3 (LHQ 3, Li et al., 2020), DIALANG diagnostic system
9 (Zhang & Thompson, 2004), word translation task, word and pseudoword reading task (based
10 on Snowling et al., 1986; Wang et al., 2019; Yang et al., 2016), phonological awareness task
11 (CTOPP; Wagner et al., 1999), digit span task (Harrington & Sawyer, 1992) and word/nonword
12 span task (Weekes, 2018). See Figure 1 for a summary of the assessment results, showing main
13 differences in proficiency-related variables across groups. Importantly, L2 proficiency level
14 was treated in the present study as a continuous variable in order to avoid issues derived from
15 dichotomizing this variable (see Maxwell & Delaney, 1993; DeCoster et al., 2011). Nonetheless,
16 with the purpose to explore the effect of L2 proficiency levels in subsequent steps of our
17 analyses, participants were equally divided into two groups according to scores obtained in
18 these tests: a high proficient English group (hereinafter HPG, including 24 participants) and a
19 low proficient English group (hereinafter LPG, including 24 participants). Notably, in the
20 present study, high L2 proficiency was defined by more extensive immersion-based exposure
21 to L2 (as indexed by LHQ 3) and larger vocabulary sizes (as indexed by DIALANG and word
22 translation task). See supplementary materials A in <https://osf.io/exnd5/> for details regarding
23 the recruitment, implementation of tests and scores obtained.
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

45 --- Insert Figure 1 about here ---
46
47

48 ***Power Considerations*** 49

50
51
52
53
54
55
56
57
58
59
60
Statistic power estimation was conducted by using the *simulate* function from the *lmer4* package (Bates et al., 2015) in R 3.6.3 version (R Core Team, 2018). A total of 1000 new datasets with n participants, based on reaction time data obtained in Kwok et al. (2017, English data on Day 1), were simulated to determine the sample size of the present study. Trials in this dataset were randomly labeled as missing trials and excluded from the data set. Then, we carried

1
2
3
4 out 1000 simulations by including the same model structure (fixed effects for Exposures,
5 Length and Stimuli Type, random intercepts for participants and items, and by-participant
6 random slopes for Exposures and Length) used in Kwok, et al. (2017). The percentage of
7 models in which significant main effects and interactions were detected (for which $p < .05$) was
8 considered as the estimation for statistical power. For $n = 22$ simulated participants, the
9 simulation including 1000 iterations returned a statistic power of 1 for main effects, two-way
10 interactions, and three-way (Exposures x Length x Stimuli Type) interaction (i.e., in all 1000
11 simulation runs, the model detected significant main effects and interactions). Note that in
12 Kwok, et al. (2017), power estimates were obtained based on the fitted model with the fixed
13 term Exposures as an integral factor. In a more traditional way, when treating Exposures as an
14 ordinal factor (with each block of exposure being a certain level), the power estimates are
15 reduced to 0.857 for the detection of the Exposures x Length x Stimuli Type three-way
16 interaction. We opted for power estimates above 0.8 since the model was estimated on data
17 after excluding outliers, and the estimated parameter should settle on the true expected value.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

33 *Materials*

34
35
36 All materials used in the study can be accessed in Supplementary materials B
37 (<https://osf.io/exnd5/>). For the training session, the same stimuli set used in Kwok et al. (2017)
38 were presented in a reading aloud task. In particular, 72 experimental stimuli, divided into 24
39 high-frequency words, 24 low-frequency words and 24 pseudowords (to function as novel
40 words) were presented. Each set contained 12 four-letters monosyllabic items and 12 seven-
41 letters bi-syllabic items. Frequency measures were taken from SUBTLEX-UK, expressed as
42 log frequency Zipf values (van Heuven et al., 2014). Pseudowords, generated by the WordGen
43 programme (Duyck et al., 2004), were orthographically and phonologically legal letter strings
44 which had one obvious, canonical pronunciation according to the standard grapheme-to-
45 phoneme correspondence rules of English. Stimuli were matched across conditions on initial
46 letters, initial phonemes, mean word frequency (for high- and low-frequency words only) and
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 the mean log bigram frequency (see Table 1). None of stimuli started with voiceless fricative
5 phonemes (“f”, “s”, “sh”, or “th”) in order to optimize voice key activation.
6
7

8
9 --- *Insert Table 1 about here* ---
10

11
12 Post-training session consisted of two tasks, namely lexical decision and word recognition
13 tasks. For the lexical decision task, a new set of 72 pseudowords was created by substituting
14 one phoneme from the 72 stimuli (high-, low-frequency words and novel words) presented in
15 the training session (in the first, middle or final position of each source word, e.g., dempton -
16 bempton). These orthographically related pseudowords were matched to the source items in
17 letter length, syllable length and mean log bigram frequency. Furthermore, another set of 72
18 unrelated pseudowords and 144 filler words, both sets non-previously presented during the
19 training, were generated by WordGen (Duyck et al., 2004). Unrelated pseudowords and filler
20 words were matched to related pseudowords in letter and syllable length, initial letter and mean
21 log bigram frequency (see Table 1 for details). Therefore, a total of 288 trials were included in
22 this task (72 related pseudowords and 72 new unrelated pseudowords, both to be responded
23 with “NO” during the lexical decision task as non-lexical items, and 144 filler words, to be
24 responded with “YES”, as lexical items).
25
26
27
28
29
30
31
32
33
34
35
36
37
38

39 For the word recognition task, the same set of 72 stimuli as presented in the training session
40 was used (namely, high-, low-frequency words and novel words). In addition, twice number of
41 fillers (non-previously trained stimuli) were selected for each experimental condition (i.e., 48
42 high-frequency words, 48 low-frequency words and 48 pseudowords) and included as “NO”
43 responses (i.e., non-previously presented during the training). Therefore, a total of 216 trials
44 were presented in the recognition task. All stimuli used in the study can be accessed in
45 supplementary materials B (<https://osf.io/exnd5/>).
46
47
48
49
50
51
52
53

54 ***Procedure***

55
56
57 Stimuli were displayed in black, lower-case letters (18-point Times New Roman font) over
58 a white background at the center of a computer screen using E-Prime software (Version 2.0,
59
60

1
2
3
4 Psychology Software Tools, Inc., Pittsburgh, PA). In the training session, stimuli were
5
6 presented across ten consecutive blocks of exposures in a random order within each block.
7
8 Participants were instructed to read each item aloud as quickly and accurately as possible. Each
9
10 trial consisted of a fixation cross displayed at the center of the screen for 1000ms, followed by
11
12 a stimulus (word or pseudoword) for 2000 ms; a blank screen was then presented for 1000ms.
13
14 A voice key incorporated into a serial response box (SR Box, Psychology Software) was used
15
16 to record participants' utterances along the reading aloud task. Both accuracy and latency
17
18 indices were recorded.

19
20
21 Following the training session, participants underwent the post-training session (lexical
22
23 decision and recognition tasks) after a 10-minute break. The session began with a lexical
24
25 decision task in which participants were instructed to decide as quickly and as accurately as
26
27 possible whether each stimulus was an English word (by pressing "J" with their right hand) or
28
29 not (by pressing "F" with their left hand). Each trial consisted of a fixation cross displayed
30
31 pseudo-randomly between 500-600ms (jitter of 100 ms with 50 ms step), followed by a blank
32
33 screen presented for 250-300ms, and subsequently by the target stimulus, which remained on
34
35 the screen until participants' response. After the lexical decision task, participants completed a
36
37 2-min arithmetic distractor task in order to minimize contamination from the previous task. In
38
39 this task, participants were required to count back from the number 269 in step of 3. Following
40
41 the distractor task, participants underwent the word recognition task. Each trial consisted of a
42
43 fixation cross displayed pseudo-randomly between 500-600ms (jitter of 100ms with 50ms step),
44
45 followed by a blank screen for 250-300ms, and subsequently by the target stimulus, which
46
47 remained on the screen until participants responded. They were instructed to decide whether
48
49 each stimulus had been presented in the initial training session (by pressing "J" with their right
50
51 hand) or not (by pressing "F" with their left hand) and to respond as quickly and accurately as
52
53 possible. Both accuracy and latency indices were recorded in the two post-training tasks.
54
55
56
57
58
59
60

Data trimming and statistical modeling

Data from five participants were eliminated due to failures to activate the voice key, leaving 43 participants (HPG: 22, LPG: 21) for further data analysis. In the training session, means of correct responses obtained for each group were 97.75% (HPG) and 96.83% (LPG), resulting at ceiling in each stimulus type across all blocks of trials (> 93%). Thus, analyses of training data were focused on naming latencies. Naming errors and no responses (6.3%), as well as data points that deviated a range of ± 2.5 standardized residual errors (model criticism, see Baayen et al., 2008) were excluded from the analysis (2.6%, $n = 753$). Afterwards, the models were refitted on the truncated dataset, including a total of 28195 data points.

