*Applied Psycholinguistics* (2023), **44**, 237–263
doi:10.1017/S0142716423000176

ORIGINAL ARTICLE

# The impact of L1 orthographic depth and L2 proficiency on mapping orthography to phonology in L2-English: an ERP investigation

Mona Roxana Botezatu

Department of Speech, Language and Hearing Sciences, University of Missouri, Columbia, MO, USA
Email: botezatum@health.missouri.edu

(Received 26 June 2022; revised 4 March 2023; accepted 7 March 2023; first published online 04 April 2023)

## Abstract

English monolinguals (Experiment 1) and first language (L1)-dominant, Spanish-English and Chinese-English bilinguals (Experiment 2), who differed in L1 orthographic depth (shallow: Spanish; deep: Chinese) and second language (L2–English) proficiency, decided whether visually presented letter strings were English words, while behavioral and EEG measures were recorded. The spelling-sound regularity and consistency of stimuli were covaried such that words had either regular/consistent (e.g., GATE) or irregular/inconsistent mappings (e.g., PINT). Irregular/inconsistent words elicited more positive P200 and less negative N400 amplitudes than regular/consistent words in monolinguals, yet only a P200 response in bilinguals. English proficiency modulated L2 reading strategies, such that bilinguals employed distinct reading unit sizes in the L2 than the L1 when L2 proficiency was low, but transferred L1 reading units to the L2 when L2 proficiency was high. ERP results suggest that high L2 proficiency may be a prerequisite to the cross-linguistic transfer of reading strategies.

**Keywords:** bilingual visual word recognition; orthographic depth; cross-linguistic transfer; P200; N400

Mapping orthography to phonology, a process that becomes highly automatic with experience, is at the core of skilled reading. In English, the ease of mapping spelling to sound may be characterized in terms of regularity (Coltheart et al., 2001), which measures whether the mappings violate grapheme-phoneme correspondence (GPC) rules (Bolger et al., 2008), and consistency (Seidenberg & McClelland, 1989), which measures whether the spellings map to multiple sounds. Irregularity and inconsistency of spelling-sound mappings have been associated with higher demands on cognitive resources in word naming (e.g., Baron & Strawson, 1976; Coltheart & Rastle, 1994; Lacruz & Folk, 2004) and lexical decision tasks (e.g., Lacruz & Folk, 2004). The cost of processing irregularity (Coltheart & Rastle, 1994) and inconsistency (Jared, 2002) has been linked to competing activations of correct and incorrect pronunciations that take time to resolve, resulting in slower and less accurate naming and recognition of words with irregular and/or inconsistent mappings (Coltheart & Leahy, 1992; Glushko, 1979). Various computational models have accounted for these effects (Coltheart et al., 1993; Plaut et al., 1996; Seidenberg & McClelland, 1989). Effects of regularity have been accounted for by the dual route cascaded model (DRC, Coltheart et al., 1993), which assumes that the phonological representations of printed words may be retrieved through a lexical-semantic pathway, but can also be assembled based on GPC rules applied to the orthographic input. The model predicts that the regularity effect occurs due to competition between the output of the lexical-semantic and assembly routes. Effect of consistency has been predicted by the DRC model to occur as a consequence of competition within the lexical route. Connectionist models of reading (Plaut et al., 1996; Seidenberg & McClelland, 1989) also account for consistency effects, assuming that phonological representations of different spelling patterns are learned through repeated exposure to a large corpus of words and are computed through a single procedure. When regularity and consistency are studied in isolation, robust effects of consistency are observed, while effects of regularity yield conflicting results (Andrews, 1982; Cortese & Simpson, 2000; Jared, 2002; Kay & Bishop, 1987). Since irregular words are often inconsistent, results may be confounded (Glushko, 1979). Critically, the magnitude of regularity and consistency effects decreases with experience, previously operationalized as chronological age (Waters et al., 1984; Weekes et al., 2006) and phonemic decoding skills (Botezatu et al., in press), as readers become more reliant on lexical/semantic processing.

For non-native English speakers, the process of computing phonology from text is further complicated by proficiency in a native language (L1) with a fully overlapping (Jared & Kroll, 2001; Schwartz et al., 2007) or partially overlapping orthography (Marian & Kaushanskaya, 2004), as bilinguals must resolve competition from both within- and cross-language pronunciation enemies (for a review of lexical access in bilinguals, see Kroll et al., 2022). Jared and Kroll (2001) were among the first to demonstrate that bilinguals experience competition from spelling-to-sound correspondences in the nontarget language. Their study showed that naming a block of filler words in the nontarget language increased competition from enemy pronunciations when naming a subsequent block of L1 or second language (L2) words. Van Leerdam et al. (2009) provided additional evidence that bilinguals experience interference from both within- and cross-language enemy neighbors when reading L2 words. The authors found that Dutch-English bilinguals activated the incorrect phonological representation of printed L2-English words when primed by auditory rimes that either derived from English (within-language) enemy neighbors, or had Dutch (cross-language) enemy neighbors and were pronounced using Dutch phonology. Furthermore, Botezatu et al. (submitted) also found a positive correlation between regularity-consistency effects in English word naming and objective measures of Spanish proficiency in English-dominant heritage speakers of Spanish.

Differences in the arbitrariness of print-to-sound mappings between bilinguals' two languages (i.e., orthographic depth; Seymour et al., 2003) further complicate the computation of phonology from orthography in the L2, as readers have already developed an optimal reading unit (i.e., phoneme, syllable, rime, whole word), or "grain size" (Ziegler & Goswami, 2005; Ziegler et al., 2001) in response to the transparency of print-to-sound mappings in their L1 (Katz & Frost, 1992; Martensen et al., 2000), which may not be the optimal reading unit for their L2. In transparent alphabetic orthographies such as Spanish, where a letter (or letter combination) is consistently mapped to one sound, readers favor small grain sizes, efficiently relying on letter-by-letter or syllable-by-syllable decoding strategies (Ziegler & Goswami, 2005; Ziegler et al., 2001). In opaque or deep alphabetic orthographies such as English, where the same letter (or combination of letters) can map to multiple sounds (e.g., the "ERE" rime in THERE [.ϵr], HERE [hır], WERE [wər]), readers may rely to a greater extent on large grain sizes (Frost, Katz & Bentin, 1987) at rime or whole word levels (Ziegler & Goswami, 2005; Ziegler et al., 2001). Chinese, which uses a logographic writing system, is also described as having a deep orthography (Lee et al., 2005) because no assembly is possible or necessary within individual graphic units (i.e., characters or morphemes), which map onto whole syllables that often constitute entire words (Liu et al., 1996; Perfetti & Liu, 2005; Perfetti et al., 2007). Most modern Chinese characters are semantic-phonetic compounds (Shu et al., 2003), which contain a semantic radical (encoding meaning) and a phonetic radical (encoding pronunciation) that enable the lexical retrieval of phonology much like in alphabetic orthographies (Perfetti & Tan, 1998; Zhang et al., 1999). Nonetheless, evidence that Chinese speakers may also access sublexical phonology (Lee et al., 2006; Liu et al., 2003) comes from studies reporting effects of regularity (i.e., operationalized as the match between the pronunciation of a Chinese character and its phonetic radical, regardless of tonal differences; see Hue, 1992) and consistency (i.e., operationalized as the relative number of friend and enemy pronunciations within an orthographic neighborhood of Chinese characters that share the same phonetic radical; see Lee et al., 2010) in Chinese character naming.

Bilinguals have been reported to transfer word reading skills from the L1 both within (Mumtaz & Humphreys, 2001) and across writing systems (Wang et al., 2005). Available evidence suggests that bilinguals who transfer reading strategies from logographic orthographies prefer using a large grain size in English, rather than relying on English GPC rules to activate phonology (Tan et al., 2003). In contrast, bilinguals who transfer word reading skills from shallow orthographies show heavier reliance on small rather than large grain sizes in English (Mumtaz & Humphreys, 2001) – a strategy that comes at the cost of resolving competition from L1 spelling-sound mappings (Schwartz et al., 2007), which often differ from L2 mappings (predominantly an issue for vowels, see Treiman et al., 1995). Furthermore, the grain size accommodation hypothesis (Lallier & Carreiras, 2018) proposes that bilingual readers of two orthographies with distinct depths may not simply transfer the grain size preferences cross-linguistically, but rather may develop a hybrid grain size between those employed by monolingual speakers of each language. The use of a hybrid grain size would predict not only an overreliance on sublexical strategies in reading a deep L2 orthography for bilingual readers of a shallow L1 orthography but also the use of larger grains in a shallower L2 orthography by readers of a deep L1 orthography. Available evidence suggests that the preferred reading strategy in the L2 may also depend on L2 proficiency. Botezatu et al. (in press) found a larger cost in naming irregular/inconsistent than regular/consistent English words in Spanish-English bilinguals with lower L2-English proficiency, suggesting that as L2 proficiency increases, bilinguals adapt their reading strategy to better fit the L2. The idea that bilinguals' reading strategies may shift with dominance is supported by Meschyan and Hernandez (2006), who reported that Spanish-English bilinguals dominant in L2-English adopted a visually driven (lexical) reading strategy in English.

## The current study

The current study exploits differences in orthographic depth between bilinguals' two languages for the purpose of determining whether bilinguals transfer reading strategies from the dominant L1 to the weaker L2 in the process of mapping print to sound in L2-English. To capture the computation of phonology from orthography, a process that is highly automatic in adult skilled readers, I employed Event-Related Potentials (ERPs) to index processing dynamics in real time. Although past work evaluating language transfer effects in reading has relied primarily on behavioral methods (for a review, see Durgunoglu & Hancin, 1992), ERPs may be sensitive to implicit L2 processing that is masked in behavioral performance (McLaughlin et al., 2004; Tokowicz & MacWhinney, 2005). Thus, ERPs may reveal effects of reading unit transfer that are not evident in behavioral measures. Studies with monolinguals have reported effects of spelling-sound regularity/consistency as a two-stage process that indexes phonological competition at sublexical (P200) and lexical (N400) levels. The P200 is sensitive to the degree of mismatch between the orthographic and phonological features of words (Kramer & Donchin, 1987). Larger P200s have been reported to irregular English words (Martin et al., 2006; Sereno et al., 1998) and to irregular (Yum et al., 2014) and low consistency Chinese characters (Hsu et al., 2009; Lee et al., 2007), reflecting competition among phonological forms. An index of lexical access (see Kutas & Federmeier, 2011, for a review), the N400, is modulated by lexical properties such as orthographic neighborhood size (Holcomb et al., 2002). Larger N400s have been reported to be regular (Yum et al., 2014) and high-consistency Chinese characters (Hsu et al., 2009; Lee et al., 2007), reflecting competition among phonological neighbors. To date, no studies have reported modulations of the N400 to regularity or consistency manipulations in English.

