Contrasting L1 profiles through early orthographic and phonological processing of written English

Gary Dicks

A thesis submitted in partial fulfilment of the requirements of
Sheffield Hallam University
for the degree of Doctor of Philosophy

August 2022

Candidate Declaration

I hereby declare that:

1. I have not been enrolled for another award of the University, or other academic or professional organisation, whilst undertaking my research degree.
2. None of the material contained in the thesis has been used in any other submission for an academic award.
3. I am aware of and understand the University's policy on plagiarism and certify that this thesis is my own work. The use of all published or other sources of material consulted have been properly and fully acknowledged.
4. The work undertaken towards the thesis has been conducted in accordance with the SHU Principles of Integrity in Research and the SHU Research Ethics Policy.
5. The word count of the thesis is 84998.

Name	*Gary Dicks*
Date	*August 2022*
Award	*PhD*
Faculty	*College of Social Sciences and Arts*
Director(s) of Studies	*Dr Jane Morgan*

Abstract

Reading is fundamental to everyday life and can appear automatic and effortless, but this seemingly simple skill involves complex interactions between neural networks within ~200ms of seeing a word. While a wealth of work has investigated when and how orthography and phonology interact during visual word recognition (VWR), much is still unclear about such early native language (L1) processing and significantly more so when considering reading a second language (L2).

To investigate this, monolingual English participants (native, alphabetic L1), late bilingual Spanish-English participants (non-native, alphabetic L1), and late bilingual Chinese-English participants (non-native, non-alphabetic L1) read English words, pseudohomophones, and pseudowords in orthographic and phonological lexical decision tasks and real English words in an orthogonal rhyme recognition task. Behavioural performance was evaluated alongside event-related potential (ERP) analysis of occipital, occipitotemporal, and frontal-central activity in the ~100ms post-stimulus timeframe (i.e., P1-O, P1-OT, and N100-FC components, respectively) as well as the ~170ms timeframe at occipitotemporal sites (N170-OT) following documented associations with early orthographic and/or phonological processing. Crucially, interpretations of analyses took the markedly different language profiles into account to examine how L1 might influence L2 processing.

Patterns of behaviour and electrophysiology showed similarities but distinguished between groups based on language profiles. Analysis highlighted evidence for early orthography-phonology mapping at occipitotemporal sites, parallel orthographic/phonological processing at frontal-central sites at ~100ms, and a lack of VWR-related occipitotemporal N170 effects across groups. Findings, overall, suggest that language profiles underlie early orthographic and phonological processing and that the three groups have distinct neural strategies for VWR linked to orthographic and phonological aspects of their language profiles. Ultimately, the outcomes support the importance of L1 properties and L1-L2 relationships (i.e., language profiles) in bilingual VWR and the increased consideration of these factors in future development of monolingual and bilingual theories and models of VWR.

Acknowledgements

Just saying "thank you" doesn't seem enough, but, first and foremost, thank you to Dr Jane Morgan and Dr Lynne Spackman, who got me back on track and continued to believe I could finish this thing throughout the numerous occasions I was not so sure. I would also like to thank, albeit from a time that seems many moons ago, Dr Anna Maria Di Betta for all her support and advice.

Speaking of support, Christy Bannister of the Graduate School should be given an award (and/or pay rise) and I will be forever grateful to her as the rock of university administration and postgraduate pastoral care.

Although he will probably never see this, I thank Mr Matthew David Rowe for the frequent feedback on my attempts at academic "work" as well as the too-often-irregular musical meetings.

Last and whatever the opposite of least is, Ms Audrey Maryse Valerie Laurent, who has been with me since the beginning and stuck around for some *unknown reason* (possibly just research purposes) – thank you, I don't know where I'd be right now without you.

Table of Contents

Table of Figures

Glossary (A-Z)

alphabetic L1	native language with an alphabetic writing system (as for Spanish-English group)
BIA/BIA+	Bilingual Interactive Activation (+) model
BIAM	Bimodal Interactive Activation Model
DRC	Dual Route Cascaded Model of Reading Aloud
EEG	electroencephalography
ERP(s)	event-related potential(s); averaged waveforms extrapolated from EEG
ESL	English as a second language
L1	native/first language e.g., English, Spanish, and Chinese, respectively, in the three experimental groups
L2	second language e.g., English for Spanish-English and Chinese-English bilinguals
lexical decision	the determination of whether a stimulus is a real word
non-alphabetic L1	native/first language with a writing system that is not alphabetic (logographic as for the Chinese-English group)
orthography	written form of a language, including writing system and rules
orthographic lexicality	distinction between real words and pseudohomophones and its effect(s)
phonology	sound form of a language, including phonotactic rules
phonological lexicality	distinction between pseudowords and pseudohomophones and its effect(s)
pseudohomophone	linguistic stimulus using legal but not legitimate orthography with legitimate legal phonology of the language e.g., *foan*
pseudoword	linguistic stimulus using legal but not legitimate orthography and phonology of the target language e.g., *fabe*
rhyme recognition	the determination of whether stimuli share a phonological rime
VWR	visual word recognition

Thesis overview

Taken for granted by the literate, the ability to read and, moreover, communicate using only lines and dots is arguably more vital today than it has ever been (Cornelissen et al., 2010). From books and bills to street signs and social media, how brains succeed so rapidly, efficiently, and consistently in finding the sounds (i.e., phonology) and meanings (i.e., semantics) associated with such markings (i.e., orthography) is, however, still not well understood (Dennis & Key, 2006). Furthermore, while there is some understanding of the roles and relative influence of orthography and phonology in reading a first/native language (L1), especially English, it is substantially less so for second language (L2) reading of English and other languages.

The overarching theme of this thesis is the influence of L1 on L2 processing in bilingual ESL (English as a second language) readers, specifically in terms of behavioural performance and the early (pre-200ms) event-related potential (ERP) timeframe of visual word recognition (VWR) processing. Such research includes exploration into the nature of orthographic and phonological processing, controversially early (~100ms) phonological activation, and the role of the N170 ERP component in VWR. The objective of the research documented in this thesis, therefore, is twofold: to investigate the nature of orthographic and phonological processing during VWR of English (in terms of behaviour and electrophysiology) and to examine any similarities and differences of VWR processing between skilled readers of English with different L1 profiles. This will be operationalized through a novel pairing of complementary perspectives (both discrete and combined orthographic and phonological tasks: orthographic/phonological lexical decision tasks and rhyme recognition task, respectively), contrasting samples of native alphabetic (English monolingual), non-native alphabetic L1 (Spanish-English late bilingual), and non-native non-alphabetic L1 (Chinese-English late bilingual) skilled readers of English.

Alongside adding to the existing understanding that will be outlined and discussed in the opening chapters, the original contributions to knowledge of this research will concern how early orthography- and phonology-related brain activity manifests within ~200ms of seeing a word within and between these populations. Crucially, this contrast extends to how L1 language profiles might impact early VWR-related brain activity within ~200ms to influence L2 reading performance. Analysis will, therefore, compare experimental manipulations within and between groups to evaluate orthographic and phonological processing through both behavioural measures (accuracy and response times) and electrophysiological measures of brain activity: ERP component latencies and amplitudes at ~100ms and ~170ms. Furthermore, the direct contrast between Spanish-English and Chinese-English late bilinguals as two groups of skilled ESL readers with fundamentally different language profiles offers another novel aspect when considered with the particularly early timeframe under investigation (pre-200ms).

The theoretical basis for this research centres on the naturally rapid rate of VWR processing and how quickly orthography and phonology can influence brain and behaviour. This involves the order and interactions of such brain responses in terms of serial or parallel processing, as found in such VWR theories as the Dual Route Cascaded model (DRC; e.g., Coltheart et al., 2001), the Bi-modal Interactive Activation Model (BIAM; e.g., Grainger & Holcomb, 2009), and parallel distributed processing (PDP) connectionist and "triangle" models (e.g., McClelland & Rogers, 2003). While these theories and much research has focused on monolingual/L1 processing, much less is understood about skilled L2 readers compared with native readers of English or, indeed, of other languages (e.g., Spanish, Chinese). The question of how a language, specifically English in the current research, is read by skilled L2 readers, therefore, is still very much open. This is especially the case when considering the ways that L1 and its properties can differ from L2, even fundamentally (e.g., writing system differences).

The research was designed on stimulus, task, and study levels to investigate when orthographic and phonological processing occurs for different linguistic stimuli (words, pseudohomophones, pseudowords, sight and sound rhymes) in variants of lexical decision task (LDT) and rhyme recognition task (RRT), along with any differences between groups. The work is novel in its contrast between three groups not previously compared in such a way in terms of experimental tasks, the perspective of the specific language profiles of the groups, and, importantly, the especially early timeframe of brain activity.

Due to a core focus of the current research being the timing of reading-related behavioural and brain responses and the high temporal resolution of EEG (electroencephalography), an ERP methodology was used for the experimental studies with simultaneous behavioural measures. Both within- and between-group comparisons were made using groups of skilled readers with different language profiles: English monolingual (native alphabetic), Spanish-English late bilinguals (alphabetic L1 and L2), and Chinese-English late bilinguals (non-alphabetic L1, alphabetic L2). In more specific terms of electrophysiology and ERPs, the research aimed to serve several key psycholinguistic inquiries concerning the largest and typically most prominent pre-200ms components in visual ERPs: the occipital and occipitotemporal P1 (positive deflection, peaking at ~100ms) and the occipitotemporal N170 (negative deflection, peaking at ~170ms). From a psycholinguistic perspective, the P1 is predominantly associated with initial visual and orthographic processing with possible links to lexicality (Dien, 2009). The N170 is also predominantly associated with visual and orthographic processing, having ties with the so-called visual word form area (Brem et al., 2009; Maurer, Brandeis, et al., 2005), but in a more deeply linguistic way than the P1, involving sublexical-lexical (i.e., letter-to-word) processing (Holcomb & Grainger, 2006), while also reflecting the foundation of an interface with phonological processes and frontal activity (Cornelissen et al., 2009; Grainger & Holcomb, 2009). ERP activity at ~100ms at frontal-central sites is also of major importance, as frontal and occipitotemporal activity at the same early ~100ms timeframe has been somewhat controversially connected with phonological activation (e.g., Klein et al., 2015; Wheat et al., 2010), which fuels the debate of how parallel the processing of orthography and phonology is during VWR. Overarching all these queries is the extent of any difference between native and ESL groups in terms of the timing and nature of any experimental effects, especially where native language profiles are relevant.

Overview of thesis structure

The theoretical background of the current research, based on both monolingual and bilingual VWR theories with evidence from previous research, will start with an overview of written language types and how they can differ orthographically and phonologically in terms of language profiles. These will be the foundations for subsequent evaluation of behavioural and ERP evidence pertaining to the timing and nature of orthographic and phonological processing when reading English before turning to a chronological account of the first ~200ms post-stimulus ERP evidence to complete the rationale. Unless pertinent, therefore, studies of later ERP components (e.g., N400), syntactic processing, sentence comprehension, or languages unrelated to the current study (e.g., Russian) are not included.

Core methodological decisions and associated issues will be outlined and discussed, using a framework of standard experimental methods (participants, design, stimuli, apparatus, and procedure). Based on these evaluations and in accordance with the overall theoretical rationale, Study 1 (employing orthographic and phonological lexical decision tasks) and Study 2 (using orthographic and phonological priming via a rhyme judgement task) will then be reported separately, forming the basis of the empirical investigation. Each study chapter will include its own literature review and method sections before, importantly, the results and discussion of findings with details specific to their respective designs within the broader area of the research topics. The thesis will close with a discussion of findings from both studies in terms of the overall aims and hypotheses, previous findings, and background literature, explaining how they contribute to current knowledge.

Chapter 1: Visual word recognition and language profiles

1.1 Language profiles, writing systems, and language types

Characteristics of a language, including its writing system and phonotactic rules, affect orthographic and phonological processing, but the extent of such influence from L1 on L2 VWR in bilinguals is largely unknown (K. I. Martin, 2017; Poeppel & Idsardi, 2022). Orthographic and phonological processing of an L2 is underpinned by the language profile (i.e., linguistic background) of the reader amongst other factors e.g., age, experience, proficiency (Yeong et al., 2014). Ultimately, the mechanisms used for reading are highly dependent on the different cognitive requirements of different writing systems (Tan et al., 2005). Consequently, language profiles are likely to have major implications for reader performance and responses, being vital for aiding interpretations of research outputs, such as the linguistic properties of Spanish and Chinese for the current research. While the current research was not designed to compare L1 and L2 reading performance within bilinguals (e.g., Spanish vs English in Spanish-English bilinguals), understanding how L1 reading skills could contribute to L2 VWR is an important and oft-overlooked factor for understanding L2 processing that serves as a main perspective of the current work. Furthermore, properties of Spanish and Chinese will help to explain any findings in terms of the reliance on orthographic and phonological processing that may relate to the ESL groups' respective native languages (Nelson et al., 2009). It is, therefore, vital to consider the extent that L2 reading uses cognitive strategies developed for L1, especially if the L1 approach facilitates or conflicts with L2 in any way, which is not unlikely when bearing in mind the various ways that languages can differ.

According to *Ethnologue: Languages of the World* (Eberhard et al., 2021), there are 7,139 living languages in the world, over half with some extent of writing system and all of which can be defined in linguistic terms. Written natural languages can differ in few or many ways, from surface features of their physical representations to deeper factors of syntax and pragmatics (Crystal, 2010; Yule, 1996). One of the main overarching factors for differences between written languages is the fundamental visual appearance of the words, which often relate to different writing systems. For instance, alphabetic languages all use a form of alphabet, whether a variant of Latin, Cyrillic, or other. This provides a finite set of symbols that permit a nigh-infinite combination of letter strings to be used as words. Logographic languages, meanwhile, use types of ideogram (logographs) that also allow a nigh-infinite number of characters and with a significantly broader set of components to do so. There are also syllabaries (as used in Japanese kana), Arabic (as used in Arabic), and syllabic alphabets (as used in Hindi), along with other less common writing systems, all of which have distinguishing features, but are not directly pertinent to the current research.

Along with the visual style, the direction of writing is another fundamental distinction between writing systems, which is more explicit in some systems than others and often requires language-specific knowledge. Alphabetic systems are left-to right and Arabic is right-to-left, but syllabic and logographic are more language-specific. Some syllabic or logographic languages, such as Japanese kana (syllabary) and Chinese (logographic), can be written vertically top-bottom or horizontally either way, but left-to-right has become more typical in recent years, especially following the mandate for left-to-right by the Chinese government in 1955 (Norman, 1988, p. 80). Therefore, for the language profiles in focus in the current research, this is one broad similarity.

From these visual/orthographic examples, written languages from different writing systems can appear to have little in common, while languages using the same writing system can also appear distinctly different due to different orthographies and scripts. It is important to note that the terms ‘writing system’ and ‘orthography’ in this context are often inaccurately used interchangeably, but ‘orthography’ should instead be considered a hyponym of ‘writing system’ in that orthography is a part of a writing system. Unlike other (conceptual) systems that convey meaning and information (e.g., mathematics, music), writing systems are specifically for language, but are also independent of language in that one writing system can cover multiple languages or one language can have multiple writing systems. English and Spanish, for instance, share a writing system, but have different orthographies that reflect differences within that writing system, including orthography-phonology mapping and the way graphemes (the smallest meaningful orthographic units of a writing system) map onto phonemes (the smallest meaningful phonological units in a language that differentiate between words). English and Chinese also differ in orthographies, but this is due to the fundamental differences in writing system (alphabetic vs logographic). Orthography is just one implementation of a writing system and different orthographies can exist for the same writing system. For instance, English and Spanish are both alphabetic, but have slightly different orthographies – Spanish uses the same Latin alphabetic, but with diacritics and the ñ character. More specifically, orthography includes details of what is termed the script, such as the spelling, punctuation, diacritics, and capitalization, which concerns the classifications of the characters used in the specific implementation of that orthography, e.g., Latin and Cyrillic alphabets, Simplified and Traditional Chinese scripts (Coulmas, 2003, p. 35).

The fundamental linguistic architectures and the core symbols used in alphabetic and logographic languages still differ dramatically, as is clear with alphabetic and syllabic systems using letters or characters to form graphemes that map onto phonemes, while logographic systems use visual characters to represent concepts. This can be illustrated with the numbers of letters, graphemes, and phonemes of the distinct orthographic and phonological profiles of English, Spanish, and Chinese. Only 26 alphabetic letters (ISO basic Latin alphabet) are required to produce over 300,000 entries in the Oxford English Dictionary (Pas et al., 2016), created with around 250 (possibly up to 1120) graphemes mapped to 40-45 phonemes (Joshi & Carreker, 2009, p. 121; Paulesu et al., 2000). Spanish uses the same Latin script as English with a few amendments to have 27 letters/digraphs with optional diacritics for 22-24 phonemes. Chinese, however, has 29 phonemes for upwards of 4762 frequent Chinese characters, as indicated in publications by the Hong Kong Education Bureau (e.g., Perfetti et al., 2012), while each Chinese syllable (vowel with optional consonants) also has one of four tones that gives Chinese an additional dimension of complexity. The distinctions and intricacies of orthography and phonology within and between these languages, however, go much deeper than the sets of symbols and sounds, which are simply used here to show the extent of the complexity. It is such patterns, interactions, and dependencies on L1 while reading L2 (English, in this case) that the current research is specifically investigating, including how L1 might influence L2 VWR and how it might differ based on L1 language profile (i.e., Spanish-English with an alphabetic L1 and Chinese-English with a logographic non-alphabetic L1).

1.2 Visual word recognition and reading

Visual word recognition (VWR) research investigates the cognitive processing involved in reading, focusing on understanding the ability to identify, interpret, and comprehend visually-presented orthographic representations of words (Frost, 1998; Rastle & Brysbaert, 2006). However, there is an important distinction to be made between VWR and reading in the context of research. Broadly, reading involves the processing of psycholinguistic properties of written language that fall into categories of orthography (visual forms), including morphology (rules of how written words are constructed), phonetics/phonology (rules underlying the sounds of words), semantics/pragmatics (meanings, context, and nonverbal cues), and syntax (rules of how sentences are structured), which can be considered the pillars of language from a linguistic standpoint (Yule, 1996). While the brain may not frame the properties of language in such categories, likely employing more “statistical" methods instead (Dehaene, 2014), these are useful metalinguistic tools that help to explain the processes at work when reading. The vital point is that these universal linguistic elements must co-operate for the language to work, much like how the psycholinguistic processes that oversee the different elements of language (orthography, phonology, semantics, syntax) must co-operate for successful language comprehension (Grainger & Holcomb, 2009). Ultimately, visual word recognition is at the core of reading, but does not necessarily require or use all psycholinguistic factors mentioned above and can be achieved with a lesser depth of processing. For instance, recognizing the type of words used in the current research, such as pseudohomophones (e.g., *cair-care*), orthographically-similar non-rhymes (e.g., *bead-dead*; *cough-dough*), and even real words as being such, may not require full access to meaning, requiring only phonological activation or even only orthographic recognition in the case of real words (Coltheart et al., 2001; McNorgan et al., 2015). Only a subset of psycholinguistic factors, (e.g., orthography in direct whole-word reading), therefore, can be sufficient to successfully recognize some linguistic stimuli, distinguishing VWR from reading in general. That said, to reduce repetition, "reading" will sometimes be used interchangeably with "VWR" henceforth.

Another important element of VWR research is the different types of linguistic stimuli used, the core of which being real words, followed by pseudowords (lexical items that are made up to look and sound like real words but are not e.g., *worb*), pseudohomophones (lexical items that have the same sound as real words but are spelt differently e.g., *werd*), and non-words (strings of letters that are not word-like and cannot be pronounced without some creativity e.g., *gtvrykb*). Real words can be defined as linguistic constructs with orthographic, phonological, semantic, and other linguistic properties that are both *legal* and *legitimate* for the respective language. In this context, legality reflects usage of the morphosyntactic and phonotactic rules and dictionary definitions of the relevant language, while legitimacy entails actual existence of the item in the language (Hauk et al., 2012). Different combinations of legal and legitimate orthography, phonology, and semantics result in different stimulus types, such as pseudowords and pseudohomophones. These are commonly used in psycholinguistic research as contrasts for real words in such exercises as lexical decision tasks (LDTs), where stimuli (real words and another stimulus type that are not true real words e.g., pseudowords) are presented one at a time, pseudorandomly, and participants indicate whether each is a real word in the target language or not.

Pseudowords, as the term implies, are essentially "fake" words (e.g., *yerg*) in that they have no direct legitimate semantic associations and only legal orthography and phonology (McNorgan et al., 2015). Pseudowords, therefore, look and sound like they could be real words, being visually, orthographically, and phonologically similar to some extent, but lacking the crucial linguistic legitimacy and semantic content (Hauk et al., 2012). Crucially, it is phonology that distinguishes pseudowords from pseudohomophones (e.g., *werd*, *foan*), which sit between real words and pseudowords with legal and legitimate phonology and semantics, but orthography that is only legal, not legitimate. This results in stimuli that look like pseudowords (and, therefore, in turn like low frequency real words), but have the same sound as real words, allowing them access to the associated semantic content. Put another way, pseudohomophones have incongruent orthography, but congruent phonology in relation to the associated legitimate lexical item e.g., the oft-used *brane* example is phonologically congruent with *brain* but orthography incongruent with it. Lastly, for completeness, true non-words (as opposed to pseudowords) can have elements of legal and legitimate sublexical orthography, such as a common bigram in the respective language e.g., -or-, but otherwise violate the linguistic rules of the language in that they have no legal or legitimate phonology or semantic associations e.g., *ybqor*. Non-words, therefore, come in the form of letter-strings and technically any other stimulus that is not pronounceable and has no meaning.

As stated earlier, reading (in a more general sense) involves all psycholinguistic factors (albeit to varying extents in different situations) and is a fuller process. This is stated again to clarify that the literature review and the current research will focus not on the broader concept of reading per se, but on theories of visual word recognition and predominantly the roles and precise timing of orthographic and phonological processes.

1.3 The importance of timing in VWR

Reading requires complex computations across neurophysiological time and space (Hauk, Davis, et al., 2006), involving dynamic interactions between different neural networks dealing with different aspects of language (Cornelissen et al., 2010; Huettig & Ferreira, 2022; Poeppel & Idsardi, 2022). Although widely accepted to happen within around half a second (Grainger & Holcomb, 2009; Harley, 2010), various evidence suggests it to be much faster. There is a multitude of psycholinguistic factors that can influence the process of word recognition, from stimulus properties of the word itself (e.g., surface properties, such as letter size and font and deeper properties, such as word frequency and familiarity) to wider elements of context and pragmatics (e.g., implicature and semantic sense). However, at least for native and proficient skilled readers, this is largely automatic and effortless (Ferrand & Grainger, 1994; Hauk et al., 2012), taking no longer than a few hundred milliseconds per word on average (Trauzettel-Klosinski & Dietz, 2012). Nevertheless, substantial orthographic and, likely, phonological processing has already occurred in this time to achieve word recognition, particularly in the first ~200ms after seeing each word, forming the foundation for further processing (Dien, 2009). As skilled readers have thousands of hours' experience in reading words, it is not unreasonable to expect rapid, automatic, stimulus-driven responses that reflect a skill learnt to such an extent (Cornelissen et al., 2010).

One of the key aims of VWR research is to answer questions about when different types of processing, such as orthographic and phonological activation, begin and conclude, as well as how they interact with one another. In this context, processing refers to cognitive activity using information either from stimuli in the real world, such as the word-forms you are currently seeing (bottom-up, stimulus-driven processing) or from memory (i.e., top-down processing), including the stored orthographic and phonological internal representations of words (Taft, 1991). After a stimulus has been perceived, this internal processing can be based on output from the same or different network (cf., serial and cascaded processing theories) or can occur concurrently in different networks working towards a shared goal (i.e., parallel processing), be that word recognition or a precursor to it (e.g., phonological activation leading to word recognition).

Questions regarding the timing of reading processes largely centre on how early the brain can identify lexical and semantic information from a visual stimulus (Hauk et al., 2012) and understanding the specific timing of cognitive actions and interactions involved in recognising written words has been described as “the holy grail” of reading-related research (Sereno & Rayner, 2003, p. 489). Visual word recognition has been stated as identifying a written word as a lexical item from an average vocabulary of ~50000 within 500ms (Grainger & Holcomb, 2009). The idea that recognition of visually presented words, from visual input via orthographic and phonological processing to semantic integration, occurs within such a relatively short length of time as 500ms could appear extraordinary. However, considering a variety of evidence from visual word recognition research, half a second to identify a single word could, instead, be deemed excessive (Klein et al., 2015; Pammer et al., 2004; Segalowitz & Zheng, 2009; Wheat et al., 2010; Woodhead et al., 2012).

The most basic gauge of how quickly words can be read is a measure of words or characters per minute (wpm, cpm) referred to simply as reading speed or reading rate (Harley, 2010; Trauzettel-Klosinski & Dietz, 2012; Yu et al., 2010). For instance, the average reading speed of ~263.16ms per English word when reading aloud from a dense block of text (Trauzettel-Klosinski & Dietz, 2012) or just ~91.6ms per English word during rapid serial visual presentation (Yu et al., 2010) show lexical access to be possible in much less than 500ms. Interestingly, native reading speeds of English (~263.16ms per word; 228±30wpm), Spanish (~275.23ms per word; 218±28wpm), and Chinese characters[1] (~235.29ms per character; 255±29cpm) are not dissimilar (Trauzettel-Klosinski & Dietz, 2012), each of which also suggesting that the initial ~200ms is critical for each native language, which extends to other languages too (see Trauzettel-Klosinski & Dietz, 2012). As shown by the sustained effects of priming in tasks that involve articulation (Klein et al., 2015), the necessary response preparation and motor activity for such behavioural responses also require time to be processed and must be accounted for in the timeframe before the response. Therefore, if average reading speeds based on reading aloud can be ~250ms (Trauzettel-Klosinski & Dietz, 2012), it strongly indicates the importance of the initial 200ms of the pre-response VWR process.

[1] While it is noted that alphabetic words and logographic characters are not directly equivalent, the comparison between alphabetic words and logographic characters makes more sense than comparing words across these language types.

Aside from the importance of the initial post-stimulus ~200ms, the main point is that lexical access (and even reading aloud) appears to require significantly less than the 500ms often cited to be the timeframe for VWR. For instance, toward the end of this timeframe, the understanding is that the N400 ERP component (peaking at ~350-500ms) represents the semantic integration of different words and concepts in context, far beyond straightforward word recognition (Kutas & Federmeier, 2009; Lau et al., 2008). Based on accepting a prerequisite of more fundamental orthographic (and, likely, phonological) processing, if such a level of comprehension can be reached by the N400 timeframe, it is reasonable to assume that such orthographic and/or phonological processing occurs before it. At the earliest end of the timeframe, effects of unconscious (i.e., masked) phonological priming have been found after presenting a word for only 67ms, while orthographic priming can occur from only 33ms (Ferrand & Grainger, 1993), suggesting that these exposures are all that are necessary for orthography and phonology, respectively, to influence neural processes. However, it is important to note that what is necessary to influence processing, as in such masked priming effects, and necessary for full processing and comprehension in broader reading contexts are not the same and different exposures in the different contexts will likely amount to different processing patterns.

The rapid speed at which the eyes skip between words when reading sentences is also an indicator of reading speed and should be considered, as it involves a ~200ms average fixation (Sereno & Rayner, 2003), interspersed with saccades as short as 100ms and typically occurring 3 or 4 times per second (Potter et al., 2014), further suggesting that up to 200ms worth of word information could be absorbed during the fixation with processing beginning at the onset of the fixation. However, a concession of ~100ms is required for the time it takes for that information to reach initial processing centres in the brain i.e., the occipital cortex (Potter et al., 2014; Sereno & Rayner, 2003). Nevertheless, this leads directly into the P1 and N170 ERP component timeframes (~80-120ms and ~130-200ms, respectively) and the timeframes to be scrutinized in the current research.