Naming latencies obtained in the training session were analyzed by generalized additive mixed-effects modeling (GAMMs), using the *mgcv* package (Wood et al., 2016) for model fitting and *itsadug* package for visualization (van Rij et al., 2017) in R environment. GAMM represents an extension to the linear modeling context that enables the application of curvilinear relationship (i.e., smooths) to predictors. In this sense, predictors can be categorized into parametric terms (equivalent to fixed and random effects) and smooth terms. Random smooths are conceptually comparable to random slopes and intercepts in the mixed-effect framework. Thus, the GAMMs approach allowed us to fit curves constrained to learning trajectory shapes as a function of parametric effects, which are essential in avoiding anti-conservative models. In the present study, we followed Pedersen et al. (2019) and used hierarchical GAMMs to understand how different smooth relationships vary between conditions. In this way, differences in the fitted curve for each term (i.e., considering one term for each level of the factor) can be formally tested for statistical significance by its comparison against the total smooth function. Thus, we conducted a maximal modeling structure including parametric effects for different types of stimuli (i.e., high-frequency, low-frequency and novel words with either long or short length) to account for possible differences in the intercept, as well as random intercepts for Participants and Items. Thin plate regression splines were applied (1) as a reference smooth to Exposure (10 blocks of exposure), (2) as a difference smooth to Exposure

1
2
3
4 conditioned on types of stimuli across groups, and (3) as a random smooth factor (the non-
5 linear equivalent to a combination between random intercepts and random slopes) across
6 Exposure by Participants. For L2 proficiency levels, DIALANG self-assessment scores were
7 chosen as an index of participants' proficiency, given the strong significant correlation between
8 these scores and the translation task scores ($\rho = 0.96$, HDI [0.95, 0.97]) and considering this
9 tool as a more ecological measurement which also assesses participants' conversational skills.
10 The interaction between Proficiency and Exposure was included by means of a tensor product
11 smooth. Further, the interaction between Proficiency and Exposure conditioned on different
12 types of stimuli was specified using the partial tensor product smooth. Thus, the model included
13 the shapes of learning curves and various L2 proficiency levels (global curves) plus word-type-
14 specific curves (word length and stimuli types) that were penalized for being close to the mean
15 function. We used Package *itsadug* (van Rij, Wieling, Baayen, & van Rijn, 2017) and *tidymv*
16 (Coretta, 2019) for visualization of the estimated differences between 1) long and short length
17 in terms of different types of stimuli (i.e., Length effect across Stimuli Type condition), 2) high-
18 and low-frequency words with long and short length (i.e., Frequency effect across Length
19 condition), and 3) high-frequency and novel words with long and short length (i.e., Lexicality
20 effect across Length condition).
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38

39 For post-training data, we fitted Bayesian linear mixed-effect models to response latencies
40 (inverse Gaussian distribution, i.e., $-1000/RT$) and Bayesian generalized mixed-effect models
41 (Bernoulli distributed with a logit link function) to accuracy data, with maximal random effects
42 structure using the package *brm* (Bürkner, 2018). See also van Viersen et al. (2022) and Casillas
43 (2020) for the application of the same combination of methodology for training and post-
44 training data in the context of orthographic learning via self-teaching. For the lexical decision
45 task, the three subsets of pseudowords related to previously trained stimuli (namely, those
46 related to high-frequency words, to low-frequency words and to novel words) were analyzed
47 separately; for each subset, the models considered the effects of Related Orthography (related
48 vs. unrelated pseudowords), L2 proficiency and their interaction. For the recognition task, the
49 models considered Stimuli Type (high-frequency, low-frequency, and novel words), L2
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 proficiency and their interaction. In both tasks, the continuous variable L2 proficiency was
5
6 standardized whereas categorical variables were encoded with sequential difference contrasts
7
8 (for 2-level predictor Related Orthography: (1/2, -1/2); for 3-level predictor Stimuli Type: (-2/3,
9
10 1/3, 1/3) and (-1/3, -1/3, 2/3), i.e., high- vs. low-frequency ~ frequency effect, and high-
11
12 frequency vs. novel ~ lexicality effect), derived using the *hypr* package 0.1.7 (Rabe et al., 2020;
13
14 Schad et al., 2020). Regularizing, weakly informative priors with a Cauchy distribution
15
16 (Gelman & Carpenter, 2020; Gelman et al., 2017) were applied to estimate predictive parameter
17
18 values. Hamiltonian Monte- Carlo sampling (Hoffman & Gelman, 2011) was carried out using
19
20 four chains to draw samples from the posterior predictive distribution. The potential scale
21
22 reduction factor (\hat{R}), which was reported for all of the models to be below 1.01, was used to
23
24 determine if models had effectively converged (Brooks & Gelman, 1998). We defined a region
25
26 of practical equivalence (ROPE) around a point null value of 0, as well as the maximum
27
28 probability of effect (MPE), to assess our hypotheses (see Kruschke & Liddell, 2018). We
29
30 applied four statistics to describe the posterior distribution: (1) the posterior mean, (2) the
31
32 highest density credible interval (HDI), (3) the ROPE, and (4) the MPE. In general, a posterior
33
34 distribution for a parameter with 95% of the HDI falling outside the ROPE and with a high
35
36 value of MPE (i.e., values close to 1) is considered compelling evidence for a given effect.
37
38 Furthermore, the relationship between individual phonological decoding efficiency (measured
39
40 as the reduction of length and lexicality effects across blocks in the training session) and the
41
42 outcomes of orthographic learning (measured as correctly recognized novel words in the word
43
44 recognition task) was investigated by means of Pearson correlation coefficient. All R code and
45
46 data obtained in the study can be freely accessed at <https://osf.io/exnd5/>.

47 48 49 **Results**

50 51 52 *Training session*

53
54
55 Table 2 shows statistical results found for data obtained in the training session. Average
56
57 naming latencies across conditions and proficiency groups can be found in the supplementary
58
59 materials B (Table SB1, <https://osf.io/exnd5/>).

1
2
3
4 --- Insert Table 2 about here ---
5
6

7 The adjust R-squared in the model resulted in 0.48, indicating it explains the 48.88% the
8 deviance. The nested model comparisons indicated that the full model with the parametric and
9 smooth terms on modality provided a significantly better fit to the data than the null model (DF
10 = 29, Difference = 5323.07, EDF =35, $p < 0.001$). The reference smooth of Exposure and
11 random factor smooth of Exposure revealed the significant effect of learning on different types
12 of stimuli and individual participants, with reduced naming latencies over the course of
13 exposures. However, the learning trajectory of words (both high- and low-frequency words)
14 differed from that of novel words, evident in the absence of significant smooth for Exposure to
15 novel words. There was also a significant interaction between Exposure and L2 proficiency,
16 indicating that learning trajectories varied as a function of L2 proficiency levels. Crucially,
17 tensor product smooth of the interaction differed significantly among types of stimuli,
18 indicating the influence of L2 proficiency was particularly higher for the learning of long novel
19 and long low-frequency words over the course of exposures.
20
21
22
23
24
25
26
27
28
29
30
31
32
33

34 In a second step, GAMM were re-computed by treating L2 Proficiency as a discrete
35 variable and then defining two groups of participants with higher (i.e., HPG) and lower (i.e.,
36 LPG) proficiency levels to further explore the effect of L2 proficiency levels (see table SB2 in
37 the supplementary materials B in <https://osf.io/exnd5/> for a summary of model fit for HPG and
38 LPG). Figure 2 plots the learning trajectories for different types of stimuli over the course of
39 exposures and the estimated differences between conditions using the *plot_diff()* function. The
40 plots corroborated the results derived from the GAMM obtained in the first and second step.
41 Specifically, a reduced impact of length effect on all types of stimuli was shown across repeated
42 exposures in the HPG. In contrast, the LPG exhibited a significant smooth for Exposure to short
43 novel words and short low-frequency words, that was not obtained for long novel words and
44 long low-frequency words. Thus, participants in the LPG benefitted from repeated exposure to
45 short novel words and short low-frequency words, whereas no trace of learning effect was
46 shown for long items, either novel or low-frequency. As plotted in Figure 2, that led to increased
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 effects of length (for novel words and low-frequency words), frequency and lexicality (for long
4 items) over the course of exposures.
5
6
7

8
9 --- Insert Figure 2 about here ---
10

11
12
13
14
15 *Post-training session*
16

17
18 --- Insert Table 3 about here ---
19

20
21 Table 3 summarizes the means of posterior probability distribution and the 95% highest
22 density interval (HDI) reported for each set of related pseudowords presented in the lexical
23 decision task. Analysis run for *pseudowords related to previously trained high-frequency words*
24 revealed a significant L2 Proficiency effect in accuracy data ($\beta = 0.47$, 95% HDI = [0.20, 0.74])
25 indicating that the higher the L2 proficiency level the higher the probability to correctly
26 categorize pseudowords as non-lexical items, independently of their orthography (see Figure 3,
27 left upper panel). Nonetheless, analysis on latency data revealed a significant interaction effect
28 between Orthographic Relation and L2 Proficiency ($\beta = 0.06$, 95% HDI = [0.02, 0.11]),
29 indicating that, although all participants were interfered, the higher the L2 proficiency level the
30 higher the interference effect showed in these participants, with longer responses during the
31 rejection of related than unrelated pseudowords (see Figure 3, right upper panel).
32
33
34
35
36
37
38
39
40
41
42
43

44 A different pattern was found in the analysis of accuracy for *pseudowords related to*
45 *previously trained low-frequency words*, showing a significant interaction between
46 Orthographic Relation and L2 Proficiency ($\beta = 0.66$, 95% HDI = [0.44, 0.86]), with higher
47 interference caused by the neighborhood (and hence more errors committed in related than
48 unrelated pseudowords) in participants with lower L2 proficiency levels; importantly,
49 participants with higher proficiency L2 levels were similarly accurate rejecting pseudowords
50 related and unrelated (see Figure 3, left middle panel). Analysis on latency data for this set
51 showed a similar interaction between Orthographic Relation and L2 Proficiency as in the
52
53
54
55
56
57
58
59
60

1
2
3
4 previous set ($\beta = 0.07$, 95% HDI = [0.03, 0.11]), with all participants being affected by the
5
6 interference but with higher interference observed in those participants with higher L2
7
8 proficiency levels (see Figure 3, right middle panel).
9

10
11 Finally, for the set of *pseudowords related to trained novel words*, the analysis conducted
12
13 on accuracy data indicated a significant interaction between Orthographic Relation and L2
14
15 Proficiency ($\beta = 0.44$, 95% HDI = [0.10, 0.96]); as in the previous set, more errors were
16
17 committed in the categorization of related than unrelated pseudowords in participants with
18
19 lower L2 proficiency levels, since those with higher proficiency showed no interference in their
20
21 accuracy (see Figure 3, left lower panel). A significant interaction between Orthographic
22
23 Relation and L2 Proficiency was also found in latency data ($\beta = -0.09$, 95% HDI = [-0.15, -
24
25 0.04]); participants with higher L2 proficiency levels tended to exhibit higher response latencies
26
27 for related than unrelated pseudowords, whereas those with lower proficiency showed no
28
29 interference in their latencies (see Figure 3, right lower panel).
30
31