## Experiment 1

Experiment 1 evaluated the time course of processing spelling-sound regularity and consistency in English monolinguals. Although effects of regularity and consistency have been traditionally evaluated in tasks that require explicit access to phonology, such as naming (Baron & Strawson, 1976; Coltheart & Rastle, 1994; Lacruz & Folk, 2004) and rhyme judgment (Bolger et al., 2008; Botezatu, in press), slower response latencies (Stanovich & Bauer, 1978) and lower accuracy rates (Schmalz et al., 2013) to irregular words and inconsistent words (Lacruz & Folk, 2004; Stone et al., 1997; Ziegler et al., 2001) have also been reported in lexical decision tasks, suggesting that phonological information may be available automatically in lexicality judgments. Lexical decision tasks have an advantage over naming tasks in ERP studies, as they eliminate undesirable movement artifacts.

**Table 1.** Mean (standard error) psycholinguistic data in English monolinguals

| Measure | English |
|---|---|
| Sight Word Reading | 94.90 (1.82) |
| Phonemic Decoding | 53.48 (2.35) |
| Word Identification | 99.48 (1.18) |
| Word Attack | 40.14 (0.72) |
| Picture Naming Latency | 802.73 (25.57) |
| Picture Naming Accuracy | 90.16 (1.64) |
| Self-Rated L1 Proficiency (/10) | 9.46 (0.13) |
| Self-Rated L2 Proficiency (/10) | 3.10 (0.48) |

As in prior studies (Andrews, 1982; Glushko, 1979; Jared, 2002; Kay & Bishop, 1987), the current work confounded regularity and consistency to generate a robust effect. It was hypothesized that if English monolinguals use a small reading unit, they would produce longer latencies and lower accuracy rates, as well as larger P200 and smaller N400 mean amplitudes to irregular/inconsistent than regular/consistent words. Alternatively, if monolinguals use a large grain size, effects of regularity/consistency may be observed in electrophysiological measures of lexical (N400) processing, but should be short-lived and therefore absent from the behavioral data.

## Method

### *Participants*

Results are presented for 21 native speakers of English (8 male; mean age = 20.9; range = 18–27 years) recruited from Pennsylvania State University, who varied in English proficiency (see Table 1) and had limited knowledge of an L2 (no exposure to an L2 prior to the age of 6, less than 4 years of formal study of an L2 at the high school or college level, and no study abroad experience). A total of 29 English monolinguals completed the study for payment, but data from 8 participants were excluded from the analysis either because they had fewer than 70% artifact-free correct response trials per condition or because they failed to maintain at least 70% accuracy for overall "yes" and "no" responses. Participants were right-handed, had normal or corrected-to-normal vision, and no history of neurological, language, or reading disorders, according to self-report.

## Materials and procedure

Participants completed two testing sessions. During the initial testing session, participants provided self-ratings of language exposure and use and performed a lexical decision task while behavioral and electrophysiological (EEG) measures were recorded. Behavioral measures of picture naming, as well as timed and untimed measures of sight word reading and phonemic decoding, were collected during a second testing session scheduled one week later. The lexical decision and picture naming tasks were presented electronically using in-house and E-Prime 2.0 software (Psychology Software Tools Incorporated, 2012), respectively, which recorded trial-level response latencies and accuracy. Each task is described below.

### Lexical decision

Effects of spelling-sound regularity/consistency were measured in a visual lexical decision task. Participants were presented with 150 words and 150 pseudowords in semi-randomized order. All items had consonant onsets and were matched on length and number of syllables (equal numbers of monosyllabic and bisyllabic trials per condition) and controlled for syllable stress. Words were matched on lexical frequency using the American National Corpus (Ide & Suderman, 2004) norms. Experimental trials were 100 English words that covaried systematically in the regularity and consistency of their spelling-sound mappings: 50 regular/consistent words (e.g., GATE) and 50 irregular/inconsistent words (e.g., PINT). Regularity was judged based on the degree to which words could be pronounced correctly following English GPC rules; consistency was judged based on rime neighborhoods of similar and contrasting pronunciations (Jared, 1997, 2002), following the Ziegler et al. (1997) norms. A word was considered to be consistent if its rime (e.g., –ATE) mapped to a single phonological representation (e.g., /_eɪt/, having rhyming neighbors, known as *friends*: DATE, FATE, RATE), but inconsistent if its rime (e.g., –INT) mapped to distinct, *enemy*, pronunciations (e.g., /_ɪnt/ as in HINT and LINT, but /_aɪnt/ as in PINT). Irregular/inconsistent words chosen for this study had the less frequent rime pronunciation. Table 2 summarizes the lexical properties of the experimental conditions.

Filler trials were 50 language ambiguous words (Spanish-English cognates with complete orthographic overlap across the two translations) and 150 pseudowords created from a set of base words with regular-consistent and irregular-inconsistent mappings by changing the onset consonant. Filler items were not considered in the current analysis.

Participants were seated in a comfortable chair in a dimly lit, sound-attenuated, and electrically shielded booth, fitted with an elastic electrode cap, and asked to perform a visual lexical decision task, which consisted of a continuous stream of 300 stimuli, with brief breaks every 100 trials. Stimuli were presented in Arial font white capital letters in the center of a black screen at a viewing distance of approximately 48 inches. The sequence of events for each trial was as follows: participants viewed a 1200 millisecond (ms) central fixation cross (+), followed by a 300 ms blank screen, a stimulus presented for 350 ms, and finally a 850 ms blank screen before the presentation of the next fixation cross. Participants were instructed not to blink or move unless the fixation cross was on the screen and to respond quickly and accurately to each stimulus by pressing one button of the response device if the stimulus was a word in English and another button if the stimulus was not a word in English. The button pressed for “yes” versus “no” responses was counterbalanced across participants. A set of 18 practice trials preceded the experimental set. Participants were given feedback about their blink performance, but not their response accuracy, after the practice trials and during breaks, if needed.

**Table 2.** Mean (standard deviation) stimulus properties

| Stimulus Characteristics | Regular/Consistent | Irregular/Inconsistent | ***p*** |
|---|---|---|---|
| Number of graphemes | 5.4 (1.2) | 5.6 (1.4) | *ns* |
| Log frequency | 1.1 (0.5) | 1.2 (0.7) | *ns* |
| Orthographic neighborhood | 4.5 (4.5) | 4.3 (4.9) | *ns* |
| Phonological neighborhood | 9.1 (8.8) | 10.7 (10.8) | *ns* |
| Familiarity | 534.3 (53.4) | 536.3 (51.1) | *ns* |
| Concreteness | 508.5 (120) | 489.4 (120.1) | *ns* |
| Imageability | 531 (97.7) | 511.9 (80) | *ns* |
| *Monosyllabic (N = 25)* | | | |
| Number of friends | 8.2 (3.9) | 1.1 (1.8) | 0.001 |
| Number of enemies | 0.4 (0.91) | 6.9 (4.6) | 0.001 |
| Summed frequency friends | 514.2 (566.6) | 154.2 (375.4) | 0.01 |
| Summed frequency enemies | 8.2 (26.5) | 657.4 (726.2) | 0.001 |
| Naming response times^ | 602.6 (33.6) | 626.3 (52.3) | 0.063 |
| Naming accuracy^ | 0.99 (0.01) | 0.95 (0.07) | 0.004 |
| Lexical Decision response times^ | 598.2 (34.9) | 626.3 (70.5) | 0.083 |
| Lexical Decision accuracy^ | 0.98 (0.02) | 0.94 (0.08) | 0.048 |
| *Bisyllabic (N = 25)* | | | |
| Number of friends syllable 1 | 8.6 (6.2) | 2.7 (4.4) | 0.001 |
| Number of enemies syllable 1 | 0.6 (1.1) | 3.7 (3.9) | 0.001 |
| Summed frequency friends syllable 1 | 712.8 (969.7) | 599.9 (1171.6) | *ns* |
| Summed frequency enemies syllable 1 | 142.1 (416.1) | 805 (1045.9) | 0.006 |
| Naming response times^ | 630.3 (49.4) | 619.9 (46.7) | *ns* |
| Naming accuracy^ | 0.99 (0.03) | 0.99 (0.02) | *ns* |
| Lexical Decision response times^ | 669.8 (88.9) | 638.5 (77.3) | *ns* |
| Lexical Decision accuracy^ | 0.95 (0.08) | 0.98 (0.03) | 0.074 |

^Values obtained from the American National Corpus norms (Ide & Suderman, 2004).

### *EEG recordings*

The EEG was recorded continuously with 29 tin electrodes mounted on an elastic electrode cap (Electro-Cap International, Eaton, OH). The electrode positions included 11 electrodes placed at standard International 10–20 system locations at left (indicated by odd numbers) and right (indicated by even numbers) hemisphere frontal (F3/F4), central (C3/C4), temporal (T3/T4), parietal (P3/P4) sites, as well as frontal (Fz), central (Cz) and parietal (Pz) midline sites. Ten additional electrodes were placed at the frontal (FPz) and occipital (Oz) poles, the left and right hemisphere fronto-central sites (FC1/FC2, FC5/FC6), and centro-parietal sites (CP1/CP2, CP5/CP6). Another eight electrodes were placed at 33% of the distance from FPz to T3/T4 (FP1'/FP2'), 67% of the distance from FPz to T3/T4 (F7'/ F8'), 33% of the distance from Oz to T3/T4 (O1'/O2'), and 67% of the distance from Oz to T3/T4 (T5'/T6').