While a substantial amount of VWR research proposes brain activity at particular post-stimulus timeframes in several key cortical areas to reflect reading-related roles, the specific timing and nature of early orthographic and phonological processing are still heavily debated (Pattamadilok et al., 2017). It is, however, clear that the initial ~200ms after seeing a word is critical for visual word recognition and, due to being a somewhat stable foundation of later processing (Cornelissen et al., 2010), this initial ~200ms is likely to be important for L2 processing too, whether in terms of initial orthographic and phonological processing or for language-dependent decision-making processes. Vitally, this is a timeframe not widely investigated in L2 readers with investigations of such early processing in bilinguals and L2 reading being under-represented in the multilingualism literature compared with such topics as syntactic processing of L2 and contrasts with L1 processing (Clahsen & Felser, 2006; Kotz et al., 2008). Combined with the controversially early effects of phonology (e.g., Klein et al., 2015; Wheat et al., 2010) and lexical access (e.g., Hauk, Davis, et al., 2006), the post-stimulus ~200ms timeframe (and therefore the P1 and N170 that reside within it) provides a window into the foundation of visual word recognition through orthographic and phonological processes, which will be described, explored, and discussed further in the following chapters.

Chapter 2: Theoretical perspectives of (monolingual) VWR

Theories of visual word recognition describe how cognitive processes and different psycholinguistic properties interact to translate sound and meaning from written language. One relatively straightforward way to conceptualize language in the mind is as a kind of mental dictionary with integrated thesaurus, glossary, and examples: a cognitive store for all known words, their constituent linguistic parts, and meanings that has been widely referred to as the mental lexicon (Harley, 2010; Taft, 1991). Each entry in the mental lexicon is associated with orthographic, phonological, semantic, and syntactic information as codes (Brysbaert, 2001; Brysbaert & Wijnendaele, 2003; Xu et al., 1999), nodes (e.g., Seidenberg & Tanenhaus, 1979), representations (e.g., Coltheart & Curtis, 1993), identities (e.g., Ehri & Wilce, 1980) or just information, depending on the choice of terminology. This concept of a mental lexicon, albeit also sometimes through different terminology, has been used as a basis for describing many psycholinguistic theories (Ferrand & Grainger, 1994; Frost, 1998; Taft, 1991).

There are currently several mainstream theoretical perspectives on how the cognitive processes involved in VWR work together for successful word comprehension. These include - but are not limited to - computational models and their founding theories, such as the Dual-Route Cascaded model of reading aloud (DRC; e.g., Coltheart et al., 2001), the Bi-modal Interactive Activation Model (BIAM; e.g., Grainger & Holcomb, 2009) as an extension of the aforementioned Interaction-Activation Model (McClelland & Rumelhart, 1981), and parallel distributed processing (PDP) connectionist and "triangle" models (e.g., McClelland & Rogers, 2003). It is important to bear in mind, however, that these models are typically centred on theories of reading aloud (as opposed to just VWR or reading silently) with the DRC being exactly that and the BIAM being based on silent reading skills mapped onto the spoken language system (Diependaele et al., 2010). Other perspectives on VWR processing are based on "VWFA" accounts (e.g., Dehaene & Cohen, 2011), the recognition potential (RP; Martín-Loeches et al., 1999), as well as evolution-based theories (e.g., Menary, 2014). There are also the highly orthographic (e.g., Hauk, Davis, et al., 2006) and early phonological (Wheat et al., 2010) accounts that underlie the current research, though none of these are mutually exclusive or divorced from the arguably more established theories (e.g., DRC, PDP, strong phonological, BIAM) that will be the primary focus of this review, starting with very brief overviews of each before delving deeper into discussing the specifics of orthographic and phonological processing during VWR.

2.1 Dual-route cascaded (DRC) model

The DRC model (Coltheart et al., 2001), as depicted in Figure 1, is a dual-route approach albeit with three routes to word recognition: grapheme–phoneme conversion, via the lexicon with semantic knowledge, and via the lexicon only (bypassing the semantic system). The lexical route uses internal orthographic representations to access phonology/semantics directly, based on the assumption that readers create and store orthographic representations when a word is learnt, and is the most frequently used route for VWR in skilled readers. The slower non-lexical route provides a method of constructing phonological representation when an internal orthographic representation is not available and/or without relying on one e.g., pseudowords. As this indirect route uses grapheme-phoneme correspondence (GPC), it is only for regular orthography and pronunciations, while operating in a serial manner, processing letters from left to right (Ziegler et al., 2001).

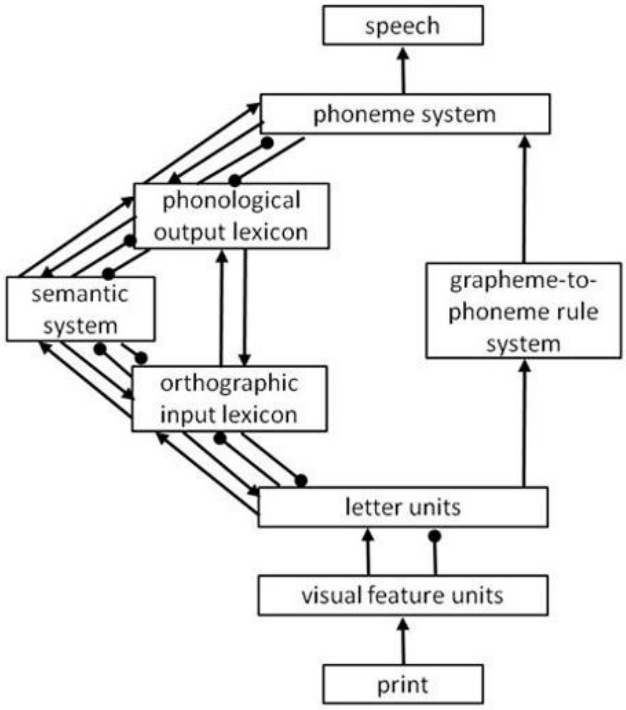

Figure 1: The Dual-Route Cascaded (DRC) model of reading aloud (Coltheart et al., 2001)

2.2 Parallel distributed processing (PDP)

Connectionist PDP models of VWR (e.g., Figure 2) are parallel processing approaches with two main paths to word recognition: a direct orthography-phonology pathway and an indirect orthography-phonology pathway via semantics (Simon & Lalonde, 2004). This appears similar on the surface to the DRC, but its implementation of processing is different. According to PDP-based VWR accounts (e.g., Plaut et al., 1996), word recognition is achieved through patterns of activation across the "triangle" of interconnected orthographic, phonological, and semantic networks using weighted input from orthography and phonology based on the input available (i.e., the linguistic information of the viewed word) with semantics accessed via phonology or directly via orthography, which also activates phonology. Considering the dual routes of DRC, PDP manages them through a single phonology-oriented route where a single grapheme (e.g., the *i* in *pint*/*mint*) is associated with two phonemes (*mint*, *pint*) based on a phonological continuum of spelling-to-sound correspondence that stretches from "rule-governed" to "exceptions" (Harm & Seidenberg, 2004). This PDP perspective, therefore, emphasizes the importance of phonology in reading and follows the notion of phonological mediation.

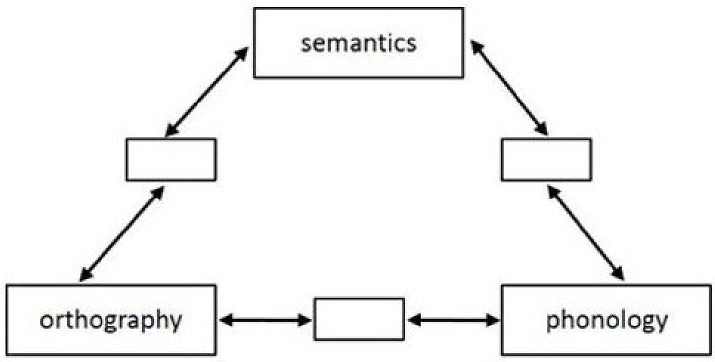

Figure 2: Connectionist triangle models (Plaut et al., 1996)

2.3 Bi-modal Interactive-Activation Model (BIAM)

Based on Interactive-Activation models (McClelland & Rumelhart, 1981), extending it to include phonological and semantic processing, the orthography-phonology mapping in the BIAM (Figure 3) has coarse and fine orthographic codes, corresponding with feature and letter/word detectors of the IA, and the semantic system can be activated by orthography or phonology (Grainger & Holcomb, 2009). Furthermore, the BIAM did not originally include an equivalent to the output phoneme in the DRC, but, for comparability between these two models, it adopted the phonological output buffer from the PDP-based CDP+ model of reading aloud (Perry et al., 2007) where integration of whole-word and sublexical phonology from the input occurs (Diependaele et al., 2010). The BIAM, then, can be seen as a kind of hybrid between dual-route and PDP approaches to VWR, though is also a distinct theory and model in its own right.

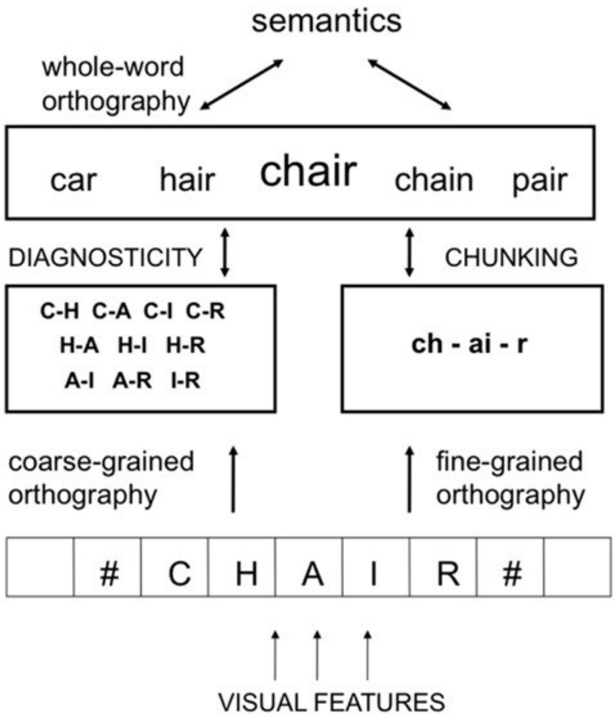

Figure 3: The Bi-modal Activation Model (Grainger & Holcomb, 2009)

2.4 Strong/weak phonological theories

Underpinning the various models of VWR are the strong and weak phonological theories, named not for the quality of the assumptions, but as conceptualizing the involvement of phonology in VWR in terms of its importance in the reading process (e.g., Frost, 2003; Rastle & Brysbaert, 2006). Strong phonological theories posit that phonology is necessary for lexical access and VWR is accomplished by extracting the phonetic representation of written input to access the lexicon with this phonological code (Frost, 1998). The strong phonological perspective divides phonological processing into two possible methods: assembled phonology and addressed phonology (Grainger & Ferrand, 1994). Assembled phonology employs a grapheme-phoneme conversion (GPC) process akin to the DRC to translate each grapheme or grapheme cluster into the relevant sounds, essentially "assembling" them, and is used for unfamiliar lexical units, implying frequent use by beginner readers (Abramson & Goldinger, 1997; Grainger & Ferrand, 1994; Van Orden, 1987). Addressed phonology refers to retrieving the sound of words and sublexical units that have been sufficiently committed to memory on a whole-word basis, used by skilled readers for high-frequency words (Grainger & Ferrand, 1994; Van Orden, 1987). This method of addressed phonology is also reminiscent of the DRC (its direct lexical route), but is critically different in its dependence and necessity for phonological processing that the direct route of the DRC does not require. The crux of the strong phonological theory is that reading is phonologically-mediated and is not possible without some form of phonological recoding, being active regardless of task necessity for phonology (Lukatela & Turvey, 1994; Luo et al., 1998; Perfetti et al., 1988).

In contrast, weak phonological theories, such as the DRC, do not give phonology such a leading or important role in VWR, instead positing that phonology can influence VWR processing (whether facilitative or inhibitive) but is not necessary. Weak phonological theories deem phonology to be sufficient, but not necessary for reading, though sometimes having a mediating role (Rastle & Brysbaert, 2006). This contrasts with strong phonological theories that essentially posit phonology as necessary and consistently activated during reading. As per dual-route and PDP models, weak phonological theories typically follow the notion of a direct orthography-semantics lexical pathway and an indirect phonologically mediated pathway that supports word recognition when necessary. Importantly, though, weak phonological theories do not deny phonological involvement, even from an early timeframe of VWR, though the weak view, like the DRC, is still inconsistent with such phenomena as fast masked phonological priming, as will be outlined later (Rastle & Brysbaert, 2006).

2.5 Contrasts between theoretical perspectives of monolingual VWR

The DRC (Figure 1), PDP (Figure 2), and BIAM (Figure 3) have similarities in that they all have direct and less direct pathways to word recognition, even though the origins, destinations, and overall definitions of the routes are different. The core difference between VWR theories is rooted in the details of the "routes" or "pathways" i.e., the means for written language to be processed and understood. For instance, dual-route theories propose distinct and somewhat dissociated lexical and non-lexical mechanisms for extracting phonology (and meaning) from visual stimuli, while connectionist/triangle theories attempt to explain the whole word recognition with a more unified mechanism. In other words, different theories may indeed support the idea of multiple routes to word recognition, but they are implemented and defined in different ways, requiring and recruiting different types and extents of information along the way. Therefore, the notion of multiple routes represents both a distinction and more common ground between theories, albeit using different terminology. The main point is that it is widely accepted that there are multiple processing pathways for reading linguistic stimuli.

Importantly, different VWR theories provide different perspectives on the reading process, but they do share important traits. For instance, these theories share the assumption that phonological activation depends on orthographic processing and can be used for lexical-semantic access, whether it is an integral part of the reading process or not (cf. strong/weak phonological theories; see Frost, 1998). Core concepts from DRC and PDP theories together with ideas from the BIAM and other key evidence (e.g., Rastle and Brysbaert's (2006) amended DRC simulation, Hauk et al. (2006), and Wheat et al. (2010)) will provide the foundation of the discussion. As the current research investigates the orthographic and phonological processes that underpin VWR, it is vital to understand how such processes are accounted for by the aforementioned theoretical perspectives, starting with several key concepts that they have in common.

The nature of processing during reading refers to the order of the (visual, orthographic, phonological, semantic, and ultimately lexical access) stages involved (Dennis & Key, 2006), if such stages even exist per se (Twomey et al., 2011), and how these different processes interact with one another (Grainger & Holcomb, 2009). The timing and nature are, therefore, not mutually exclusive and investigations into any interactions often coincide with questions of timing, resulting in attempts to determine whether lexical and semantic information is retrieved in parallel, whether reading occurs in stages, or it is a combination (Hauk et al., 2012). On the surface, accounts of neurophysiological activity during reading can give the impression that different VWR-related processes work in stages, following a serial path from visual to orthographic to phonological and finally semantic to achieve lexical access (Twomey et al., 2011). However, there is evidence to suggest that some processes, such as top-down phonological and semantic, likely work in parallel along multiple pathways simultaneously (McClelland & Rogers, 2003), while others, such as initial visual and orthographic, do work strictly in series (Grainger et al., 2006). Evidence of phonological mediation during VWR, for instance, such as stronger facilitation of real word target responses from pseudoword primes than real word primes (e.g., Lukatela & Turvey, 1994) and the classic homophone effect where homophones inhibit semantic processing (e.g., Van Orden, 1987), shows early involvement of phonology and that the initial orthographic-phonological processing path is not purely serial (as will be discussed later). Indeed, the well-documented easier and faster reading of regular words (e.g., *mint*) than irregular words (e.g., *pint*) is taken as evidence for lexical/direct and non-lexical/indirect routes resulting in a conflict from the irregular word input that needs to be resolved (Pattamadilok et al., 2015). This conflict (and resolution) suggests that the two routes are not fully independent, but parallel and interactive for some stimuli and with the possibility of one route confounding the other. Such resolution takes time, hence the process taking longer.

One of the main challenges for theories and models of VWR is to account for fast phonological effects, such as from masked phonological priming (Diependaele et al., 2010). The requirement of a VWR model to account for both pronunciation of irregular words (including pseudowords and unknown real words) and the rapid phonological processing evidenced through masked phonological priming, dubbed the fast-phonology test (Diependaele et al., 2010), has been shown to be highly problematic for the DRC model simply due to being too fast for it to accurately account for VWR behaviour. In contrast, this is achieved by the BIAM using rapid parallel orthographic-phonological processing (Diependaele et al., 2010). Ultimately, the BIAM assumes that the mapping between letter inputs and input phonemes (as used for auditory comprehension as opposed to speech production) is rapid i.e., fast phonology, while the DRC does not and, therefore, does not consistently work where rapid phonological processing occurs in real-world VWR e.g., masked phonological priming using pseudohomophones (Wheat et al., 2010). For instance, the DRC requires the signal to be processed through its output phoneme layer and fed back in to explain such a pseudohomophone priming effect on a related real word target (Diependaele et al., 2010), which is a much longer process than the interactivity between input phonemes and lexical phonology, as suggested in the BIAM. Such fast phonology can also be achieved by the DRC (Rastle & Brysbaert, 2006), but only if lexical decisions in VWR are based on phonological activation as opposed to whole-word orthographic input, which is not always the case (Diependaele et al., 2010). Furthermore, there is contention about how this amendment to the DRC that allows it to work with fast phonology would account for inhibitory effects from true real word homophone (not *pseudo*homophone) primes (Diependaele et al., 2010). Ultimately, the DRC assumption of minimal impact of phonology on silent reading is strongly challenged by rapid phonological activation and does not fit with the strong phonological argument that phonology is necessary for semantics (McNorgan et al., 2015; Rastle & Brysbaert, 2006).

Another specific issue for models of VWR is the time it can take to find the correct pronunciation of irregular words, pseudowords, and unknown real words (Diependaele et al., 2010). The DRC model, for example, has been shown to "fail" to account for pronunciations of pseudowords as documented from human participants (Pritchard et al., 2012). This indicates serious issues for modelling pseudoword reading, specifically for the nonlexical route and its mechanism of grapheme-phoneme conversion, suggesting that pseudowords and, by extension, unknown real words require alternative computation than the model currently allows.

The BIAM explains conflicts between orthography and phonology, such as when reading an orthographically-similar non-rhyme (e.g., *mint-pint*), through the use of a fine-grained orthographic code that activates and interacts with phonology in parallel with a coarse-grained orthographic code that employs more direct orthographic-semantic processing akin to the lexical pathway of the DRC (Grainger & Holcomb, 2009). Whereas orthographic-phonological processing happens using output phonemes (i.e., language production) according to the DRC, letters/graphemes are mapped to input phonemes in the BIAM. Following the "reading aloud" perspective of the DRC model, its phonological lexicon works for output phonology, neglecting so-called input phonology. In any case, phonology is part of linguistic input and psycholinguistic processing, but to what extent, in which contexts, and whether it plays a different role in different readers (e.g., L1 and L2) is still up for debate and a topic the current research aims to address.

While the DRC (Coltheart et al., 2001) and BIAM (Grainger & Holcomb, 2009) do not entirely agree about the timing of phonological activation and its relation to orthographic processing, both posit a largely serial relationship between orthographic and phonological processing. However, evidence of early and potentially parallel phonological activation is no longer uncommon (e.g., Ashby, 2010; Klein et al., 2015; Pammer et al., 2004; Wheat et al., 2010). In order to investigate the serial and/or parallel processing of orthographic and phonological processing, both must be required by the task, which should also allow orthography and phonology to facilitate or inhibit processing and to interact with one another as per such theories as the BIAM states they can. Furthermore, it should be possible to observe the relative strength of orthographic and phonological processes. The rhyme judgement task employed in the current research is precisely such a task, as it entails the active processing of both orthography and phonology with no overt need for semantics. When both orthographic and phonological rime-processing are required (e.g., in rhyme judgement), rime-based phonological priming typically facilitates responses (Coch et al., 2008), but its interactivity with orthographic congruency and how processing of the two elements works is not clear and will be addressed in the current research.

This chapter has so far discussed how perspectives of VWR account for the relationship between phonological and orthographic processing. The research documented in this thesis, however, is also concerned with this relationship in terms of bilingual L2 readers of a language.

Chapter 3: Theoretical perspectives of bilingualism

The Bilingual Interactive Activation models BIA/BIA+ (Dijkstra & van Heuven, 2002; van Heuven & Dijkstra, 2010) and the more recent Multilink model (Dijkstra et al., 2019) provide the most relevant and robust details due to their focus on bilingual word recognition and processing, as opposed to language translation e.g., Inhibitory Control (IC) model (Green, 1986) or language production e.g., Revised Hierarchical Model (RHM; Kroll & Stewart, 1994). However, just as the BIA/BIA+ models were combinations of prior theories, Multilink is a combination of characteristics from BIA/BIA+ and RHM, as well as WEAVER/WEAVER++ (Word Encoding by Activation and VERification; Roelofs, 1997, 2014), but features pertinent to the visual domain and reading will be the focus. Due to the overlaps between learning/developing bilinguals and late bilinguals, the BIA-d (Grainger et al., 2010) is also of note.

3.1 BIA/BIA+ and Multilink

Centring mainly on the BIA/BIA+ as the cornerstone of many current outlooks on bilingual VWR, there are broadly considered to be two subsystems (see Figure 4), one for linguistic processing e.g., word recognition and one for associated non-linguistic strategy, known in BIA/BIA+ terminology (and adopted by Multilink) as the word identification system and the task/decision system, respectively (Dijkstra et al., 2019; van Heuven & Dijkstra, 2010). The task/decision system is involved in non- and extra-linguistic strategies, participant expectancies, and the contexts of the linguistic material being processed, informed by non-linguistic contextual knowledge as well as by the word identification system itself. The word identification system, however, provides the features most pertinent to the current research.

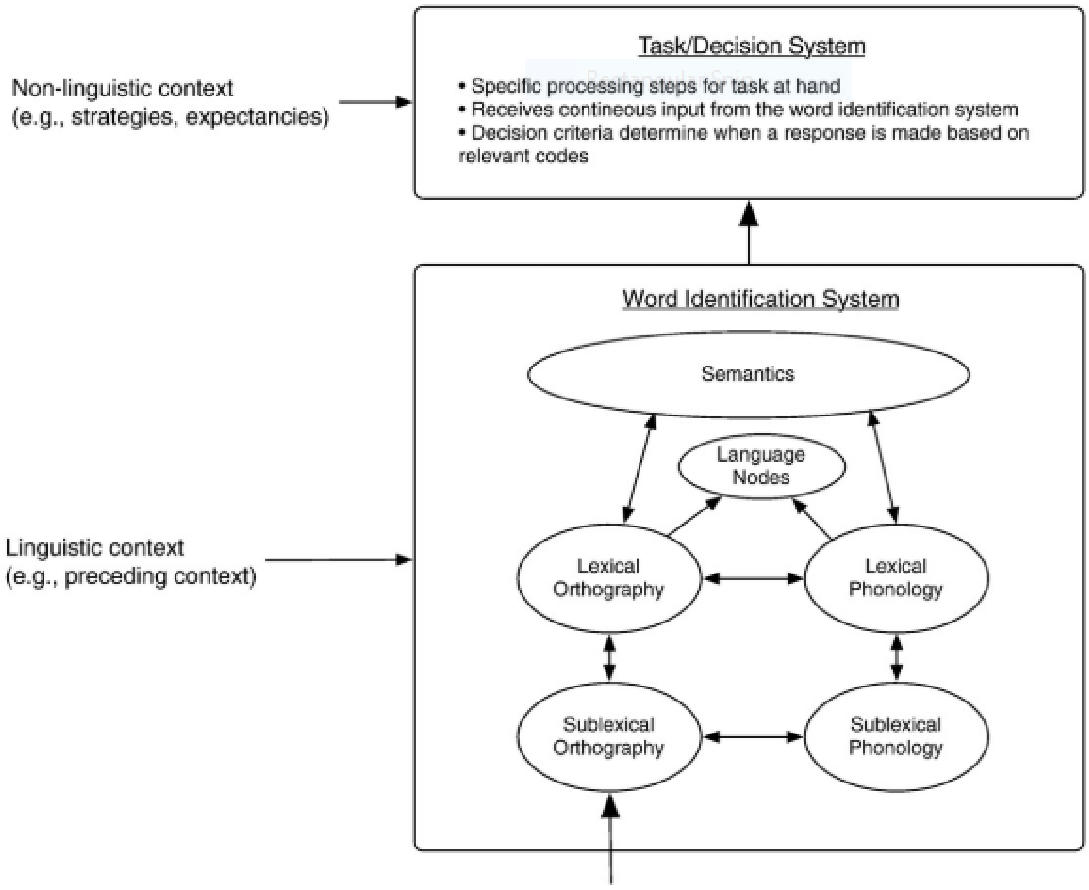

Figure 4: The BIA+ model (van Heuven & Dijkstra, 2010)

The word identification system comprises two key features of BIA/BIA+/Multilink-based VWR theory: an integrated lexicon and language non-selective access. The lexicon is integrated in that L1 and L2 are not stored and understood as two entirely different sets of words and rules, but as one unified entity (van Heuven & Dijkstra, 2010). This integrated lexicon incorporates the words, concepts, and linguistic constructs of both languages, which can be connected across languages where viable e.g., vocabulary, though there may also be language-specific information e.g., syntactic rules (Dijkstra et al., 2019). Building on the premise of a single integrated lexicon, language non-selective/parallel access backgrounds the importance of language membership (i.e., which language is being read), keeping options for either language open, at least during initial VWR processing (van Heuven & Dijkstra, 2010). In contrast, the main alternative is language selective (serial) access, where word identification processes would be limited to the language of the stimulus and require language membership to be recognized very early on and as a required step before additional processing can occur, has comparatively very little evidence to support it (Dijkstra et al., 2019).

The orthographic, phonological, and semantic representations involved in the word identification system interact dynamically at various points during processing (van Heuven & Dijkstra, 2010), which is not dissimilar from the interactivity and feedback mechanisms described by such theories as the BIAM (Grainger & Holcomb, 2009). There are other similarities to the BIAM (Grainger & Holcomb, 2009) and DRC (Coltheart et al., 2001) too, such as the BIA/BIA+ positing that a visual linguistic stimulus activates sublexical orthography, while whole-word orthography and sublexical phonology can be activated simultaneously (dependent on proficiency, which will be discussed further later), allowing orthographic and phonological dual routes to word recognition and semantics. Furthermore, activation is also cascaded between orthographic, phonological, and semantic stages, with the latter two also occurring after orthographic processing (van Heuven & Dijkstra, 2010).

One especially important extension of the monolingual/native theories is that the orthographic and phonological processes will also recall the relevant language membership information ("language nodes" in Figure 4). However, this information theoretically does not influence word identification processes, as there is no feedback to stages of lexical activation (van Heuven & Dijkstra, 2010). The assumptions of an integrated lexicon, language non-selective access, and parallel L1/L2 activation form the foundation of the current understanding of bilingual language processing and, therefore, this research and any findings it may produce.

Another critical aspect of the BIA/BIA+ models is the temporal delay assumption, which, put simply, says that L2 activation typically lags behind L1 (Duyck, 2005). This feature of the BIA/BIA+ posits that activation of L2 phonological and semantic codes/representations are delayed compared with L1 codes/representations in the same linguistic context (Brysbaert et al., 2002). More precisely, it relates to thresholds of activation, factoring in word frequency and the proficiency of the bilingual, where L2 activation of low frequency or lesser-known L2 words will be delayed compared with L1 activation. Evidence of this on a neural level has been shown using semantic activation and the N400 (Moreno & Kutas, 2005), the latency of which was observed to be longer for L2 than L1 in bilinguals and longer than in monolinguals of the bilinguals' L2 e.g., Spanish-English bilinguals and English monolinguals. Furthermore, language proficiency and age of exposure (cf. early/late bilinguals) were concluded to be important factors for the discrepancy in N400 latency and, by extension, the level of temporal delay that contributes to L2 processing (Moreno & Kutas, 2005).

The BIA and BIA+ models have thus far been largely discussed as one and, while the BIA+ is a development of the preceding BIA model, there are several key distinctions. Perhaps the most important is that the BIA model permits interactions between subsystems, but the nature of the BIA+ is one of bottom-up processing, which precludes the task/decision subsystem from affecting the word identification subsystem (van Heuven & Dijkstra, 2010). Further to this emphasis on bottom-up and relative exclusion of top-down processing in the BIA+, it also does not permit any influence of language membership on the word identification subsystem, inhibitory or otherwise, while the BIA does allow language membership to inhibit the word identification subsystem. In contrast, participant expectancies (in the sense of top-down contextual cues e.g., task instructions) do not affect word activation in the BIA, but the BIA+ permits the possibility of such expectancies influencing the task/decision subsystem (van Heuven & Dijkstra, 2010). Following the focus on L1 profiles in ESL bilinguals and the potential influence of linguistic context (in terms of orthographic or phonological lexical decisions or orthographic and phonological priming between word pairs) in the current research, these distinctions of bottom-up and top-down processing between BIA and BIA+ are of particular interest.