32
33 --- *Insert Figure 3 about here* ---
34
35

36
37 Thus, across all sets of pseudowords, those participants with higher L2 proficiency levels
38
39 (right end of the graphs) showed no interference in their accuracy to correctly categorize
40
41 pseudowords (either related or unrelated) as non-lexical items, but a clear cost in their speed to
42
43 respond to those pseudowords orthographically related to known stimuli (either more or less
44
45 familiar). Conversely, participants with lower L2 proficiency levels (left end of the graphs)
46
47 exhibited a clear interference in their accuracy to correctly reject pseudowords, particularly for
48
49 those related to less familiar and newly trained words, whereas no cost was observed in their
50
51 response latencies to pseudowords related with the novel words previously trained.
52

53
54 --- *Insert Table 4 about here* ---
55

56
57 Table 4 summarizes the means of posterior probability distribution and the 95% highest
58
59 density interval (HDI) reported for the recognition task. Analysis on accuracy data revealed
60

1
2
3
4 interactions between Frequency (i.e., high-frequency vs. low frequency words) and L2
5 proficiency ($\beta = -0.45$, 95% HDI = $[-0.73, -0.18]$) and between Lexicality (i.e., high-frequency
6 vs. novel words) and L2 Proficiency ($\beta = -0.47$, 95% HDI = $[-0.73, -0.20]$), indicating that both
7 variables were modulated by the L2 proficiency level (see Figure 4, left panel). Thus, whereas
8 participants with higher proficiency levels exhibited a significantly high recognition of all
9 stimuli previously trained, the recognition performance observed in lower proficient bilinguals
10 was poorer for low-frequency words and around chance level for novel words (see Figure 4,
11 left panel). Analysis on latency data indicated no modulation of L2 proficiency, with response
12 times modulated by Frequency (i.e., longer reaction times for low- than for high-frequency
13 words) and by Lexicality (i.e., longer reaction times for novel than for low-frequency words),
14 independently on L2 proficiency level (see Figure 4, right panel).
15
16
17
18
19
20
21
22
23
24
25
26

27 --- Insert Figure 4 about here ---
28
29

30 Regarding the relationship between individual phonological decoding abilities and
31 orthographic learning outcomes obtained in the recognition task, medium and positive
32 correlations between the reduction of the effect of length ($r = 0.45$, $p = .003$) and lexicality (r
33 $= 0.68$, $p < .001$) and the recognition accuracy were obtained, indicating that better recognition
34 performance on novel words was associated with sharper decline in the effects of length and
35 lexicality across exposures.
36
37
38
39
40
41
42

43 Discussion

44
45
46 The aim of the present study was to investigate the impact of distant L1-L2 reading
47 strategies in the process of L2 orthographic learning. While the evidence for the cost of a distant
48 L1 orthography in L2 word processing is relatively robust, its influence on L2 orthographic
49 learning through a self-teaching mechanism remains unclear, with studies showing inconsistent
50 results that lead to incomplete conclusions. We tested L2 self-teaching in Chinese-English
51 bilinguals, one of the most orthographically distant L1-L2 pairs, and made use of various fine-
52 grain manipulations both during and after the training with novel words. Our results confirm
53
54
55
56
57
58
59
60

1
2
3
4 the hypothesis that L1-L2 variation negatively impacts the process of L2 orthographic learning,
5
6 with L2 proficiency modulating the detrimental engagement in L1 holistic processing that is
7
8 likely at the base of such disadvantage. In what follows, we discuss the differential pattern
9
10 exhibited by Chinese-English bilinguals as a function of L2 proficiency, both during training
11
12 and post-training sessions.

13
14
15 Findings obtained during the training phase indicated that successful phonological
16
17 decoding of novel L2 words allowed higher proficient Chinese-English bilinguals to form novel
18
19 orthographic representations. That was reflected in the progressive reduction of length,
20
21 frequency and lexicality effects throughout the course of training (see Figure 2, upper panels),
22
23 as well as significantly high recognition rates of trained novel words (see Figure 4, left panel).
24
25 These results are consistent with those found in native English speakers, showing the rapid
26
27 formation and strengthening of orthographic representations that enable the transition from sub-
28
29 lexical, serial decoding to direct recognition of novel and low-frequency L2 word-forms
30
31 (Maloney et al., 2009; Kwok et al., 2017; Kwok & Ellis, 2014; 2015; Weekes, 1997). Moreover,
32
33 these participants exhibited a significant interference effect in the lexical decision task during
34
35 the categorization of neighbor pseudowords related to recently trained novel words, indicating
36
37 the integration of these novel words into the mental lexicon and their competing activation
38
39 during the categorization of related pseudowords (Bermúdez-Margaretto et al., 2022; Bowers
40
41 et al., 2005; Leminen et al., 2016; Merks et al., 2011). Importantly, such interference was
42
43 observed at latency rather than at accuracy level, indicating the precise but time-consuming
44
45 categorization of pseudowords orthographically related to previously trained stimuli, likely
46
47 based on letter-by-letter decoding. Therefore, bilinguals with well-developed L2 orthographic
48
49 systems exhibit efficient phonological decoding skills that enable them to engage in effective
50
51 L2 processing mechanisms that, in turn, lead to the establishment of new orthographic
52
53 representations. That was further confirmed by the positive correlation between the reduction
54
55 of length and lexicality effects and the recognition rates for trained novel words. Notably, the
56
57 interference exhibited by these participants in their response latencies was observed across all
58
59 sets of pseudowords, regardless of their similarity to either those of high- and low-frequency or
60

1
2
3
4 newly learnt words; this pattern demonstrates the capability of these participants to activate
5
6 lexical knowledge, either very familiar representations or those recently learnt during a very
7
8 short exposure (i.e., ten repetitions).
9

10
11 However, results obtained in individuals with lower L2 proficiency levels were in sharp
12
13 contrast with the pattern observed in their higher proficiency peers, showing no reduction but
14
15 increased length effect in novel and low-frequency words and increased frequency and
16
17 lexicality effects in long items over the course of the training (see Figure 2, lower panels), as
18
19 well as by-chance recognition of trained novel words (see Figure 4, left panel). Moreover, in
20
21 the lexical decision task, lower L2 proficiency participants showed no interference effect on
22
23 response latencies but on accuracy during the categorization of neighbor pseudowords related
24
25 to previously trained novel words. Thus, these stimuli were accepted as L2 words at a
26
27 significantly higher rate than unrelated pseudowords, most likely due to a coarse analysis of the
28
29 visual word-shape and weaker sensitivity to intra-word structures. Interestingly, the
30
31 interference pattern caused by orthographic neighborhood varied gradually in these participants
32
33 depending on the orthographic relation (namely, whether pseudowords were related to high-,
34
35 low-frequency or novel words previously trained). Thus, when presented with pseudowords
36
37 related to high-frequency words, these participants showed interference in their response
38
39 latencies with no cost in their accuracy levels; this result indicate their capability to activate
40
41 representations of highly frequent, well-known stimuli, as well as to accurately reject these
42
43 pseudowords, likely following a visual processing strategy facilitated by the frequency and
44
45 strong representation of neighbor stimuli. Nonetheless, although these participants were
46
47 capable to activate lexical representations of low-frequency words, as reflected in their
48
49 interference in latencies during the categorization of pseudowords related to these stimuli, they
50
51 failed to correctly reject them, since a decoding strategy is needed to successfully categorize
52
53 these stimuli due to their low-frequency –rather than the holistic reading strategy that they were
54
55 most likely using. Furthermore, these participants showed no sign of interference in their
56
57 response times when presented with pseudowords related to recently learnt novel words, but
58
59 only in their accuracy to correctly categorize them; this indicates the representation of these
60

1
2
3
4 words was too weak to be activated and to cause interference in response times, and hence that
5
6 the experience with these stimuli was too short to allow their efficient storage in the lexicon.
7
8 Therefore, this gradual pattern of interference exhibited in these participants (that decreases in
9
10 latency but increases in accuracy depending on the frequency of the neighbor stimuli exerting
11
12 the interference) reflects both their unstable capability to learn and activate lexical knowledge
13
14 as well as their poor ability to efficiently use an L2 decoding-based reading strategy.
15
16

17
18 Considering together the results obtained for low proficient participants, our findings
19
20 indicate the cost of L1-L2 variation to efficiently store new L2 orthographic representations,
21
22 led by a L1-based visual analysis of the novel word-form structure. Such cost was particularly
23
24 evident during the training of long novel and long low-frequency words; contrary to short novel
25
26 and short low-frequency words, these long items exhibited no reduction of reading latencies
27
28 across the training, causing increased effects of lexicality and frequency for long items (as well
29
30 as increased length effect in novel and low-frequency words) in lower proficient bilinguals.
31
32 These effects suggest that the analysis of the physical structure of words to map orthographic
33
34 information into phonology is limited by word length (Ehri, 2014; Rosenthal & Ehri, 2008),
35
36 with L2 proficiency level at the base of such constraint. Notably, L2 proficiency is characterized
37
38 here by the level of exposure to English reading and vocabulary, suggesting an experience-
39
40 based source of activation. In individuals with higher proficiency, lower-level perceptual
41
42 processing of letters was likely accessible and optimized over exposures, resulting in the
43
44 parallel co-activation of phonological and orthographic representations of each novel (and low-
45
46 frequency) word after the training. Conversely, the expertise with letter combinations appeared
47
48 to be less developed in bilinguals with relatively poor L2 perceptual experience, who indeed
49
50 performed in a manner that was consistently sensitive to length structure across repetitions.
51
52

53
54 The pattern of results found in the present study is in agreement with previous research
55
56 focused on cross-linguistic mechanisms underlying L2 word recognition (e.g., Akamatsu, 2002;
57
58 Wang et al., 2003; Wang & Koda, 2005; Wong et al., 2011), which show that English bilinguals
59
60 with logographic L1 background exhibited a higher sensitivity to whole-form, visual encoding