Electrodes were also placed over the left (A1) and right (A2) mastoids, with the left mastoid serving as the reference. The right mastoid was recorded to determine if there was any differential mastoid activity. Horizontal eye movements were measured with an electrode lateral to the right eye (HE), and vertical eye movements/ blinks were monitored with an electrode below the left eye (LE). Impedances were reduced to less than 5 kilo-ohms (kΩ) for the scalp and mastoid electrodes and less than 10 kΩ for the eye electrodes. The EEG was amplified by a SA Bioamplifier system with a bandpass of 0.1 to 40 hertz (Hz) and sampled continuously at a rate of 200 Hz. A 15-Hz low pass filter was applied to each individual participant's data prior to grand-averaging the data for presentation. This filtering step was not applied to the data used for the statistical analyses.

### Picture naming

Expressive vocabulary was assessed using a published picture naming task (Hoshino & Kroll, 2008). Participants named 36 black-and-white line drawings in English. Each trial began with a fixation sign, which was replaced with a 500 ms blank screen at the press of a button, followed by a randomly selected line drawing presented until response onset and ended with a 200 ms blank screen. A set of 20 practice trials preceded the experimental set. Picture naming performance was scored in terms of number and average latency of correct response trials.

### Sight word reading

The efficiency of sight word reading was assessed in English using both timed (i.e., speeded) and untimed measures. Speeded sight word reading was assessed in English using the Sight Word Efficiency subtest of the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999), which measured the number of words of increasing difficulty (out of 104) read aloud correctly within 45 seconds. Untimed sight word reading was assessed in English using the Word Identification subtest of the Woodcock Reading Mastery Tests – Revised (WRMT-R; Woodcock, 1987), which measured the number of words of increasing difficulty (out of 105) read aloud correctly.

### Phonemic decoding

Phonemic decoding was assessed in English using both timed (i.e., speeded) and untimed measures. Speeded phonemic decoding was assessed in English using the Phonemic Decoding Efficiency subtest of the TOWRE (Torgesen et al., 1999), which measured the number of nonwords of increasing difficulty (out of 63) read correctly within 45 seconds. Untimed phonemic decoding was assessed in English using the

Word Attack subtests of the WRMT-R (Woodcock, 1987), which measured the number of nonwords of increasing difficulty (out of 45) read correctly.

### *Language exposure and use*

Participants completed a paper-and-pencil language history questionnaire (Li et al., 2020), providing demographic and language experience information, as well as self-ratings of L1 and L2 proficiency. A proficiency self-rating score was computed for each language by averaging self-ratings of listening, speaking, reading, and writing.

### *Data analysis*

Study materials, data, and analysis code are available at https://osf.io/rk3xv/. The study design and analyses were not pre-registered. The analyses were implemented in R version 3.6.0 (R Development Core Team, 2016; http://cran.us.r-project.org/) using the ez package version 4.4-0 (Lawrence, 2016). Behavioral and EEG measures were used to evaluate performance on the lexical decision task. Correct response latencies recorded within the 200–2000 ms time window post-stimulus onset, accuracy rates, and mean amplitudes elicited by the regular/consistent and irregular/inconsistent conditions were analyzed using repeated measures analyses of variance (ANOVA). Trials on which the response was more than 3 standard deviations above participant or item means were removed from the response latency and accuracy analyses, as these very slow responses likely reflected a distinct response strategy and would be confounded by decay of the stimulus from working memory.

The raw EEG signal was time-locked to stimulus onset. Recordings were examined for muscle artifact and eye movements/blinks. Contaminated trials were removed from the analyses, along with incorrect response trials. Artifact rejection involved a two-stage process. The computer initially applied a standard algorithm to reject artifact-filled trials (i.e., trials containing vertical or horizontal eye movements exceeding 1500 A/D units from 100 ms before to 900 ms after word onset), and then, the resulting output was manually checked and validated. To ensure that averages were based on artifact-free data, parameters of the standard algorithm were adjusted when needed. After excluding trials with errors and artifacts, 12.2% of trials from the regular/consistent condition and 14.6% of trials from the irregular/inconsistent condition were rejected. The percentage of rejected trials did not differ significantly between experimental conditions, $t(20) = -1.52, p = .144$. Separate ERPs to items in the critical conditions were averaged offline for each participant at each electrode site, aligned to a 100 ms pre-stimulus baseline (i.e., −100 ms to 0 ms pre-stimulus onset) and ending 900 ms post-stimulus onset. Based on visual inspection of the waveforms and prior reports, the following time windows were selected to evaluate the mean amplitude elicited by regular/consistent relative to irregular/inconsistent words in native English speakers: 150–250 ms (P200) and 250–450 ms (N400). Repeated measures ANOVAs were performed separately for the midline electrode sites (FPz, Fz, Cz, Pz, Oz) and for each of three concentric rings of lateral electrode sites (inner circle: FC1/FC2, C3/C4, CP1/CP2; middle circle: F3/F4, FC5/FC6, CP5/CP6, P3/P4; outer circle: FP1/FP2, F7/F8, T3/T4, T5/T6, O1/O2). Lateral electrode

**Table 3.** Mean (standard error) lexical decision latencies and accuracy rates to English words with regular-consistent versus irregular-inconsistent mappings in English monolinguals

| Measure | Regular-Consistent | Irregular-Inconsistent |
|---|---|---|
| Response Latencies | 595.97 (14.11) | 590.83 (14.93) |
| Response Accuracy | 95.63 (0.0001) | 96.89 (0.001) |

sites also included a factor of hemisphere (left versus right). Analyses were divided into these regions to allow for description of the topographic distribution of the effects, considering the relative positions of the electrodes. Each set of sites allows for comparison of effects from anterior to posterior, and across the sets, effects can be described along the dimension of medial (midline) to lateral (outer circle) (see Guo et al., 2012; Holcomb et al., 2005; West & Holcomb, 2000 for other studies of language processing using this analysis method). For each set of analyses, factors of interest were regularity/consistency, electrode site, and hemisphere (if appropriate).

Only regularity/consistency main effects and interactions of regularity/consistency with the other variables (electrode, hemisphere) were reported, since main effects of electrode site or hemisphere, or interactions of only those terms reflect topographic differences in ERP patterns separate from the conditions that were manipulated in the current experiment. All results are reported as significant at the .05 level, unless otherwise noted. For comparisons with more than two levels, the Greenhouse-Geisser correction for nonsphericity was applied; uncorrected degrees of freedom and corrected *F*- and *p*-values were reported. Significant interactions were followed up with simple-effects analysis, and/or post hoc Bonferroni-corrected *t* tests were appropriate to better characterize the effects. Corrected *p*-values are also reported for the Bonferroni-corrected post hoc *t* tests.

## Results

### *Behavioral results*

Incorrect response trials (representing 5.8% of the data) were first excluded from the response latency data. Then, absolute outliers (trials with responses below 200 ms and above 2000 ms, representing 0.4% of the data) and relative outliers (trials with responses 3 standard deviations slower than each participant or item mean response latency, representing 2.5% of the data) were excluded from the behavioral data. Response latencies and accuracy rates to words with regular/consistent and irregular/inconsistent mappings are shown in Table 3. English monolinguals were equally fast, $F(1, 20) = 0.939$, $p = .344$, and accurate, $F(1, 20) = 1.593$, $p = .221$, in responding to regular/consistent and irregular/inconsistent English words.

### *Electrophysiological results*

Figure 1 shows the ERP waveforms elicited by regular/consistent versus irregular/inconsistent words at all 29 scalp electrode sites. The waveforms elicited during the task showed a series of peaks consistent with previous visual word recognition studies. Specifically, there was a clear negative peak around 100 ms (N100), followed by a

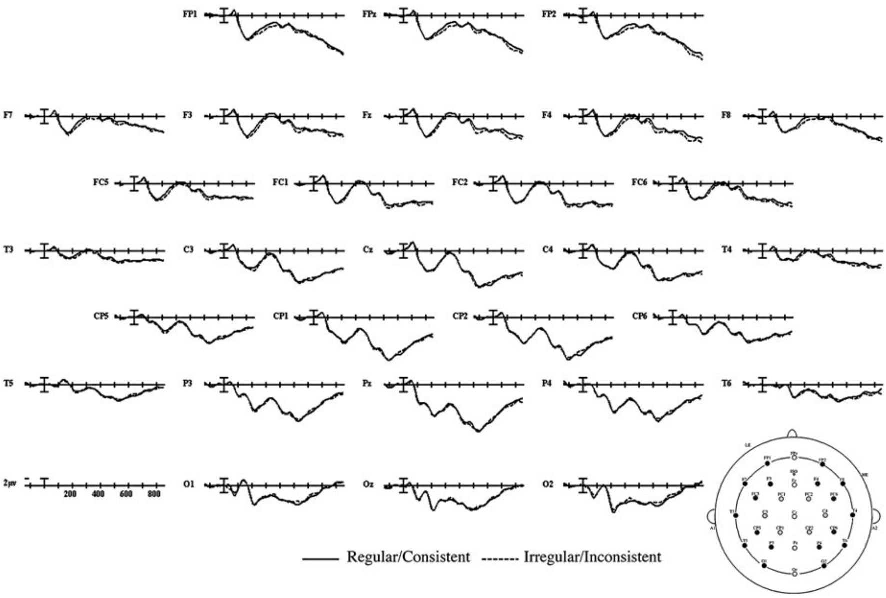

**Figure 1.** Grand Average Event-related Potentials to Regular/Consistent (e.g., GATE) and Irregular/Inconsistent (e.g., PINT) Words in English Monolinguals. Data are Plotted in the −100 to 900 ms Time Window Post-stimulus Onset at all 29 Electrode Sites. Negative is Plotted Up.

positive peak around 200 ms (P200), a negative peak around 400 ms (N400), and a positive peak around 600 ms (LPC). The regularity/consistency effect was notable as a subtle mean amplitude difference between conditions, such that irregular/inconsistent words showed a trend toward a more positive waveform relative to regular/consistent words at some sites in each analysis window. However, the sites that showed the effects varied across components (P200, N400), and the differences between the conditions were small, as supported by the statistical results reported below.