Another version or extension of the BIA/BIA+ that is especially relevant to the current research is the Developmental Bilingual Interactive-Activation model or BIA-d (Grainger et al., 2010), due to its focus on L2 VWR and processing in late bilinguals, albeit those with L1 and L2 that share an alphabetic writing system (e.g., Spanish-English late bilinguals). As the name implies, the BIA-d centres on bilingual development, specifically from adult late bilingual learners operating along the language production-oriented principles of the RHM (Kroll & Stewart, 1994) to skilled L2 readers that the BIA more appropriately describes. Using L1 and L2 whole word representations (forms) and the semantic concept associated with them (meaning), the framework (as shown in Figure 5) incorporates the development of connectivity from standard monolingual form-meaning mapping to the initial learning of the L2 word. From there, development continues along stages of L2 vocabulary acquisition that strengthen the connections between the L2 representation and both L1 word and the semantic concept. The BIA-d model links this connectivity development to better L2 control in terms of the ability to inhibit L1 activation while processing L2 input (Grainger et al., 2010). While such inhibition exists in the BIA/BIA+, accomplished by language nodes (van Heuven & Dijkstra, 2010), complete inhibition of L1 during L2 processing is not possible (Brysbaert & Dijkstra, 2006; Dijkstra et al., 2019; van Heuven & Dijkstra, 2010), only damage control to minimize L1 involvement and facilitate L2 VWR (Grainger et al., 2010). Nonetheless, the BIA-d presents a welcome and necessary extension to the BIA/BIA+ and bilingualism theory in general with its acknowledgement of the development and language backgrounds of bilingual readers. The chief limitation of this BIA-d account in the context of the current research, however, is that it relates to late bilinguals learning an L2 with the same alphabet as L1. It is, therefore, not directly applicable to bilinguals with fundamentally different L1 and L2 in terms of alphabetic or writing system (e.g., Chinese-English bilinguals with a mix of logographic L1 and alphabetic L2), though a similar premise could work for other bilingual combinations. As with other elements of bilingualism theory and research, however, further research is required with such non-alphabetic populations (Van Heuven & Wen, 2018).

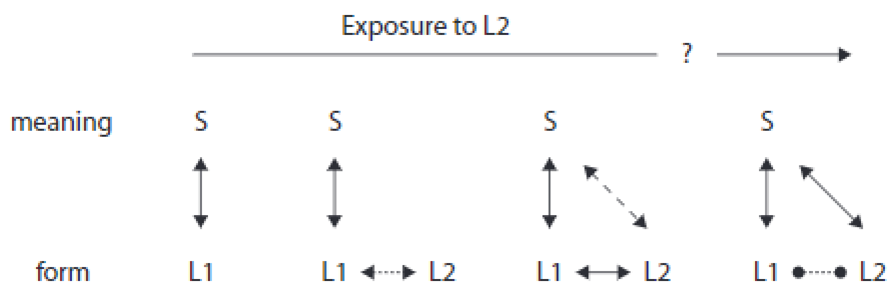

Figure 5: Proposed framework for uniting RHM and BIA (Grainger et al., 2010)

Lastly in this overview of bilingualism theories, it is worth considering the lexicon of a monolingual reader as tantamount to an integrated lexicon of a bilingual reader, a perspective recently taken by Multilink model of monolingual and bilingual comprehension and production (Dijkstra et al., 2019). The idea of an integrated lexicon involves the different languages being treated as part of the same whole lexicon, distinguished in some way by language identity, similar to how social/cultural norms might separate a lexicon (van Heuven & Dijkstra, 2010). However, this might only hold true for vocabulary while syntax/grammar could prove problematic, possibly requiring separate resources for the different systems (Brysbaert & Dijkstra, 2006; Brysbaert & Duyck, 2010). Nonetheless, conceptualizing an integrated lexicon with language non-selective access essentially as an extension of a monolingual lexicon (as opposed to an addition to a monolingual lexicon) paves the way for VWR theories based on monolingual/native processing to be applied to L2 reading (Brysbaert & Duyck, 2010). Indeed, a major recommendation for investigating bilingualism is to first consider how theories and features from models of native/monolingual processing can be applied and adapted if necessary (Brysbaert & Duyck, 2010, p. 359; Dijkstra et al., 2019). The consideration and implementation, therefore, of both monolingual and bilingual language processing together is perhaps the most important and pertinent aspect of the Multilink model (Dijkstra et al., 2019). To an extent, this happened with the BIA in terms of using existing monolingual models to develop a bilingual model (Brysbaert & Duyck, 2010), but Multilink takes it further by implementing both monolingual and bilingual processing in a single model. However, this is not to say that Multilink can simply replace BIA/BIA+, but its promising performance compared with IA (McClelland & Rumelhart, 1981), and BIA/BIA+ alongside implementation of word frequency, length, and interlingual similarity as lexical factors and L2-proficiency as a reader-oriented factor make it a valid contender and an important milestone for language research (Dijkstra et al., 2019).

The framework of Multilink is based on the same local-connectionist network principles as the BIA/BIA+, but incorporates aspects of RHM and WEAVER++ as well as other new features in order to cover both comprehension and production (Dijkstra et al., 2019). Therefore, for the purposes of the current research, only parts pertaining to comprehension and, where possible, VWR will be discussed. As can be seen in the layers and interconnectivity of elements in Figure 6, Multilink itself also recruits a very familiar architecture and flow of lexical processing (Dijkstra et al., 2019), not unlike that of DRC, BIAM, or BIA/BIA+. Language input has its orthography evaluated first, allowing it to be linked to phonological representations and semantic concepts, as well as language membership. Importantly, however, the architecture still lacks mechanisms for sublexical processing, phonological onset timing, and details of other potentially parallel processes, though the authors acknowledge this (Dijkstra et al., 2019). Unfortunately for the current work, it is such sublexical and phonology-related processes that are most pertinent, though this does provide an opportunity for the current research to contribute to the development of Multilink and other models.

As is understandably typical for bilingualism models, Multilink has an emphasis on word-level interlingual factors, such as cognates, which are vital to be addressed by any bilingualism theory. However, along with the missing sublexical and phonological onset processing, the RHM- and vocabulary-focused side of Multilink leaves such aspects of bilingualism as language background, relevant L1 VWR skills, and similarities between L1 and L2 somewhat under-represented. Nonetheless, the interactive word-level system and its connections that are all bidirectional and variable in strength are, akin to the BIA/BIA+, overseen by a task/decision system, which is especially important because the output of the Multilink model is task-dependent (Dijkstra et al., 2019). As in the BIA/BIA+, the task/decision system provides some overarching top-down control of processing and output based on non-linguistic factors e.g., task instructions, though somewhat differently from the BIA/BIA+ implementations, it can also inspect linguistic factors, such as orthographic/phonological/semantic activation levels and language membership, to better specify the output according to the task requirements (Dijkstra et al., 2019). With this in mind, this part of Multilink is more like the top-down nature of BIA than BIA+ and could be extended to include language background factors.

The aforementioned capacity to account for L2 proficiency is an important and pertinent aspect of Multilink, as L2 proficiency is a key factor in behavioural performance during bilingual VWR and, therefore, the cognitive strategies and brain activity that underlies that performance (Fitzpatrick & Izura, 2011; Hulstijn, 2011). However, it is arguably much more individualized than Multilink allows, so its inclusion is laudable, but it appears to require significant refinement and much less reliance on assumptions that are very generalized while also still being specific to Dutch-English. Furthermore, this specificity or restriction to Dutch-English bilinguals is one of the main shortcomings of Multilink at present, at least in the context of the current research. Dutch-English bilinguals represent a population that does not directly fit with either Spanish-English or Chinese-English due to the comparable orthographic depth and same alphabetic profile of Dutch compared with English. However, many of the principles and features are still relevant already to other populations, especially due to the combined monolingual/bilingual approach, and this will hopefully be developed further in future.

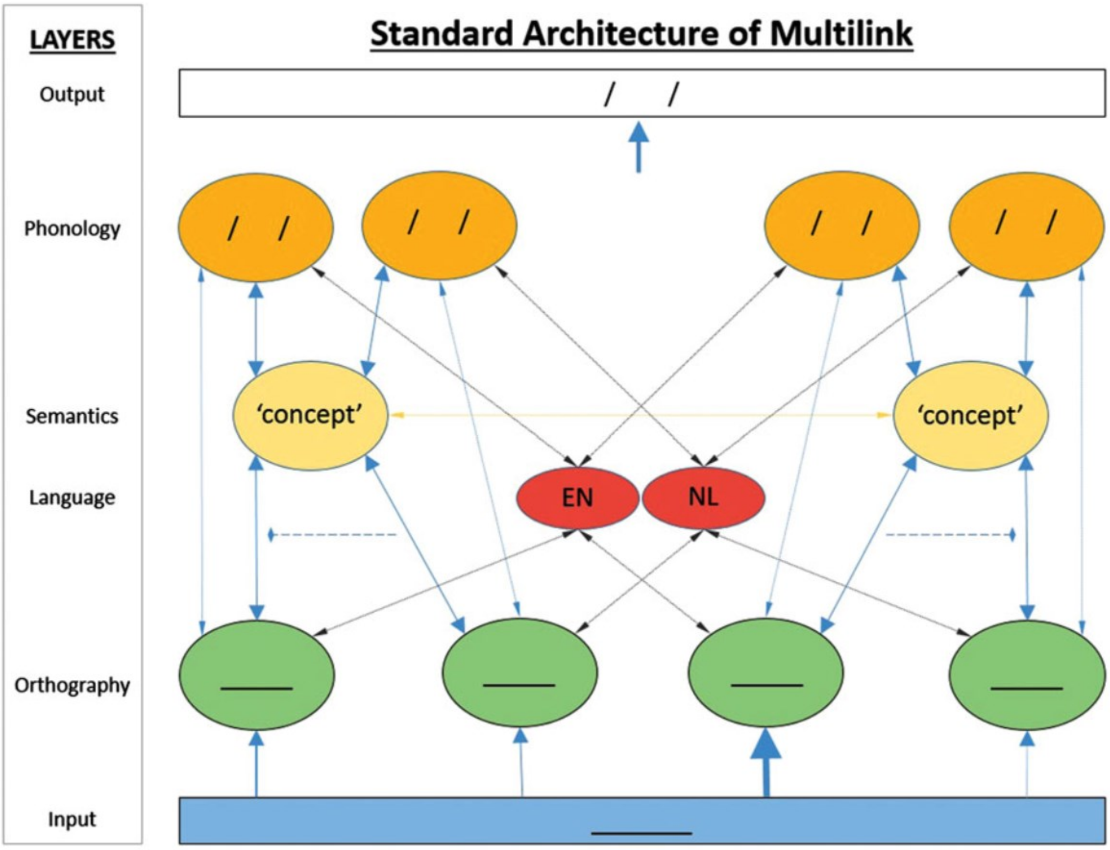

Figure 6: Multilink lexical processing architecture (Dijkstra et al., 2019)

3.2 L2 reading strategy

One of the core questions amidst VWR, L2 acquisition, and bilingualism literatures (and, consequently, central to the current research) concerns how a second, other, or non-native language is read and, specifically, how similarly or differently the routes to word recognition operate in L1 and L2 readers of a language (see Brysbaert & Dijkstra, 2006; Guo et al., 2009; Huettig & Ferreira, 2022). Native and bilingual L2 processing of the same target language are often concluded to be comparable in many ways, using similar processing strategies (Diependaele et al., 2011; Lemhöfer et al., 2008) and possibly increasing in similarity with proficiency (Kim et al., 2017). For instance, it is reasonable to accept that bilinguals with L1 and L2 that share a writing system and/or orthography (e.g., Spanish-English bilinguals) use L1 skills for L2 processing. However, various different patterns have been observed in terms of timeframe of activation and N170 lateralization, questioning any normal reading strategy (Huettig & Ferreira, 2022). How such differences manifest is much less clear for less proficient and late bilinguals (Amora et al., 2022; Maurer, Brandeis, et al., 2005; Yum & Law, 2021). It is also much less clear how the patterns of L2 processing may differ between bilinguals with L1 and L2 from different orthographies and/or writing systems or between ESL readers with different L1 language types (e.g., alphabetic L1 vs non-alphabetic L1), such as native readers of a logographic language reading an alphabetic language (e.g., Chinese-English bilinguals).

Considering effects across different types of bilinguals, it has been suggested that the consistency of orthography (cf. orthographic depth) in L1 influences language acquisition and production of L2 orthography (Paulesu et al., 2021). However, the difference in L1 orthographic depth between native English, Spanish-English, and Dutch-English groups showed no impact on reading performance (Diependaele et al., 2011; Verdonschot et al., 2012). Such findings could support the idea that L2 processing in bilinguals relies on direct, whole-word processing more than indirect, language-specific conversion rules (Clahsen et al., 2010), but might only be supported when L1 and L2 are at least broadly similar, resulting in much of the L2 processing relying on L1 techniques. Furthermore, differences in L2 processing based in any way on L1 properties could be dependent on stimuli and their similarity to L1. For instance, native readers of an alphabetic language with shallow orthography, such as Spanish and Italian, might be expected to be faster at reading pseudowords due to the natural requirement of grapheme-phoneme conversion skills. Evidence of this was observed with Italian-English participants, who represent the same ESL alphabetic L1 population as Spanish-English bilinguals (due to using an alphabetic script with a similarly shallow orthography), reading aloud both Italian- and English-derived pseudowords more accurately and faster than English monolingual participants (Paulesu et al., 2000). Due to the differences between English and Italian orthographies, the orthographic representations of the different English- and Italian-derived pseudoword types should also provide clues to their lexical status. However, this was only found to benefit the Italians, possibly suggesting that English participants, at least initially, attempted to read the pseudowords via a direct whole-word pathway, while the Italians recruited their theoretically superior grapheme-phoneme skills to access accurate pronunciation faster.

During L2 VWR, there is evidence to suggest that orthography-phonology mappings of both L1 and L2 are activated in parallel, perhaps even simultaneously (Brysbaert et al., 1999; Oppenheim et al., 2018; Van Wijnendaele & Brysbaert, 2002; Wu & Thierry, 2010), thus allowing interlingual phonological priming and supporting the notion of a single integrated lexicon described earlier (Dijkstra et al., 2019; van Heuven & Dijkstra, 2010). Such parallel automatic phonological activation during bilingual VWR is a prime factor in the proposition that brain activity, processing, and reading behaviours will differ between bilingual L2 reading and monolingual L1 reading of the same language (Brysbaert & Dijkstra, 2006), e.g., Spanish-English or Chinese-English and native English, respectively. Indeed, with this automaticity of phonological coding inherent in VWR, monolingual or bilingual (Brysbaert & Dijkstra, 2006), incorrect phonology can be activated in bilinguals with L1 and L2 that use the same or similar orthographic system due to different orthography/phonology mappings between the languages (e.g., /oo/ and /ou/ in English and French, respectively), which may or may not have the same phonological representations. The extent that this impacts bilinguals with L1 and L2 of different language systems e.g., Chinese-English bilinguals with a logographic L1 and alphabetic L2 is, however, much less well understood and requires more research (Van Heuven & Wen, 2018).

The relative complexity of orthography-phonology mapping in English means that reading new words, pseudowords, and pseudohomophones can be challenging, especially for L2 readers. The need to use an indirect approach to read English pseudowords and pseudohomophones requires that readers have some form of grapheme-phoneme conversion skills, which may not be the case based on skills required and acquired for their L1, as is the case for Chinese (Seymour et al., 2003). While native Spanish readers have roots in grapheme-phoneme conversion as part of their L1, native Chinese readers do not (Ellis et al., 2004). This lack of overt phonological cues found in logographic scripts portrays Chinese as a particularly visual/orthographic language, implying that a phonological route would not be learnt through L1 acquisition. Expert readers of Chinese can extract phonological information from radicals in the orthographic representations, but, unlike grapheme-phoneme conversion in alphabetic languages, this is not a tool that learner or skilled readers can necessarily use to pronounce new or unknown (L1 or L2) words (W.-J. Kuo et al., 2004). Therefore, the route used by native Chinese readers to read English may not be directly orthographic-phonological in the sense expected of alphabetic scripts, instead being a more direct visual pathway.

In a particularly relevant contrast between late-bilingual Spanish-English (alphabetic L1) and Japanese-English (non-alphabetic L1) with native English as a control group, more "false positive" errors (incorrect responses answered in the affirmative) were made to English homophones than to orthographic controls across all groups (Ota et al., 2010a). This suggests that L1 and L2 readers might employ similar strategies that depend on the target language e.g., GPC for new irregular words in English, including L2 readers with alphabetic and non-alphabetic native languages and therefore somewhat regardless of language profile. This has been previously proposed as the explanation for native-like L2 processing in bilinguals in studies observing comparable VWR behaviour between L1 and L2 readers (e.g., Diependaele et al., 2011; Lemhöfer et al., 2008). However, the proficiency of the participants in these cited studies allowed them to perform at native-like levels, which could have masked any underlying differences and ones that behavioural measures alone may not be sufficient to detect (e.g., Timmer & Schiller, 2012). Furthermore, care should be taken when making inferences from errors alone, as the similarity of error patterns across L1 and L2 readers is only evidence of the more difficult homophone stimuli consistently eliciting errors. Similar patterns of correct responses and brain activation would more solidly support the idea of L1 and L2 readers generally using the same reading strategies. Therefore, deeper measures, such as EEG/ERP, a wider variety of tasks that examine correct responses, and contrasts between bilinguals of different proficiencies are needed to investigate these questions further.

Language proficiency, through development and experience (Fitzpatrick & Izura, 2011; Hulstijn, 2011), could be key to processing requirements for reading L2. For instance, late bilinguals, defined here as individuals who did not learn L2 from an early age (e.g., before the critical period), have less time with L2 and less exposure to it than L1, particularly within the critical period (Hulstijn, 2011). This is coupled with any similarities or differences between L2 and L1, many of which can influence L2 processing in facilitative or inhibitory ways depending on L1 and L2 language types and the extent of similarities or differences (Maurer et al., 2008). Although the orthography and phonology of different languages do not necessarily lend themselves to form a framework for L2 reading (e.g., Chinese orthography for learning English), this is especially the case for late bilinguals, who not only have more experience with L1, but possibly use it as a framework for L2 when possible (Hulstijn, 2011). The neural basis of any such differences, including when and how the signal path diverges from native orthographic and phonological processing, requires further investigation. Despite such behavioural and developmental reasons for understandable performance differences (Fitzpatrick & Izura, 2011), bilingual models of VWR have traditionally not addressed proficiency (van Heuven & Dijkstra, 2010). As mentioned earlier, however, Multilink (Dijkstra et al., 2019) can be tuned to emulate different L2 proficiencies, which are strongly linked to its concept of "frequency-dependent resting level activation" (Dijkstra et al., 2019), showing proficiency is posited to be a significant factor.

Ultimately, the characteristics of the salient target language must play a role in determining the way it is processed and so there are likely to be some similarities between L2 readers regardless of L1 (Tan et al., 2005). It is, however, reasonable to accept that L2 word recognition is influenced by and at least partly based on L1-related VWR skills being used as a foundation (M. Wang et al., 2003), though how they manifest in L2 processing may not be the same across languages and interlingual influences may work both ways (Brysbaert & Dijkstra, 2006).

3.3 Interlingual influences

Intrinsic to the integrated lexicon of BIA, BIA+, and Multilink models of bilingual word processing, it is accepted that L1 influences L2 processing (van Heuven & Dijkstra, 2010; Dijkstra et al., 2018). With both L1 and L2 being part of one integrated lexicon, it follows that lexical entries between languages can interact similarly to how lexical entries within a language can (van Heuven & Dijkstra, 2010), including during word identification (Brysbaert & Duijck, 2010). As a key feature of language non-selective access (e.g., van Heuven, Dijkstra, & Grainger, 1998), information of language membership is not required to access the integrated lexicon and the result is parallel activation of similar/related lexical items from both L1 and L2 (van Heuven & Dijkstra, 2010). The practicality of this bilingual system can appear lacking though. When a visually presented word also activates orthographically or phonologically similar or neighbouring words from L1 and L2, a process of inhibition, such as a task-dependent decision process (Dijkstra et al., 2018), is theoretically required to avoid interlingual conflicts and related performance errors (van Heuven & Dijkstra, 2010). However, it has been shown that bilinguals are often unable to successfully control the processing in the word identification system (van Heuven & Dijkstra, 2010), resulting in processing conflicts between L1 and L2 that cannot be avoided (van Heuven, Schriefers, Dijkstra, & Hagoort, 2008). L2 visual word recognition is, therefore, influenced by L1 (Brysbaert & Dijkstra, 2006), as the existing language (L1) network forms a foundation for L2 acquisition, providing L1 skills where applicable and assimilating the L2 (Perfetti et al., 2010). To what extent, in what circumstances, and how L2 processing might differ based on language profile, however, still requires much investigation. In particular, the timing of VWR processes in the brain during L2 reading, especially when taking the language profile of the L2 reader into account and at early timeframes, has not been widely addressed in the literature. The time-course of processing during reading is key to a number of questions about how L1 and L2 readers successfully achieve VWR and, therefore, needs to be understood (Sereno & Rayner, 2003).

L1-oriented brain activations have been observed during L2 reading, strongly supporting the idea that L1 reading skills influence L2 reading (Kim, Liu, & Cao, 2017). Interlingual phonological priming between L1 and L2, for instance, strongly supports a significant role of L1 in L2 processing (van Heuven et al., 2008), even to the extent of suggesting that L2 reading consistently but unconsciously activates L1 and that L2 is read using L1, if only as a mediator (e.g., Wu & Thierry, 2010). Interlingual phonological priming has also been shown to occur in both directions between L1 and L2 to comparatively equal extents (Van Wijnendaele & Brysbaert, 2002), which parallel activation in an integrated lexicon also explains (van Heuven & Dijkstra, 2010). Such interlingual influence also extends to word form neighbours of both L1 and L2, which have also been found to activate the L1 and L2 counterparts in parallel (Dijkstra et al., 2018), providing further support for both languages being activated in parallel (van Heuven & Dijkstra, 2010). Furthermore, phonological properties of L1 have been shown to influence L2 phonological processing of English using homophones and near-homophones in both Japanese-English (alphabetic L1) and Arabic-English (non-alphabetic L1) bilinguals (Ota et al., 2009). This directly supports the idea that L1 phonology can influence L2 VWR in bilinguals with an alphabetic L1 (e.g., Spanish-English) or a non-alphabetic L1 (e.g., Chinese-English) when reading English, which feeds into how the L1 influence on L2 processing could differ between bilinguals with different language profiles, such as whether it is a property of L2 processing when L1 and L2 are distinct (e.g., English and Chinese) or whether it is supported when L1 and L2 are linguistically closer (e.g., English and Spanish).

On the surface, it might reasonably be expected that psycholinguistic processing in bilinguals who are equally proficient across L1 and L2 would be equivalent when reading either language. Indeed, bilinguals have been shown to use native-like processing strategies for reading L2 (Diependaele et al., 2011; Lemhöfer et al., 2008), using the same brain networks as processing L1 (van Heuven & Dijkstra, 2010), and relying on whole-word processing in L2 reading (Clahsen et al., 2010). However, any language-specific differences between L1 and L2 would still need to be considered, as distinctions between languages (e.g., orthographic depth) have been shown to influence both behavioural performance and brain activity (Jamal et al., 2012; Luke et al., 2002; Meschyan & Hernandez, 2006). The extent of reliance on language-dependent VWR strategies while reading L2 and the extent that L1- or L2-based processes are involved, especially for late bilinguals with a dominant L1, are still relatively unknown though. Psycholinguistic processing of the same L2 (e.g., English) between different groups of equally proficient bilinguals (e.g., Spanish-English and Chinese-English) could be the same as one another, driven by the target language or could be entirely different from other bilinguals and from native readers. However, a scenario in which reading L2 uses expertise from both L2 and L1 is arguably more likely than relying on L1 or L2 skills exclusively (Jamal et al., 2012), especially when considering how different L1 and L2 can be linguistically, but it also requires much more investigation.

3.4 Summary

These opening chapters have introduced what written language of different systems and types can entail and given an overview of the main theoretical perspectives on monolingual and bilingual VWR, respectively, providing descriptions of the key overarching factors. The next chapter will take these further into the specifics of how processing orthography and phonology are understood, including how they can interact during VWR, such as through phonological mediation and the process of orthography-phonology mapping.

Chapter 4: Orthographic and phonological processing in VWR

4.1 Orthography in VWR

As the linguistic foundation of written language and necessary for any VWR processing to occur at all, it seems fitting to start the breakdown of orthographic and phonological processing during VWR with orthography. Orthographic processing is, using terminology from the Interactive-Activation (IA) model (McClelland & Rumelhart, 1981), separated into the concepts of feature detectors (for visual features of letters e.g., roundness of a letter *O*), letter detectors (orthographic identification of letters based on visual features), and word detectors (overall orthographic recognition of lexical items based on orthographic processing). In the context of VWR, these cover a wealth of factors potentially involved in orthographic evaluation, such as number of letters, number of graphemes, whole-word visual familiarity, n-gram frequency, orthographic content and letter order, word shape, visual clarity, letter size, letter spacing, letter constancy, and font style to name a few. Each of these variables contributes to word recognition to varying extents and on different levels, at different times, and in different combinations to constitute orthographic processing, which involves both visual and psycholinguistic properties (Seghier et al., 2014). It should be noted, however, that several of these factors are more visual than orthographic in nature, such as visual familiarity, visual clarity, and word shape.

The distinction between visual processing and orthographic processing in the context of VWR and reading is an important one. Both are inherently related in VWR with orthography providing the rules of a language's writing system and therefore the framework for the physical visual representations of words (Perfetti & Liu, 2005).

However, orthography does not account for all visual features of linguistic stimuli (e.g., lines, word size, colour) and so the initial stage of word recognition involves such non-linguistic visual features through pre-lexical processing that precedes orthographic processing, underpinning the whole word identification and comprehension process (Grainger & Holcomb, 2009). Orthographic processing, meanwhile, refers to the encoding of the visual forms of a language, including letters/characters, letter/consonant clusters and graphemes, and arguably even word shapes and fonts. Essentially, initial visual processing in the context of VWR is pre-lexical activation of visual features that are not necessarily linguistic and orthographic processing entails visual processing in a psycholinguistic framework (Grainger & Holcomb, 2009; Seghier et al., 2014).

Considering the dynamic nature of psycholinguistic processing and potentially infinite combinations of finite linguistic parameters (Fischer-Baum et al., 2014), it is worth considering orthographic processing with a "parametric approach" (Tagamets et al., 2000). Instead of assuming discrete differences when psycholinguistic parameters differ, as shown in orthographically differing stimuli (e.g., words, pseudowords, letter/symbol strings), orthographic familiarity is considered along a continuum on which words, pseudowords, symbol strings (e.g., ¢µ€$), letter strings (e.g., yhgdr), and false fonts (e.g., ʅƉlꞩɫꭗ) reside based on how legal and/or legitimate they are within a language or context. Moreover, orthographic familiarity can be considered on a continuum in terms of smaller linguistic units than full word-like stimuli (e.g., letters, morphemes), which could be the objects of statistical learning strategies of language (cf. Dehaene, 2014). In other words, statistical knowledge of orthography can be described and defined in terms of characters (including letters, punctuation, diacritics etc.), n-grams (e.g., bigrams, trigrams), and morphemes, as well as whole words (Hirshorn et al., 2016).

Based on the concept of open bigrams (Grainger & van Heuven, 2003) and pairs of adjacent and non-adjacent letters (Grainger & Holcomb, 2009), it could be that orthographic processing is not necessarily detecting letters to compute their identity (and position), but detecting grapheme, n-grams, or "chunks" (Grainger & Holcomb, 2009). Rather, orthographic processing is more likely to be a combination of the two: letter identification for first and last letters, and grapheme/chunk retrieval dependent on first and last letters i.e., all potential chunks surrounding them. This entails that the mental English alphabet far exceeds 26 units and even ~40 graphemes in a practical sense due to its significant orthographic depth and irregularity, instead including all viable (legal and legitimate) letter cluster combinations of formal alphabetic units. Furthermore, the serial/parallel processing debate in terms of orthography can refer to whether letter identification, grapheme detection, or another method of orthographic perception takes precedence, but the key question in focus is the timeframe and serial/parallel processing of orthographic and phonological elements, respectively.