1
2
3
4 of words and less reliance on phonological information in comparison to native English
5
6 speakers or bilinguals with alphabetic backgrounds. Importantly, our findings reveal that the
7
8 transfer of an L1 logographic reading strategy affects not only how bilinguals process L2
9
10 alphabetic input, but also how they acquire new L2 orthographic information. Therefore, the
11
12 current study extends these findings to L2 self-teaching mechanisms and highlights the critical
13
14 role of language proficiency in modulating the impact of cross-linguistic variation in L2
15
16 orthographic learning. Future investigations should further confirm these findings by including
17
18 various bilingual groups with distinct L1 backgrounds. Moreover, these studies might also
19
20 consider using both reaction times (RTs) and articulation times (TAs), as is traditional in
21
22 developmental word learning studies (e.g., Suárez-Coalla et al., 2016). Indeed, TAs are
23
24 particularly suitable for testing reading performance in transparent L2 orthographies, since
25
26 these scores are more susceptible to serial decoding processes that might last after response
27
28 word onset (Davies et al., 2013). In this sense, the use of TAs might have been particularly
29
30 sensitive to evaluate whether distant L1-L2 learners with lower L2 proficiency levels in an
31
32 alphabetic L2 language engage in phonological decoding processes during the training or, on
33
34 the contrary, follow a L1 strategy based on holistic word analysis. Future L2 word learning
35
36 studies might overcome this limitation in the present study by evaluating both RTs and ATs.
37
38

39
40 Furthermore, the underperformance observed in lower proficient bilinguals in this study
41
42 brings important remarks to consider at theoretical, practical, and experimental levels. Thus,
43
44 although all participants had comparable, high rates of naming accuracy during the training task,
45
46 there was no trace of a learning effect on long novel and low-frequency words in those
47
48 participants with lower proficiency. This result is only partially consistent with the self-teaching
49
50 framework (Share, 1995, 1999, 2008), since it implies that accurate phonological decoding of
51
52 L2 novel words through a brief period of exposure does not allow distant L1-L2 bilinguals,
53
54 particularly those with lower L2 proficiency, to create and strengthen orthographic
55
56 representations for these stimuli. The underperformance seen in these participants throughout
57
58 the training and post-training sessions is most likely a by-product of the L1 Chinese
59
60 orthographic background that importantly, might persist over the course of further exposures,

1
2
3
4 hence being seriously detrimental for the learning of L2 alphabetic scripts. This suggests that
5
6 L2 exposure solely might not be the most optimal strategy to ensure L2 orthographic learning
7
8 in distant L1-L2 bilinguals, calling for specific L2 teaching interventions focused on the lack
9
10 of experience with L2 low-level perceptual processes. This is in line with previous statements
11
12 (e.g., Nassaji, 2014) indicating that L2 instructional methodologies should be aware of the
13
14 consequences of cross-linguistic variance, in particular the difficulties caused by L1-L2
15
16 divergence. Moreover, our findings highlight the relevance of considering the naming latencies,
17
18 together with accuracy rates, as a more sensitive marker of orthographic learning –less
19
20 influenced by at-ceiling performance effects– both collected during the ongoing process under
21
22 study and at the post-training phase. Previous studies addressing L2 self-teaching and focusing
23
24 exclusively on post-training data have reported accuracy rates indicative of successful
25
26 orthographic learning in distant L1-L2 bilinguals (e.g., Li, Wang et al., 2021); however, the
27
28 present data offer a more precise analysis of the L2 self-teaching process and its modulation by
29
30 cross-linguistic variation, hence pointing to the pitfall of restricting the evaluation to offline
31
32 accuracy measures to fully understand orthographic learning processes.
33
34

35
36 To conclude, the present study contributes to the understanding of L2 orthographic
37
38 acquisition, elucidating the influence of distant L1-L2 reading strategies and the modulatory
39
40 role of L2 proficiency in this process. Importantly, our results suggest that the self-teaching
41
42 mechanism goes beyond an accurate decoding over a sufficient number of exposures to
43
44 successfully acquire new L2 orthographic representations. In this regard, the present work
45
46 highlights the importance of efficiency over accuracy in phonological decoding for less skilled
47
48 L2 learners. Implications at the practical level suggest that teaching methods for learning
49
50 alphabetic L2 should emphasize the enhancement of low-level perceptual processes that in turn
51
52 would help the engagement in grapheme-to-phoneme reading strategies.
53

54 **Disclosure statement**

55
56
57 No potential conflict of interest was reported by the author(s).
58
59
60

Funding

This research was supported by a grant from the Spanish Ministry of Science and Innovation (project number PID2020-114246GB-100) awarded to University of La Laguna, and by funds from the Regional Government of the Canary Islands, (Consejería de Economía, Conocimiento y Empleo, CEI Canarias-ULL, SD-19/02).

Data Accessibility Statement

The data and materials from the present experiment are publicly available at the Open Science Framework website: <https://osf.io/exnd5/>

References

- Akamatsu, N. (2002). A similarity in word-recognition procedures among second language readers with different first language backgrounds. *Applied Psycholinguistics*, 23(1), 117–133. <https://doi.org/10.1017/S0142716402000061>
- 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(4), 858–877. <https://doi.org/10.1017/S0305000917000514>
- 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(4), 390–412. <https://doi.org/10.1016/J.JML.2007.12.005>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
- Ben-Yehudah, G., Hirshorn, E. A., Simcox, T., Perfetti, C. A., & Fiez, J. A. (2019). Chinese-English bilinguals transfer L1 lexical reading procedures and holistic orthographic coding to L2 English. *Journal of Neurolinguistics*, 50, 136–148. <https://doi.org/10.1016/J.JNEUROLING.2018.01.002>

- 1
2
3
4 Berghoff, R., McLoughlin, J., & Bylund, E. (2021). L1 activation during L2 processing is
5 modulated by both age of acquisition and proficiency. *Journal of Neurolinguistics*, 58,
6 100979. <https://doi.org/10.1016/J.JNEUROLING.2020.100979>
7
8
9
10 Bermúdez-Margaretto, B., Kopytin, G., Myachykov, A., Fu, Y., Pokhoday, M., & Shtyrov, Y.
11 (2022). Bilingualism and acquisition of novel written words: the impact of phonological
12 conflict between L1 and L2 scripts. *Psychological Research*, 1, 1–20.
13
14 <https://doi.org/10.1007/S00426-021-01529-Y>
15
16
17
18
19 Bosse, M. line, Chaves, N., Largy, P., & Valdois, S. (2015). Orthographic learning during
20 reading: the role of whole-word visual processing. *Journal of Research in Reading*,
21 38(2), 141–158. <https://doi.org/10.1111/J.1467-9817.2012.01551.X>
22
23
24
25
26 Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of
27 novel word learning on the identification of visually similar words. *Cognition*, 97(3),
28 B45–B54. <https://doi.org/10.1016/J.COGNITION.2005.02.002>
29
30
31
32
33 Bowey, J. A., & Muller, D. (2005). Phonological recoding and rapid orthographic learning in
34 third-graders' silent reading: A critical test of the self-teaching hypothesis. *Journal of*
35 *Experimental Child Psychology*, 92(3), 203–219.
36
37 <https://doi.org/10.1016/J.JECP.2005.06.005>
38
39
40
41
42 Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative
43 simulations. *Journal of computational and graphical statistics*, 7(4), 434-455.
44
45
46
47 Bürkner, P. C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *R*
48 *Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>
49
50
51
52 Casillas, J. V. (2020). The longitudinal development of fine-phonetic detail: Stop production
53 in a domestic immersion program. *Language Learning*, 70(3), 768-806.
54
55
56
57 Cheng, Y. L. (2017). English Orthographic Learning in Chinese-L1 Young EFL Beginners.
58 *Journal of Psycholinguistic Research*, 46(6), 1453–1470.
59
60 <https://doi.org/10.1007/S10936-017-9507-X/TABLES/4>

- 1
2
3
4 Chung, S. C., Chen, X., Commissaire, E., Krenca, K., & Deacon, S. H. (2019b). Testing the
5 self-teaching hypothesis in second language reading. *Writing Systems Research*, 0(0), 1–
6 11. <https://doi.org/10.1080/17586801.2019.1625478>
7
8
9
10 Chung, S. C., Chen, X., & Geva, E. (2019a). Deconstructing and reconstructing cross-
11 language transfer in bilingual reading development: An interactive framework. *Journal*
12 *of Neurolinguistics*, 50, 149-161.
13
14
15
16
17 Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route
18 cascaded model of visual word recognition and reading aloud. *Psychological Review*,
19 108(1), 204–256. <https://doi.org/10.1037/0033-295X.108.1.204>
20
21
22
23
24 Coretta, S. (2019). tidymv: Plotting for generalised additive models [Computer software 988
25 manual]. Retrieved from <https://github.com/stefanocoretta/tidymv> (R package 989
26 version 2.0.0)
27
28
29
30
31 Cunningham, A. E. (2006). Accounting for children’s orthographic learning while reading
32 text: Do children self-teach? *Journal of Experimental Child Psychology*, 95(1), 56–77.
33 <https://doi.org/10.1016/J.JECP.2006.03.008>
34
35
36
37
38 Cunningham, A. E., Perry, K. E., Stanovich, K. E., & Share, D. L. (2002). Orthographic
39 learning during reading: examining the role of self-teaching. *Journal of Experimental*
40 *Child Psychology*, 82(3), 185–199. [https://doi.org/10.1016/S0022-0965\(02\)00008-5](https://doi.org/10.1016/S0022-0965(02)00008-5)
41
42
43
44
45 Cuppini, C., Magosso, E., & Ursino, M. (2013). Learning the lexical aspects of a second
46 language at different proficiencies: A neural computational study. *Bilingualism:*
47 *Language and Cognition*, 16(2), 266–287. <https://doi.org/10.1017/S1366728911000617>
48
49
50
51 de Jong, P. F., & Share, D. L. (2007). Orthographic Learning During Oral and Silent Reading.
52 *Scientific Studies of Reading*, 11(1), 55–71. <https://doi.org/10.1080/10888430709336634>
53
54
55
56 van Daal, V.H. & Wass, M. (2017). First- and Second-Language Learnability Explained by
57 Orthographic Depth and Orthographic Learning: A “Natural” Scandinavian
58
59
60

1
2
3 Experiment, *Scientific Studies of Reading*, 21:1, 46-59.