In the 150–250 ms (P200) window, there was a significant interaction between regularity/consistency and hemisphere only over the inner circle: $F(1,\ 20) = 5.057$, $p = .036$. Although post hoc tests failed to reach significance, the pattern of results shows a left hemisphere distribution of the effect over the inner circle, with more positive P200 amplitudes elicited by irregular/inconsistent than regular/consistent words.

In the 250–450 ms (N400) window, there was a trend toward a significant interaction between regularity/consistency and electrode over the middle: $F(3,\ 60) = 3.292$, $p = .066$ and outer circles: $F(4,\ 80) = 2.478$, $p = .098$. Although these interactions failed to reach significance, the pattern of results shows a more negative N400 amplitude elicited by regular/consistent relative to irregular/inconsistent words over frontal (F3/F4, F7/F8) and fronto-central (FC5/FC6) electrode sites.

## Discussion

Experiment 1 examined effects of regularity/consistency in English word recognition in monolinguals. Behavioral measures revealed no differences between regular/

consistent and irregular/inconsistent English words. Electrophysiological measures revealed subtle amplitude differences over the P200 and N400 components, with trends suggesting more positive waveforms to irregular/inconsistent words over the P200 and more negative waveforms to regular/consistent words over the N400.

The pattern of behavioral results observed in monolinguals is consistent with previous behavioral studies that found no effects of spelling-sound regularity and consistency in native English speakers in lexical decision (Coltheart et al., 1979; Seidenberg et al., 1984). The lack of spelling-sound regularity effects in lexical decision has been linked to faster access of the orthographic lexicon relative to the assembly route in proficient readers, such as college-aged students (Besner, 1990). When regularity effects are detected in lexical decision, they have been associated with the presence of low-frequency words with low bigram frequencies in the stimulus list (Seidenberg et al., 1984). The results of the current study are not surprising, considering that participants were college-aged English monolinguals with good reading skills and that the lexical decision task did not contain a manipulation of lexical frequency.

The pattern of results observed in the ERP data (i.e., more positive P200 and less negative N400 amplitudes to irregular/inconsistent relative to regular/consistent words), although not largely significant, replicates the ERP signature previously reported in response to spelling-sound regularity (Yum et al., 2014) and consistency (Hsu et al., 2009; Lee et al., 2007) in Chinese monolinguals. Together with the behavioral results, the ERP measures bring additional support to the argument that college-aged English monolinguals are not reliant on the regularity and consistency of English words in lexical decision. Regularity and consistency effects are also found to be absent in lexical decision in Chinese monolinguals (Yum et al., 2014), suggesting that skilled readers may bypass phonology in lexical decision.

## Summary of results

Experiment 1 revealed that English words with irregular/inconsistent mappings elicited an ERP signature of more positive P200 and less negative N400 amplitudes than words with regular/consistent mappings in English monolinguals performing a lexical decision task. Critically, the effects were short-lived, disappearing before a behavioral response was made.

## Experiment 2

Experiment 2 investigated the time course of processing spelling-sound regularity/consistency in English (a deep orthography) in non-native speakers with distinct L1 orthographic depths (i.e., shallow: Spanish; deep: Chinese). It was predicted that bilinguals transfer distinct grain sizes from the L1 to the L2 depending on L1 orthographic depth. Since Spanish-English bilinguals can successfully rely on small reading units in L1-Spanish, they were predicted to use a small reading unit when reading L2-English words, evidenced by larger P200 and smaller N400 amplitudes for words with irregular/inconsistent mappings. In contrast, since Chinese-English bilinguals can successfully rely on large reading units in L1-Chinese, they were predicted to transfer the preference for a large reading unit when reading L2-English words, evidenced by a difference between regular/consistent and irregular/consistent conditions at the lexical level (N400), but not at the sublexical level (P200). In other words, if bilinguals' transfer word reading strategies from the L1 to the L2 as a function of L1 orthographic depth, or adopt a hybrid grain size as proposed by the grain size accommodation hypothesis (Lallier & Carreiras, 2018), then Chinese-English bilinguals should experience reduced sensitivity to regularity/consistency in L2-English relative to Spanish-English bilinguals. It is also possible that L2 proficiency may drive the observed effects. If L1–L2 transfer occurs at lower levels of L2 proficiency, when bilinguals are not yet attuned to the degree of consistency in the L2 print-to-sound mappings and experience most interference from the L1, then bilinguals with lower English proficiency should transfer L1 reading units to the L2, whereas bilinguals with higher English proficiency should adapt a large grain size to better fit the L2–English, approaching native-like performance. If transfer occurs at higher levels of L2 proficiency, then bilinguals should adopt a distinct reading unit in the L2 at lower levels of L2 proficiency and should transfer their L1 reading unit preference once they have achieved higher levels of L2 proficiency. Alternatively, if bilinguals are observed to differ from the monolingual group in similar ways, regardless of proficiency level, then results may simply reflect general strategy differences in L2 readers.

## Method

### *Participants*

Results are presented for two groups of non-native English speakers recruited from Pennsylvania State University: (1) 22 Spanish-English bilinguals (9 male; mean age = 24.1; range = 18–32 years) and (2) 18 Chinese-English bilinguals (3 male; mean age = 21.6; range = 18–26 years) with no consistent exposure to English before age of 6 and whose L1 was the primary language of instruction in school through the eighth grade and was still their dominant language based on self-ratings and picture naming accuracy measures. However, performance on word reading and phonemic decoding measures revealed that participants were stronger in L2-English as a consequence of living and studying at the university level in an L2-English environment (see Table 4). The two groups differed on L2-English proficiency, with Chinese-English bilinguals performing more poorly than the Spanish-English bilinguals on L2-English measures of word reading, picture naming accuracy, and self-ratings.

A total of 50 bilinguals completed the study, but data from 7 Spanish-English to 3 Chinese-English bilinguals were excluded because they either failed to produce at least 40% correct, artifact-free responses per condition or to maintain at least 60% accuracy for overall "yes" and "no" responses. All of the other participant qualification criteria were identical to those used in Experiment 1.

### *Materials and procedure*

Experiment 2 used the same stimuli, experimental paradigm, EEG equipment/parameters, and exclusion criteria as Experiment 1. The initial testing session

**Table 4.** Mean (standard error) psycholinguistic data in bilinguals

| | Spanish-English Bilinguals | | | Chinese-English Bilinguals | | |
|---|---|---|---|---|---|---|
| Measure | L1-Spanish | L2-English | ***p*** | L1-Chinese | L2-English | ***p*** |
| Sight Word Reading | 79.62 (3.26) | 88.86 (1.84) | .019 | 83.18 (2.74) | 78 (2.51) | *ns* |
| Phonemic Decoding | – | 44.71 (2.27) | – | – | 37.41 (2.16) | – |
| Word Identification | 74.62 (0.50) | 96.52 (0.89) | .001 | 86.94 (3.06) | 92.88 (1.03) | .081 |
| Word Attack | 32.24 (0.38) | 35.67 (1.46) | .033 | – | 35.12 (1.25) | – |
| Picture Naming Latency | 906 (57.54) | 827 (54.15) | *ns* | 767 (63.93) | 788 (60.61) | *ns* |
| Picture Naming Accuracy | 99.18 (0.39) | 86.26 (1.89) | .001 | 95.09 (1.21) | 75.99 (3.13) | .001 |
| Self-Rated Proficiency (/7) | 6.76 (0.11) | 6.05 (0.14) | .001 | 6.93 (0.04) | 5.29 (0.15) | .001 |

was identical to Experiment 1. The only difference was that during the second testing session, behavioral measures of picture naming, sight word reading, and phonemic decoding were not only administered in English (L2), but also in the L1 (Spanish or Chinese). Participants were tested in L1 first and L2 last to avoid dominant language inhibition following weaker-language use (Misra et al., 2012).

### *Picture naming*

Expressive vocabulary in L1 and L2-English was assessed using the same picture naming task described in Experiment 1. Participants were randomly assigned one of two equivalent sets of 36 black-and-white line drawings to name in L1 or L2, respectively. In each language, picture naming performance was scored in terms of number and average latency of correct response trials.

### *Sight word reading*

The efficiency of sight word reading was assessed in L2-English using both timed (i.e., Sight Word Efficiency) and untimed (i.e., Word Identification) measures, as described in Experiment 1. Speeded sight word reading was tested in L1-Spanish using an equivalent measure of the TOWRE (i.e., a list of 104 Spanish words of increasing length) developed by Miller et al. (2006) and in Mandarin Chinese using a list of 92 Chinese nouns increasing in difficulty (T. Guo, personal communication, May 18, 2010). Each version of the task measured the number of words of increasing difficulty read aloud correctly within 45 seconds. Untimed sight word reading was measured in Spanish using the Word Identification subtest of the Pruebas de Aprovechamiento-Revisada (Woodcock et al., 2004), composed of 76 Spanish words of increasing length, and in Mandarin Chinese using a second in-house test developed for the current study, composed of 92 Chinese nouns increasing in difficulty. Each version of the task measured the number of words of increasing difficulty read aloud correctly.

#### *Phonemic decoding*

Phonemic decoding was tested in English only using the Phonemic Decoding subtest of the TOWRE (Torgesen et al., 1999) and the Word Attack subtest of the WRMT-R (Woodcock, 1987) described in Experiment 1.

#### *Language exposure and use*

Participants completed a comprehensive paper-and-pencil language history questionnaire (Li et al., 2020), providing information about L2 learning experience and use. A proficiency self-rating score was computed for each language by averaging self-ratings of listening, speaking, reading, and writing.