4.2 Phonology in VWR

One of the most significant topics in VWR concerns the extent that phonology is involved in VWR, referring to the way phonological activation works to facilitate or inhibit comprehension, including if it always occurs regardless of context, if it always occurs but is differentially influential, or if it is a dynamic process that only occurs under particular circumstances (Fischer-Baum et al., 2014). The question of why phonological activation occurs during reading is also noteworthy and can largely be attributed to the phonological nature of natural human language and the process of "mapping" visually presented words onto known phonological internal constructs when learning to read. This is strongly related to the notion of phonology as a universal component of the reading process and the ontogenetic primacy of phonology in language generally (Grainger & Ferrand, 1994; Slobin, 2003), as natural languages are primarily phonological and speech-based (Ferrand & Grainger, 1994; Frost, 1998). Furthermore, the importance of phonology does not appear to be specific to English and is grounded in this phonological foundation of natural human languages (Diependaele et al., 2010). Although not surprising, it is important to note that research has shown phonological effects in languages other than English, such as French (Ferrand & Grainger, 1994), Chinese (Xu et al., 1999), Dutch (Brysbaert, 2001), even Serbo-Croatian (Lukatela & Turvey, 1990)and Hebrew (Gronau & Frost, 1997). This has been taken to suggest the involvement of phonological activation in reading or, at least, some phonological effects to be universal (Brysbaert & Wijnendaele, 2003). Considering that writing systems, first and foremost, represent spoken language and not meaning directly (Perfetti & Liu, 2005), it follows that phonology is firmly placed in the reading process. While phonological activation, though not the phonology itself, could be universal, the means of achieving it from a visually-presented stimulus is probably not universal, being dependent on such factors as reading proficiency, native language, and stimulus language, as well as the ever-present factor of context. The basis of written language on sounds and the apparent universality of that is enough to suggest that reading acquisition is essentially the mapping of orthographic codes onto existing phonological ones (Ehri & Wilce, 1980; Grainger & Holcomb, 2009).

Phonology having such a supporting and significant purpose in VWR is also reinforced by research into beginner readers and the acquisition of reading skills, which underlines the importance of phonological awareness and the phonics method of teaching in reading skill acquisition (Ehri & Wilce, 1980; National Reading Panel, 2000; Pogorzelski & Wheldall, 2005). This extends to bilingual development too, as learnt phonology of known language(s) has been shown to influence the learning and comprehension in additional language acquisition (Mulík & Carrasco-Ortiz, 2021). Furthermore, phonological impairment from both specific learning disorders, such as dyslexia, and psychophysical disorders, such as hearing loss and deafness, are associated with issues for reading performance (Classon et al., 2013). If phonological processing was not functionally useful, this would not be the case, suggesting that phonological activation in skilled readers is not epiphenomenal and possible functions of phonology in the mind include supporting lexical access and helping comprehension (cf., phonological mediation), as well as reinforcing short-term memory (Classon et al., 2013; Leinenger, 2014). The propositions, therefore, that beginner readers would prosper from prior and enhanced phonology and that skilled readers can also benefit from an internal phonological activation are supported by the phonological basis of written language, the importance of phonology for reading skill acquisition, and the presence of subvocalization during reading. Conversely, there are still assertions that rapid phonological processing might not serve a functional purpose (Rastle & Brysbaert, 2006), leading to the open question concerning the extent of its use in different paradigms and the ways in which it can facilitate or even inhibit lexical access. However, it seems unlikely that phonological activation during reading is just a non-functional by-product. This also fits with the notions of language as predominantly phonological, of written language being based on speech (that precedes it in terms of both development and evolution of all natural languages), and the precept of language comprehension being based on triggering existing knowledge of the linguistic input (Dehaene, 2014; Snowling & Hulme, 2005).

Some of the clearest evidence for phonological involvement during VWR has been asserted to come from effects of phonology in paradigms that do not, in theory, explicitly require phonological processing or when the effects of phonological processing have a detrimental impact on VWR performance and reading comprehension (Frost, 1998). In such circumstances, such as reading visual tongue-twisters (e.g., "Take a taste of tender turtle"; McCutchen & Perfetti, 1982), reading without phonological activation would arguably lessen conflicts and increase efficiency, but the brain appears to process phonology anyway, so the conflicts occur (Alario et al., 2007; Ferrand & Grainger, 1994). Phonological effects in circumstances when it is not required provide evidence of phonological mediation and involvement in VWR, suggesting a level of phonological processing to be present during silent reading, somewhat irrespective of the purpose of the activity. Such phonological effects can be used to refute theories of phonology as backgrounded (i.e., weak phonological theories), but support theories that foreground phonology and its consistent activation during VWR (i.e., strong phonological theories). However, an apparent lack of phonological activation when it is required for the task could support the reliance on the orthographic/direct pathway and, depending on the evidence, that orthographic activation is stronger than phonological, at least in some circumstances, strongly indicating a dynamic, but imperfect context-, stimulus-, and task-dependent system that, while often accurate, does not always get the balance right (Fischer-Baum et al., 2014).

Importantly, phonological priming effects have been shown to be relatively unaffected when the context of the phonological (re)coding is problematic or even inhibitive (Brysbaert & Dijkstra, 2006), such as pseudohomophone/pseudoword lexical decisions in a phonological lexical decision task and orthographically congruent non-rhymes in a rhyme judgement task. This suggests that phonological activation and, indeed, the reliance on phonology for VWR cannot be consciously controlled through objective- or task-based reading strategies (Brysbaert & Dijkstra, 2006). Such phonological priming effects are, therefore, prelexical and, accepting that prelexical phonological activation is a vital factor in VWR generally, it follows that it would also significantly impact bilingual VWR (Brysbaert & Dijkstra, 2006), especially when bilinguals' L1 and L2 use the same or similar orthography e.g., Spanish-English bilinguals.

Dynamic processing of orthographic and phonological input also fits with the matter of task-dependence and that observing effects in one paradigm does not necessarily mean they will (or will not) be present in another (Rastle & Brysbaert, 2006). For instance, phonological processing in word-naming is more necessary (McNorgan et al., 2015) and more consistent (Ashby et al., 2009) than in lexical decision and other silent reading activities, so processing extents and differences would likely be different between tasks, which has been proposed as a potential issue for using phonology-dependent tasks to investigate phonology (Frost, 1998). However, differences in the extent of phonological activity are likely due to the requirement of phonology, such as the need for overt phonology in naming that is not necessarily present for lexical decision, which just highlights the dynamic nature of VWR and, provided appropriate care is taken in interpretation, does not invalidate any approach to VWR research. While evidence of phonological processing during VWR from experiments that do not explicitly require it would be ideal, it is still not fully understood what constitutes phonological processing in the brain or how it works. Considering the variety of evidence, it is perhaps an understatement to say that phonology has the capacity to be influential to word recognition, whether in terms of facilitation and positive priming or inhibition that leads to confusion and performance errors, but further investigation is required to understand the whole role of phonology in VWR. In particular, the relative strengths of orthography and phonology during VWR in different contexts (e.g., orthographic lexical decision, rhyme judgement) are still relatively unknown, as well as how phonological processing and the aforementioned interaction with orthography works in L2 readers with different L1 language types (e.g., Spanish-English and Chinese-English late bilinguals).

4.2.1 Phonological mediation

While there is a compelling argument that phonology is a necessary part of VWR and more vital than some (e.g., weak phonological) theories typically allow, phonology is still not collectively deemed essential to VWR and the possibility of a balanced usage of the orthographic and indirect routes is not always embraced (Luo et al., 1998). For instance, lexical access might only be indirect or phonologically-mediated when the stimuli are low frequency words, otherwise unfamiliar (e.g., non-words), or just difficult to read (Xu et al., 1999). Therefore, skilled readers might rarely use phonology in VWR, instead using the direct orthographic-lexical pathway unless the stimulus or task demands the involvement of phonology (Van Orden, 1987). Early beginner readers, however, will rely heavily on the phonological route because initially and until they become more skilled, all language-like inputs are low frequency words, unfamiliar, or just difficult to read (Abramson & Goldinger, 1997).

The mental manifestation of phonological mediation – whether discussed in terms of internal phonology, phonological recoding, subvocalization (Baron, 1973; Leinenger, 2014), or "inner speech" – is essentially a translation to the internal articulation of a word for it to be heard in the mind (Vygotsky, 1986, p. 225) and an integral part of the speech planning process (Wheeldon & Levelt, 1995). It is described here as a translation and not tethered to reading processes because such subvocalization also occurs when thinking and does not require visual presentations. It is, therefore, a translation from one cognitive code (whether orthographic or not) to the internal phonological representation. This is distinct from phonetic-acoustic articulation, overt speech, and the associated explicit motor function that is required for reading aloud but not necessary for silent reading (Abramson & Goldinger, 1997; Brysbaert & Wijnendaele, 2003). The salience of subvocalization can vary from reader to reader and word to word, being most reported for more difficult and unfamiliar words (Brysbaert, 2001). However, it is conceivable that subvocalization is automatic and always happens, but is simply not always salient enough to be consciously realized, perhaps only occurring in certain circumstances, such as when the stimuli or task demand it. The existence of subvocalization in the minds of both beginner and skilled readers, however, does suggest phonology to have a prominent purpose in VWR.

One classis example of this internal phonological recoding and its effect on reading comprehension is readers "hearing" a homophone of a word semantically related to a category, such as responding in the affirmative to the word *brake* being a word related to the category *car* when the visual stimulus was actually *break* (Van

Orden, 1987). This effect shows automaticity and clear reliance on internal phonology, at least in some contexts. Such errors also occur in skilled readers and with high-frequency as well as low-frequency words, despite the homophone only having the phonology and not also the orthographic that is linked with the semantic content of the category (Van Orden & Kloos, 2005). The phonological similarity between the stimuli (the homophones in this case) leads to poorer behavioural performance and suggests that phonological processing can occur even when it inhibits accurate performance (Frost, 1998). It is reasonable to accept such findings as evidence for phonological activation, while noting the strength of that phonological activation and its ability to override orthographic and even lexical/semantic processing outcomes. Importantly, it could suggest automatic phonological activation and that, when available, it will be relied upon over and above other psycholinguistic codes, which makes sense in terms of VWR being evolutionarily and perhaps cognitively based on speech (Dehaene, 2014; Menary, 2014; Snowling & Hulme, 2005). The same pattern found in English readers of phonology-based semantic priming has also been observed in Dutch readers (Drieghe & Brysbaert, 2002), proposing that such an early phonological mechanism might be similar across readers of similarly orthographic languages (e.g., alphabetic languages). Likewise, phonological priming effects were found to be similar between native English monolinguals and Dutch-English bilinguals, which could again be accounted for by the orthographic similarities between English and Dutch (Timmer & Schiller, 2012). However, it remains to be seen how this extends to other bilingual groups with an alphabetic L1 (e.g., Spanish-English) and certainly to bilinguals with a non-alphabetic L1 (e.g., Chinese-English).

In research contexts (commonly through priming tasks), phonological mediation is observed when phonology is activated by orthographic input before lexical access is achieved, thus showing the involvement of phonology facilitating VWR. It provides one explanation for how word recognition is still possible when more than just the physical visual representations are required to easily recognise a word, such as with irregular words (e.g., *knight*), loan words (e.g., *cul-de-sac*), phonologically incongruent homographs (heteronyms, e.g., *bass* for low frequency/pitch and *bass* for the fish), and even words with pronunciations dependent on context or co-text (e.g., abstract; Perea & Carreiras, 2008). By extension, it could also explain successful recognition in cases of "orthographic poverty" when words are modified or distorted, such as in orthographically and phonologically similar pseudowords (e.g., *tesk*, *tecks*), orthographically odd pseudohomophones (e.g., *tekst*), and transposed-letter stimuli (e.g., txet, anwser; Frankish & Turner, 2007). Furthermore, phonological priming effects where target identification is facilitated by pseudohomophones i.e., not real words, such as through masked priming of a real-word target e.g., *FLOOR* by a pseudohomophone prime e.g., *flore* (in contrast with a pseudoword control prime e.g., *flop*), shows that the phonological activation occurs due to sublexical grapheme-phoneme correspondence as opposed to any word-level lexical processing (Brysbaert & Dijkstra, 2006).

Lexical access during VWR has also been shown to be influenced by both homophones and pseudohomophones of semantically-related words e.g., *main* and *mayn* would both facilitate reading *lion* due to their phonology being the same as *mane* (Lukatela & Turvey, 1994). Word-naming responses to a target word are expected to be faster following semantically-related words than controls (i.e., facilitative semantic priming), but this is based on the actual semantically related word being presented, which is not the case with these homophones and pseudohomophones. Nonetheless, semantic activation and lexical access are clearly facilitated by the phonology of such stimuli (Lukatela & Turvey, 1994; Van Orden & Kloos, 2005). The pseudohomophone evidence is especially important because pseudohomophones, unlike homophones and other real words, do not have legitimate orthography, so would have no whole-word orthographic representation in memory to activate either the corresponding phonological or the semantic representation of the real-word counterpart (i.e., the whole-word orthography alone of *mayn* cannot activate the phonology or semantics of *mane*). This could suggest that phonology can at least influence word recognition earlier than or in parallel with orthography (Leinenger, 2014), even if it does not consistently do so in normal, less task-dependent circumstances.

Some of the strongest evidence for phonological mediation in visual word recognition comes from priming experiments, especially the masked priming paradigm (e.g., Bijeljac-Babic et al., 1997; Brown & Hagoort, 1993; Brysbaert, 2001; see also Rastle & Brysbaert, 2006). Using such stimuli, effects associated with phonology have been proposed to occur earlier than effects associated with orthography (Hauk, Patterson, et al., 2006; Wheat et al., 2010), which does not tally with a purely serial account (if it is accepted, as is reasonable, that orthography must be processed before phonological processing can begin). However, rapid phonological activation and parallel processing could explain such activity, while also concreting the role of phonology in VWR.

Support for phonological mediation from results of word-naming tasks have also suggested that phonological processing can take priority over direct orthographic routes (Ashby, 2010; Lukatela & Turvey, 1994), though this could just be a side effect of task-dependence and context, again proposing dynamic processing (Van Orden & Kloos, 2005). The importance and possible primacy of phonological mediation poses an important proposition for models of reading aloud, implying that silent reading could involve different mechanisms and not simply be the precursor to reading aloud without the motor and articulatory actions. Indeed, the observation of phonological but not orthographic priming effects on behavioural performance proposes that phonological processing takes precedence when the task involves an overt performance of the phonological code (Timmer & Schiller, 2012). This is perhaps to be expected and, generally, it shows how the task at hand can influence processing (cf. top-down processing). Considering this comparison in terms of a silent reading task that does not necessarily require phonological activation (lexical decision) and a reading aloud task that does require explicit and overt phonology (word naming), one way to build upon it would be to explore phonology in silent reading tasks that require it explicitly such as in the phonological lexical decision task, pseudohomophone processing, rhyme recognition employed in the current research.

4.3 The mapping of orthography and phonology

There is little question that phonology holds a place in the reading process, but its importance, purpose, and position in the time-course of that process are still unclear. Investigations into such things often focus on whether phonological activation is pre-lexical or post-lexical, referring to a time before or after finding the entry in memory, respectively (Leinenger, 2014). Pre-lexical phonological activation entails direct orthography-phonology conversion, while post-lexical activation suggests that an orthographic-lexical pathway has been used and the phonological representation is activated from it. Some theories posit that orthographic input is rapidly converted to internal phonological representations on which the word recognition process is then based (Rastle & Brysbaert, 2006). Indeed, it has been suggested that early, pre-lexical phonological activation is automatic in the sense that it is rapid, involuntary and, as Perfetti, Bell, and Delaney (Perfetti et al., 1988) put it, "nonoptionally" (Ashby, 2010; Ferrand & Grainger, 1994; Gronau & Frost, 1997; Lukatela & Turvey, 1994). If this is the case - even if it is also epiphenomenal (Abramson & Goldinger, 1997) - the question of whether phonology is necessary for lexical access is moot, as it is an involuntary prerequisite in the reading process. The questions that are still very much open concern the extent of involvement in both different contexts (e.g., phonology-dependent and orthography-driven) and different populations (e.g., ESL bilinguals with other alphabetic or non-alphabetic native languages), including how orthographic and phonological processes interact during VWR and what timecourse they take.

The orthographies of alphabetic (e.g., English, Spanish) and logographic (e.g., Chinese) languages work differently, much of which relates to the information that can be obtained from the visual representations alone. This is of major importance when considering L2 reading and the influence of L1 skills, such as the way native logographic readers process alphabetic scripts as in Chinese-English bilinguals. Orthography-phonology mapping, therefore, presents a clear contrast between reading alphabetic languages and reading logographic scripts. This link between orthography and phonology represents one of the most significant questions of how reading is achieved, especially by L2 readers with an L1 that does not have such direct connections (e.g., Chinese-English bilinguals).

Grapheme-phoneme correspondence can vary dramatically between written languages, which manifests itself in how phonology maps onto orthography and the extent of the relationship between graphemes (orthography) and phonemes (phonology). While the influence of phonology on VWR can be considered in terms of strength (i.e., weak vs strong phonological accounts, as mentioned earlier), the relationship between orthography and phonology in a language can be labelled in terms of orthographic depth, being shallow or deep or anywhere in between and referring to a scale that indicates how closely the visual representations are to their phonetic ones, involving the regularity of orthographic vowel and n-gram pronunciation (Harley, 2010; Katz & Frost, 1992). For example, the English word *kite* is more directly related to its phonetic representation than the English word *knight*, which is highly irregular due to the silent *k*, *gh* consonant cluster, and lack of word-final *e* (to inform the vowel shape). Albeit with many regular words that use relatively clear grapheme-phoneme conversion rules, the English language has a higher proportion of irregular words than many other languages, described as "quasi-regular" (Plaut et al., 1996), and is, therefore, considered to have a deep orthography.

Orthographic depth is primarily a factor of languages with direct links between orthography and phonology (e.g., grapheme-phoneme conversion rules that describe grapheme-phoneme correspondence), as found in alphabetic languages, such as English and Spanish (Ellis et al., 2004). According to the orthographic depth hypothesis (Katz & Frost, 1992), shallow orthographies should be easier to read, especially when the writing system involves phonological cues, which relates to the importance of phonological recoding in different languages. This is demonstrated through English being less phonology-dependent than Serbo-Croatian, but more phonology-dependent than

Hebrew (Frost et al., 1987). Furthermore, languages with shallow orthographies should allow learning how to read aloud and spell more quickly than languages with deeper orthographies (Ellis et al., 2004). Children from many European countries were found to be proficient in reading familiar words and pseudowords before the end of the first school year with word reading accuracy of >90% in all but Portuguese, French, Danish, and, markedly, English (Seymour et al., 2003). These languages all have deeper, more opaque orthographies than the languages for which >90% accuracy was achieved (e.g., Spanish and even German). Furthermore, the rate of learning English in particular is worse than twice as slow as other orthographies (Ellis et al., 2004). However, unlike Spanish orthography, for example, which can generally be trusted for grapheme-phoneme conversion links, English is orthographically deep, hence English children finding it more difficult to read pseudowords, pseudohomophones, and low frequency words (Perfetti & Liu, 2005). While other factors are involved in differences between learning and reading different languages (such as differences between education systems and pedagogies beyond the scope of the current research), orthographic depth clearly has a significant impact (Paulesu et al., 2021). This is a particular point of interest for L2 learners and readers, especially ones with orthographically and/or phonologically distinct native languages (e.g., Chinese-English bilinguals), as the profound effect of the writing system and orthography is likely even more profound when reading new L2 words, L2-based pseudowords, pseudohomophones, and low frequency words.

The importance of bilinguals' native language for processing an alphabetic L2 often relates directly to the orthographic depth hypothesis (Katz & Frost, 1992) or, at least, the relationship between orthography and phonology. The continuum between shallow and deep orthographies is analogous to the extent of reliance on phonological information compared with visual/orthographic information, which further distinguishes between English, Spanish, and Chinese native languages profiles. Grapheme-to-phoneme conversion, for instance, in languages with deep orthographies e.g., English is potentially more difficult, especially for L2 learners with an orthographically shallow L1 (e.g., Spanish) or an L1 with a very differently complex orthography (i.e., without GPC, e.g., Chinese), at least until it is sufficiently practiced (Seymour et al., 2003).

Spanish and English are alike as far as using a largely similar alphabet, though orthographic depth differs considerably with Spanish being much more phonological and open to more straightforward grapheme-phoneme mapping, making it easier to learn (Seymour et al., 2003). In contrast with English, Spanish has a relatively shallow orthography with relatively few irregular words and almost one-to-one grapheme-phoneme correspondence. It can, therefore, be considered an especially phonological language that naturally encourages orthography-phonology pathways. Despite the difference in orthographic depth, however, Spanish is relatively close to English on the Language spectrum and, due to using an alphabetic writing system and largely similar orthography, objectively much closer than Chinese. It remains to be seen, though, whether these similarities are an advantage or a double-edged sword in practice: the prior experience and expertise with a very similar alphabetic script could provide an advantage for Spanish-English bilinguals reading English, though the different dependency on phonology in English will also play a major role in orthographic-phonological processing. For instance, the visual/orthographic similarities could increase the difficulty of reading English low frequency words as well as pseudowords and pseudohomophones in particular, which all essentially force the reader to resort to a grapheme-phoneme conversion process. Processing Spanish words phonologically is likely the easiest route due its nature, but doing the same in English can be more cumbersome (e.g., irregular words), albeit sometimes essential to determine intended meaning (e.g., homographs). However, considering Spanish as a very phonological or even “phonetic” language with a shallow orthography, it follows that native Spanish readers will automatically take a phonological approach to English VWR, especially if L2 processing uses L1 foundations, as the current research is investigating.

Languages with alphabetic (e.g., English, Spanish) or syllabic (e.g., Japanese kana) writing systems typically have orthographies with direct links to phonology, which can vary in strength due to orthographic depth and phonological transparency, while logographic languages (e.g., Chinese) simply do not have such orthographies. Ordinarily, Chinese would not be considered to have "orthographic depth" due to its logographic (non-alphabetic) writing system and the lack of direct relationship between its characters and their pronunciations. If considered in such terms, though, Chinese (and other logographic languages) could be described as mirroring a deep orthography, even conceptualized as bottomless, due to the visual/orthographic focus and lack of direct phonological cues in such written forms (Ellis et al., 2004).

Approaching Chinese phonologically would be largely fruitless in most cases, especially without substantial expertise, due to its visual/orthographic basis and general lack of direct phonological markers (Perfetti et al., 2010; Tan et al., 2005). Instead, Chinese phonology is retrieved from orthographic input through a direct (addressed phonology) pathway or an indirect (assembled phonology) pathway (Tan et al., 2005). It has been suggested that addressed and assembled pathways to phonology in Chinese VWR roughly correspond to the direct lexical and non-lexical/indirect routes for alphabetic scripts (e.g., DRC), but implemented in distinctly different ways (Lu et al., 2011). For instances, assembled phonology is distinct from grapheme-phoneme conversion due to there being no sub-syllabic links or reliable cues for pronunciation in Chinese (Perfetti & Liu, 2005), resulting in the divergent and undefined orthography-phonology relationship (Lu et al., 2011). However, the majority of Chinese characters do include a phonetic component that provides an indication of pronunciation along with the semantic radical that specifies meaning (Hoosain, 1991; L.-J. Kuo et al., 2015; Williams & Bever, 2010). Some expert readers of Chinese may use the statistical information expressed through phonetic radicals as an indirect sublexical route to pronunciation (Dehaene, 2014; W.-J. Kuo et al., 2004; Lee et al., 2004). The statistical information contained in phonetic radicals is essentially an indication of frequency and familiarity, as opposed to direct grapheme-phoneme correspondences (Mechelli et al., 2005). This draws a broad comparison with grapheme-phoneme conversion, but also somewhat of a parallel with alphabetic scripts in that frequency and regularity determine pronunciation of otherwise undefined combinations of letters, such as the vowel cluster -*ea*- being /i:/ in English unless defined in an exception word (e.g., *bear*). As characters are recognised via a direct orthography-semantic route, however, this apparently phonological outcome from sublexical evaluation still likely has a semantic basis (W.-J. Kuo et al., 2004; Perfetti & Liu, 2005). Reading such a logographic language, therefore, entails recognising whole characters, resembling a direct lexical-semantic route of VWR that essentially bypasses or does not require phonology. This also proposes phonology to be a by-product, echoing the debates of weak or strong phonological theories of reading alphabetic scripts.

Although phonology is consistently activated (through addressed phonology) when reading Chinese, it appears to be at least backgrounded, perhaps being unnecessary, and not as important as the focus on orthography (Perfetti et al., 2010), which is understandable considering the strongly visual and phonologically implicit nature of the language. However, it can lead to greater potential for orthographic confusion in terms of relying on orthography when phonological processing is needed (e.g., reading phonologically-incongruent homographs/rhymes). Conversely, less phonological coding might entail less potential for confusion from phonology, but it ultimately leads to increased difficulty for phonological tasks that require explicit phonological processing (M. Wang et al., 2003; Zhan et al., 2013). For instance, presentations of "mixed pseudohomophones" (one of two graphical components is replaced with a homophonic component; an English example might be *phish*) and "pure pseudohomophones" (both graphical components are replaced with homophonic components; an English example might be *kace*) resulted in a lack of phonological activation in native Chinese readers to the pure pseudohomophones, mirroring non-word responses (Zhan et al., 2013). These findings demonstrate the skills to process the strong visual/orthographic nature of Chinese that suggest processing English by Chinese-English bilinguals occurs through learning English words as whole units and reading them in the same direct fashion as skilled native readers and as Chinese characters are read (Ellis et al., 2004). Alternatively, some form of grapheme-phoneme conversion route could be learnt to decipher alphabetic words, which often require orthography-phonology links to be read.

The lack of regularity and increased orthographic depth is a point about English that could prove problematic for both ESL groups considering its high sound-symbol relationship beside the shallow orthography of Spanish and the relative lack of phonological cues in Chinese. For instance, Chinese-English bilinguals (non-alphabetic L1 orthography) have been shown to make more errors than Korean-English bilinguals (alphabetic L1 orthography with orthographic depth comparable to English) in a phoneme deletion task using English stimuli (M. Wang et al., 2003). Moreover, many of the errors made by the Chinese-English group were phonologically incorrect but orthographically acceptable, suggesting that phonological stimuli/tasks in an alphabetic L2 can be problematic for bilinguals with a logographic L1. Such evidence also suggests that reading non-alphabetic scripts, such as Chinese, use and even rely on orthographic information more than phonological information due to the visual/orthographic nature of their L1 (M. Wang et al., 2003), especially when processing alphabetic scripts, such as English.

Whether in terms of orthographic depth or orthography-phonology mapping, the relationship between orthography and phonology offers a critical contrast between languages. Its manifestation, including how phonological cues in scripts can drastically differ between languages, is likely to affect L2 reading performance and processing, which the current research aims to examine.

4.4 The temporal nature of orthographic and phonological processing

One crucial question concerning the nature of both orthographic and phonological processing is whether phonology directly follows some form of orthographic activation or whether phonology can be processed in parallel with other orthographic processes. Somewhat overarching the main theories of VWR as discussed earlier is this conceptualisation of processing within and between orthography and phonology as "overlapping" (Grainger & Holcomb, 2009), "cascaded" (DRC; Coltheart et al., 2001), and even "parallel" (PDP; McClelland & Rogers, 2003). Some orthographic processing is required for phonological processing to begin and overlap before further orthographic processing proceeds. The notion that reading involves cascaded processes with interactive elements has been an intrinsic part of many VWR theories for some time (Barber & Kutas, 2007; Dien, 2009; Grainger & Holcomb, 2009; Pulvermüller et al., 2009; Sereno et al., 2003). Essentially, while one process must logically hand off to another at some point, it appears it is more like a relay race than the flow chart often ironically used to illustrate such models.