4
5 <http://dx.doi.org/10.1080/10888438.2016.1251437>
6
7

8 Davies, R., Rodríguez-Ferreiro, J., Suárez, P. & Cuetos, F. (2013). Lexical and sub-lexical
9 effects on accuracy, reaction time and response duration: Impaired and typical word and
10 pseudoword reading in a transparent orthography. *Reading and Writing: An*
11 *Interdisciplinary Journal*, 26(5), 721–738. <https://doi.org/10.1007/s11145-012-9388-1>
12
13
14

15 Deacon, S. H., Pasquarella, A., Marinus, E., Tims, T., & Castles, A. (2019). Orthographic
16 processing and children's word reading. *Applied Psycholinguistics*, 40(2), 509–534.
17
18 <https://doi.org/10.1017/S0142716418000681>
19
20
21

22 DeCoster, J., Gallucci, M., & Iselin, A. M. R. (2011). Best practices for using median splits,
23 artificial categorization, and their continuous alternatives. *Journal of experimental*
24 *psychopathology*, 2(2), 197-209.
25
26
27

28 Duyck, W., Desmet, T., Verbeke, L. P. C., & Brysbaert, M. (2004). WordGen: A tool for
29 word selection and nonword generation in Dutch, English, German, and French.
30
31 *Behavior Research Methods, Instruments, & Computers* 36:3, 36(3), 488–499.
32
33 <https://doi.org/10.3758/BF03195595>
34
35
36

37 Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2021). What are the top 200 most spoken
38 languages. *Ethnologue: Languages of the world*.
39
40
41

42 Eberhard, D. M., Gary F. S. & Charles, D. F. (eds.). (2020). *Ethnologue: Languages of the*
43 *World*. 23rd edition. Dallas, Texas: SIL International. Online version:
44
45 <http://www.ethnologue.com>
46
47
48

49 Ehri, L. C. (2014). Orthographic Mapping in the Acquisition of Sight Word Reading, Spelling
50 Memory, and Vocabulary Learning. 18(1), 5–21.
51
52 <https://doi.org/10.1080/10888438.2013.819356>
53
54
55
56
57
58
59
60

- 1
2
3
4 Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal*
5 *Learning and Verbal Behavior*, 12(6), 627–635. <https://doi.org/10.1016/S0022->
6 [5371\(73\)80042-8](https://doi.org/10.1016/S0022-5371(73)80042-8)
7
8
9
10 Fu, Y., Wang, H., Guo, H., Bermúdez-Margaretto, B., & Domínguez Martínez, A. (2020).
11 What, Where, When and How of Visual Word Recognition: A Bibliometrics Review.
12 *Language and Speech*, 23830920974710. <https://doi.org/10.1177/0023830920974710>
13
14
15
16
17 Gelman, A., Simpson, D., & Betancourt, M. (2017). The Prior Can Often Only Be
18 Understood in the Context of the Likelihood. *Entropy*, 19(10), 555.
19 <https://doi.org/10.3390/E19100555>
20
21
22
23
24 Gelman, A., & Carpenter, B. (2020). Bayesian analysis of tests with unknown specificity and
25 sensitivity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(5),
26 1269-1283.
27
28
29
30
31 Ginestet, E., Valdois, S., Diard, J., & Bosse, M. L. (2020). Orthographic learning of novel
32 words in adults: effects of exposure and visual attention on eye movements. *Journal of*
33 *Cognitive Psychology*, 32(8), 785–804. <https://doi.org/10.1080/20445911.2020.1823987>
34
35
36
37
38 Hamada, M., & Koda, K. (2008). Influence of First Language Orthographic Experience on
39 Second Language Decoding and Word Learning. *Language Learning*, 58(1), 1–31.
40 <https://doi.org/10.1111/J.1467-9922.2007.00433.X>
41
42
43
44 Hamada, M., & Koda, K. (2010a). The Role of the Phonological Loop in English Word
45 Learning: A Comparison of Chinese ESL Learners and Native Speakers. *Journal of*
46 *Psycholinguistic Research*, 40(2), 75–92. <https://doi.org/10.1007/S10936-010-9156-9>
47
48
49
50
51 Hamada, M., & Koda, K. (2010b). The Role of Phonological Decoding in Second Language
52 Word-Meaning Inference. *Applied Linguistics*, 31(4), 513–531.
53 <https://doi.org/10.1093/APPLIN/AMP061>
54
55
56
57
58
59
60

- 1
2
3
4 Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading:
5 cooperative division of labor between visual and phonological processes. *Psychological*
6 *review*, *111*(3), 662.
7
8
9
10 Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill.
11 *Studies in second language acquisition*, *14*(1), 25-38.
12
13
14
15 Hoffman, M. D., & Gelman, A. (2011). The No-U-Turn Sampler: Adaptively Setting Path
16 Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1593–
17 1623. [https://doi.org/10.48550/](https://doi.org/10.48550/arxiv.1111.4246)
18 [arxiv.1111.4246](https://doi.org/10.48550/arxiv.1111.4246)
19
20
21
22 Hu, CF. (2013). Predictors of reading in children with Chinese as a first language: a
23 developmental and cross-linguistic perspective. *Reading and writing*, *26*, 163–187.
24 <https://doi.org/10.1007/s11145-012-9360-0>
25
26
27
28
29 Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing,
30 estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic*
31 *bulletin & review*, *25*(1), 178-206. <https://doi.org/10.3758/s13423-016-1221-4>
32
33
34
35 Kwok, R. K. & Ellis, A. W. (2014). Visual word learning in adults with dyslexia. *Frontiers in*
36 *Human Neuroscience*, *8*(5). [https://doi.org/10.3389/FNHUM.2014.00264/](https://doi.org/10.3389/FNHUM.2014.00264/ABSTRACT)
37 [ABSTRACT](https://doi.org/10.3389/FNHUM.2014.00264/ABSTRACT)
38
39
40 Kwok, R. K., Cuetos, F., Avdyli, R., & Ellis, A. W. (2017). Reading and lexicalization in
41 opaque and transparent orthographies: Word naming and word learning in English and
42 Spanish. *Quarterly Journal of Experimental Psychology*, *70*(10), 2105–2129.
43 <https://doi.org/10.1080/17470218.2016.1223705>
44
45
46
47
48
49 Kwok, R. K. & Ellis, A. W. (2015). Visual word learning in skilled readers of English.
50 *Quarterly Journal of Experimental Psychology*, *68*(2), 326–349.
51 <https://doi.org/10.1080/17470218.2014.944549>
52
53
54
55
56 Leminen, A., Kimppa, L., Leminen, M. M., Lehtonen, M., Mäkelä, J. P., & Shtyrov, Y.
57 (2016). Acquisition and consolidation of novel morphology in human neocortex: A
58 neuromagnetic study. *Cortex*, *83*, 1–16
59
60

- 1
2
3
4 Li, L., Marinus, E., Castles, A., Hsieh, M. L., & Wang, H. C. (2021). Semantic and
5
6 Phonological Decoding in Children's Orthographic Learning in Chinese. *Scientific*
7
8 *Studies of Reading*, 25(4), 319–334. <https://doi.org/10.1080/10888438.2020.1781863>
9
- 10
11 Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An
12
13 enhanced tool for assessing multilingual experience. *Bilingualism: Language and*
14
15 *Cognition*, 23(5), 938–944. <https://doi.org/10.1017/S1366728918001153>
16
- 17
18 Li, Y., Li, H., & Wang, M. (2020). Orthographic learning via self-teaching in Chinese: The
19
20 roles of phonological recoding, context, and phonetic and semantic radicals. *Journal of*
21
22 *Experimental Child Psychology*, 199, 104913.
23
24 <https://doi.org/10.1016/j.jecp.2020.104913>
25
- 26
27 Li, Y., Wang, M., & Espinas, D. (2021). Orthographic learning through self-teaching among
28
29 learners of English as a second language. *Reading and Writing*, 34(5), 1295–1320.
30
31 <https://doi.org/10.1007/s11145-020-10115-4>
32
- 33
34 Li, Y., Wang, M., & Espinas, D. (2022). Self-teaching new words among English language
35
36 learners. *Reading and Writing*, 1-20.
37
- 38
39 Maloney, E., Risko, E. F., O'Malley, S., & Besner, D. (2009). Short article: Tracking the
40
41 transition from sublexical to lexical processing: On the creation of orthographic and
42
43 phonological lexical representations. *Quarterly Journal of Experimental Psychology*,
44
45 62(5), 858-867.
46
- 47
48 Martens, V. E. G., & de Jong, P. F. (2006). The effect of word length on lexical decision in
49
50 dyslexic and normal reading children. *Brain and Language*, 98(2), 140–149.
51
52 <https://doi.org/10.1016/J.BANDL.2006.04.003>
53
- 54
55 Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical
56
57 significance. *Psychological bulletin*, 113(1), 181.
58
59
60