### ***Data analysis***

The same artifact-rejection procedure was used as in Experiment 1. After excluding trials with errors and artifacts, 16.3% of trials from the regular/consistent condition and 17.2% of trials from the irregular/inconsistent condition were rejected. The percentage of rejected trials did not differ significantly between experimental conditions, $t(39) = -0.74$, $p = .463$. The following time windows were selected to evaluate the electrophysiological signature of the regularity/consistency effect in bilinguals: 175–300 ms (P200) and 300–450 ms (N400). These windows are slightly later than those used in Experiment 1, but they fit the patterns observed in the bilingual data better, and it is not surprising that non-native English speakers may require slightly longer processing times in the task (e.g., Ivanova & Costa, 2008; Sullivan et al., 2018). Experiment 2 used the same analyses as Experiment 1, except that L1 orthographic depth, operationalized categorically as group (Spanish-English versus Chinese-English bilinguals), and English proficiency, operationalized continuously as Word Attack scores were entered in the analyses as between-subject variables. Data were analyzed using the lme4 package version 1.1–11 (Bates et al., 2016). Regularity/consistency, group, and English proficiency main effects and interactions of regularity/consistency with the other variables (group, English proficiency, electrode, and hemisphere) are reported.

## Results

### ***Behavioral results***

Incorrect response trials (representing 13.5% of the data) were first excluded from the response latency data. Then, absolute outliers (trials with responses below 200 ms and above 2000 ms, representing 0% of the data) and relative outliers (trials with responses 3 standard deviations slower than each participant or item mean response latency, representing 2.9% of the data) were excluded from the behavioral data. Response latencies and accuracy rates to words with regular/consistent and irregular/inconsistent spelling-sound mappings are shown in Table 5. Bilinguals were equally fast, $F(1, 39) = 1.21$, $p = .273$, and accurate, $F(1, 38) = 2.68$, $p = .102$, in responding to regular/consistent and irregular/inconsistent English words. There was a significant main effect of group on response latencies: $F(1, 39) = 6.27$, $p = .016$ and accuracy: $F(1, 39) = 8.08$, $p = .009$. The results indicate that Chinese-English bilinguals were overall slower and less accurate in recognizing English words relative to Spanish-English bilinguals.

**Table 5.** Mean (standard error) lexical decision latencies and accuracy rates to English words with regular-consistent versus irregular-inconsistent mappings in bilinguals

| | Spanish-English Bilinguals | | Chinese-English Bilinguals | |
|---|---|---|---|---|
| Measure | Regular-Consistent | Irregular-Inconsistent | Regular-Consistent | Irregular-Inconsistent |
| Response Latencies | 746.96 (43.19) | 748.30 (44.69) | 919.47 (48.19) | 904.03 (49.85) |
| Response Accuracy | 89.99 (1.32) | 91.60 (1.33) | 83.81 (2.00) | 85.70 (2.12) |

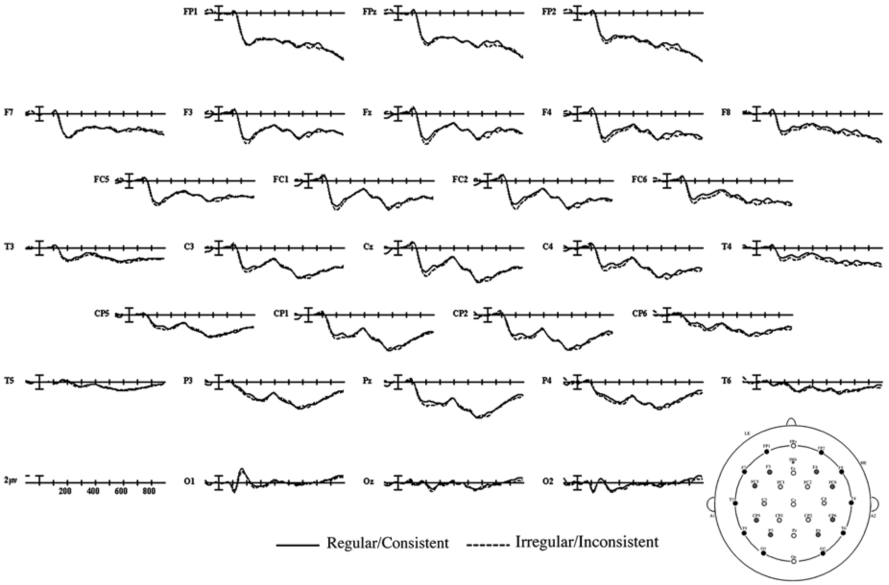

**Figure 2.** Grand Average Event-Related Potentials to Regular/Consistent (e.g., GATE) and Irregular/Inconsistent (e.g., PINT) Words in Spanish-English Bilinguals. Data are Plotted in the −100 to 900 ms Time Window Post-stimulus Onset at all 29 Electrode Sites. Negative is Plotted Up.

### *Electrophysiological results*

Figures 2 and 3 show the ERP waveforms elicited by regular/consistent versus irregular/inconsistent words at all 29 scalp electrode sites in the Spanish-English and Chinese-English bilingual groups, respectively. The waveforms elicited during the task showed a series of peaks consistent with previous visual word recognition studies and Experiment 1 results. Specifically, there was a clear negative peak around 100 ms (N100), followed by a positive peak around 200 ms (P200), a negative peak around 400 ms (N400), and a positive peak around 600 ms (LPC). The regularity/consistency effect was characterized by a larger P200 mean amplitude to irregular/inconsistent words and a pattern of smaller N400 mean amplitudes to irregular/inconsistent words that failed to reach significance. The P200

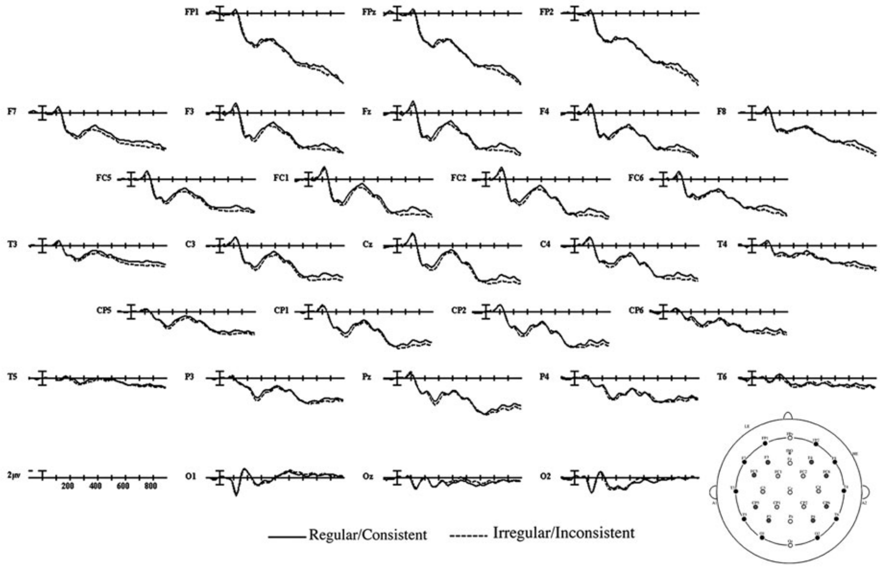

**Figure 3.** Grand Average Event-related Potentials to Regular/Consistent (e.g., GATE) and Irregular/Inconsistent (e.g., PINT) Words in Chinese-English Bilinguals. Data are Plotted in the −100 to 900 ms Time Window Post-stimulus Onset at all 29 Electrode Sites. Negative is Plotted Up.

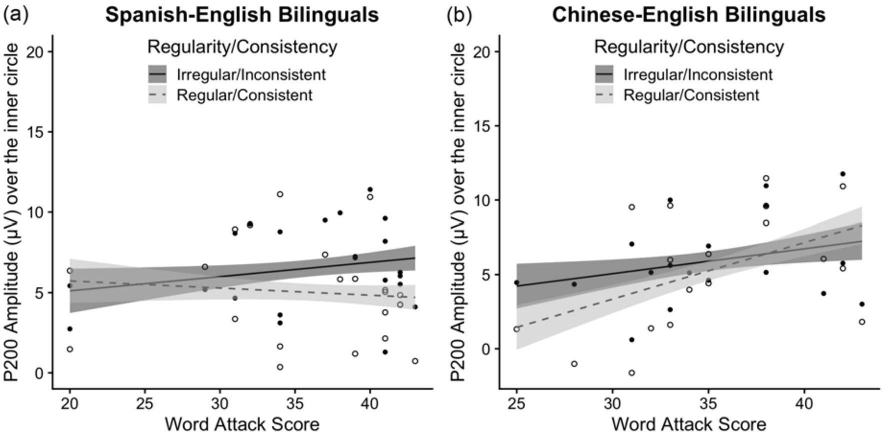

**Figure 4.** Continuous Model Fits of P200 Mean Amplitudes over the Inner Circle to English Words with Irregular/Inconsistent Relative to Regular/Consistent Spelling-Sound Mappings in Spanish-English Bilinguals (Panel A) and Chinese-English Bilinguals (Panel B) as a Function of Word Attack Scores. Shaded Areas indicate Standard Errors.

modulations were larger in Spanish-English bilinguals with high English proficiency (Figure 4A) and Chinese-English bilinguals with low English proficiency (Figure 4B). Differences between the waveforms for the critical conditions are described below.

In the 175–300 ms (P200) time window, there was a significant main effect of regularity/consistency over each set of electrode sites: midline: $F(1, 40) = 8.16$, $p = .027$; inner: $F(1, 40) = 8.43$, $p = .024$; middle: $F(1, 40) = 9.68$, $p = .014$ and outer circles: $F(1, 40) = 8.74, p = .014$. The result reveals that in both bilingual groups, irregular/inconsistent words elicited more positive mean amplitudes relative to regular/consistent words over the P200. Furthermore, there was a significant interaction among regularity/consistency, group, and Word Attack scores over the inner: $F(1, 40) = 7.70$, $p = .033$ and middle circles: $F(1, 40) = 8.08$, $p = .028$ and a trend toward significance over the midline: $F(1, 40) = 6.37, p = .063$ and outer circle: $F(1, 40) = 6.01, p = .061$. Post hoc tests showed that the effect of regularity/consistency over the P200 was larger for Spanish-English bilinguals with high English proficiency over the inner: $F(1, 22) = 3.45$, $p = .077$ and middle circles: $F(1, 22) = 4.64$, $p = .043$. In contrast, the effect of regularity/consistency over the P200 was larger for Chinese-English bilinguals with lower English proficiency over the inner: $F(1, 18) = 4.53$, $p = .047$ and middle circles: $F(1, 18) = 3.80$, $p = .067$.