Several perspectives, such as the DRC and BIAM, share the concepts of overlapping processes and of phonology generally following orthography. However, it is possible that not all orthographic properties are required for phonology to be activated or that different properties or combinations of properties work together in different contexts. Different levels of reading performance in different situations based on task demands suggests the latter is most likely and that the reading process is dynamic (Fischer-Baum et al., 2014). Considering this variance in reading performance and variety of psycholinguistic effects (and the various interactions between them), it is not unreasonable to accept VWR as a dynamic process dependent on variable contextual cues and, whether in series or in parallel, a combination of perceptual, orthographic, phonological, semantic, and pragmatic information being required for successful VWR (Laszio & Federmeier, 2014). Dynamic processing in reading essentially entails that different possible routes to VWR can be more or less efficient dependent on the stimuli, task, and context at hand (Cecere et al., 2017). This is where the idea of interactions between adjacent processes (e.g., orthographic, phonological) as not entirely serial or distinctly parallel becomes important.

Processing during reading likely involves "sequential overlapping steps" (Holcomb & Grainger, 2006, p. 1631), a key phrase that somewhat embodies the BIAM (Grainger & Holcomb, 2009). It suggests that processes are not purely serial or parallel, but involve some serial steps and other processes in parallel along with interactions between them at several points in the time-course where orthographic and phonological processes “cooperate” (Grainger & Holcomb, 2009). Furthermore, different processes may not only operate in a feed-forward, stage-wise fashion, but also work with feedback mechanisms (Twomey et al., 2011), such as the proposition in the DRC that feedback from phonological analysis to orthographic processing can occur (Coltheart et al., 2001). Communication between cognitive processes could, therefore, be bidirectional and, if the feedback is distributed in time, it can influence word recognition (Grainger & Holcomb, 2009). Lexical and phonological routes (e.g., of dual-route models), while both based on the same low-level perception, theoretically interact through reciprocal connections that can feed backwards as well as forwards (Dien, 2009), using bi-directional and bi-modal activation (cf. BIAM; Grainger & Holcomb, 2009; Twomey et al., 2011). The reading process could, therefore, be not entirely serial and not entirely parallel either, instead involving a dynamic combination of serial and parallel processing in terms of stimulus, task, and overall comprehension context-dependence (Fischer-Baum et al., 2014; Twomey et al., 2011). Indeed, processes required to occur consecutively could overlap alongside potentially competitive parallel processes, resulting in lexical access ultimately being achieved by the strongest and most efficient line of processing, as indicated in the BIAM (Grainger & Holcomb, 2009;

Holcomb & Grainger, 2006). However, deeper research into the relative strengths of different potentially serial or parallel processes (e.g., orthographic and phonological) is required to determine if this is the case and, indeed, whether the relationship changes based on different contexts and the readers themselves (cf. language profiles).

There is no question that initial bottom-up visual processing must take place before further bottom-up and/or top-down lexical processing can begin, as it is the fundamental basis of visual word recognition. The question, then, concerns the orthographic and/or phonological activity immediately following the initial visual pre-linguistic processing. As orthography is concerned with language-related properties of visual stimuli and, therefore, has a closer relationship with them than the somewhat arbitrary associations of phonology and semantics, it is a logical and more palpable assumption that the processing of orthography, following the initial visual evaluation, occurs before and not after phonology (Grainger & Holcomb, 2009). Indeed, orthographic priming effects have been shown to occur earlier than phonological ones (Grainger et al., 2006). However, phonological effects have also been found to occur earlier than orthographic effects (e.g., Wheat et al., 2010), indicating that the extent of orthographic processing required before further processing can occur and, importantly, how orthographic activation does not always or entirely occur before phonological activation are still unclear. Furthermore, in a study of both orthographic and phonological priming effects, orthographic priming effects were detected in ERPs but not in behavioural measures while phonological priming effects were found in both behavioural and ERP results (Timmer & Schiller, 2012). This shows the bottom-up nature of orthographic processing and associated effects being somewhat subtler than phonological effects, but also suggests that the strength of orthographic priming might not be enough to influence performance on a behavioural level or that phonological processing could sufficiently suppress it. There are, therefore, competing views on how serial or parallel and how interactive orthographic and phonological processes are during visual word recognition.

The nature of interactions between the orthographic and phonological is a key point where theories diverge, typically based on how serial or parallel the processing is of these elements. In terms of VWR beginning with visual and orthographic letter form decoding before moving on to phonology (or more directly to semantics) to achieve word recognition, an element of serial processing is commonly accepted though (C. D. Martin et al., 2006). However, recent findings have questioned the exclusivity of this, proposing at least some parallel processing of orthographic and phonological codes (e.g., Cornelissen et al., 2009). The debate of serial and parallel processing is directly associated with the speed of different processes, including findings of concurrent, conflicting, or overlapping orthographic and phonological activations, some of which will be discussed in this section and others in the following chapter on timing.

Developmentally (i.e., from beginner to skilled reader), phonological encoding from orthographic input has been described as following a trajectory: from slow and serial (contrary to rapid phonological activation) to fast and parallel, becoming less serial and more parallel as reading skills are acquired and improved (Alario et al., 2007). This ties in with the somewhat accepted premise that beginner readers process words serially through assembled phonology, while skilled readers access the lexicon directly, processing orthography and phonology in parallel and using addressed phonology, as proposed in the DRC model (Coltheart et al., 2001). It should, however, be noted that the distinction between slow serial and fast parallel hypotheses somewhat ignores the possibility of fast serial processing, which is not unreasonable when considering development and reading acquisition as activities that essentially improve the performance of reading i.e., from slow and serial to fast and serial. Indeed, as an alternative to this notion of serial processing becoming more parallel as reading skills develop (Alario et al., 2007), reading acquisition could simply be the speeding up of serial processing and skilled readers could, therefore, just be much faster serial processors of words than beginner readers (Van Orden & Kloos, 2005). This also highlights a metalinguistic distinction that may not exist psychologically: orthographic and phonological processing could be two parts of the same whole with reading being non-linguistic in the sense of processing a picture or ideogram (Coltheart, 2005). This would entail the physical visual representation of a word being immediately processed in terms of its lines, shapes, dots, and strokes (Treiman & Kessler, 2005), similar to how a simple line drawing might be processed. This would arguably be a simpler albeit currently less substantiated view of reading in which physical orthographic representations, the actual written words, are recognized directly, akin to how other visual stimuli, such as pictures, are perceived in the visual domain (Van Orden & Kloos, 2005). In this case, phonology would be either activated directly from orthography or via semantic activation when the meaning of the "image" is obtained (and lexical access is achieved). However, such a theory is not out of line with direct lexical pathways to VWR and fuelling such direct pathways are lexical frequency and the extent of language exposure, which contribute significantly to VWR when processing visual/orthographic input (Mechelli et al., 2005). This is especially true for learners and L2 readers and is perhaps due to the statistical way that the brain might encode language on both word and language levels (Dehaene, 2014). Visual and orthographic familiarity are therefore vital factors for word recognition and phonological familiarity is also likely to be significant due to its primacy in language.

Although serial processing of phonology remains a possible explanation for some contexts of VWR (Coltheart et al., 1999), increasing evidence of early phonological activation based on the distinction between the slow serial and fast parallel hypotheses strongly suggests phonological processing occurs rapidly and in parallel with orthographic and other psycholinguistic processing (e.g., Klein et al., 2015; Pammer et al., 2004; Wheat et al., 2010). Comparisons of onset- and rime-based phonological priming, for instance, can provide insight into the slow-serial/fast-parallel debate, but they also present further contention in terms of whether one is stronger than the other. Slow serial processing entails rime-based priming to be slower and weaker than onset-based priming, Phonological onset priming (e.g., *dog-doll*) and rime priming (e.g., *fog-dog*) have been found to similarly facilitate behavioural responses (Alario et al., 2007), suggesting that phonological processing of onset and rime occur in parallel and fast parallel processing might not discriminate between onset- and rime-based phonological priming or that, at least for such relatively short words, onsets and rimes are not separate entities in internal phonology. Investigating a phonological effect based on word rimes, therefore, would provide insight into whether phonological activation is slow and serial or fast and parallel.

In stark contrast to long-standing and prevalent theories of visual word recognition that posit a more serial nature of processing (visual/orthographic before phonological and semantic), recent findings have increasingly supported parallel orthographic and phonological processing and for phonological processing that occurs much earlier than previously thought, sometimes even before timeframes associated with orthographic processing (e.g., Klein et al., 2015; Wheat et al., 2010). Taking this idea further, the notion of orthographic and phonological processes needing to cooperate to achieve word recognition, as foregrounded in the BIAM (Grainger & Holcomb, 2009), is particularly important. However, it appears to be somewhat flouted by the consideration of parallel orthographic and phonological routes as competitive (Coltheart et al., 2001; e.g., Coltheart & Curtis, 1993; Rastle & Coltheart, 1998; Zorzi, 2010). Routes, pathways, codes, and nodes need to coordinate for optimal word recognition and conflicts between them can result in processing delays and performance degradation (Grainger & Holcomb, 2009). Although the difference between their performances is often based on speed (Meyer et al., 1974; Xu et al., 1999), the notions of processes being competitive and cooperative relate to connectionist theories (McClelland & Rogers, 2003; Plaut et al., 1996; Zorzi, 2010), but also link to the idea of contextually dynamic activation in that the “winner” is at least partly dependent on the stimuli and task at hand (Fischer-Baum et al., 2014).

4.4.1 Accounts of ERP timing

Despite the volumes of research into VWR, accounts of when and how different forms of linguistic information are processed in the brain are still often incomplete, unclear, or non-specific (Cornelissen et al., 2010). Findings that different reading-related activities occur within particular post-stimulus timeframes have been vital. For instance, descriptions of orthographic processing occurring before 250ms and lexical-semantic effects predominantly beginning at 300ms post-stimulus provide a useful general framework (Grainger & Holcomb, 2009; Hauk et al., 2012). However, much neural activity, including one or multiple processes, can happen within these phases. In terms of VWR, 100-200ms post-stimulus is one of the most critical timeframes and the question of what happens within that initial ~200ms, leading up to lexical and semantic access, is still wide open (Cornelissen et al., 2010).

With some functional and temporal overlap, several distinct ERP components are associated with reading-related processes: the occipital P1 at ~100ms (P1-O), right occipitotemporal P1 at ~100ms (P1-OT), P150-Cz/VPP at ~150ms, left occipitotemporal N170 at ~170ms (N170-OT), the N2-P3 complex from ~180ms, the recognition potential at ~250ms, the MFN/N300/PMN/P2 between 200ms and 400ms, and N400 at ~350-500ms provide a temporal and topographic map of the visual word recognition process from initial visual processing to semantic integration (Dien, 2009). However, even this more detailed timecourse with more specific associations with language processes is arguably too simplistic and does not accommodate the nuances of VWR. Nevertheless, the overlapping processing timeframes are a vital aspect, showing a relation to the sequential cascading processes of cognitive theories discussed earlier (e.g., the DRC, Coltheart et al., 2001; and especially the BIAM, Grainger & Holcomb, 2009).

Another account posits analysis of sublexical orthography (i.e., letters and letter clusters) occurs at ~150-200ms (cf. VPP, N170), lexical processing at ~325ms (~300-400ms) and orthography to whole-word mapping at ~350ms before semantic and conceptual integration occurs at ~400ms (Hauk et al., 2012). Although it does not disregard phonological processing, this timeline focuses on orthography and the timeframes appear too conservative and regimented, especially considering the reading speeds mentioned earlier and that some processes may overlap. Furthermore, the stage-wise feed-forward nature of these accounts does not accommodate possible bidirectional and feedback processes (Twomey et al., 2011).

4.4.1.1 ~100ms

As discussed previously, the orthography-phonology-semantics "feedforward" perspective based on serial processing is still a common trope in VWR theories (Twomey et al., 2011), which is understandable due to the need to perceive visual (and, arguably, orthographic) elements before top-down internal (phonological, semantic) representations can be involved. Indeed, according to fast masked phonological priming studies (e.g., the classic Ferrand & Grainger, 1993; Ziegler et al., 2001), phonological activation is ~20-30ms after orthographic activation (Alario et al., 2007). Despite such accounts alongside ERP activity at ~100ms (e.g., P1; Dien, 2009) being strongly associated with low-level visual perception (Dien, 2009) and a specialization to print being observed in the P1 timeframe (Tong et al., 2016), left IFG activation associated with phonological processing has been found at ~125-130ms (Cornelissen et al., 2010), phonology has been reported to be accessed by ~100ms (Wheat et al., 2010), subphonemic priming has been observed at ~80ms (Ashby, 2010). Importantly, however, orthographic processing covers a range of (psycho)linguistic elements and findings of orthographic processing at these early post-stimulus timeframes do not refute the possibility of such early phonological processing, even if some element of orthography does require processing first. It is not questioned that some bottom-up orthographic processing must occur before top-down phonological activation, but parallel processing would permit phonological processing to begin before orthographic processing is complete. Furthermore, the unusually early phonological activation, some of which is controversial, is reasonable when considering simply how quickly and efficiently we read. Therefore, processing of phonology could occur in parallel with, after some, and even before the remaining orthographic elements are processed, so further evidence for or against such early phonological processing (<200ms, ~100ms even) is needed.

Early processing in visual word recognition is broadly represented by activity up to and around ~100ms during reading (Dien, 2009). This timeframe is associated with the initial, pre-lexical analysis of raw visual features (Grainger & Holcomb, 2009), which permits lexical processing in the case of linguistic stimuli (Dien, 2009). In terms of ERP measures of electrophysiology, this includes the P1 component at occipital and occipitotemporal sites, along with its negative-deflecting frontal-central counterpart (N100-FC), and the P150/VPP at central/vertex sites. The P1 (~100ms) and P150 (~150ms) are particularly sensitive to orthographic factors, such as word length, visual familiarity, n-gram frequency, and lexicality (Holcomb & Grainger, 2006; Sereno et al., 1998). Due to the early time-point, such experimental effects in language studies can often be interpreted in terms of physical visual features, such as visual familiarity, word length, word size, and word shape. While these properties are not necessarily linguistic exclusively, effects associated with true psycholinguistic properties have also, somewhat controversially, been found in this early timeframe, such as sublexical phonological processing (Ashby, 2010) and phonological judgement (Wheat et al., 2010) in frontal-central areas.

Sensitivities to linguistic forms have also been observed at the 80-150ms timeframe from various other locations, including both left and right occipital, left occipitotemporal areas, as well as left frontal-central (Cornelissen et al., 2010). As the left frontal-central and left occipitotemporal activations (possibly related to early N170 activity) were not found to differentiate real words from consonant strings or false-fonts (Cornelissen et al., 2010), such activity could reflect letter-level sublexical processing, while occipital (P1-related) responses reflect initial visual familiarity effects. It must, however, be noted that activity associated with frontal-central regions is not linked with visual form and, instead, its linguistic associations lie with phonology and speech production. This, therefore, is primarily of interest for its consequence for the serial/parallel debate concerning orthographic and phonological processing and how phonology is involved in early VWR processing. Overall, it enquires about whether any activity in this early timeframe is linguistic in nature, how early it becomes linguistic, and which regions are involved and to what extent. For instance, if the P1 is as directly physical as some literature might suggest, especially considering its presence is likely involuntary (Luck et al., 1994), a lack of differences in P1 could, therefore, be due to the matching of pre-lexical (i.e., visual), sub-lexical and early lexical parameters (e.g., bigram frequency), as such visual and coarse orthographic tuning processes drive the amplitudes of such early components (Grainger & Holcomb, 2009).

4.4.1.1.1 Occipital and occipitotemporal P1 components

Maximal at occipital and occipitotemporal sites within an 80-120ms timeframe, the P1 is usually the most pronounced ERP response to visual stimuli around its typical latency of ~100ms. The occipital P1 occurs prominently and involuntarily to any physical visual stimulus, reflecting low-level perception and is not linguistic in nature. However, effects of linguistic manipulations have been found on both its amplitude and latency, leading to debate regarding how linguistic properties might modulate it (Hauk et al., 2009). It is most likely that low-level visual properties, such as orthographic word length, size, and luminance, could influence this occipital P1, but these factors are not linguistic per se, instead being visual properties, even if they are manifestations of language, such as word length or frequency (Hauk & Pulvermüller, 2004). Therefore, if linguistic factors do influence processing as early as ~100ms, it is perhaps more likely to be in occipitotemporal or frontal-central areas and not on the occipital P1. However, as will be discussed in the rationale for Study 1, lexicality effects on the P1 have been reported to exist (Sereno et al., 1998).

4.4.1.1.2 Frontal-central activity

In a similar timeframe as the occipital/occipitotemporal P1, studies have reported early brain responses to phonology at ∼100ms at left frontal-central sites (Cornelissen et al., 2009; Klein et al., 2015; Pammer et al., 2004; Wheat et al., 2010; Woodhead et al., 2012). As discussed previously, however, ~100ms post-stimulus is more typically associated with the beginnings of orthographic processing, but the different topography, potential relationship to the left inferior frontal gyrus (and surrounding areas), and the evidence itself is compelling.

Left-lateralized frontal-central activity to words has elsewhere been shown as stronger to low-frequency words and pseudowords than to high-frequency words (Fiebach et al., 2002), showing a linguistic association and that a role in phonological encoding and recoding is highly likely (Cornelissen et al., 2010). It has also been shown that stimuli with legitimate phonology, such as real words, elicit a stronger response in left frontal-central areas to stimuli without direct phonological cues, such as faces (Cornelissen et al., 2009). Such sensitivity to legitimate linguistic stimuli and the evidence of stronger activation to low frequency words and pseudowords than to high frequency real words in this region (e.g., Fiebach et al., 2002), along with its activation in both lexical and phonological decision tasks (Heim et al., 2005), suggests that left frontal-central activity plays at least a supporting role in processing phonology during reading (i.e., grapheme-phoneme conversion).

If left frontal-central activation reflects the management of phonological information (Carreiras et al., 2014), it could feed back to or bidirectionally communicate with areas active for visual/orthographic processing, such as occipitotemporal areas and the prominent sites for linguistic N170 effects, in order to support successful VWR. Sublexical phonological processing and phoneme activation in the inferior frontal gyrus are connected with occipital and occipitotemporal/temporoparietal cortices used for sublexical orthographic processing, including letter and grapheme recognition (Klein et al., 2015). Functional connections between left occipitotemporal and left frontal areas during reading have been found (Kujala et al., 2007), which has been observed to words, pseudowords, and letter strings, but not false-fonts (Mechelli et al., 2005), while top-down feedback from left frontal-central to left occipitotemporal areas has been observed within 200ms for words, but not false fonts (Woodhead et al., 2012). The activation strengthens with reading acquisition, correlating positively with occipitotemporal responses to linguistic stimuli and showing a strong bidirectional reading-related relationship between inferior frontal areas and posterior temporal and inferior parietal cortices in the left hemisphere (Thiebaut De Schotten et al., 2014).

However, such network connectivity depends heavily on phonological information being processed and available at such a relatively early time-point.

In a masked priming experiment using MEG (magnetoencephalography; a technique on par with EEG for temporal resolution), Wheat et al. (2010) used real English word TARGETS following orthographically-matched pseudohomophones (e.g., *brein-BRAIN*), orthographically-matched pseudowords (e.g., *broin-BRAIN*), and unrelated pseudowords (e.g., *lopus-BRAIN*) as primes. Activity in left frontal-central areas being strongest for pseudohomophone priming compared with orthographic priming was concluded to show that phonology is accessed within 100ms (Wheat et al., 2010). Activation in these areas has been found previously for pronounceable letter strings in visual LDT (Pammer et al., 2004) and "passive" reading (Cornelissen et al., 2009), which strongly supports its role in phonological processing during reading. This is, therefore, compelling evidence, though also somewhat controversial due to the very early timing relative to other studies (Grainger & Ziegler, 2011).

Evidence of visual, orthographic, and phonological processing occurring within and around ~100ms post-stimulus relates to the fast-parallel hypothesis and that not only word processing is rapid, but different elements are achieved in a similar timeframe before integrating to achieve lexical access and comprehension. This could also explain how both orthographic and phonological effects have been found in the same P1 and N170 timeframes, including how some evidence can be interpreted differently (e.g., Wheat et al., 2010).

4.4.1.2 ~100-200ms

Electrophysiological response to visually presented words between the offset of the P1 and ~200ms post-stimulus often relates to the occipitotemporal N170 ERP component. Peaking at ~170ms, the N170, also called N1 and even N200 (Coch & Meade, 2016), is typically the second prominent component along the visual ERP timeline after the P1, often being the most prominent in terms of magnitude (i.e., amplitude and duration). Maximal to words at occipitotemporal sites, and possibly generated in the fusiform gyrus or superior temporal sulcus (Itier & Taylor, 2004), the N170 appears to be predominantly visual in nature, but an association with VWR in particular is undeniable (Amora et al., 2022; Gibbons et al., 2022; Maurer, Brandeis, et al., 2005; F. Wang & Maurer, 2017, 2020; Yum & Law, 2021).

N170 amplitude is reportedly sensitive to a variety of psycholinguistic effects, including traditional lexicality, orthographic lexicality, phonological lexicality and pseudohomophone effects, orthographic repetition, phonological priming, lexical frequency, and linguistic identity. Importantly, N170 effects have shown that readers' brains can distinguish between different linguistic stimulus types within ~200ms (Pattamadilok et al., 2015). For instance, N170 amplitudes have been reported as larger to words than pseudowords (Mahé et al., 2012; Maurer, Brandeis, et al., 2005), larger for word-like stimuli than for false fonts (Hasko et al., 2013), and larger to words than symbols regardless of task (F. Wang & Maurer, 2017). However, N170 amplitudes have also been reported as larger to consonant strings (McCandliss et al., 1997) and pseudowords (Compton et al., 1991) compared to familiar words, as well as larger to low frequency words compared to high frequency words (Hauk & Pulvermüller, 2004; Sereno et al., 1998). The sensitivity of the N170 is demonstrated further through effects observed to linguistic stimuli without the overt attention of the participant (Bentin et al., 1999), linking it to word superiority and the automaticity of language processing (Carreiras et al., 2014). Meanwhile, however, there have also been studies reporting no significant difference in N170 amplitudes between conditions, such as between words, pseudowords and consonant strings (Hauk, Davis, et al., 2006).

In between a lexicality effect or a lack thereof, more specific studies have found the N170 to be larger to non-word letter strings than to symbol strings (Appelbaum et al., 2009; Bentin et al., 1999; Coch & Meade, 2016; Helenius et al., 1999; Mahé et al., 2012; Maurer, Brandeis, et al., 2005), as well as dots and shapes (Eulitz et al., 2000), and even alphanumeric symbols (Bentin et al., 1999). These findings all relate to the notion of letter string processing occurring at 150-200ms post-stimulus (Hauk et al., 2012), the N170 reflecting bigram analysis (Dien, 2009), and the N170 response indexing real orthography (Simon et al., 2004). These findings also follow on from the P1/P100 being involved with visual-based "low-level perception" (Dien, 2009), implying the N170 reflects a deeper level of this same process.

Due to this range of observations of the N170, there is still much more to learn about how it reflects language-specific orthographic processing and to what extent, as well as whether phonological or other top-down processes influence it (not dissimilar to the debate surrounding the so-called VWFA). Effects on the latency of the N170 are markedly less clear and less reported than N170 amplitude effects (Maurer et al., 2008), but ERP components also differ in topography and their maximal location on the scalp. It is this dimension that appears to distinguish the N170 to words from the N170 to faces and other objects. Broadly, faces and objects have a more right-lateralised or bilateral

N170 response, while linguistic stimuli primarily elicit a left-lateralized N170 (Maurer, Brandeis, et al., 2005), which will henceforth be the N170 under scrutiny here.

4.4.1.2.1 Visual/orthographic expertise

While the VWR literature typically describes the N170 as an index of early orthographic fine-tuning during reading (Coch & Meade, 2016), there is still uncertainty about the extent that orthographic processing is reflected too. Evidence leans towards the N170 being predominantly an index of visual and orthographic processing, perhaps reflecting a direct orthographic route to word recognition (Coch & Meade, 2016; Pattamadilok et al., 2015). However, it is not clear what about visual/orthographic stimuli its neural generators respond to or whether this orthographic sensitivity is related to visual familiarity (Hasko et al., 2013).

The N170 during VWR has been shown to reflect expertise for writing systems that link graphic symbols to phonological representations and is not limited to alphabetic writing systems or native language (Maurer et al., 2008). It has also been highlighted as an index of reading proficiency and print expertise (Amora et al., 2022), so reasonably expected to be smaller in late bilingual ESL readers. However, reading mechanisms are dependent on the different cognitive requirements of different writing systems, so it is not yet clear specifically how the processing that underlies N170 activity contributes to L2 reading when L1 and L2 systems are different (Tan et al., 2005), but it is likely that the N170 is a key index of orthography-phonology mapping (Dien, 2009).

In terms of written language, familiarity is a function of frequency with a strong positive correlation between them (Tanaka-Ishii & Terada, 2011). Effects of word frequency have been shown on the N170 amplitude, being smaller to high frequency words than low frequency (Hauk & Pulvermüller, 2004). Such findings of word frequency effects, where high frequency words lead to a decrease in N170 amplitudes compared with low frequency words (Assadollahi & Pulvermuller, 2003; Hauk & Pulvermüller, 2004; Sereno et al., 2003), suggest that the N170 is sensitive to familiarity, as more familiar words require less effort, which is reflected in the decreased amplitude. Considering the link between word frequency and familiarity, this again relates to the left occipitotemporal neural source of the N170, which is strongly associated with visual expertise and "VWFA" accounts (Brem et al., 2009; Itier & Taylor, 2004; Maurer, Brandeis, et al., 2005).

Left-lateralized N170 responses to English words and more bilateral responses to novel character strings (for which visual expertise is not possible or unlikely) were observed in both English monolinguals and Japanese-English bilinguals (Maurer et al., 2008). This shows that the left-lateralization and specificity of the N170 is not limited to native language and further supports the association with visual expertise while providing some evidence of similarity between L1 and L2 readers. Furthermore, monolingual English readers (with no working knowledge of Japanese scripts) did not respond a typical left-lateralized linguistic N170 to Japanese syllabic or logographic characters (Maurer et al., 2008), while participants with no knowledge of Chinese presented with inverted Chinese characters show no behavioural or electrophysiological effects of orientation as would be typically found in Chinese readers (Fu et al., 2012). These findings support theories of visual expertise by showing how a lack of expertise also influences the N170 from an alternate perspective. However, the left-lateralization of the N170 to written language appears to be dependent on expertise of the script of a particular language, which has been shown to be based on script familiarity and not just familiarity with particular visual forms (Maurer et al., 2008). This suggests that the left-lateralized N170 marks specialized visual form processing, corresponding with VWFA theories and sub-lexical, letter-level processing. However, while the N170 is typically left-lateralized for English individuals reading English, it is more bilateral in Chinese individuals reading Chinese, but less clear for ESL readers and whether their L1 can make a difference. These different interpretations illustrate the necessity for further neurophysiological research for more precise timing and lateralization information about these processes.

4.4.1.2.2 N170 and phonology?

While the N170-related psycholinguistics literature typically posits the N170 as more visual/orthographic in nature, the apparent context-dependence of the component suggests this not to be the whole story: reports are mixed on whether the N170 is an index of orthographic or phonological processing or, indeed, a potentially dynamic combination of the two (Yum & Law, 2021). Visual/orthographic processing does appear to be dominant, but an additional query into the nature of the N170 is whether it is associated with any stage of phonological processing. For instance, pairs of words and pseudowords with varying orthographic and phonological relatedness were found to elicit significantly different responses from one another and modulate left occipitotemporal N170 activity, reflecting a continuum of orthographic and phonological relatedness (Holcomb & Grainger, 2006). This suggests the left occipitotemporal N170 is associated with orthographic processing as well as the mapping of phonology to orthography, perhaps being part of the interface between orthographic and phonological processing (Grainger & Holcomb, 2009) or between bottom-up and top-down influences (Twomey et al., 2011).

Unlike the visual representations of words, phonological and semantic psycholinguistic features of words do not exist as physical properties and so integration of bottom-up visual/physical with top-down non-visual psycholinguistic properties is required for VWR (Twomey et al., 2011). This is demonstrated in such classic psycholinguistic effects of word superiority where letters are better perceived as part of real words than non-words (Coch & Mitra, 2010; McClelland & Rumelhart, 1981), of pseudohomophones where the legitimate phonology of pseudohomophones impact reading performance (e.g., Braun et al., 2009; Briesemeister et al., 2009), and of lexicality where real words are processed more efficiently than other linguistic stimuli (e.g., pseudowords, pseudohomophones; McNorgan et al., 2015; Twomey et al., 2011). Top-down processes theoretically influence the nature of the response for a given stimulus by constraining the potential activation beforehand to the most likely based on encyclopaedic knowledge, context, and statistical processing (Dehaene, 2014).