- 1
2
3
4 Merkx, M., Rastle, K., & Davis, M. H. (2011). The acquisition of morphological knowledge
5 investigated through artificial language learning. *Quarterly Journal of Experimental*
6 *Psychology*, 64(6), 1200–1220.
7
8
9
- 10
11 Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word
12 recognition tasks: Where are they?. *Journal of Experimental Psychology: General*,
13 118(1), 43.
14
15
16
- 17
18 Nation, K., Angell, P., & Castles, A. (2007). Orthographic learning via self-teaching in
19 children learning to read English: Effects of exposure, durability, and context. *Journal of*
20 *experimental child psychology*, 96(1), 71-84.
21
22
23
- 24
25 Nassaji, H. (2014). The role and importance of lower-level processes in second language
26 reading. *Language Teaching*, 47(1), 1–37. [https://doi.org/DOI:](https://doi.org/DOI:10.1017/S0261444813000396)
27 10.1017/S0261444813000396
28
29
- 30
31 Pae, H. K., Sun-A Kim, Mano, Q. R., & Kwon, Y. J. (2017). Sublexical and lexical
32 processing of the English orthography among native speakers of Chinese and Korean.
33 *Reading and Writing*, 30(1), 1–24. [https://doi.org/10.1007/S11145-016-9660-](https://doi.org/10.1007/S11145-016-9660-X/TABLES/2)
34 [X/TABLES/2](https://doi.org/10.1007/S11145-016-9660-X/TABLES/2)
35
36
37
38
- 39
40 Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized
41 additive models in ecology: an introduction with mgcv. *PeerJ*, 7, e6876.
42
43
- 44
45 Perfetti, C., Cao, F., & Booth, J. (2012). Specialization and Universals in the Development of
46 Reading Skill: How Chinese Research Informs a Universal Science of Reading.
47 *Scientific Studies of Reading*, 7(1), 5–21. <https://doi.org/10.1080/10888438.2012.689786>
48
49
50
- 51
52 Pritchard, S. C., Coltheart, M., Marinus, E., & Castles, A. (2018). A Computational Model of
53 the Self-Teaching Hypothesis Based on the Dual-Route Cascaded Model of Reading.
54 *Cognitive Science*, 42(3), 722–770. <https://doi.org/10.1111/COGS.12571>
55
56
57
58
59
60

- 1
2
3
4 Rabe, M. M., Vasishth, S., Hohenstein, S., Kliegl, R., & Schad, D. J. (2020). *hypr*: An R
5 package for hypothesis-driven contrast coding. *Journal of Open Source Software*, 5(48),
6 2134. <https://doi.org/10.21105/JOSS.02134>
7
8
9
10 Ricketts, J. (2011). Research Review: Reading comprehension in developmental disorders of
11 language and communication. *Journal of Child Psychology and Psychiatry*, 52(11),
12 1111–1123. <https://doi.org/10.1111/J.1469-7610.2011.02438.X>
13
14
15
16
17 Rosenthal, J., & Ehri, L. C. (2008). The mnemonic value of orthography for vocabulary
18 learning. *Journal of educational psychology*, 100(1), 175.
19
20
21
22 Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori
23 contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110,
24 104038. <https://doi.org/10.1016/J.JML.2019.104038>
25
26
27
28
29 Schwartz, M., Kahn-Horwitz, J., & Share, D. L. (2014). Orthographic learning and self-
30 teaching in a bilingual and biliterate context. *Journal of Experimental Child Psychology*,
31 117(1), 45–58. <https://doi.org/10.1016/J.JECP.2013.08.008>
32
33
34
35
36 Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading
37 acquisition. *Cognition*, 55(2), 151–218. [https://doi.org/10.1016/0010-0277\(94\)00645-2](https://doi.org/10.1016/0010-0277(94)00645-2)
38
39
40
41 Share, D. L. (1999). Phonological Recoding and Orthographic Learning: A Direct Test of the
42 Self-Teaching Hypothesis. *Journal of Experimental Child Psychology*, 72(2), 95–129.
43
44 <https://doi.org/10.1006/JECP.1998.2481>
45
46
47
48 Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental
49 onset of self-teaching. *Journal of Experimental Child Psychology*, 87(4), 267–298.
50
51 <https://doi.org/10.1016/J.JECP.2004.01.001>
52
53
54
55 Share, D. L. (2008). Orthographic Learning, Phonological Recoding, and Self-Teaching.
56 *Advances in Child Development and Behavior*, 36, 31–82.
57
58 [https://doi.org/10.1016/S0065-2407\(08\)00002-5](https://doi.org/10.1016/S0065-2407(08)00002-5)
59
60

- 1
2
3
4 Smith, A. C., Monaghan, P., & Huettig, F. (2021). The effect of orthographic systems on the
5 developing reading system: Typological and computational analyses. *Psychological*
6 *Review*, 128(1), 125–159. <https://doi.org/10.1037/rev0000257>
7
8
9
10 Snowling, M., Stackhouse, J., & Rack, J. (1986). Phonological dyslexia and dysgraphia—A
11 developmental analysis. *Cognitive Neuropsychology*, 3(3), 309–339.
12
13
14 Suárez-Coalla, P., Álvarez-Cañizo, M., & Cuetos, F. (2016). Orthographic learning in
15 Spanish children. *Journal of Research in Reading*, 39(3), 292–311.
16
17 <https://doi.org/10.1111/1467-9817.12043>
18
19
20
21 Stein, M., Federspiel, A., Koenig, T., Wirth, M., Lehmann, C., Wiest, R., Strik, W., Brandeis,
22 D., & Dierks, T. (2009). Reduced frontal activation with increasing 2nd language
23 proficiency. *Neuropsychologia*, 47(13), 2712–2720.
24
25 <https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2009.05.023>
26
27
28
29 van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A
30 new and improved word frequency database for British English. *Quarterly Journal of*
31 *Experimental Psychology*, 67(6), 1176–1190.
32
33 <https://doi.org/10.1080/17470218.2013.850521>
34
35
36
37
38 van Rij J, Hendriks P, van Rijn H, Baayen RH, Wood SN. (2017). Analyzing the Time
39 Course of Pupillometric Data. *Trends in Hearing*. 23. doi:10.1177/2331216519832483
40
41
42
43 van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2017). itsadug: Interpreting Time
44 Series and Autocorrelated data Using GAMMs (version 2.3). [https://CRAN.R-](https://CRAN.R-project.org/package=itsadug)
45 [project.org/package=itsadug](https://CRAN.R-project.org/package=itsadug)
46
47
48
49
50 van Viersen, S., Protopapas, A., Georgiou, G. K., Parrila, R., Ziaka, L., & de Jong, P. F.
51 (2022). Lexicality effects on orthographic learning in beginning and advanced readers of
52 Dutch: an eye-tracking study. *Quarterly Journal of Experimental Psychology*, 75(6),
53 1135-1154. <https://doi.org/10.1177/17470218211047420>
54
55
56
57
58
59
60

- 1
2
3
4 Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (1999). *Comprehensive test*
5
6 *of phonological processing: CTOPP*. Austin, TX: Pro-ed.
7
- 8
9 Wang, H. C., Castles, A., Nickels, L., & Nation, K. (2011). Context effects on orthographic
10
11 learning of regular and irregular words. *Journal of Experimental Child Psychology*,
12
13 *109*(1), 39–57. <https://doi.org/10.1016/j.jecp.2010.11.005>
14
- 15 Wang, M., & Koda, K. (2005). Commonalities and Differences in Word Identification Skills
16
17 Among Learners of English as a Second Language. *Language Learning*, *55*(1), 71–98.
18
19 <https://doi.org/10.1111/J.0023-8333.2005.00290.X>
20
21
- 22 Wang, M., Koda, K., & Perfetti, C. A. (2003). Alphabetic and nonalphabetic L1 effects in
23
24 English word identification: a comparison of Korean and Chinese English L2 learners.
25
26 *Cognition*, *87*(2), 129–149. [https://doi.org/10.1016/S0010-0277\(02\)00232-9](https://doi.org/10.1016/S0010-0277(02)00232-9)
27
28
- 29 Wang, C., Yang, Z., Cao, F., Liu, L., & Tao, S. (2019). Sound integration in native Chinese
30
31 speakers learning English: Brain fails in automatic responses but succeeds with more
32
33 attention. *Cognitive Neuroscience*, *10*(2), 100-116.
34
- 35 Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and
36
37 writing on word knowledge. *Studies in Second Language Acquisition*, *27*(1), 33–52.
38
39 <https://doi.org/10.1017/S0272263105050023>.
40
41
- 42 Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied linguistics*,
43
44 *28*(1), 46-65. <https://doi.org/10.1093/applin/aml048>.
45
46
- 47 Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming
48
49 latency. *Quarterly Journal of Experimental Psychology*, *50*(2), 439-456.
50
51
- 52 Weekes, B. S. (2018). Learning written word vocabulary in a second language: Theoretical
53
54 and practical implications. *Bilingualism: Language and Cognition*, *21*(3), 585–597.
55
56 <https://doi.org/10.1017/S1366728917000141>
57
58
59
60

- 1
2
3
4 Wong, A. C. N., Bukach, C. M., Yuen, C., Yang, L., Leung, S., & Greenspon, E. (2011).
5
6 Holistic Processing of Words Modulated by Reading Experience. *PLOS ONE*, 6(6),
7
8 e20753. <https://doi.org/10.1371/JOURNAL.PONE.0020753>
9
- 10
11 Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing Parameter and Model Selection for
12
13 General Smooth Models. *Journal of the American Statistical Association*, 111(516),
14
15 1548–1563.
16
17 https://doi.org/10.1080/01621459.2016.1180986/SUPPL_FILE/UASA_A_1180986_SM
18
19 0752.PDF
20
- 21
22 Yang, J., Shu, H., McCandliss, B. D., & Zevin, J. D. (2013). Orthographic influences on
23
24 division of labor in learning to read Chinese and English: Insights from computational
25
26 modeling. *Bilingualism: Language and Cognition*, 16(2), 354-366.
27
- 28
29 Yang, Z., Wang, C., Feng, L., Li, S., & Tao, S. (2016). Unsuccessful letter–sound integration
30
31 in English reading by native Chinese speakers: evidence from an event related potentials
32
33 study. *Science Bulletin*, 61(24), 1855–1864. <https://doi.org/10.1007/s11434-016-1078-y>
34
- 35
36 Yang, J., Zevin, J. D., Shu, H., McCandliss, B. D., & Li, P. (2006, January). A “triangle
37
38 model” of Chinese reading. In 28th annual conference of the cognitive science society,
39
40 Mahwah, NJ.
41
- 42
43 Zhang, S., & Thompson, N. (2004). DIALANG: A Diagnostic Language Assessment System
44
45 (review). *The Canadian Modern Language Review / La Revue Canadienne Des Langues*
46
47 *Vivantes*, 61(2), 290–293. <https://doi.org/10.1353/CML.2005.0011>
48
- 49
50 Ziegler, J. C., Perry, C., & Zorzi, M. (2014). Modelling reading development through
51
52 phonological decoding and self-teaching: implications for dyslexia. *Philosophical*
53
54 *Transactions of the Royal Society B: Biological Sciences*, 369(1634).
55
56 <https://doi.org/10.1098/RSTB.2012.0397>
57
58
59
60

Figure Captions

Figure 1. *Radar plot of the preliminary tests data*

Note. Each point represents the mean score (from 0 to 1) obtained for each group across the tests used. The tests used included: Language History Questionnaire (LHQ 3.0, Proficiency + Immersion), DIALANG diagnostic system (preliminary-level test), word translation task (WTT), word (WRT +)/pseudoword (WRT -) reading tasks, phonological awareness task (PAT), and digit (DST)/word (WST +)/nonword (WST-) span tasks. Proficiency level, measured through DIALANG test, was treated as a continuous variable across the whole population in the following models.