In the 300–450 ms (N400) time window, the effect of consistency did not reach significance. However, the pattern of results revealed a hint of a more negative N400 amplitude to words with regular/consistent spelling-to-sound mappings.

## Discussion

Experiment 2 examined effects of regularity/consistency in L2-English word recognition in bilinguals who varied in L1 orthographic depth and L2 proficiency. Regularity/consistency effects were not found on response latencies or accuracy data. In the ERP record, irregular/inconsistent words elicited more positive P200 amplitudes than regular/consistent words in both bilingual groups, and these modulations were larger in Spanish-English bilinguals with high English proficiency and in Chinese-English bilinguals with low English proficiency.

The absence of an effect of regularity/consistency on response latencies in bilinguals replicates Experiment 1 results and previous reports in the monolingual literature (Besner, 1990; Coltheart et al., 1979; Seidenberg et al., 1984), suggesting that like monolinguals, bilinguals may be able to extract information about the lexicality of a letter string directly from orthography. However, consistent with previous reports (Sebastian-Gallés et al., 2006), lexical decision in the L2 was overall slower and less accurate for the less proficient group (i.e., Chinese-English bilinguals), likely as a consequence of structural distance and lower exposure-based word frequency (for a review of “the weaker links hypothesis,” see Gollan et al., 2008). The null effect of regularity/consistency observed on response latencies may suggest that behavioral measures are not sensitive enough to identify implicit processing of regularity/consistency in the weaker L2 (McLaughlin et al., 2004).

The more positive P200 amplitudes elicited by irregular/inconsistent relative to regular/consistent words in both bilingual groups replicate the P200 response to regularity previously reported in monolingual speakers of English (Martin et al., 2006; Sereno et al., 1998) and Chinese (Yum et al., 2014) and to consistency previously reported in monolingual speakers of Chinese (Hsu et al., 2009; Lee et al., 2007). Results extend these findings to non-native speakers, replicating the signature of more positive P200 amplitudes to irregular/inconsistent English words. In contrast to previous reports (Hsu et al., 2009; Lee et al., 2007; Yum et al., 2014), the manipulation of spelling-sound consistency and regularity did not modulate the N400 amplitude in bilinguals. This may suggest an overreliance on sublexical strategies in bilinguals reading in the weaker L2 and possibly the absence of a fully developed neighborhood of friend and enemy pronunciations in this language. Additionally, contrary to the prediction, the ERP signature associated with the processing of spelling-sound regularity/consistency did not fundamentally differ between bilingual groups, when the groups were compared as a whole. This result suggests that regardless of L1 orthographic depth, non-native English speakers were sensitive to regularity/consistency in a deep alphabetic L2 orthography and employed a sublexical strategy. However, when bilinguals in each group were divided based on English proficiency (operationalized as Word Attack scores, a measure of phonemic decoding ability), a different pattern of results emerged. Specifically, Spanish-English bilinguals with higher phonemic decoding skills in L2-English produced a larger P200 response to irregular/inconsistent than regular/consistent words in L2-English. In contrast, Chinese-English bilinguals with lower phonemic decoding skills in L2-English produced a larger P200 response to irregular/inconsistent than regular/consistent words in L2-English, suggesting more difficulty in resolving competition among phonological forms (Martin et al., 2006), while more proficient Chinese-English bilinguals did not produce an electrophysiological response to regularity/consistency. This pattern of results may suggest that bilinguals with a shallow L1 orthography approach reading a deep L2 orthography with a larger grain at lower levels of L1 proficiency, but employ a smaller reading unit as their L2 proficiency increases. In contrast, bilinguals with a deep L1 orthography may approach reading a relatively more transparent L2 orthography with a small grain size, but increase their reading unit as their L2 proficiency goes up.

## Summary of results

Experiment 2 revealed that a more positive P200 amplitude was elicited by irregular/inconsistent relative to regular/consistent L2-English words in bilinguals, regardless of differences in L1 orthographic depth. The ERP response was larger in Spanish-English bilinguals with higher L2-English proficiency and in Chinese-English bilinguals with lower L2-English proficiency.

## General discussion

The study investigated the computation of phonology from orthography in native English speakers (Experiment 1) and two groups of non-native English speakers with distinct L1 orthographic depths and L2 proficiency (Experiment 2) for the purpose of determining whether differences in L1 orthographic depth and L2 proficiency modulate bilinguals' transfer of word reading strategies from the dominant L1 to the weaker L2. No differences were found in behavioral data.

However, the ERP record revealed that monolinguals produced more positive P200 and less negative N400 amplitudes in response to irregular/inconsistent relative to regular/consistent L1-English words, whereas bilinguals produced a more positive P200 amplitude in response to irregular/inconsistent relative to regular/consistent L2-English words. The ERP response was more robust in bilinguals, but it did not differ as a function of L1 orthographic depth. Critically, the magnitude of the P200 response was larger in Spanish-English bilinguals with higher L2-English proficiency and in Chinese-English bilinguals with lower L2-English proficiency.

The results of the current investigation support the claim that reading experience and skill modulate sensitivity to spelling-sound regularity and consistency (Stuart & Masterson, 1992; Waters et al., 1984) and extend it to account for differences in English proficiency between native and non-native English speakers and between groups of non-native English speakers who vary in L1 orthographic depth. The study suggests that the ERP response to irregularity/inconsistency in spelling-sound mappings occurs 25–50 ms later in non-native than native English speakers, but may be more robust in non-native speakers than in monolinguals. This finding is supported by studies reporting that the effect of consistency is more robust in groups with lower reading skill (Waters et al., 1984; Weekes et al., 2006). The P200 response has been previously shown to be larger in poorer decoders (Martin et al., 2006), suggesting that the larger P200 response in non-native English speakers may reflect poorer phonemic decoding skills relative to native English speakers (see Tables 1 and 4 for differences in English phonemic decoding and other measures of English proficiency between the monolingual and bilingual groups). Since bilinguals are completing this task in the weaker L2, which is also a deep alphabetic orthography, it is not surprising that they would experience difficulty in selecting among competing phonological forms.

Within the bilingual group as a whole, the ERP response to irregularity/inconsistency did not change as a function of differences in L1 orthographic depth. However, when considering differences in L2-English proficiency at each level of L1 orthographic depth, a different pattern of results emerged. In Chinese-English bilinguals, the magnitude of the P200 response to irregularity/inconsistency was large at low levels of L2-English proficiency, but were nearly absent at high levels of L2-English proficiency. This suggests that Chinese-English bilinguals' approach L2-English word reading with a small reading unit at low levels of L2-English proficiency, possibly as a consequence of structural or script differences between the two languages, which may be linked to the overall lower L2-English proficiency in Chinese-English bilinguals relative to Spanish-English bilinguals, but adopted a larger reading unit at higher levels of proficiency, where their performance was similar to that of English monolinguals. The similarity in performance between Chinese-English bilinguals with higher English proficiency and English monolinguals may be explained by the fact that both groups have a deep L1 orthography. This pattern is also consistent with studies showing evidence that bilinguals' reading strategies shift to better fit the L2 as L2 proficiency increases (Botezatu et al., in press; Meschyan & Hernandez, 2006). Botezatu et al. (in press) found that the magnitude of regularity/consistency in L2-English word naming decreased at higher levels of L2-English proficiency, suggesting that bilinguals who had achieved higher L2 proficiency may have become more efficient in resolving competition from within- and cross-language phonological neighbors (Jared & Kroll, 2001) or may have become more reliant on lexical mediation in the L2 to reduce the effects of cross-language competition (Nosarti et al., 2010). However, Spanish-English bilinguals showed the opposite pattern of performance, producing a P200 response to irregularity/inconsistency at high levels of L2-English proficiency, but not at low levels of L2-English proficiency. This result is inconsistent with the pattern observed in English monolinguals and Chinese-English bilinguals in this study and elsewhere (Botezatu et al., in press; Nosarti et al., 2010; Waters et al., 1984; Weekes et al., 2006) and may suggest that Spanish-English bilinguals are reliant on lexical mediation in the L2 at low levels of L2 proficiency, but employ a smaller reading unit at higher levels of proficiency. It is possible that the presence of Spanish-English cognates in the filler stimuli may have boosted competition from L1 spelling-sound mappings in this group, causing them to rely on lexical mediation at low levels of English proficiency in order to reduce the interference effects from L1 competitors, which should be largest at lower levels of L2 proficiency. However, this pattern of performance is consistent with that reported by Botezatu et al. (submitted), who found that heritage speakers of Spanish showed an increase in the magnitude of the regularity/consistency effect in English word naming as a function of higher Spanish proficiency and related this pattern to bilinguals' frequent code-switching experience, which may be associated with a cooperative (as opposed to competitive) use of the two languages (Beatty-Martínez et al., 2020).