Along with the relationship with visual expertise, the left-lateralized N170 has been suggested to represent the expertise of skilled readers of writing systems that link graphic symbols to phonological representations (Maurer et al., 2008), further proposing it to be a critical component in orthography-phonology mapping and not just visual/orthographic processing (Dien, 2009). It appears to be more specific to writing systems with the capacity for grapheme-phoneme conversion and direct links between the written and phonological forms of the language (M. Wang et al., 2003). Indeed, the prominence of the N170 has been shown to positively correlate with the extent of grapheme-phoneme correspondence in the language (Maurer & McCandliss, 2007). However, left-lateralized N170 effects have also been observed to Chinese logographic characters that do not have such a grapheme-phoneme connection (Fu et al., 2012), showing that such N170 responses are not limited to alphabetic writing systems. This seems to support the association with visual processing due to the visual nature of Chinese itself (Lin et al., 2011), though readers of logographic scripts also report experiencing phonological recoding (M. Wang et al., 2003), meaning that N170 effects to such a language without direct grapheme-phoneme relationships is not sufficient evidence for a purely visual or orthographic nature of the N170.

It should be noted that phonological involvement with the N170 cannot yet be ruled out due to the phonological basis of Language and the inherent phonological representations of linguistic stimuli. An alternative direct visual-phonology or orthography-phonology route is a possible explanation, though further investigation of orthographic and phonological interactions is required for such a conclusion. Furthermore, elements of the evidence for visual expertise often overlap with phonological involvement. For instance, the observation of a broad and extreme effect of lexicality on N170 responses that differentiated between real English words (with defined legal and legitimate phonology) and character strings (with no clear phonological representation) in proficient readers of English (Maurer et al., 2008). This is not to say that this was a phonological effect, but there was a phonological difference between conditions, as can often be found to be the case with the evidence for the N170 as a predominantly visual component. Nevertheless, considering the contention surrounding the N170 in the psycholinguistic literature, the strong basis in visual processing and extent of visual, orthographic, and phonological processing potentially involved, as well as the undisputed importance of ERP activity over left occipitotemporal areas within ~200ms of seeing a word (Cornelissen et al., 2010), the N170-OT is a prime target for electrophysiological VWR research.

4.4.2 L2 VWR in bilingual brains

Significantly less is known about brain activity during L2 processing in bilinguals than L1 processing in bilinguals and, certainly, than native VWR processing in general, especially concerned with the early pre-200ms post-stimulus period that is the focus of the current research (Poeppel & Idsardi, 2022; Yeong et al., 2014). However, similar patterns of brain activity across both hemispheres in occipital, occipitotemporal, and frontal areas during VWR have been observed between native readers of English (alphabetic, deep orthography), Spanish (alphabetic, shallow orthography), and Chinese (logographic), as well as Hebrew (alphabetic, deep orthography) (Rueckl et al., 2015). These findings support a universal basis of reading that is independent of language or language type, writing system, or orthography. Considering the contrasts and complexities between different languages (see §1.1), it is somewhat surprising that such similarities exist. Pertinent questions about bilingualism, therefore, concern specifically how these similarities and any differences manifest between L1 and L2 readers of a language (Brysbaert & Wijnendaele, 2003; Perfetti & Liu, 2005). For instance, there are similarities between native English readers reading English, native Spanish readers reading Spanish, and even native Chinese readers reading Chinese, but where any similarities lay between native English readers reading English and Spanish-English bilinguals or Chinese-English bilinguals also reading English is still unclear. Although the current research is not directly focused on the language selective/non-selective access debate (e.g., Dijkstra & van Heuven, 2002), the system that cognitively distinguishes the languages of bilinguals is highly relevant.

Different networks for VWR based on language has been shown in native readers, but it is not clear how it manifests in bilinguals reading their L2 (Cao et el., 2017), including how both the L2 and native languages might impact L2 processing. There is increasing support for the notion that the same neural networks are used for both languages (Perani & Abutalebi, 2005), as opposed to L1 and L2 being processed in different areas of the bilingual brain. Highlighted by the assertion that conflict between L1 and L2 cannot be avoided (van Heuven et al., 2008) and theorized as the integrated lexicon discussed earlier (van Heuven & Dijkstra, 2010), the understanding is that L2 processing relies on L1 expertise and understanding L2 relies on associations with L1 (Alvarez et al., 2003). This is supported further by the finding that brain activity associated with L1 is consistently found also during L2 processing in ESL bilinguals (Kim, Liu, & Cao, 2017) Furthermore, brain activation common to both L1 and L2 has been argued to be stronger during L2 processing, especially in late bilinguals and/or bilinguals with low L2 proficiency (Indefrey, 2006; van Heuven & Dijkstra, 2010).

As mentioned in Chapter 3:, proficiency is likely to be a key factor. For instance, the extent of day-to-day usage and exposure to L2 has been shown to affect brain activation during L2 reading (Perani & Abutalebi, 2005; Vingerhoets et al., 2003). Furthermore, brain activation associated with L1 processing also found during L2 reading could depend on proficiency in that higher proficiency resulted in more L1-related activation (Kim et al., 2017). However, the difference(s) in activation based on proficiency are unclear and observations of additional activity when processing L2 compared with L1 are not consistent (e.g., van Heuven & Dijkstra, 2010; Vingerhoets et al., 2003).

While it might seem that L1-based L2 processing would be found more in beginner/learner bilinguals (i.e., lower proficiency), the notion that proficiency is a factor in L1-oriented activity during L2 reading also speaks to the BIA+ and Multilink argument that L1 and L2 are part of an integrated lexicon. In this case, the lexicon would become more integrated as proficiency increases, but the stronger/dominant L1 would take the lead during L2 processing. The premise of L1 being dominant even in L2 processing is especially pertinent to the current research, due to using late bilingual groups (Spanish-English and Chinese-English) with such distinct L1 profiles (alphabetic, shallow orthography and logographic, deep orthography, respectively) that are matched for proficiency and experience of English, allowing direct investigation into the possible effects of L1 on L2 processing.

Observations of interlingual involvement when reading L2 relate strongly to the pertinent question of whether L2 reading is achieved through accommodation (i.e., acquiring and using new skills for L2), assimilation (i.e., reading like L1), or a combination of both dependent on the L1 and the linguistic context (Perfetti et al., 2010). Specifically, it is not clear whether native speakers of non-alphabetic languages would employ similar methods as used for L1 reading or whether they would adapt to using more "alphabetic" methods, such as a dual route approach that is generally accepted for alphabetic languages, but not clear for others (Lu et al., 2011). For instance, Chinese-English bilinguals have shown a "Chinese-like strategy" (i.e., assimilation) when reading English (Perfetti et al., 2010). Based on similar brain activity for L1 and L2 reading, it has been proposed that Chinese-English bilinguals largely use L1 strategies for reading English that are grounded in the visual and orthographic processing pertinent to reading logographic scrips (M. Wang et al., 2003), which suggests that the skills required to read Chinese are also at least relevant if not sufficient for reading English.

Assimilation to Chinese-based reading strategies in Chinese-English bilinguals reading English was also concluded from ERP observations of right-lateralized (Tong et al., 2015) or bilateral occipitotemporal N170 activation, as often found when reading Chinese (Nelson et al., 2009). However, left hemisphere dominance, as is typical for English and alphabetic reading, was not found, suggesting that Chinese-English bilinguals assimilate to English when reading English (as opposed to accommodating and recruiting left hemisphere regions). Further, it has been shown that Chinese-English bilinguals exhibit bilateral occipital and occipitotemporal activation for both Chinese and English (Perfetti et al., 2010). Such bilateral occipital activation for reading Chinese characters, as indicated by low resolution electromagnetic tomography (LORETA), that is typically left-lateralised for high-frequency English words represents a potential distinction between reading alphabetic and logographic languages that is pertinent to the current research (Liu & Perfetti, 2003). Furthermore, the right hemisphere has been found to be more involved in reading Chinese than reading alphabetic scripts, including switches in activation from left to right hemisphere occipital areas (Lu et al., 2011; cf. Tan et al., 2005). Considering the formal linguistic differences between languages and the psycholinguistic processes involved based on this metalinguistic perspective, precisely how skills overlap is not clear. However, considering written language in a less linguistic and more symbolic, graphic, or even figurative sense, especially with Chinese orthography having such a visual bias, it could simply be that the aforementioned

Chinese-like strategy involves the advanced spatial and configural processing that logographic scripts support.

In contrast with the Chinese-English right-hemisphere activity, brain activity of Spanish-English bilinguals reading English stimuli has been found to be more pronounced in left frontal regions and more widespread across left hemisphere networks overall than to Spanish stimuli (Jamal et al., 2012). Such a difference could be due to the substantial difference in orthographic depth and phonological transparency between the languages, especially considering the phonological associations of left frontal areas (Klein et al., 2015; Pammer et al., 2004; Wheat et al., 2010). The wider spread of left hemisphere activity could relate to the additional processing required for L2 reading, while lesser widespread activation to Spanish reflects more efficient brain responses from tighter integration of necessary networks due to it being the native language.

Differences in lexical processing and reading performance between L1 and L2 are often attributed to differences in orthographic depth and phonological transparency between languages (Jamal et al., 2012). This could also relate to how L2 processing deviates between different populations of its readers (e.g., Spanish-English and Chinese-English bilinguals reading English). Considering how orthography and phonology are associated in alphabetic compared with logographic languages, particularly in terms of the relationship being inherently intertwined for alphabetic scripts, it is not unreasonable to consider that processing that connects orthography and phonology would be distinctly different between the two language types (Lin et al., 2011). Despite the differences, the neural patterns of phonological processing when reading alphabetic languages (e.g., English, Spanish) and logographic scripts (e.g., Chinese) have been shown to overlap (Tan et al., 2005), such as similar regions being found to activate for subvocal rehearsal, including grapheme-phoneme conversion (where applicable) and phonology-orthography feedback pathways to both alphabetic words and Chinese characters. Ultimately, however, regions associated with phonological processing have been shown to differ significantly, while regions associated with lexical/addressed and non-lexical/indirect/assembled phonology appear to be distinct according to the language type being natively processed (Tan et al., 2005). Furthermore, the neural networks active during reading are partly language-dependent with the more prominent pathway depending on the orthographic depth of the language (Cao et el., 2017). The use of different neural networks dependent on the target language imply different strategies and processing pathways that reflect the distinctly different techniques of each L1 language type to access phonology (and semantics) from the L2 orthographic input (Nelson et al., 2009). Essentially, processes in L2 readers may focus on what is necessary and what can facilitate comprehension in terms of L1 skills and L2 knowledge to achieve word recognition, which is highly relevant to the current research due to the variable dependence on orthographic and phonological processing.

4.5 Summary

This chapter outlined how orthographic and phonological processes (and interactions between them) during VWR are understood from a cognitive perspective. Evidence of the pre-200ms neurophysiology underlying these processes in monolingual and bilingual readers was also presented, which is acutely associated with the methodology used by the current research.

Chapter 5: Methodology review

The divergence of methodological factors is proposed to be the primary explanation for the various inconsistencies in empirical findings (Proverbio & Adorni, 2008), especially with different theories often being attached to different methods (Laszlo & Federmeier, 2014). In addition to the possible variation in data acquisition and analysis techniques, differences between stimulus constraints, participant control, and ERP analysis techniques are among the main causes (Glezer & Riesenhuber, 2013), while task demands between studies have also been cited for such discrepancies (Hasko et al., 2013; Pattamadilok et al., 2015). Based on the importance of methodology that these sentiments endorse, this chapter will review the methods and methodological issues relevant to the current research in terms of participants, design, stimuli (and other materials), apparatus, and procedure), especially where a choice was required. For instance, the inclusion criteria for the participants, constraints on the stimuli, and the combined behavioural/ERP approach all require justification for their specifications. Importantly, sections will focus on general methodological details that are common across both studies with information specific to each study presented in their own respective chapters.

5.1 Participants

It is acknowledged that language backgrounds, histories, and knowledge can never be truly matched and, therefore, that within-participant variability is inevitable.

Nevertheless, several aspects can be measured and controlled to support better comparisons both within and between groups.

5.1.1 Group specifications

Participants were either native (monolingual; n=20) or L2 readers of English: Spanish-English bilingual (n=20) or Chinese-English bilingual (n=20). The three groups will henceforth be referred to interchangeably as the English or native alphabetic, Spanish or non-native alphabetic L1, and Chinese or non-native non-alphabetic L1, respectively. The Spanish and Chinese groups will also be referred to collectively as “ESL” or "non-native" participants or groups.

English monolinguals were defined as individuals with English as their native and only language, being as monolingual as is possible in a multilingual world, following the notion that the majority of people have at least some knowledge of a second language (Brysbaert & Dijkstra, 2006). The Spanish-English bilinguals and Chinese-English bilinguals (with Mandarin or Cantonese as their variant of Chinese) were late bilinguals with English as their only other language, used prominently in daily life at the time of testing due to living full-time in the UK. Individuals with significant knowledge of any other unrelated languages were not included, though it was acceptable for Chinese-English participants to have knowledge of both Mandarin and Cantonese.

All participants met the language inclusion criteria and were deemed representative of their respective experimental group: the Spanish-English participants represent a population of L2 English readers with an alphabetic first language (non-native alphabetic L1 group) and the Chinese-English participants represent a population of L2 English readers with a non-alphabetic first language (non-native non-alphabetic L1 group).

5.1.1.1 Why English as the target language?

On a global scale, there are more people that understand English as a second language than as a first (Brysbaert & Dijkstra, 2006), 978.2 million according to *Ethnologue* (Eberhard et al., 2021), warranting further study into how English is processed as L2, which is arguably even more pertinent than only investigating reading English as a native language. According to the Office of National Statistics (2013), for instance, the foreign-born population of the UK in 2011 had increased by 53.3% over the previous decade to approximately 7.5 million. This highlights the substantial number of non-native individuals coming to work and/or study in the UK, many of whom could benefit from improvements to English language education (i.e., ESOL/EFL/ESL) through a better understanding of L2 processing. Literacy and education of English as an additional language is therefore a valid applied purpose for the focus on English in this research. In terms of literacy education in particular, tests involving reading familiar words and pseudowords showed that children from most European countries were proficient in reading before the end of the first school year with word reading accuracy of >90% in all but the languages with deeper, more opaque orthographies, such as Portuguese, French, Danish, and especially English (Seymour et al., 2003). Overall, the rate of learning English is worse than twice as slow as languages with other types of orthography (Ellis et al., 2004), providing motivation to focus on English and how it is understood by L2 readers. By learning more about how English is known by its non-native readers, this could eventually contribute to understanding how English is acquired.

While vast, the second language literature more typically focuses on comparisons between L1 and L2 and the debate of language selective or non-selective access, whereas the current study approaches L1 and L2 processing from a different perspective, using a single target language (English) to investigate differences in processing between L1 and L2 reading. Vitally, this research has also been designed to test how two groups of L2 readers of English with two different native languages (Spanish and Chinese, respectively) process English compared with one another and a control group of native English readers. The within-participants comparisons of experimental conditions will enable observation of how native readers of different languages each process orthography and phonology when reading English. Meanwhile, the between-participants comparisons of the same experimental conditions across the native language groups will allow observation of any differences in the timecourse and nature of processing the orthography and phonology of English between readers with different native language backgrounds. Investigating VWR in bilingualism has not previously been approached with this perspective, these bilingual groups (and the way their language profiles interrelate), and the staggered differences of the native languages involved (English, Spanish, Chinese).

Underlying these theoretical and applied reasons for the focus on English as the target language, the native language of the researcher is English and the location of the data collection phase of the research was Sheffield (England) were also significant practical considerations. These are included for transparency, but they also allowed more achievable goals for recruiting participants with the target language as L2.

5.1.1.2 Why Spanish and Chinese as ESL groups?

The Spanish-English group represents L2 readers of English with an alphabetic L1, while the Chinese-English group represents L2 readers of English with a non-alphabetic L1. While it must be appreciated that these languages and the groups representing them are only samples of their respective language categories, the comparison between language types adds a valuable layer to the research, allowing observation of how readers with different L1 types process a common alphabetic script.

Spanish-English and Chinese-English groups also represent L2 late bilingual readers of English, but each overlap with the other two groups in important ways. The Spanish-English group overlaps with the English group through having an alphabetic native language and much of the orthography and phonology. The Chinese-English group, meanwhile, does not share these traits with either group, but does share the non-native perspective of English with the Spanish-English group. Spanish-English bilinguals were chosen for the non-native alphabetic L1 group due to the prominent status of the Spanish language worldwide and an element of opportunity sampling provided an advantage over other potential groups (e.g., French-English bilinguals). Furthermore, the difference in orthographic depth between English and Spanish provides insight into the importance of L1 dependency (Seymour et al., 2003), which is extended further when considering the complexity of orthography-phonology mapping of a non-alphabetic language and comparisons involving the Chinese-English group. The global status of the language was also a factor in selecting Chinese-English bilinguals for the non-native non-alphabetic L1group. The decision was also based on a combination of the increasing number of Chinese individuals in the UK (Office of National Statistics,

2012) and the amount of literature focused on Chinese (and Chinese/English contrasts) compared with other non-English and non-alphabetic languages, along with its status as a strong exemplar of such a language.

Alongside the theoretical differences between alphabetic/non-alphabetic and native/ESL groups and potential orthography and phonology differences within groups, the core applied rationale for focusing on these language profiles was based on them being the top three most widely used languages in the world and the clearly L2-heavy ratio of English. Based on figures from *Ethnologue* (Eberhard et al., 2021), the three languages with the most native users are Chinese (including Mandarin and Cantonese) with over one billion native speakers (~1006 million), accounting for ~12.78% of the global population, followed by Spanish with 471.4 million native speakers (~5.99%), and then English with 369.9 million native speakers (~4.69%). Along with Hindi (600 million), these three languages also have the highest totals of L1 and L2 speakers with Spanish at 543 million, Chinese (including Mandarin and Cantonese) at 1.2 billion, and, lastly, English with 1.35 billion (Eberhard et al., 2021). As stated previously, English is more common as a second language than a first (Brysbaert & Dijkstra, 2006), being by far the most-used L2 with 603 million bilinguals having English as a second language (190 million Chinese as L2, 94m Spanish as L2). Native speakers of Spanish and Chinese, respectively, with an L2 of English are, therefore, prime target groups for the current work.

5.1.2 Participant control

Language profiles, including proficiency, was a significant factor for controlling groups and ensuring meaningful comparisons. Several mechanisms were used to screen participants before their formal involvement in the studies based on research-related inclusion criteria and to confirm sufficient English language proficiency to satisfactorily complete the tasks.

Firstly, bilingual participants were required to have an International English Language Testing System (IELTS) score of 6.5 or higher, following the typical entry requirements by UK universities for non-native readers of English, or equivalent evidence of English language proficiency, such as the Test of English as a Foreign Language (TOEFL; 6.5 IELTS = 79-93 TOEFL; http://www.ets.org/toefl/institutions/scores/compare/) or entry to Level 7 higher education and demonstration of sufficient real-world English communication skills. Prior to further participation in the studies, participants were asked to complete a revised version of the Language History Questionnaire (LHQ; Li et al., 2014), which was designed to document participants' language usage, history, and general linguistic ability. All are vital factors for any research focused on language, particularly when groups with different native languages are involved. Documenting relevant details of these factors was necessary to allow sufficient representation and understanding of each experimental group and was initially evaluated prior to any experimentation. The primary purpose of the LHQ was to screen participants for eligibility in terms of general proficiency as well as to allow control over the extent of bilingualism and when learning English began. Only Spanish and Chinese applicants with profiles as late bilinguals were invited to take part in the studies. Late bilinguals were defined as being raised with their native language as their only prominent language and not learning English from an early age, instead acquiring it later (e.g., in school or when coming to live in the UK) and only having more recent immersion in the language (e.g., international students). Participants were accepted as late bilinguals if they matched the above definition and no arbitrary cut-off was used. Furthermore, the information sheet, all correspondence about the study, and the instructions before and during the experiment sessions were all in English. Reading, understanding, and then following this information properly were also monitored to gauge proficiency, as they require a good level of real-world English language comprehension. Lastly, a spoken phonology test using a random selection of pseudowords and pseudohomophones was conducted after the experimental tasks (see §5.2.1.3 for more details). This provided additional support and a baseline for participants' proficiency in skills necessary to complete the experimental tasks (i.e., orthographic and phonological processing using English phonotactic rules).

5.2 Overview of participants

Native English monolingual, Spanish-English bilingual, and Chinese-English bilingual individuals from Sheffield Hallam University (SHU) and University of Sheffield (UoS) were invited to take part in the studies by e-mail, in person, through online advertisements, and physical posters. Participants were offered two hours' worth of university credits ("PsyCreds" for SHU undergraduate students) or £20 worth of "Love2Shop" High Street vouchers (www.highstreetvouchers.com) for their participation.

The planned sample size of 60 (n=20 per group) was mainly based on counterbalancing requirements (specified in 5.3.2 and study-specific chapters), practical considerations of EEG recording procedure (e.g., timeframe per participant), and being larger and more robust than is typical for ERP studies, though power analyses were conducted per study (see study-specific chapters). To obtain this sample size, 64 individuals were recruited and tested, including one additional Spanish-English, one Chinese-English, and two native English participants who replaced prior participants due to technical issues. The same individuals participated in both studies.

The following information is based on self-reports by the participants through a combination of the pre-study language history questionnaire and consent form (see §5.1.2 for more information). Participants were aged 18-44 (M=24.56, SD=6.40) and Table 1 shows the age distribution within groups. All participants had normal or corrected-to-normal vision and no reported language, learning, or neurological disorders (e.g., dyslexia, reading disorder).

Table 1: Descriptive statistics of ages per group

	Mean (SD) age	Age range
English	20.82 (4.31)	18-31
Spanish-English	29.29 (6.30)	19-44
Chinese-English	23.70 (5.43)	19-38

The sample had a female majority within each group (English=60%, Spanish-English=70%, Chinese-English=80%), but was as balanced as the accessible recruitment methods and availability of suitable participants allowed. Although some specific differences in semantic processing between male and female participants have been shown with regard to language processing (Wirth et al., 2007), there is currently no known evidence for gender/sex differences concerned with the early orthographic or phonological processes associated with the current research. Furthermore, other studies have shown no overall differences between the sexes in right-handed individuals (e.g., Galin et al., 1982).

All participants reported right hand dominance, which was supported by the results of a revised version of the Edinburgh Handedness Inventory (EHI; Oldfield, 1971; Veale, 2014). The mean laterality quotient (MLQ) scores of the English group (M=4.83, SD=0.17), Spanish-English group (M=4.8, SD=0.16), and Chinese-English group (M=4.64, SD=0.23) were similar between groups and sufficiently close to 5 (fully right-handed). Use of only right-handed participants is typical in EEG/ERP research due to the potential difficulties of analysis and interpretation that stem from differences in lateralisation between right-handed and left-handed individuals (Galin et al., 1982).

All participants were provided with all necessary information and answers to any questions before the experiments. Informed consent documents (see Appendix A and B) were completed in line with the institutional ethics approval (see Appendix C). Answers to any other questions and a full debrief (see Appendix D) were provided afterwards.

5.2.1 Control tests

The following measures and tests were included to better monitor participant inclusion criteria, as well as to increase understanding of the participants involved and providing data in the current research. Therefore, these measures will also be used in study-specific and overall analyses to provide an extra dimension to aid interpretation and conclusions of core behavioural and ERP results.

5.2.1.1 English proficiency

Based on the means of EFL proficiency (measured with IELTS scores) not significantly (or numerically) differing between Spanish-English (M=6.83, SD=0.49) and Chinese-English (M=6.83, SD=0.57), $t(38)<0.001$, $p>.999$, MD=0, the ESL groups were well matched for English proficiency.

5.2.1.2 English:L1 usage ratios in ESL groups

English:L1 usage ratios were calculated from self-reported language usage scores from the LHQ (see Appendix E) for both ESL groups. As these usage ratios did not significantly differ between Spanish-English (M=0.53, SD=0.16) and Chinese-English (M=0.53, SD=0.08) groups, $t(38)=0.05$, $p=.961$, MD=0.002, the ESL groups were well matched for their everyday usage of English relative to their L1.

5.2.1.3 Phonology post-test

In order to provide some validation of participants' knowledge of English phonology separately from the experimental tasks, as well as to help explain any anomalies in the results, a phonology test was administered that required participants to read aloud a random selection of 50 pseudowords and pseudohomophones (from a pool of 100 taken from the stimulus sets of Study 1; see Appendix F). Participants were scored on whether they pronounced the pseudowords and pseudohomophones as designed and expected. The English group (M=48.5, SD=2.28) scored significantly higher than the Spanish-English group (M=44.05, SD=4.67), $t(27.58)=3.83$, $p<.001$, MD=4.45, and the Chinese-

English group (M=35.6, SD=10.56), $t(20.77)=5.34, p<.001$, MD=12.9, while the Spanish-English group (M=44.05, SD=4.67) scored significantly higher than the Chinese-English group (M=35.6, SD=10.56), $t(26.17)=3.27, p=.003$, MD=8.45.

The English participants were understandably the most proficient, as they are native readers of English with no significant language conflict. The Spanish-English participants have a strong phonological profile, but the test stimuli were still based on their L2, which explains the difference from English readers. Chinese-English participants, however, have a primarily visual/orthographic and almost entirely different language profile from both English and Spanish-English participants and the outcome of the test reflects this through the relative difficulty of orthographic-phonological processing of English-based stimuli. Overall, these results give a broad overview of the language profiles involved in the current research, suggesting that behavioural performance and possibly the underlying electrophysiological activity could relate to these results.

5.3 Design

Both Study 1 and Study 2 used mixed factorial designs with the same 3-way between-participants variable (L1 language profile), but different within-participants variables to meet the specific aims of the respective study. The between-participants variable, language profile, was operationalized through participants' native language in three distinct groups: native alphabetic (English monolingual participants), non-native alphabetic L1 (Spanish-English bilingual participants), and non-native non-alphabetic L1 (Chinese-English bilingual participants). The within-participants variables differed between studies and were operationalized through the respective task methodologies and the properties of the visually-presented single-word stimuli, all of which will be described and discussed in the study-specific chapters (5.9 and Chapter 7:). The overall designs, however, both employed a combined behavioural and EEG/ERP approach where simultaneous behavioural tasks were instrumental in providing a purpose for the participant, focusing and maintaining their attention (as opposed to their EEG being recorded passively). The dependent variables for both studies, therefore, were accuracy and response times (RTs) from the behavioural tasks along with area amplitudes and 50% fractional area latencies from P1-O, P1-OT, N100-FC, and N170-OT ERP timeframes (see §5.3.3 for technical specifications). Analysis of behavioural and ERP measures, respectively, will be used to complement one another.

As the timing of brain activity is at the heart of this research, a methodology that directly measures neural activity with high temporal resolution was essential. Behavioural measures alone and even neuroimaging techniques such as fMRI, PET, and fNIRS are not capable of discerning between different processes in such precise timeframes as is necessary with word recognition. ERP methodologies, however, offer a clinical way to attempt answering questions of the relative strengths, timings, and significance of different types of processing due to their time-locked nature and the fine temporal resolution of EEG (Dien, 2009). For instance, orthographic priming can be detected more easily in ERPs than through behavioural measures (Timmer & Schiller, 2012), making the EEG/ERP approach well-suited to the current research. By no means does this imply that behavioural measures are redundant, though, as they provide an objective record of performance on tasks at trial level per participant and, therefore, indication of participant, trial, and task validity, as well as insight into the way brain activity manifests in more real-world scenarios. Response time data is used for comparisons of overt task performance and accuracy data is necessary to allow correct responses to be analysed alone and incorrect responses to be discarded or analysed separately (Picton et al., 2000). Furthermore, error-related negativities in the electrophysiological data could impact on the results if incorrect responses were also included in the averaging. Therefore, analyses would use only data from correct responses, so the data describes how participants respond when they are accurate and not a mixture with when they think they are correct but wrong.