Figure 2. *Results obtained in the training session*

Note. For each high and low proficiency group, both nonlinear learning curves for different types of stimuli as well as the estimated difference of length, frequency and lexicality effects are plotted.

Figure 3. *Results obtained in the lexical decision task for pseudowords orthographically related to previously trained stimuli as a function of L2 proficiency level (from upper to lower panels: pseudowords related to high-frequency words, to low-frequency words and to novel words)*

Note. Thin lines represent 300 sample draws from the posterior distribution for each word type, whereas thick lines illustrate uncertainty (95% HDI) around the posterior medians. Dashed lines represent the intercept of the model.

Figure 4. *Probability of a correct response and naming latency obtained in the recognition task for each stimuli type as a function of L2 proficiency*

Note. Thin lines represent 300 sample draws from the posterior distribution for each word type, whereas thick lines illustrate uncertainty (95% HDI) around the posterior medians. Dashed lines represent the intercept of the model.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

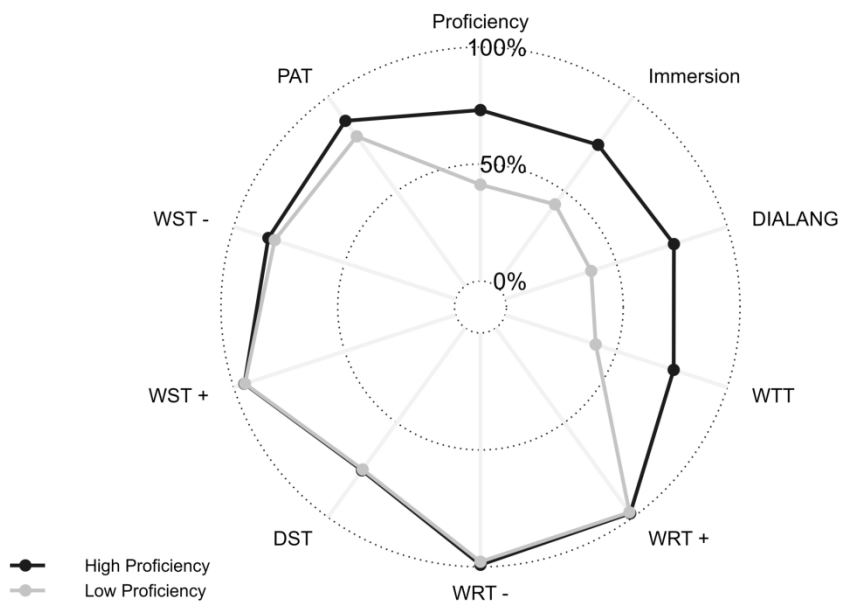


Figure 1. Radar plot of the preliminary tests data

Note. Each point represents the mean score (from 0 to 1) obtained for each group across the tests used. The tests used included: Language History Questionnaire (LHQ 3.0, Proficiency + Immersion), DIALANG diagnostic system (preliminary-level test), word translation task (WTT), word (WRT +)/pseudoword (WRT -) reading tasks, phonological awareness task (PAT), and digit (DST)/word (WST +)/nonword (WST-) span tasks. Proficiency level, measured through DIALANG test, was treated as a continuous variable across the whole population in the following models.

101x76mm (600 x 600 DPI)

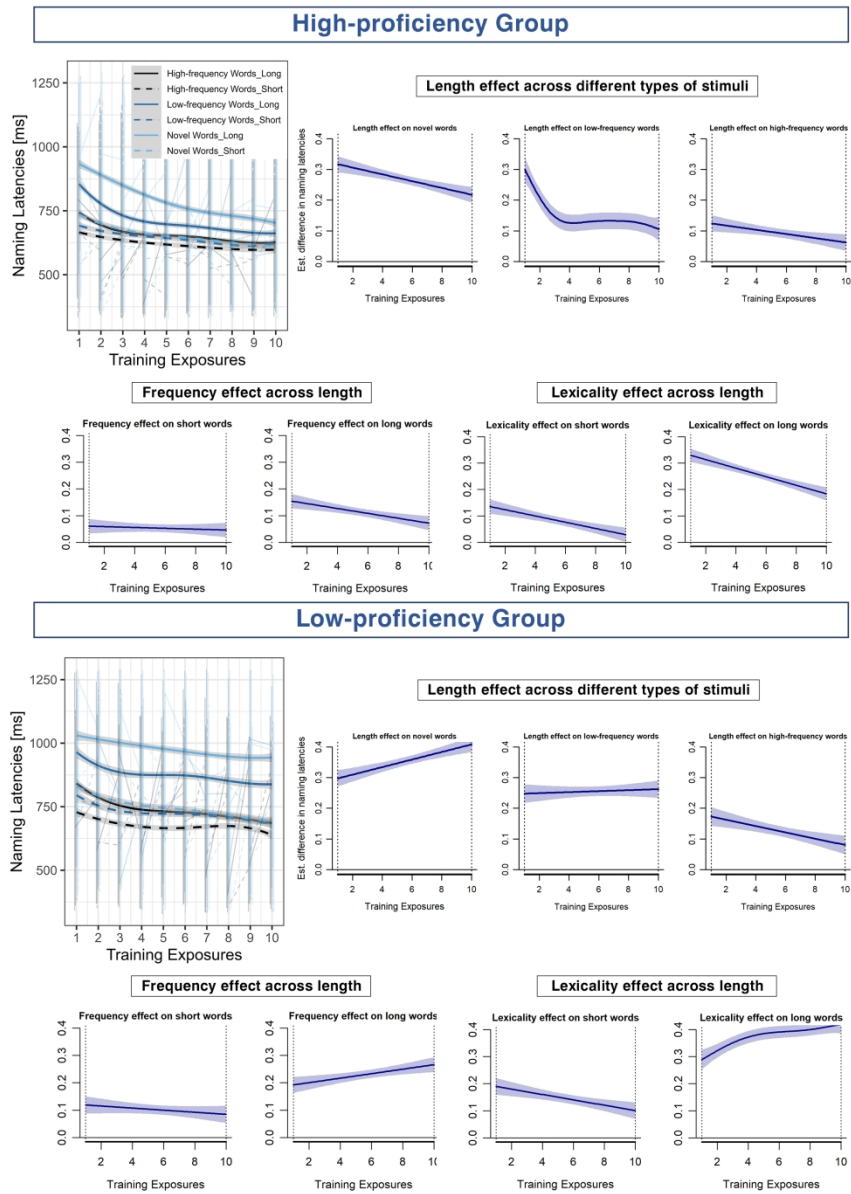


Figure 2. Results obtained in the training session
 Note. For each high and low proficiency group, both nonlinear learning curves for different types of stimuli as well as the estimated difference of length, frequency and lexicality effects are plotted.

434x598mm (144 x 144 DPI)

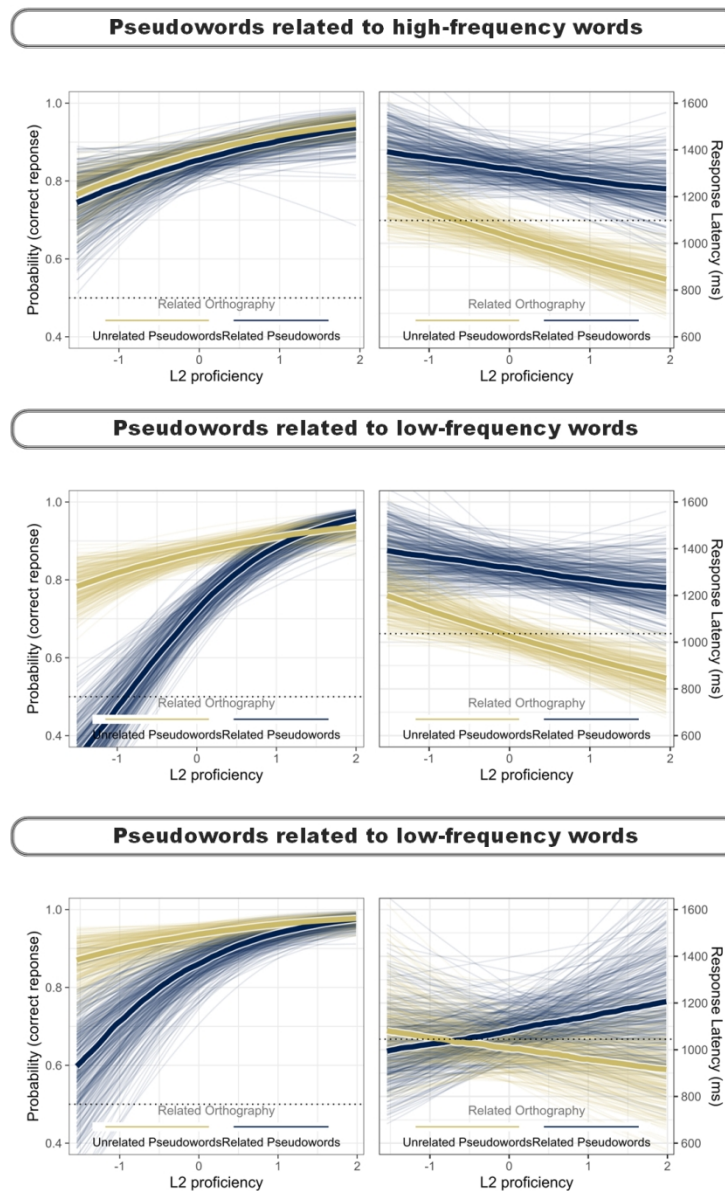


Figure 3. Results obtained in the lexical decision task for pseudowords orthographically related to previously trained stimuli as a function of L2 proficiency level (from upper to lower panels: pseudowords related to high-frequency words, to low-frequency words and to novel words)

Note. Thin lines represent 300 sample draws from the posterior distribution for each word type, whereas thick lines illustrate uncertainty (95% HDI) around the posterior medians. Dashed lines represent the intercept of the model.