The distinct performance in the two bilingual groups as a function of L1 orthographic depth and L2 proficiency may suggest that bilinguals may use the transparency of their L1 orthography as a reference point for gauging the level of transparency of their L2 orthography. When English proficiency was low, Spanish speakers may have employed a large grain size in L2-English, perhaps expecting a high degree of irregularity/inconsistency, whereas Chinese speakers employed a small grain size in English, apparently expecting regularity/consistency. It should be noted that both reading strategies could be (and were) used effectively. Thus, when L2 proficiency was high, both groups of L1-dominant bilinguals appeared more likely to transfer reading strategies from the L1 to the L2 as appropriate, suggesting that bilinguals gauge the level of transparency found in their L2 orthography before transferring reading strategies from the L1 to the L2. This interpretation is supported by the contrastive analysis hypothesis (Lado, 1957; Stockwell et al., 1965), which postulates that cross-linguistic transfer between the L1 and L2 is modulated by the differences and similarities between the two languages. Differences between the two languages create interference effects leading to errors (i.e., *negative transfer*), whereas similarities create facilitation (i.e., *positive transfer*). The results of the current study may suggest that contrastive reading strategies may be adopted in the L2 when proficiency is low to minimize the effects of interference from the stronger L1, whereas positive transfer occurs when L2 proficiency is high. Consequently, differences between the L1 and L2 spelling-sound mapping systems may be more salient for low-proficiency L2 speakers, whereas similarities between the two systems may become more salient for higher-proficiency L2 speakers.

In conclusion, the present investigation provides evidence that L2 proficiency modulates L1-to-L2 transfer of reading units in L1-dominant bilinguals. The finding that bilinguals with high, but not low, L2 proficiency transfer reading strategies from the L1 to the L2 is important because it suggests that cross-linguistic transfer may be an indicator of high rather than low L2 proficiency. Furthermore, this finding suggests that bilinguals' adaptation of reading strategies to better fit to the L2 implies an understanding of the similarities between the two orthographies and the recruitment of those reading units that are common across the two languages. Additionally, it is important to note that this evidence was revealed by examining the ERP record, while no differences were captured by behavioral measures. This suggests that behavioral measures might not be sensitive enough to identify implicit processing of regularity/consistency in the weaker L2.

### *Limitations and future directions*

The current study has a number of limitations. A relatively small sample of participants was tested and some of the effects observed revealed statistical trends. Future studies should attempt to replicate the current findings with larger participant samples. Additionally, effects of spelling-sound regularity/consistency were evaluated in a lexical decision task, which does not require explicit access to phonology. Future ERP studies should compare effects of regularity/consistency in lexical decision and delayed naming tasks to determine whether effects are larger in tasks that require explicit access to phonology, as reported in behavioral studies. Another limitation is that the lexical decision task included some filler items that were language ambiguous, which may have overtly co-activated Spanish-English bilinguals' L1. Future studies should investigate whether the effects are replicated when the task only includes language-specific stimuli. In the current study, the two bilingual groups differed in both L1 orthographic depth (i.e., shallow: Spanish; deep: Chinese) and script (i.e., alphabetic: Spanish; logographic: Chinese). To determine whether the interaction between bilingual group and proficiency was driven by differences in orthographic depth, not script, future research should compare performance between bilinguals whose L1 writing systems differ on orthographic depth, but not script. Also, given that the current study evaluates performance in two bilingual groups studying in an L2-speaking environment, the pattern of results reported fails to capture changes associated with the initial contact with the L2. Future studies should employ longitudinal designs to evaluate the association between within-subject change in L2 proficiency and L2 word reading strategies adopted by bilingual readers of shallow L1 versus deep L1 orthographies. This would enable researchers to follow changes in L2 word reading approaches from the initial contact with the L2 until high levels of L2 proficiency have been reached, while also considering the effect of L2 immersion and L1 re-immersion experiences.

**Supplementary material.** To view supplementary material for this article, please visit https://doi.org/10.1017/S0142716423000176

**Replication package.** The study materials, data, and code required to replicate the study and all analyses in this article are available at https://osf.io/rk3xv/

**Acknowledgments.** This research was supported by a Doctoral Dissertation Improvement Grant from the National Science Foundation (BCS-1052926) and by a graduate student research endowment from the College of Health and Human Development at Pennsylvania State University. I would like to thank Leah Galluzzi for her help in programming and running a pilot version of the experiment. I would also like to thank Maya Misra for helpful suggestions in the writing of this manuscript.

**Conflicts of interest.** The author declares none.

## References

**Andrews, S.** (1982). Phonological recoding: Is the regularity effect consistent? *Memory and Cognition*, **10**, 565–575. https://doi.org/10.3758/BF03202439

**Baron, J., & Strawson, C.** (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, **2**(3), 386–393. https://doi.org//10.1037//0096-1523.2.3.386

**Bates, D., Maechler, M., Bolker, B., Walker, S., Bojesen Christensen, R. H., Singmann, H., Dai, B., Grothendieck, G., & Green, P.** (2016). *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–11*. Retrieved from https://cran.r-project.org/web/packages/lme4

**Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Tamargo, R. E. G., & Kroll, J. F.** (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. **46**(6), 1022–1047. https://doi.org/10.1037/xlm0000770

**Besner, D.** (1990). Does the reading system need a lexicon. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension Processes in Reading*, (pp. 73–99). Lawrence Erlbaum Associates, Inc.

**Bolger, D. J., Hornickel, J., Cone, N. E., Burman, D. D., & Booth, J. R.** (2008). Neural correlates of orthographic and phonological consistency effects in children. *Human Brain Mapping*, **29**(12), 1416–1429. https://doi.org/10.1002/hbm.20476

**Botezatu, M. R.** (in press). Does MOED rhyme with FRUIT? An event-related potential study of cross-language rhyming. *Neuroreport*.

**Botezatu, M. R., Misra, M., & Kroll, J. F.** (in press). Proficiency in a second language influences processing of print-to-sound mappings. *Linguistic Approaches to Bilingualism*.

**Botezatu, M. R., Peterson, S., & Garcia, D.** (submitted). Consistency matters: How language experience influences spelling-sound competition dynamics in the dominant language. *Frontiers in Psychology*.

**Coltheart, M., Besner, D., Jonasson, J. T., & Davelaar, E.** (1979). Phonological encoding in the lexical decision task. *The Quarterly Journal of Experimental Psychology*, **31**(3), 489–507.

**Coltheart, M., Curtis, B., Atkins, P., & Haller, M.** (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, **100**(4), 589.

**Coltheart, M., & Rastle, K.** (1994). Serial processing in reading aloud: Evidence for dual route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 1197–1211.

**Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J.** (2001). DRC: Route Cascaded model of visual word recognition and reading aloud. *Psychological Review*, **108**, 204–256.

**Coltheart, V., & Leahy, J.** (1992). Children's and adults' reading of nonwords: Effects of regularity and consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**(4), 718.

**Cortese, M. J., & Simpson, G. B.** (2000). Regularity effects in word naming: What are they? *Memory & Cognition*, **28**(8), 1269–1276.

**Durgunoglu, A. Y., & Hancin, B. J.** (1992). An overview of cross-language transfer in bilingual reading. *Advances in Psychology*, **83**, 391–411.

**Glushko, R. J.** (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, **5**(4), 674–691.

**Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C.** (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, **58**(3), 787–814. https://doi.org/10.1016/j.jml.2007.07.001

**Guo, T., Misra, M., Tam, J. W., & Kroll, J. F.** (2012). On the time course of accessing meaning in a second language: An electrophysiological and behavioral investigation of translation recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **38**(5), 1165.

**Holcomb, P. J., Grainger, J., & O'Rourke, T.** (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, **14**, 938–950. https://doi.org/10.1162/089892902760191153

**Holcomb, P. J., Reder, L., Misra, M., & Grainger, J.** (2005). The effects of prime visibility on ERP measures of masked priming. *Cognitive Brain Research*, **24**(1), 155–172.

**Hoshino, N., & Kroll, J. F.** (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, **106**(1), 501–511.

**Hsu, C., Tsai, J., Lee, C., & Tzeng, O. J.** (2009). Orthographic combinability and phonological consistency effects in reading Chinese phonograms: An event-related potential study. *Brain and Language*, **108**, 56–66.

**Hue, C.-W.** (1992). Recognition processes in character naming. In H.-C. E. Chen & O. J. L. Tzeng (Eds.), *Language processing in Chinese. Advances in psychology* (pp. 93–107). North-Holland.

**Ide, N., & Suderman, K.** (2004). The American National Corpus First Release. Language Resources and Evaluation Conference (lrec), Lisbon.

**Ivanova, I., & Costa, A.** (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, **127**(2), 277–288. https://doi.org/10.1016/j.actpsy.2007.06.003

**Jared, D.** (1997). Spelling-sound consistency affects the naming of high-frequency words. *Journal of Memory and Language*, **36**, 505–529.

**Jared, D.** (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language*, **46**, 723–750.

**Jared, D., & Kroll, J. F.** (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language*, **44**(1), 2–31. https://doi.org/10.1006/jmla.2000.2747

**Katz, L., & Frost, R.** (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67–84). Elsevier.

**Kay, J., & Bishop, D.** (1987). Anatomical differences between nose, palm, and foot, or, the body in question: Further dissection of the processes of sub-lexical spelling-sound translation. In M. Coltheart (Ed.), *Attention and Performance XII: The psychology of reading* (pp. 449–469). Erlbaum.

**Kramer, A. F., & Donchin, E.** (1987). Brain potentials as indices of orthographic and phonological interaction during word matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **13**, 76–86. https://doi.org/10.1037//0278-7393.13.1.76

**Kroll, J. F., Bice, K., Botezatu, M. R., & Zirnstein, M.** (2022). On the dynamics of lexical access in two or more languages. In L. Gleitman, A. Papafragou, & J. Trueswell (Eds.), *Handbook of the Mental Lexicon*. (pp. 583–597). Oxford University Press.

**Kutas, M., & Federmeier, K. D.** (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, **62**, 621–647.

**Lacruz, I., & Folk, J. R.** (2004). Feedforward and feedback consistency effects for high-and low-frequency words in lexical decision and naming. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, **57A**, 1261–1284.

**Lado, R.** (1957). *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press.

**Lallier, M., & Carreiras, M.** (2018). Cross-linguistic transfer in bilinguals reading in two alphabetic orthographies: The grain size accommodation hypothesis. *Psychonomic Bulletin & Review*, **25**(1), 386–401.

**Lawrence, M. A.** (2016). *Package 'ez': Easy analysis and visualization of factorial experiments*. https://cran.r-project.org/package = ez.