Both behavioural and EEG/ERP methodologies are widely used in psycholinguistic research, particularly for investigating timing-related questions of individual cognitive processes associated with language. Therefore, the use of a combined behavioural and EEG/ERP approach in this research allows integration and comparison with a wealth of previous studies, which is vital to be able to contribute to the existing knowledge on orthographic and phonological processing.

5.3.1 Complementary study designs

To investigate different levels of orthographic and phonological processing, different stimulus types and experimental paradigms are required. Studies in the current research were, therefore, designed with parallel contrasting but complementary task methodologies. The orthographic and phonological variations of the classic lexical decision task of Study 1 contrasts with the rhyme recognition paradigm with orthogonal orthographic and phonological manipulations of Study 2, but both provide insight into the same overall questions of orthographic and phonological processing in L1 and L2 readers of English.

Study 1 and Study 2 were both designed to operationalize and observe early pre-200ms orthographic and phonological processing in L1 and L2 readers of English. Through different stimulus and task implementations, each provides a distinct perspective that complements the other and provides wider scope than each alone allows. Study 1 focuses on orthography and phonology separately across two tasks (orthographic and phonological lexical decision tasks) using different stimulus types (real words, pseudohomophones, pseudowords) presented individually to operationalize orthographic/phonological processing in terms of orthographic and phonological lexicality. Study 2 approaches orthographic and phonological processing through direct comparisons of real word stimulus pairs in orthogonal manipulations orthographic/phonological congruency within one orthographic/phonological priming task (rhyme recognition task).

The use of both a two-task (Study 1: oLDT & pLDT) and a single-task approach (Study 2: RRT) allows observation of similar and related processes from perspectives different from one another and the literature (e.g., masked priming). Due to the oLDT and pLDT having different cognitive requirements and different stimulus types for the experimental manipulations of orthography and phonology, comparing across them to contrast orthographic and phonological processing shows only task-driven effects (Twomey et al., 2011), which may also reflect differences in strategy, limiting observations of relative timing between orthographic and phonological processes (Q. Zhang et al., 2009). The use of different tasks for different processes allows separated investigation of orthographic/phonological processing and other psycholinguistic phenomena (i.e., orthographic and phonological lexicality). For valid comparisons of the relative timecourses of different types of processing, however, the methodology must permit direct comparison between them (Ferrand & Grainger, 1994). To accomplish this and build upon Study 1, the orthogonal design of the RRT used in Study 2 involves all experimental manipulations within the same trial blocks of a task that uses the same instructions throughout: all four permutations of phonological (P) and orthographic (O) congruence (+) and incongruence (-) between the prime and target of visually-presented stimulus pairs (Classon et al., 2013; Weber-Fox et al., 2003). In practice, this means that all conditions of the independent variable(s) will be presented within the same task. In terms of analysis, this allows direct comparison of conditions (Ferrand & Grainger, 1994; Luck, 2004). Supporting valid comparisons in this way helps to better see how orthography and phonology may contrast, conflict, or cooperate. Using the same task and same instructions also relates to eliciting the same top-down processes, as well as the same bottom-up processes through the same balanced stimuli for each orthographic and/or phonologically focused condition, theoretically leaving only the stimulus-based experimental manipulations of orthography and phonology to influence responses.

As outlined earlier, real words have all psycholinguistic properties, while pseudohomophones and pseudowords have only a subset of these. Study 1 requires and uses pseudohomophones and pseudowords to observe the contrast between visual familiarity and phonology (van der Mark et al., 2009). However, the limited psycholinguistic values and the infrequency of such non-words in the real world can also be a disadvantage as reading them is limited to demonstrating grapheme-phoneme conversion (Ellis et al., 2004) and pseudowords can still activate semantic processes (Nation & Cocksey, 2009). Furthermore, the involvement of real words, pseudohomophones, and pseudowords highlights the theoretical differences between processing different types of linguistic stimuli. For instance, a direct lexical route will theoretically be used for relatively high frequency real words, while both pseudohomophones and pseudowords will require a grapheme-phoneme conversion route. The crux is that the comparison between real words (RW) and pseudohomophones (PH1) in the oLDT differ in more than stimulus-level legitimate orthography, but how they are theoretically processed too.

The use of only real words (and no pseudowords or pseudohomophones) for all conditions in Study 2 helps to control potential contamination from different processing routes and top-down phonology/semantics (Luck, 2004), while also allowing exploration of validity and reliability between designs through respective measures of orthographic and phonological processing in Study 1 and Study 2. This will inform the relationship strength between processing in the different stimuli and task circumstances e.g., the measure of real word responses in the oLDT and the real words in the RRT or the predominantly phonological effects of the pLDT and the orthographically incongruent rhyme condition in the RRT.

Due to the automaticity and relative simplicity of VWR and reading (Hauk et al., 2012), the simpler the task in VWR studies, the more ecologically valid and attributable to reading processes any findings can be argued to be. It is also especially important for tasks to be as easy as possible for L2 participants, while also obtaining as clear output as possible from native readers for their data to serve as the best control for ESL group comparisons. Lastly, using the same task, the same instructions, real (relatively high frequency) simple words for all experimental manipulations, allowing direct comparison between them, enhance ecological validity and are recommended when using EEG/ERP due to its sensitivity to external and extraneous variables, therefore increasing the impact of any findings (Luck, 2004).

5.3.2 Counterbalancing

The order of tasks in Study 1 and Study 2 was fully counterbalanced within groups, as well as within and between studies, and stimuli were presented pseudorandomly within each task for each participant. Response bias was minimised by using equal numbers of trials for each condition in each task (van der Mark et al., 2009). Further study-specific details about counterbalancing will be provided in the study-specific chapters (5.9 and Chapter 7:).

5.3.3 The ERP approach

The current research investigates the timing and nature of orthography- and phonology-related brain activity within ~200ms after seeing a linguistic stimulus in order to contribute evidence to the behavioural and electrophysiological VWR literature and to complement the existing body of neuroscience work. Behavioural response times, recorded separately or alongside neuroimaging techniques, provide vital indications of real-world effects, but are still not sufficient to discern between different neural processes, especially in such short and early timeframes as is necessary with VWR (Timmer & Schiller, 2012). In terms of brain activity, while fMRI studies have provided a breadth of information about the nature and neuroanatomy of reading, the temporal resolution of fMRI does not afford it insight into the specific timing of reading processes

(Pattamadilok et al., 2015). Neuroimaging methods, such as fMRI, are characteristically limited in temporal resolution compared with spatial resolution and are therefore not appropriate for investigating the absolute or relative timing of brain activity (Hauk et al., 2012). Furthermore, many neuroimaging studies are methodologically unable to directly contrast their data with behavioural data, sometimes because the temporal resolution of the neuroimaging method is incompatible with behavioural measures, other times simply because the behavioural data is not collected simultaneously (van Heuven & Dijkstra, 2010). The combination and integration of brain and behaviour, however, is a vital and often overlooked perspective in language research, hence the commitment to use such a perspective in the current research. The ERP approach, however, is particularly appropriate for investigating time-based phenomena due to the high (potentially sub-millisecond) temporal resolution of EEG, which is the primary reason for its usage in the current research.

Due to the specific focus on timing in this research and the nature of recording event-related potentials, details of stimulus presentation timing are critical. ERP components are typically based on time-windows that are broadly accepted, but not entirely or necessarily consistent from study to study e.g., 80-120ms for P1, 130-210ms for N170. Therefore, more precise timings of between-factor processing are not usually forthcoming, despite the possibility of millisecond (and even sub-millisecond) precision. The use of peak latency measures could deceptively appear to allow such specific timing with its single time-point, but the methodological, statistical, and physiological issues with this approach are now well-documented and with the technological advances of recent decades, researchers no longer have to rely on outmoded measurements that they may previously have had no choice but to use.

Variable stimulus onset asynchrony (SOA) was used in both studies to introduce controlled jitter into the timing of the tasks. This allows cleaner responses to be recorded by helping to prevent participants from providing anticipated responses that can occur based on identically timed cycles of trials. Further study-specific details about timing will be provided in the study-specific chapters (5.9 and Chapter 7:).

Where the EEG/ERP approach is limited and how it differentiates from neuroimaging techniques (e.g., fMRI, PET, fNIRS) concerns its lower spatial resolution and the accompanying inability to accurately specify the neuroanatomical origins of the neural signals it documents. In other words, EEG can detect and record the necessary signals, but cannot deduce where in the three-dimensional headspace they came from, which is known as the inverse problem (see Ryynanen et al., 2004). Due largely to this inverse problem based on the limited spatial resolution of EEG, using an ERP methodology precludes direct inferences about the neuroanatomical locations of neural generators and associated experimental effects. It is acknowledged that magnetoencephalography (MEG) is capable of capturing valid temporal and spatial measures, but it is also substantially more expensive and was not a viable financial or practical option here. Therefore, with all factors of theory, method, finance, and access considered, the ERP approach is the best option to answer questions about entwined cognitive processes in time and, therefore, the most appropriate methodology for the current research.

5.3.3.1 Event-related potentials

Event-related potentials (ERPs) are derived directly from continuous electroencephalogram (EEG) by averaging the data (per electrode site/cluster) of multiple discrete timeframes (often called epochs) that represent experiment trials of a specific condition e.g., all the presentations of a real word in an LDT. However, due to the relatively small amplitudes of electrophysiological activity and the high sensitivity of EEG (not only to legitimate brain activity, but extraneous noise), relatively high numbers of trials are required to attain a suitably strong signal-to-noise ratio (SNR) for a valid ERP output (Luck, 2005). ERP data collection is, therefore, largely a battle between the specific signal that is the part of the EEG recording attuned to the cognitive property of interest and the potentially damaging noise made from background brain activity, non-brain electrophysiological activity (e.g., EOG), and environmental interference (e.g., 60Hz line noise). More trials resulting in more statistically powerful averages is widely accepted as a key solution, though special care should always be taken at the recording stage. In some ways, this is not different from collecting behavioural data (e.g., response times), which also need multiple trials to obtain statistical power, though the case of ERP data has more serious consequences. Without sufficient trials, SNR, and clean waveforms, measurements can struggle with validity and may not accurately reflect the cognitive property of interest.

Early pre-200ms electrophysiological brain responses are reasonably stable, especially compared with later processing, so can provide a good measure of early processing, contributing less to the variability often seen between studies and participants (Hauk et al., 2012). This stability is partly attributable to the sensory processing of physical stimulus attributes, but also to the automaticity developed from performing an action such as reading words so frequently over many years (Cornelissen et al., 2009). Interactions between such sensory processing in the visual system and language-related processes could begin as early as ~60ms (Assadollahi & Pulvermuller, 2003), but certainly occurring within the first 200ms post-stimulus (Cornelissen et al., 2009). Considering reports of different psycholinguistic effects occurring at various points across the timeframe (Cornelissen et al., 2009), however, it is unclear specifically which processes are involved in these interactions and when. For instance, phonological effects have been found around 100ms (Ashby, 2010; Wheat et al., 2010), while lexical-semantic manipulations have been interpreted to influence ERP waveforms at 168ms (Segalowitz & Zheng, 2009) and 200ms (Hauk & Pulvermüller, 2004; Moscoso del Prado Martín et al., 2006). However, effects from ERPs support orthographic activation as occurring prior to phonological activation (Timmer & Schiller, 2012) with fast masked phonological priming studies proposing phonological activation to be approximately 20-30ms behind orthographic activation (Ferrand & Grainger, 1993). Furthermore, processing of different reading-related factors could follow different timecourses (Amsel, 2011), perhaps dynamically dependent on stimulus and task. It is important, therefore, to note a fundamental principle of (serial and parallel) processing in terms of ERP activity: brain activity associated with late components/subcomponents will sometimes rely on earlier activity and the processes the earlier components/subcomponents reflect, while earlier processes may rely on stimulus properties directly or other cognitive processes. Consequently, the understanding of these early neural processes and their timeframe is vital for a better understanding of the interactions between subsequent processes necessary for accurate word recognition. For all contexts, though, further investigation of the initial ~200ms of electrophysiological activity is required to better understand this early processing and the foundation of later processing in both L1 and L2 readers.

5.3.3.2 ERP measurements

As with standard behavioural analysis, ERP analysis is based on data from each participant that is submitted for appropriate statistical analysis (e.g., paired-samples *t*-tests between conditions within a group). Unlike such definitive behavioural measures as response time, however, there are various options for measuring ERP waveforms (and the numerical data used to represent them), which are the root of some methodological issues of ERP analysis.

Most commonly, ERPs are described in terms of components (e.g., P1, N170) that have accrued typical time windows (e.g., 80-120ms for P1, 130-210ms for N170) to denote their predominant period of activation. Furthermore, ERP components referred to with the same name often have different unstandardized time windows between studies. As an example, effects on a "P150" component based on a 100-160ms timeframe and an "N200" based on analyses of a 160-200ms timeframe have been reported (Coch & Mitra, 2010), which typically and drastically overlap with the visual P1 peaking at ~100ms and the onset of the subsequent major negative deflection, the N170. The same time windows are neither used consistently between studies, which marks one key issue, nor necessarily valid for all participants in all situations with some components being more variable in time than others. However, even if time windows were chosen a priori or the same ones were used in every ERP analysis, there is still the question of the method of measuring in terms of obtaining numerical data that can be subjected to statistical analysis. For instance, ERP waveforms can be measured in a variety of ways for both amplitude and latency, using peak, mean, fractional area, and other calculations.

Peak measurements of amplitude and latency were the norm, but this was largely due to lack of computational power to calculate more complex measurements, such as mean and positive/negative area amplitudes. Alongside more contemporary options that go beyond component-based analysis, such as the mass univariate approach (Groppe et al., 2011), the measurement options for components are positive/negative area (depending on component deflection) and mean for amplitude, alongside fractional area for latency (e.g., 50% area latency or 25% for onset approximation). Area measures are typically more robust, being based on more than a single data point (as in peak measures), resulting in being generally less sensitive to noise and less sensitive to the choice of timeframe than, for example, peak measures. Furthermore, while area measures are not a direct descriptor of the shape of ERP components, they can provide an indication of the component height and breadth as opposed to just the highest amplitude (as in peak measures). As with any method, though, each have their strengths and weaknesses, but the only reasons to use peak measurements are for direct replication or methodological review, while mean- and area-based are both preferred in any other case (Luck, 2014).

Mean amplitude is sensitive to overlapping components and can be essentially confounded if there is significant activity in opposite polarity e.g., negative amplitudes in the timeframe for measuring a positive component, as the required measurement will be partially cancelled out. Using signed area amplitudes (i.e., positive or negative area amplitudes) can be preferable to overcome this when measures of the component in question are expected to be only positive or negative amplitudes e.g., large components, such as P1 and N170, as signed area amplitude measures essentially ignore any amplitudes of the opposite polarity e.g., offsets of previous or onsets of following peaks/components, avoiding the cancellation that occurs for mean amplitude. This also means that the measurement timeframe can be wider than for mean or other amplitude measures (as long as it does not encroach the next peak/component of the same polarity). By their nature, signed area amplitudes have a minimum of zero and so are more sensitive to noise than mean amplitudes, which can be problematic when comparing across groups with different signal-to-noise ratios (e.g., due to different numbers of correct trials viable for averaging), but this can be addressed with appropriate data processing (in terms of EEG/ERP e.g., filtering, artefact correction as well as statistical data cleaning e.g., treatment of outliers) and can be preferable overall to the complication of cancellation in mean amplitudes. Using signed area amplitudes also better coincides with the usage of (50%) fractional area latencies to measure the ERP latencies and eliminates the effects of latency jitter (e.g., within participants) for monophasic components (e.g., P1 and N170).

The outcome is that there are numerous choices, many with significant advantages or generally acknowledged shortcomings, but no dominant consensus about which should be used, what each specifically mean in terms of neural processing, and what relationship they have between one another. The consistent presence and prominence of the P1 and N170 along with their relative stability due to their early appearances (Cornelissen et al., 2010), in visual paradigms allows closer focus on manipulations rather than eliciting the components (Luck, 2005), which provide advantages for the current research, not least in terms of measurement choice. Such nature of the P1 and N170 components alongside the aforementioned points about robustness and sensitivity informed the selection of area amplitude and fractional area latency as the principal measures of ERPs in the current research, using a 50% fractional area to approximate peak amplitudes without using peak measures.

In addition to potential issues of ERP measurement, time windows, and SNR, variability within EEG data and ERP measures is also acknowledged. The within-participants variability of the location, size, and even response of a component or region is one possible reason for the inconsistencies between studies supposedly investigating the same cognitive processes with the same stimulus types (Glezer & Riesenhuber, 2013). While measuring from groups will still ultimately involve the potential issues of within-participant variability, clustering electrodes from the higher density 128-channel EEG used in the current research can help to alleviate problems.

5.4 Stimuli

The scientific method demands control of stimulus parameters. Analysis cannot discriminate between the uncontrolled variable(s) and the experimental variable and so cannot be relied upon to accurately determine the locus of any findings. Ideally, all stimulus parameters would be balanced across conditions/sets/tasks (where applicable) or, at least, all *known* stimulus parameters would be. In practice, however, suitable stimulus sets can be difficult to create, but it is crucial that they are appropriate for the experiment, being its core and the way participants interface with the task. Any other factors that may be extraneous can be confounding and should also be controlled where possible and reasonable.

There is a wealth of psycholinguistic factors that can have potentially confounding effects on psycholinguistic/VWR responses. Indeed, factors of lexical equivalence (i.e., how words can be similar) are so numerous that stimulus control is a problem that can only realistically be minimized and not eliminated (Van Orden & Kloos, 2005). However, many studies only control what might be considered a bare minimum of psycholinguistic variables, such as just number of syllables, orthographic length, and lexical frequency (e.g., Pattamadilok et al., 2017). Instead, it is especially vital that linguistic stimuli are tightly controlled across as many psycholinguistic factors as possible, whether they are the specific focus of the research or not and certainly if they can potentially influence responses, as not doing so can impact on the validity and reproducibility of findings (Izura et al., 2011).

One practical issue concerns which psycholinguistic factors can be controlled with more or less leniency, as some flexibility is typically required just to allow a sufficient number and appropriate range of stimuli for the purpose of the experimental design. It is, therefore, duly noted that ideal linguistic stimulus sets are rarely possible to form without relaxing the control of some parameters (Van Orden & Kloos, 2005). For instance, a common problem is attaining sufficient numbers of stimuli, which is particularly challenging in ERP studies due to methodological demands for significantly more trials than in, for example, a purely behavioural experiment. Nonetheless, strict control and quality for stimuli is necessary to be able to accurately interpret results later, further highlighting the importance of identifying which parameters can be flexible and least likely to influence results.

Stimulus set creation was a major challenge in the current research design. Besides the ERP method itself requiring substantial numbers of trials for sufficient signal-to-noise ratio due to ERPs being especially sensitive to changes in stimulus properties (hence being a strong and widely used method for such research), there were several reasons for this challenge. Namely, the experimental manipulations in both studies are directly operationalized through the English (or English-based) lexical stimuli, making stimulus selection especially critical. While English is a language with a vast and varied vocabulary, not all English words were viable for inclusion (even before considering potential interlingual issues). This was due to them either not passing tight inclusion criteria (based on various psycholinguistic variables that are known to influence processing – see following section) or not being viable for one of the experimental conditions. For instance, relatively few English words have an orthographically rime-matched non-rhyme counterpart e.g., *mint-pint*, as required for rhyme recognition tasks (such as the one used in Study 2). Furthermore, the relative novelty of the behavioural tasks being used with English monolingual and late bilingual ESL participants also demanded additional attention that stimuli were appropriate for all groups in terms of not being too difficult for ESL participants (but not being too easy for native readers).

Creating stimulus sets to be used with ESL bilinguals, especially those with an alphabetic L1 (e.g., Spanish-English), without significant interlingual conflict as well as meeting the demands of the behavioural tasks (where the nature of the English lexical items contributes directly to the experimental conditions) was especially challenging. While there is little direct overlap between English and Chinese that might influence Chinese-English bilinguals' reading performance (especially on an orthographic level), interlingual influence in bilinguals with L1 and L2 of the same or similar type (e.g., both Germanic/alphabetic as in Dutch-English) is well-documented. The inherent linguistic interlingual similarities between English (the target language) and Spanish (the L1 of one experimental ESL group), therefore, added another layer of complexity to stimulus creation.

Following evidence for an integrated nonselective access model of bilingualism (van Heuven & Dijkstra, 2010; Dijkstra et al., 2018), L2 counterparts of L1 words are activated in parallel in bilinguals with L2 being unconsciously translated automatically (on at least phonological and semantic levels), albeit to varying extents dependent on the stimulus. While stimulus control is essential, suppression of L1 translations during L2 reading is posited not to be possible due to this automaticity of bilingual processing and not feasible due to the overlapping phonemic profiles of some languages (especially in the case of etymologically related languages using the same system, such as English and Spanish). However, such suppression (imagining it possible) would arguably be unnatural and not truly representative of how bilingual readers read L2 anyway, leaving a predicament in limiting interlingual factors so artificially (as they will never be fully eliminated). Therefore, it is reasonable to accept that stimuli of the same language type as the ESL bilingual's L1 (e.g., alphabetic as in a Spanish-English bilingual) will always and consistently trigger both languages, albeit to varying extents based on interlingual similarities and other psycholinguistic contexts/factors. Consequently, it is not an exaggeration to claim that no amount of control could ever eliminate interlingual influence in bilinguals with languages of the same type, at least not without severely damaging ecological validity, something that is already a methodological issue in lab-based psycholinguistic research. Furthermore, all stimuli cannot be cross-checked and cross-referenced with all possible translations, derivations, and connotations in all potential participants. For instance, it is impossible to know the extent of phonological overlap on a per-participant level and/or for which stimuli automatic unconscious L2-L1 translation occurs e.g., not all Chinese-English readers with equivalent L2 proficiency will experience L1 phonological priming when seeing an English stimulus pair that have phonological similarities in their Chinese translations. Ultimately, controlling stimuli and tasks for as many psycholinguistic factors as reasonably possible, maintaining the L2 context and minimizing L1 involvement, and checking item-level data are all that can be done.

Due to the needs of this research to be specific about orthographic and phonological characteristics of stimuli and any psycholinguistic factors that can potentially influence VWR, it is recommended to be as rigorous in stimulus selection as possible. For full transparency, the method section of each study chapter will highlight any potentially problematic stimuli and explain why each was accepted for inclusion (over and above their properties for operationalizing the experimental task).

Essentially, in addition to usual complexities and requirements of an experimental stimulus set, the various ways in which languages can overlap and that bilingual readers can be influenced by cross-language lexical properties (as discussed in §3.3), the research questions and the tasks designed to address them added further complexity on the stimulus constraints for creating appropriate and meaningful stimuli. Therefore, several allowances had to be made in the creation of the stimulus sets, resulting in a small number of stimuli that are acknowledged to be potentially problematic and slightly less representative of the cognitive processes they are being used to operationalize. However, all stimuli will be re-evaluated through item analyses, taking behavioural responses into account, with potentially problematic items being especially scrutinized and only the strongest and most representative samples being used in the overall analyses.

5.4.1 Psycholinguistic variables

Many psycholinguistic variables should be controlled in research using word-based stimuli due to their capacity to influence processing, which is especially the case when using sensitive neurophysiological methods. However, it is important to note that for some variables it is sufficient for them to be matched (e.g., lexical frequency), some would ideally also be maintained at specific a priori values (e.g., orthographic length, word class), whereas others should simply be minimized (e.g., semantic relatedness).

All of the following psycholinguistic variables have the potential to have a significant impact on visual word recognition performance and therefore the potential to skew interpretations and conclusions of those performances if not sufficiently controlled. The criteria used in this research allow studying a precise set of psycholinguistic properties that will assist analysis and support interpretation of the data. They also help to limit potential confounds from cognitive processes not directly related to the purposes of the research, such as unwanted priming and attempts to integrate the stimuli with one another (i.e., from phrasal associations and semantic relatedness).

5.4.1.1 Orthographic and phonological length

Due to the focus on visual and particularly orthographic processing, it was necessary to control such an overt physical property as orthographic length. However, due to EEG/ERP methodological constraints that dictate a requirement of significantly more stimulus presentations for usable and meaningful analyses than in, for example, a purely behavioural experiment, sufficient numbers of stimuli with the necessary level of control over other factors were not possible using only 4-letter, 3-phoneme words, so the stimulus pool was increased to include 5-letter and 4-phoneme words too. This was acceptable, as orthographic word length is one psycholinguistic parameter that has been shown to allow some flexibility in terms of stimulus creation. More specifically, a single letter of difference in orthographic word length does not typically lead to word length effects (Jalbert et al., 2011). There is also evidence that any such effects of similar word length, such as comparing conditions with means of 6.5 and 8.5 letters, disappear in adulthood anyway (Acha & Perea, 2008).

Orthographic length facilitates recognition at 3-5 letters, has no effect at 5-8 letters, and inhibits performance at 8-13 letters (New et al., 2006). This could mean that the 4-letter words and perhaps some of the 5-letter words might facilitate responses just based on their orthographic length, while some of the 5-letter words will have no length effect. However, the proportion of 4- and 5-letter words as well as 3- and 4-phoneme words was equal between conditions and this controlled mix of word lengths offers the benefit of minimising any facilitative effects of the 4-letter stimuli, while also reducing potential visual repetition effects and visual coding (Baddeley et al., 2002).

Sometimes, despite the substantial control and detail of stimuli, individual pairs are not matched on highly relevant factors for rhyme recognition, such as orthographic length (number of letters) or phonological length (number of phonemes), as in *chair*/*bear*, *tale*/*snail* (e.g., Grossi et al., 2001; MacSweeney et al., 2013), resulting in some pairs not being matched for orthographic word length, while others were. It is important to note that the word lengths were not altered or manipulated between base words and pseudohomophones i.e., the pseudohomophone counterpart of a base word has the same number of letters and phonemes in the same structure as the base word.

5.4.1.2 Word class

Single words are rarely seen in isolation and syntactic (along with semantic) constraints surround word recognition processes. This highlights some potential issues surrounding syntactic constraints that may even be important to single-word processing, such as word class varying within the same stimulus set, which are not limited to just using nouns and verbs or simple forms, but ranging from singular count nouns that would ordinarily be preceded by a determiner (e.g., article or quantifier) to past tense verbs that are typically preceded by the particle to, a pronoun, adverb, or auxiliary verb. For example, the stimulus pair *jazz-has* not only manipulates the orthography-phonology relationship between prime and target as intended for a rhyme condition (Bitan et al., 2009), but word class differs between them, resulting in a different set of psycholinguistic values attributed for each. These rules create implicit syntax for many word forms that may or may not impact processing times. However, as there is currently no evidence to reject the premise that syntactic constraints substantially influence the way isolated words (as well as words in sentential contexts) are processed, it is sensible to attempt controlling for these factors.

Simple nouns are typically chosen for their ease of comprehension and the focused elements of processing can be observed more clearly (Almeida & Poeppel, 2013), while mixing the word class within experiments when it is not part of an experimental manipulation can and likely will cause unnecessary confusion for the participant. The even mix of singular count and mass nouns (i.e., no plural forms) within each condition also allows a mechanism for keeping attention without compromising syntactic boundaries and the potential confound of syntactic inhibition from mixing word classes.

5.4.1.3 Phonological/phoneme structure

The use of count nouns with a CVC phoneme structure allows the pre-consonant indefinite article 'a' to be consistently implicit, thus controlling potential effects of syntax. It has also been reported that the impact on word recognition processes is different between vowels and consonants (Carreiras et al., 2009), further supporting the constant phoneme structure across all stimuli.

5.4.1.4 Vowel types

Any vowel sounds that were deemed potentially difficult for non-natives to produce were avoided. Due to restrictions in the grapheme-phoneme mappings of English (e.g., many monophthongs do not have a phonotactically legal graphemic alternative), the pseudohomophone sets (PH1 and PH2, respectively) included a more restricted usage of vowels than the real word and pseudoword conditions. However, there was still a healthy mixture of vowel types within the pseudohomophone conditions.

5.4.1.5 Frequency and familiarity

High frequency, highly familiar words were used for several reasons. Most important is that low frequency, unfamiliar words increase difficulty, which is an especially important factor to minimize here due to the involvement of L2 readers of English. English word frequency has been found to influence behavioural and electrophysiological reading performance in French-English bilinguals (who are comparable in language profile to the Spanish-English participants), but ERP effects were not observed until ~300ms (Peeters et al., 2013), after the timeframe under scrutiny in the current research. Higher frequency, more familiar, and easily comprehensible theoretically allowed fairer comparison between L1 and L2 speakers of English: using words that L2 readers may not recognise or even that are only known by native readers would conflate findings absolutely and negate the purpose of the study.