194x318mm (144 x 144 DPI)

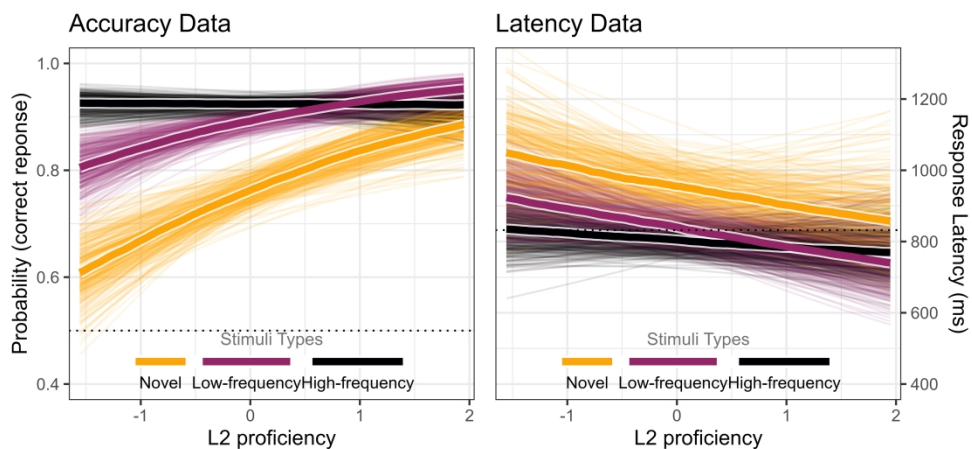


Figure 4. Probability of a correct response and naming latency obtained in the recognition task for each stimuli type as a function of L2 proficiency

Note. Thin lines represent 300 sample draws from the posterior distribution for each word type, whereas thick lines illustrate uncertainty (95% HDI) around the posterior medians. Dashed lines represent the intercept of the model.

203x96mm (600 x 600 DPI)

Table 1

Mean word frequency (Zipf) and mean log bigram frequency (SDs shown in brackets) matched across the experimental conditions for stimuli presented in the training and post-training sessions

Training Session						
	Four letters			Seven letters		
	High-Freq.	Low-Freq.	Pseudo.	High-Freq.	Low-Freq.	Pseudo.
Word frequency	5.12 (0.20)	3.79 (0.31)	—	5.02 (0.24)	3.65 (0.47)	—
Bigram frequency	3.38 (0.15)	3.32 (0.22)	3.38 (0.12)	3.38 (0.10)	3.30 (0.19)	3.39 (0.09)
Post-training Session – Lexical Decision Task						
Bigram frequency	Related Pseudowords			Unrelated Pseudowords		
	3.53 (0.16)			3.55 (0.08)		

Table 2

Summary of generalized additive mixed-effect model obtained in the training session

Parametric coefficients		Estimate	SE	t
(Intercept)		-1.61	0.03	-47.058 ***
Low-frequency words (Short)		0.08	0.03	2.917 **
Novel words (Short)		0.11	0.03	4.16 ***
High-frequency words (Short)		0.11	0.03	4.048 ***
Low-frequency words (Long)		0.29	0.03	10.566 ***
Novel words (Long)		0.42	0.03	15.574 ***
Approximate significance of smooth terms:		EDF	Ref. DF	F
Reference smooth:	Exposure	2.059	2	10.966 ***
	L2 proficiency	1	1	14.171 ***
Difference smooth: (Exposure)	High-frequency (Short)	4.09	6	9.262 ***
	Low-frequency (Short)	3.381	6	4.6 ***
	Novel (Short)	2.550e-05	6	0
	High-frequency (Long)	4.020	6	2.597 ***
	Low-frequency (Long)	4.39	6	2.575 **
	Novel (Long)	1.269e-04	6	0.426
	Tensor product smooth	Exposure : Proficiency	5.90	23
Difference smooth: (Exposure by Proficiency)	High-frequency (Short)	4.278e-01	18	0
	Low-frequency (Short)	4.099e-05	18	0
	Novel (Short)	3.063	18	0.405 *
	High-frequency (Long)	1.391	18	0.155
	Low-frequency (Long)	3.515	18	1.217 ***
Random smooth: Participants (word types)	Novel (Long)	3.942	18	3.761 ***
	Participants (word types)	185.2	251	24.924 ***
	Random factor smooth: Participants (exposure)	39.85	42	20945.79 ***
Random intercepts		EDF	Ref. DF	F
Participants		39.18	41	23755.56 ***
Items		62.72	66	21.566 ***

Note. The model fit Word Length, Stimuli Type and Proficiency as a function of Exposure. $R^2 = .48$; deviance explained = 48.88%; $n = 28195$. EDF = effective degrees of freedom; EDF >1 indicates the non-linear learning trajectory; Ref. DF = reference degrees of freedom.

Table 3

Summary of the posterior distribution modelling for response accuracy and latencies obtained in the lexical decision task

Pseudowords related to high-frequency words					
Model 1: Accuracy Data					
Population Level	Median	HDI	% in ROPE	MPE	\hat{R}
Intercept	1.85	[1.54, 2.17]	0	1	1
Related Orthography	-0.14	[-0.54, 0.29]	0.16	0.75	1
L2 Proficiency	0.47	[0.20, 0.74]	0	1	1
Interaction	-0.04	[-0.44, 0.34]	0.2	0.58	1
Model 2: Latency Data					
Population Level	Median	HDI	% in ROPE	MPE	\hat{R}
Intercept	-0.98	[-1.04, -0.92]	0	1	1
Related Orthography	0.26	[0.22, 0.31]	0	1	1
L2 Proficiency	-0.06	[-0.11, 0.00]	0.12	0.97	1
Interaction	0.06	[0.02, 0.11]	0.02	1	1
Pseudowords related to low-frequency words					
Model 3: Accuracy Data					
Population Level	Median	HDI	% in ROPE	MPE	\hat{R}
Intercept	1.43	[1.13, 1.73]	0	1	1
Related Orthography	-0.96	[-1.19, -0.71]	0	1	1
L2 Proficiency	0.73	[0.46, 0.99]	0	1	1
Interaction	0.66	[0.44, 0.86]	0	1	1
Model 4: Latency Data					
Population Level	Median	HDI	% in ROPE	MPE	\hat{R}
Intercept	-1.03	[-1.08, -0.97]	0	1	1
Related Orthography	0.15	[0.11, 0.20]	0	1	1
L2 Proficiency	-0.05	[-0.10, 0.00]	0.14	0.97	1
Interaction	0.07	[0.03, 0.11]	0	1	1
Pseudowords related to novel words					
Model 5: Accuracy Data					
Population Level	Median	HDI	% in ROPE	MPE	\hat{R}
Intercept	2.27	[1.84, 2.78]	0	1	1.01
Related Orthography	-0.89	[-1.48, -0.28]	0	1	1
L2 Proficiency	0.72	[0.36, 1.06]	0	1	1
Interaction	0.44	[0.10, 0.96]	0.01	0.99	1
Model 6: Latency Data					
Population Level	Median	HDI	% in ROPE	MPE	\hat{R}
Intercept	-1.00	[-1.14, -0.86]	0	1	1
Related Orthography	-0.05	[-0.11, 0]	0.14	0.97	1
L2 Proficiency	0.06	[-0.09, 0.19]	0.22	0.76	1
Interaction	-0.09	[-0.15, -0.04]	0	1	1

Note. The table includes posterior medians, the 95% of the highest density credible interval (HDI), the percentage of the HDI within the region of practical equivalence (ROPE), the maximum probability of effect (MPE), and the potential scale reduction factor (\hat{R}).

Table 4*Posterior distribution modelling for response accuracy and latencies obtained in the recognition task*

Model 1: Accuracy Data						
Population Level	Median	HDI	% in ROPE	MPE	\hat{R}	
Intercept	1.94	[1.73, 2.17]	0	1	1	
Frequency	0.36	[0.05, 0.65]	0	0.99	1	
Lexicality	1.33	[1.02, 1.61]	0	1	1	
L2 Proficiency	0.3	[0.11, 0.50]	0	1	1	
Frequency : L2 Proficiency	-0.45	[-0.73, -0.18]	0	1	1	
Lexicality : L2 Proficiency	-0.47	[-0.73, -0.20]	0	1	1	
Model 2: Latency Data						
Population Level	Median	HDI	% in ROPE	MPE	\hat{R}	
Intercept	3.66	[2.09, 4.97]	0	1	1	
Frequency	-4.03	[-5.69, -2.53]	0	1	1	
Lexicality	-4.58	[-6.39, -2.96]	0	1	1	
L2 Proficiency	0.51	[-0.92, 1.98]	0.22	0.77	1	
Frequency : L2 Proficiency	-0.46	[-1.75, 0.74]	0.43	0.78	1	
Lexicality : L2 Proficiency	-0.51	[-1.94, 0.79]	0.12	0.79	1	

Note. The table includes posterior medians, the 95% of the highest density credible interval (HDI), the percentage of the HDI within the region of practical equivalence (ROPE), the maximum probability of effect (MPE), and the potential scale reduction factor (\hat{R}).