**Lee, C., Huang, H., Kuo, W., Tsai, J., & Tzeng, J. L. O.** (2010). Cognitive and neural basis of the consistency and lexicality effects in reading Chinese. *Journal of Neurolinguistics*, **23**, 10–27. https://doi.org/10.1016/j.jneuroling.2009.07.003

**Lee, C., Tsai, J., Chan, W., Hsu, C., Hung, D. L., & Tzeng, O. J. L.** (2007). Temporal dynamics of the consistency effect in reading Chinese: An event-related potentials study. *Neuroreport*, **18**, 147–151.

**Lee, C., Tsai, J., Su, E. C., Tzeng, J. L. O., & Hung, D. L.** (2005). Consistency, regularity, and frequency effects in naming Chinese characters. *Language and Linguistics*, **6**, 75–107.

**Lee, C., Tsai, J. L., Huang, H. W., Hung, D. L., & Tzeng, O. J.** (2006). The temporal signatures of semantic and phonological activations for Chinese sublexical processing: an event-related potential study. *Brain Research*, **1121**, 150–159. https://doi.org/10.1016/j.brainres.2006.08.117

**Li, P., Zhang, F., Yu, A., & Zhao, X.** (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, **23**(5), 938–944.

**Liu, I., Chen, S., & Sue, I.** (2003). Regularity and consistency effects in Chinese character naming. *Chinese Journal of Psychology*, **45**, 29–46. http://search.proquest.com/docview/620299941?accountid = 13158

**Liu, I., Wu, J. T., & Chou, T. L.** (1996). Encoding operation and transcoding as the major loci of the frequency effect. *Cognition*, **59**, 149–168. https://doi.org/10.1016/0010-0277(95)00688-5

**Marian, V., & Kaushanskaya, M.** (2004). Mapping written input onto orthographic representations: The case of bilinguals with partially overlapping orthographies. *Proceedings of the Annual Meeting of the Cognitive Science Society* (p. 26).

**Martensen, H., Maris, E., & Dijkstra, T.** (2000). When does inconsistency hurt? On the relation between phonological consistency effects and the reliability of sublexical units. *Memory and Cognition*, **28**, 648–656. https://doi.org/10.3758/BF03201254

**Martin, F. H., Kaine, A., & Kirby, M.** (2006). Event-related brain potentials elicited during word recognition by adult good and poor phonological decoders. *Brain and Language*, **96**, 1–13. https://doi.org/10.1016/j.bandl.2005.04.009

**McLaughlin, J., Osterhout, L., & Kim, A.** (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience*, **7**(7), 703–704. https://doi.org/10.1038/nn1264

**Meschyan, G., & Hernandez, A. E.** (2006). Impact of language proficiency and orthographic transparency on bilingual word reading: An fMRI investigation. *Neuroimage*, **29**, 1135–1140. https://doi.org/10.1016/j.neuroimage.2005.08.055

**Miller, J. F., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L., & Francis, D. J.** (2006). Oral language and reading in bilingual children. *Learning Disabilities Research & Practice*, **21**, 30–41. https://doi.org/10.1111/j.1540-5826.2006.00205.x

**Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F.** (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language*, **67**(1), 224–237. https://doi.org/10.1016/j.jml.2012.05.001

**Mumtaz, S., & Humphreys, G. W.** (2001). The effects of bilingualism on learning to read English: Evidence from the contrast between Urdu-English bilingual and English monolingual children. *Journal of Research in Reading*, **24**, 113–134. https://doi.org/10.1111/1467-9817.t01-1-00136

**Nosarti, C., Mechelli, A., Green, D. W., & Price, C. J.** (2010). The impact of second language learning on semantic and nonsemantic first language reading. *Cerebral Cortex*, **20**(2), 315–327. https://doi.org/10.1093/cercor/bhp101

**Perfetti, C. A., & Liu, Y.** (2005). Orthography to phonology and meaning: Comparisons across and within writing systems. *Reading and Writing*, **18**, 193–210.

**Perfetti, C. A., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., & Tan, L.** (2007). Reading in two writing systems: Accommodation and assimilation of the brain's reading network. *Bilingualism: Language and Cognition*, **10**, 131–146. https://doi.org/10.1017/S1366728907002891

**Perfetti, C. A., & Tan, L. H.** (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24**, 101–118. https://doi.org/10.1037//0278-7393.24.1.101

**Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K.** (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, **103**(1), 56.

**Psychology Software Tools Incorporated**. (2012). E-Prime. Retrieved from http://www.pstnet.com

**R Development Core Team**. (2016). *R: A language and environment for statistical computing.* In R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

**Schmalz, X., Marinus, E., & Castles, A.** (2013). Phonological decoding or direct access? Regularity effects in lexical decisions of grade 3 and 4 children. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, **66**, 338–346.

**Schwartz, A. I., Kroll, J. F., & Diaz, M.** (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes*, **22**(1), 106–129. https://doi.org/10.1080/01690960500463920

**Sebastian-Gallés, N., Rodríguez-Fornells, A., de Diego-Balaguer, R., & Díaz, B.** (2006). First-and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience*, **18**(8), 1277–1291.

**Seidenberg, M. S., & McClelland, J.** (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, **96**, 523–568. https://doi.org/10.1037//0033-295X.96.4.523

**Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K.** (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, **23**(3), 383–404. https://doi.org/10.1016/s0022-5371(84)90270-6

**Sereno, S. C., Rayner, K., & Posner, M. I.** (1998). Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport*, **9**, 2195–2200.

**Seymour, P. H. K., Aro, M., & Erskine, J. M.** (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, **94**, 143–174. https://doi.org/10.1348/000712603321661859

**Shu, H., Chen, X., Anderson, R. C., Wu, N., & Xuan, Y.** (2003). Properties of school Chinese: Implications for learning to read. *Child Development*, **74**, 27–47. https://doi.org/10.1111/1467-8624.00519

**Stanovich, K. E., & Bauer, D. W.** (1978). Experiments on the spelling-to-sound regularity effect in word recognition. *Memory and Cognition*, **6**(4), 410–415.

**Stockwell, R. P., Bowen, J. D., & Martin, J. W.** (1965). *The grammatical structures of English and Spanish* (Vol. **4**). University of Chicago Press.

**Stone, G. O., Vanhoy, M., & Van Orden, G. C.** (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language*, **36**, 337–359. https://doi.org/10.1006/jmla.1996.2487

**Stuart, M., & Masterson, J.** (1992). Patterns of reading and spelling in 10-year-old children related to prereading phonological abilities. *Journal of Experimental Child Psychology*, **54**, 168–187. https://doi.org/10.1016/0022-0965(92)90034-4

**Sullivan, M. D., Poarch, G. J., & Bialystok, E.** (2018). Why is lexical retrieval slower for bilinguals? Evidence from picture naming. *Bilingualism: Language and Cognition*, **21**(3), 479–488.

**Tan, L. H., Spinks, J. A., Feng, C., Siok, W. T., Perfetti, C. A., Xiong, J., Fox, P. T., & Gao, J.** (2003). Neural systems of second language reading are shaped by native language. *Human Brain Mapping*, **18**, 158–166. https://doi.org/10.1002/hbm.10089

**Tokowicz, N., & MacWhinney, B.** (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An Event-Related Potential investigation. *Studies in Second Language Acquisition*, **27**(02), 173–204.

**Torgesen, J. K., Wagner, R. K., & Rashotte, C. A.** (1999). *Test of word reading efficiency*. PRO-ED.

**Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D.** (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, **124**, 107–136. https://doi.org/10.1037/0096-3445.124.2.107

**Van Leerdam, M., Bosman, A. M., & de Groot, A. M.** (2009). When MOOD rhymes with ROAD: Dynamics of phonological coding in bilingual visual word perception. *The Mental Lexicon*, **4**(3), 303–335.

**Wang, M., Perfetti, C. A., & Liu, Y.** (2005). Chinese–English biliteracy acquisition: Cross-language and writing system transfer. *Cognition*, **97**, 67–88. https://doi.org/10.1016/j.cognition.2004.10.001

**Waters, G. S., Seidenberg, M. S., & Bruck, M.** (1984). Children's and adults' use of spelling-sound information in three reading tasks. *Memory and Cognition*, **12**, 293–305.

**Weekes, B. S., Castles, A. E., & Davies, R. A.** (2006). Effects of consistency and age of acquisition on reading and spelling among developing readers. *Reading and Writing*, **19**, 133–169. https://doi.org/10.1007/s11145-005-2032-6

**West, W. C., & Holcomb, P. J.** (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience*, **12**(6), 1024–1037.

**Woodcock, R. W.** (1987). *The Woodcock reading mastery tests—Revised*. American Guidance Service.

**Woodcock, R. W., Muñoz-Sandoval, A. F., McGrew, K. S., & Mather, N.** (2004). *Batería III Woodcock-Muñoz*. Riverside Publishing.

**Yum, Y. N., Law, S.-P., Su, I.-F., Lau, K.-Y. D., & Mo, K. N.** (2014). An ERP study of effects of regularity and consistency in delayed naming and lexicality judgment in a logographic writing system. *Frontiers in Psychology*, **5**, 315.

**Zhang, S., Perfetti, C. A., & Yang, H.** (1999). Whole word, frequency-general phonology in semantic processing of Chinese characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **25**, 858–875. https://doi.org/10.1037//0278-7393.25.4.858

**Ziegler, J. C., & Goswami, U.** (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, **131**(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

**Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M.** (2001). Identical words are read differently in different languages. *Psychological Science*, **12**(5), 379–384. https://doi.org/10.1111/1467-9280.00370

**Ziegler, J. C., Stone, G. O., & Jacobs, A. M.** (1997). What is the pronunciation for –ough and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, & Computers*, **29**, 600–618. https://doi.org/10.3758/BF03210615

---

**Cite this article:** Botezatu, MR. (2023). The impact of L1 orthographic depth and L2 proficiency on mapping orthography to phonology in L2-English: an ERP investigation. *Applied Psycholinguistics* **44**, 237–263. https://doi.org/10.1017/S0142716423000176