It is also very important to note that other psycholinguistic effects can be attenuated through the control of frequency and familiarity. For instance, high-frequency words have been shown to render effects of regularity, word neighbourhood, and visual complexity non-significant (Coltheart et al., 2001), while word length effects dissipate with word familiarity (Alario et al., 2007), and high-frequency words are much less problematic than low-frequency words in terms of repetition effects (Almeida & Poeppel, 2013). Along with its potential effect on the route used to gain lexical recognition (Coltheart et al., 2001), these factors strongly advocate frequency to be a vital property to control, especially when considering the brain as functioning with statistical information (e.g., Dehaene, 2014).

Word stimuli are often erroneously split into high and low frequency conditions by dividing the continuous variable of lexical frequency at some arbitrary point and without definition (e.g., Fischer-Baum et al., 2014; Glezer et al., 2016). Where applicable, therefore, all stimuli were relatively high-frequency (based on British CELEX values) and not arbitrarily split, instead being statistically matched between conditions

(Almeida & Poeppel, 2013). Any difficulties stemming from frequency or familiarity were minimised and the intention was for all participants, native or L2 English readers, to reach ceiling effects for performance where possible.

5.4.1.6 Age of acquisition

Age of Acquisition (AoA) is a measure of when words are initially encountered and learnt during reading skill development in childhood, literally referring to the age of children when a word is usually first encountered and learnt. It is important to acknowledge and control AoA in VWR research, as it can affect how quickly and efficiently words are read. For instance, *penguin* will typically be learnt and used at a much earlier age than *albatross*, so *penguin* will be processed more quickly and efficiently than *albatross* after both have been learnt, even in skilled adult readers. While it is yet to be seen how native-based AoA measures work with L2 readers and L2 acquisition, AoA of L2 in late bilinguals has been highlighted as a confound for observing how proficiency can affect brain activation from L1 and L2 (van Heuven & Dijkstra, 2010). For instance, L2 proficiency has been shown to affect brain activity associated with semantics, while L2 AoA affects grammar-related areas (Wartenburger et al., 2003). This is related to L2 usage and exposure, which have also been shown to influence brain activation during L2 processing (Perani and Abutalebi, 2005; van Heuven & Dijkstra, 2010; Vingerhoets et al., 2003), but has been measured by self-report and accounted for in the analysis. It is, therefore, a useful complement to measurements used to balance stimulus sets. It provides an indication that conditions are matched on a level of general lexical complexity, something that could be very important in observing psycholinguistic effects in L2 readers. Therefore, AoA was not only controlled through the use of relatively high frequency short words, matched on many factors (including concreteness and imageability), which is deemed sufficient for controlling AoA for the adult skilled readers from L1 and L2 groups (Izura et al., 2011), stimulus sets within each study were also matched on the Bristol norms for AoA (Stadthagen-Gonzalez & Davis, 2006).

5.4.1.7 Bigram frequency

Bigram analysis has been suggested to be a vital part of visual word recognition and that its effects occur within 200ms of seeing a word (Hauk et al., 2008). This falls within the critical time period being investigated by this research, requiring bigram frequency to be matched between conditions. Bigram frequency also contributes to visual familiarity, which is of particular significance in visual word recognition studies, especially when the focus is on such initial and early responses. It should also be noted that mean token measurements of bigram and biphone frequency were used for comparison due to being more appropriate for word identification tasks (Knight & Muncer, 2011).

5.4.1.8 Orthographic and phonological neighbourhoods

Due to the focus on orthographic processing, orthographic neighbourhood, using Coltheart's metric of orthographic neighbourhood size (Coltheart's n), was one of several orthography-related variables to be controlled, especially as it has been shown to influence lexicality decisions (Proverbio & Adorni, 2008).

Phonological neighbourhood effects can occur for pseudowords similarly to low frequency real words, while responses to high frequency words are not significantly influenced, presumably because the effect of frequency essentially masks or subsumes that of neighbourhood (Coltheart et al., 2001). Controlling phonological neighbourhood between conditions, therefore, was necessary and helps to increase comparability between stimulus types.

5.4.1.9 Semantic relatedness

As phonological and semantic processing are directly linked and separating them is problematic (Brunswick, 2010, p. 87), semantic activation must also be balanced across conditions as much as possible. Semantic processing, albeit in different forms and to different extents, arguably percolates through every level of visual word recognition. This is not unexpected, as per the ultimate purpose of reading, but it does have implications for methodologies not directly investigating semantic processing. Therefore, in order to observe effects of orthography and phonology more directly with confidence that they are such and not significantly driven by semantics, a targeted approach is required.

Semantic associations were avoided as much as possible within and across conditions in each task and no significant semantic string, recurrent semantic theme, or recognisable semantic field was present in the stimulus sets. Latent Semantic Analysis (LSA) is a research tool that can provide a context-based measure of semantic association between words, phrases, and documents (Landauer et al., 1998). Using the LSA scale based on general word usage up to university age, it was used to support the manual checking of semantic association between all stimuli by highlighting any strongly associated (>0.8) pairings to be reassessed. Any semantically associations suggested to be too strong by either manual checking or LSA were either discarded or separated into different trial blocks.

5.4.1.10 Imageability

Imageability and concreteness have been shown to affect behavioural responses and brain activity, distinguishably influencing characteristics of ERPs, such as the response potential (Martín-Loeches et al., 2001). Therefore, as imageability is not a focus of this research, stimulus sets were matched in terms of imageability, using values from the MRC database (Coltheart, 1981).

5.4.1.11 Phrasing

Due especially to the potential conflict when involved in prime-TARGET pairs (e.g., Landi & Perfetti, 2007), syntactic and phrasal associations, when multiple words form an extended or different meaning (e.g., *bat* and *man* are separate nominal entities, but compounded, they form *batman*), were limited as much as possible within all stimulus sets.

5.4.1.12 Interlingual factors

In addition to the already numerous psycholinguistic factors outlined in the previous sub-sections that must be controlled within the target language, bilingual participants' native languages must be considered in terms of potential cross-linguistic/interlingual conflicts i.e., factors influencing word processing between L1 and L2. Different languages can overlap linguistically in various orthographic, phonological, semantic, and/or syntactic ways, which increases with the relatedness of the languages. Spanish has much greater overlap with English

As interactions between L1 and L2 cannot be avoided (van Heuven, Schriefers, Dijkstra, & Hagoort, 2008), cognitive traces of a native language stopped early in development can still be found despite not consciously knowing the language (Pierce et al., 2015), and stimuli similar to words in L1 of the ESL groups can potentially lead to phonological priming effects when processing L2 (Carrasco-Ortiz et al., 2012a; Wu & Thierry, 2010), interlingual factors were a major concern in creating stimulus sets. Interlingual homographs, interlingual homophones, and cognates between English and the ESL participants' native languages were minimized as much as reasonably and practically possible by checking all stimuli against relevant lexical databases and dictionaries, as well as consulting several native speakers of the ESL groups' native languages who did not take part in the main study to further check for real-world and potential colloquial usages. However, the similarity of English and Spanish due to their largely shared alphabetic system means that interlingual crossover can never be entirely avoided without compromising the usefulness of the stimuli, which is especially pertinent, as this similarity of languages and shared language system was a significant reason for using Spanish-English bilinguals in this research. Therefore, some carefully considered allowances will be made when vetting interlingual items: conjugated forms in Spanish and loan words from English will be permitted, items of notably low frequency in Spanish will not be excluded.

One of the most important points about the design using a wholly English context is that the focus of the work is the processing of English (and only English) at orthographic and phonological levels (semantic processing is not a focus, though semantic factors are controlled). Although neither local, sentence, nor global contexts have been reliably found to influence the automatic initial processing of word stimuli in both bilinguals' languages (van Heuven & Dijkstra, 2010), the wholly English global context (in preparation and presentation of the experiment) could work to "regulate the selection of lexical representations" (van Heuven & Dijkstra, 2010), minimizing cross-linguistic conflicts and errors. The context is English reading and the tasks require English cognition and responses: participants will not be included in analysis without sufficient accuracy in responses and item-level responses will be excluded if they are outliers. Therefore, in order to further minimise any effects of potential interlingual factors and to focus the responses of ESL participants as much as possible, it was made clear that the experiment was being conducted in English and was not directly concerned with their native languages, but with their processing of English specifically.

5.4.2 Stimulus criteria

All stimuli across both studies 1 and 2 were presented visually in bold, black, lower case, **Courier New** font (42pt) on a silvery grey background (E-Prime colour="gray"). This milder contrast between font and background colour (when compared, for example, to black text on a white background) was used to reduce participant fatigue. The larger size was used simply to make the stimuli more easily and immediately readable. With regard to these technical details, however, there does not yet seem to be a consensus from previous research and so decisions were made based on attempting to make the stimuli as clear and easy-to-read as possible for all participants.

5.4.3 Stimulus creation

The MRC Psycholinguistic Database (Coltheart, 1981) was used to generate a large pool of real English nouns that were monosyllabic, three phonemes long, with a consonant-vowel-consonant (C-V-C) structure. This pool was then manually filtered further, discarding inappropriate, archaic, and uncommon words, to leave a usable pool of semantically simple English nouns, including no technical jargon, explicitly dialectal/regional words (including American English words and words biased toward British English[2]), and words with accent-dependent phonology (e.g., *bath*, *grass*, *bus*), following evidence that inner speech resembles the reader's own accent (Filik & Barber, 2011). The resulting pool was used for the real words (RW condition) and as base words for the two pseudohomophone sets (PH1 and PH2) and the pseudoword set (PW) in Study 1, as well as for all stimuli in Study 2, as will be described in respective study-

[2] While the native English participants would arguably have a predominantly British English vocabulary, American English plays a significant part in the acquisition of English for native speakers of many other languages, including Spanish and Chinese, due to the reach of American television and film.

specific sections. In terms of either their base words or stimulus forms, stimuli were not repeated across conditions or tasks.

5.4.4 Stimulus control

While real words, by definition, have values for all psycholinguistic variables, pseudowords and pseudohomophones (as used in Study 1) have only a subset of these properties. These limited psycholinguistic values and the infrequency of such non-words in the real world can affect ecological validity, as reading such stimuli is limited to demonstrating grapheme-phoneme conversion (Ellis et al., 2004) and pseudowords can still activate semantic processes (Nation & Cocksey, 2009). Furthermore, using only real words for all conditions (and no pseudowords or pseudohomophones) also helps to control potential contamination from top-down phonology and semantics (Luck, 2004). However, the usefulness or even necessity in particular psycholinguistic paradigms is unavoidable and undeniable (e.g., lexical decision tasks). Nevertheless, pseudowords do share legitimate psycholinguistic properties with real words on a sub-lexical level (graphemes, phonemes, n-grams) to which psycholinguistic variables, such as frequency, familiarity, and regularity, can also be applied (Hauk, Davis, et al., 2006). The number of letters, bigram and biphone frequency, and orthographic and phonological neighbourhoods can also be measured for pseudowords, though their lack of semantic content makes measures of whole word frequency, familiarity, and imageability impossible. In the case of the pseudohomophones and pseudowords for Study 1, therefore, the criteria apply to the base word.

5.4.4.1 Statistical support

The current research focused on matching conditions across the following measures of psycholinguistic factors: orthographic length (i.e., number of letters), phonological length (i.e., number of phonemes), mean frequency (per million, British English), bigram and biphone frequency (mean token values), orthographic neighbourhood (Coltheart's n), phonological neighbourhood, imageability (MRC; Coltheart, 1981), familiarity (subjective frequency; Balota et al., 2001), and semantic relatedness (using Latency Semantic Analysis; Landauer et al., 1998). The CELEX database (Baayen et al., 1995) and MRC database (Coltheart, 1981) via N-Watch (Davis, 2005) were used to obtain all stimulus property values except for semantic relatedness, which were retrieved from the LSA website (http://lsa.colorado.edu/). Furthermore, overt semantic and syntagmatic links were minimised as much as possible throughout the sets.

5.4.4.2 Scale-based assessment

In order to further bolster the strength and validity of the stimulus sets, scale-based variants of the orthographic and phonological lexical decision tasks (Study 1) and rhyme recognition task (Study 2) were created using the same pools of stimuli. These were completed online (via Qualtrics: www.qualtrics.com) by individuals who did not take part in the main studies (details can be found in the study-specific chapters). These surveys not only support the balancing of stimuli, but support their use as exemplars in the respective conditions. Summaries of the results and comparison tests are presented in the stimuli section of the relevant study-specific chapter.

5.5 Apparatus

Stimuli were presented to participants on a 22” LCD monitor (LG L226WTQ-PF) using E-Prime (version 2.0.10.353; Psychology Software Tools Inc.). The display ran at 1680x1050 resolution with a 60Hz refresh rate (16.67ms refresh duration) to which stimulus presentations and related event markers were synchronised. Time-locked event markers for conditions and individual trials were transmitted from E-Prime on the stimulus presentation computer to asalab (ANT Neuro) on the EEG recording computer via the EEG amplifiers using parallel port communication. Both E-Prime and asalab recorded participants' behavioural responses, which participants made by pressing buttons with their right hand on a handheld gamepad (Logitech Precision). This was used over a more typical button-based response box because it was deemed to be more familiar to participants in general.

5.5.1 EEG apparatus

During the experiments, participants were situated in an RF-shielded and sound-insulated Faraday cage (www.wardray-premise.com) to help minimise extraneous noise in the EEG recordings and in the room. All EEG data were recorded using two cascaded Refa 72-channel amplifiers (http://www.tmsi.com/products/systems/item/refa) using waveguard™ caps (www.ant-neuro.com) with 129 actively shielded Ag/AgCl electrodes (128 channels plus ground) arranged according to the 5% electrode system (see Figure 7 for electrode layout). The ANT-Neuro and waveguard™ system is gel-based (using conductive gel between the electrodes and scalp), which has several advantages over other systems, such as being less prone to bridging and affording a consistent electrode configuration across participants. EEG was recorded directly to hard disk using asalab 4.7.12 (Advanced Source Analysis laboratory; www.ant-neuro.com) from a cascaded 127-channel setup (using two linked 64-channel amplifiers, sacrificing one channel as a shared reference) through a SynFi fiber-to-USB converter. When recorded, all EEG data was processed and averaged to ERPs using ASA 4.9.3 (www.ant-neuro.com) before the relevant data exported for statistical analysis using SPSS (v26; IBM).

The use of high-density 128-channel EEG affords several advantages over the more typical 64-channel system. The main benefit is the greater spatial resolution in terms of having twice as many electrodes to cover the same area, which allows greater precision for results per region of scalp, improving the quality and signal-noise ratio for electrode clusters and theoretically enhancing source localisation (Ryynanen et al., 2004). It is, however, recognised that EEG/ERP methodologies are generally not the primary choice for investigations into the source(s) of neural generators and thus the deeper analysis in this research is still focused on the temporal order of processing. Despite this main focus and that the increased density does not directly affect the temporal resolution, having access to more sites and therefore more data per region of interest is a significant advantage. The additional channels help to cover each region more completely, as well as scalp areas not covered by 64-channel setups, resulting in more electrodes per region (for clustering) and less critical data loss. Furthermore, by using relevant clusters of electrodes, the increased density can improve the signal-to-noise ratio considerably, especially when using the average reference (Dien, 1998).

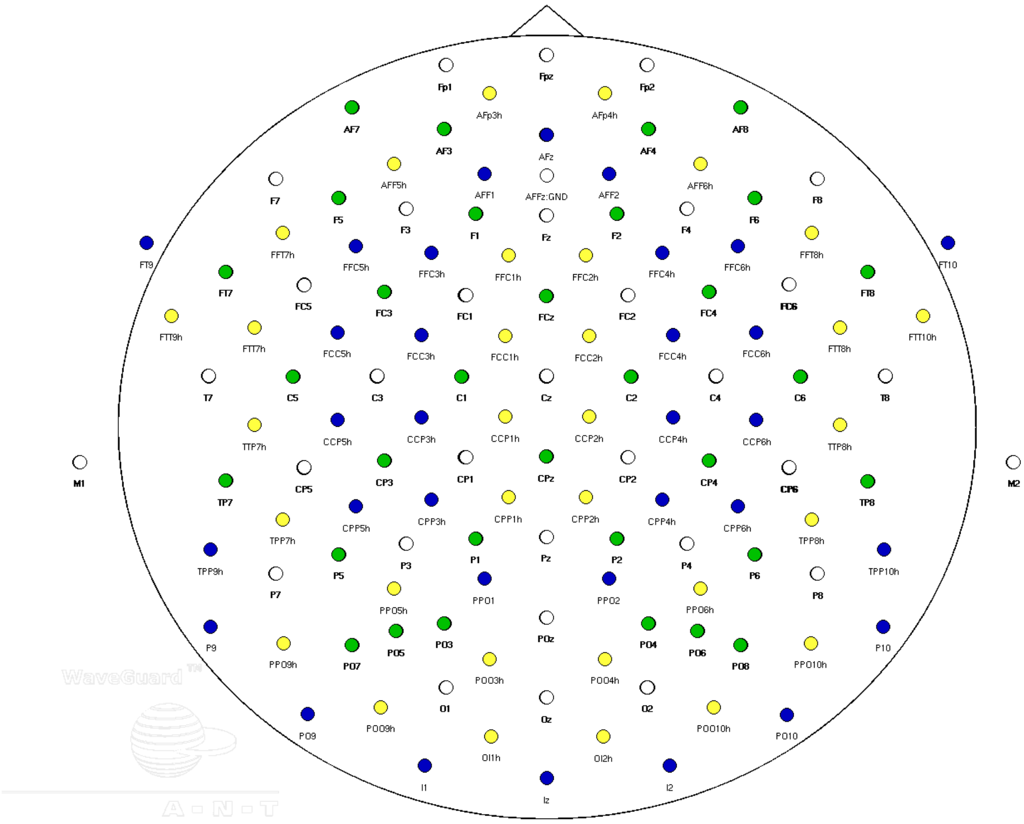

Figure 7: ANT Neuro waveguard™ 128-channel cap electrode layout

5.6 Procedure

Prior to any experimental sessions, the participant information sheet was supplied to potential participants along with a revised version of the Language History Questionnaire (LHQ; Li et al., 2014; see Appendix E and §4.2.1.2 for details), which was completed and used to aid eligibility checks. Eligible individuals were scheduled for their experimental session at a time that suited them, while participants whose LHQ did not meet the inclusion criteria were politely declined from taking further part in the study.

In the experimental sessions, the participant information sheet (see Appendix A) was provided again and both a consent form (see Appendix B) and a revised version of the Edinburgh handedness inventory was completed (see Appendix G). These were followed by a reminder of anonymity, confidentiality, and the right to withdraw from the study, as well as another chance to ask any questions before the EEG system was configured with the participant. For data acquisition, participants were seated comfortably in a sound-attenuated and dimly lit room at a suitable distance from the screen with checks made that they could easily and effortlessly read the on-screen stimuli. Each participant completed all tasks of Study 1 and Study 2 within the same EEG recording session, but the order of tasks was fully counterbalanced within each study and each group. Further task details can be found in the study-specific chapters (5.9 and Chapter 7:).

Before each task, participants were informed in spoken English by the researcher as well as via on-screen instructions what was expected of them to do. This formed part of a strategy to emphasise the English basis of the research and, in particular, that the participants were to be required consider all stimuli as English words. This emphasis toward English stimuli and an English-based task was an attempt to minimise any effects of unidentified or colloquial interlingual homographs present in the stimuli. However, if any stimuli could be perceived as interlingual homographs, the stimulus control limits this to a very small and likely inconsequential proportion of the trials overall.

Although the behavioural tasks across both studies can be described as answering a yes/no question (e.g., "Is the stimulus a real word?" and "Do these words rhyme?" in the RRT), neither task was described this way in order to promote positive responses (e.g., positively choosing the answer as "they are different, non-rhyming words" as opposed to "no, they do not rhyme" in the RRT). It is realised, however, that this is tantamount to yes/no and also that researchers can never be sure of how a participant actually thinks and considers such tasks. Further details about the instructions and other procedural information specific to each task can be found in the following study-specific chapters (5.9 and Chapter 7:).

With regard to the response method, 56 published LDT studies (spanning from 1995 to 2014; see Appendix H for references) were systematically reviewed in order to determine the best method in terms of which hand or finger and device to be used. However, there was no clear consensus and often no reasons provided for why one method was used instead of another. Therefore, participants were asked to provide responses with their dominant right hand and the button used for responses per task was counterbalanced within each task and group. Blinking and body movements (other than to press a response button) during the tasks were strongly discouraged, particularly when stimuli were displayed, though it was made clear that there would be ample breaks within and between each task. To allow the breaks within each task (approximately every 3-4 minutes), each task was split into trial blocks that were identical in terms of the number of trials, conditions, instructions, and timings used in each (see study-specific chapters for more details).

Prior to each experimental task, all participants completed a block of practice trials, which included four exemplar trials per condition that were not used in the experimental task (see study-specific chapters for details). If performance was deemed poor in the practice trials or participants were still unsure about what the task entailed after completing the practice, participants were asked to repeat the practice until both researcher and participant were satisfied. However, this was only necessary once across all 64 participants and the second practice made things clearer.

Within each study, all participants saw the same sets of stimuli per task, but each in a different, pseudorandomized order, following recommended presentation guidelines (e.g., Hauk, Davis, et al., 2006). Stimuli across all tasks were presented on screen for 500ms, in line with other visual word recognition research involving bilingual participants (e.g., Guo et al., 2009; Wu & Thierry, 2010). This relatively short presentation duration also promotes the initial automatic response from participants and one that does not involve, for example, reading the stimulus multiple times.

Due to the importance of timing in ERP experiments as well as to avoid offset responses and overlap (Luck, 2005), there are several time-based recommendations that were followed in the design and procedure of these studies. Namely, all tasks employed a variable stimulus onset asynchrony (SOA) by pseudorandomly jittering the interstimulus and intertrial intervals (ISI and ITI, respectively). Details of timing and other procedural details specific to each task can be found in the study-specific chapters. Sufficient time and the pseudorandomised jitter between stimuli are also necessary to help minimise overlap and offset artefacts in averaged ERP waveforms (Luck, 2005). Taken together with the careful control of the stimuli (see §5.4.2), these details of stimulus presentation allow the experiment to focus more precisely on effects of orthography as opposed to more general visual processing and the potential confounds of repeating visually similar stimuli.

When all tasks had been completed, participants were asked to complete the post-study phonology test (outlined in §5.1.2). Finally, participants were debriefed and given the opportunity to ask any further questions.

5.6.1 EEG acquisition

Participants were fitted with the most suitably-sized waveguard™ cap, taking care to position the vertex (electrode Cz) halfway between nasion and inion and halfway between the left and right preauricular points. Full-band DC EEG data was recorded from two amplifiers in a cascaded configuration to obtain the high-density 128-channel EEG. Overall, it was considered that 512Hz was sufficient for the purpose of this research and to analyse the EEG as ERPs. Using a sampling rate of 512Hz, allowing 512 samples per second, is not unusual of such ERP research, as it allows a good balance between stability during recording and the high temporal resolution for which EEG was used. Future research should consider increasing the recording sampling rate to at least 1000Hz to better take advantage of the millisecond precision that EEG is capable of, provided that an analysis approach is used that embraces it.

Impedances were typically kept below 10kΩ and always lower than the 25kΩ recommended in the manufacturer guidelines. Due to the dual-amplifier cascaded configuration, either one channel (namely, channel 95, electrode Iz) or an additional, external electrode (depending on which provided better signal quality) was used as the ground for the second amplifier while the first amplifier used the standard ground from the cap (electrode GND). The left mastoid (electrode M1) served as the online reference for the whole system, shared between both amplifiers, resulting in a total of 127 channels of recorded EEG.

5.6.2 EEG/ERP processing

5.6.2.1 Pre-processing

Using ASA 4.9.3, the EEG was first visually inspected for any potentially problematic regions of data to ensure the recordings were viable for pre-processing and analysis. EEG was re-referenced offline using a whole-head average reference, which was used due to being more reliable and less biased overall in terms of electrode/region bias and the zero potential line, particularly when using a high-density montage (Dien, 1998), and test-retest reliability analysis has shown it to be beneficial (Dien, 2017). In terms of processing the EEG data in preparation for analysis, these following steps were performed in the same order and with the same settings (where appropriate and possible) for all participant data.

Recordings were first treated with a high-pass filter (half-power, 0.01Hz cutoff, 24dB/octave slope) to remove the DC offset and help improve the signal-to-noise ratio (SNR; Acunzo et al., 2012). Based on findings that high-pass filtering with cutoffs of 0.3Hz and above can increase artefacts and reduce statistical power (Tanner et al., 2015), while also distorting ERP onset times that are pertinent to the current research (Acunzo et al., 2012), using a 0.01Hz cutoff was recommended (Acunzo et al., 2012). Furthermore, the software guidelines state that using a high-pass cutoff >0.02Hz can result in discontinuities in the data (www.ant-neuro.com). As a general rule, a 12 or 24 dB/octave slope is advised for both high-pass and low-pass filtering (Luck, 2014), though it has also been suggested that the slightly steeper rolloff slope is recommended for high-pass filters with such low frequency cutoffs (Widmann et al., 2015), hence the choice of 24db/octave. Both high-pass and the subsequently applied low-pass filters used in processing the data were zero-phase, FFT (Fast Fourier Transform) Butterworth filters, which work bidirectionally to avoid phase shift (Tanner et al., 2015).

Low-pass filtering (half-power, 30Hz cutoff) with a 12dB/octave slope, increasing only as high as 36db/octave where necessary (e.g., excessive 50Hz line noise). Low-pass filtering was used for this purpose because it is almost always preferable and less destructive to the data than using a notch filter (Luck, 2014). It has been suggested that low-pass filtering should be reserved until after ERP measurements are taken and used only to make the waveforms look clearer for publication (Luck, 2014). However, the artefact correction procedure required cleaner data with less high frequency noise to work more accurately and without creating artefacts of its own, so this trade-off of best practices was accepted and low-pass filtering was applied.

Along with the previous high-pass filtering, this low-pass filtering created the desired asymmetrical band pass of 0.01-30Hz, attenuating potentially problematic higher frequency noise. The continuous data were then checked again for any anomalies and any channels with remaining excessive noise were either disabled or replaced using interpolation with neighbouring channels where appropriate. Interpolation was only performed when the rogue channel was in relative isolation, having good neighbouring channels to rely on, and not to attempt replacing more than one channel in the same cluster. The EEG was then segmented according to the experimental conditions into 700ms epochs, including a 100ms pre-stimulus baseline.

Prototypical ocular artefacts (e.g., eyeblinks) were then marked manually to allow a first pass of artefact correction to be performed based on values extrapolated from visual inspection of the data (typically +/-100µv following normal EEG signal range). The artefact correction procedure in ASA uses principal component analysis to separate brain signal from artefacts based on topography (Ille et al., 2002). Following the best efforts to correct ocular artefacts, the EEG data was visually inspected again for uncorrected and other artefacts (including significant movement or channel artefacts), which were marked manually and subjected to subsequent passes of artefact correction, as necessary. Automatic artefact detection (also typically based on amplitudes of +/-100µv) was also run to catch any other spikes that manual inspection missed. Any trials residing within segments with poor signal-to-noise ratio were disabled and not used in further analyses to prevent adverse effects on the averaged waveforms. All trials associated with correct behavioural responses were averaged together per condition, excluding any with overlapping and uncorrected artefacts (Landi & Perfetti, 2007). Although high-pass filtering can result in sufficiently corrected baselines (Tanner et al., 2016), automatic baseline correction was applied to ensure the waveforms for each condition begin as close to the zero line as their signal-to-noise ratios permit (Tanner et al., 2016). The averages from each participant were then averaged together to provide grand averages of each condition. Weighted averaging and grand averaging were used to compensate for the inevitable discrepancy of admissible trial counts between groups and conditions.

For ERP measurement, the intention was to use a priori time windows specific to each component of interest, but it is essential that these time windows are appropriate for the data of each participant and that the components of interest reside within these windows. Therefore, the proposed time windows were checked against the averaged waveforms of all individual participants. Following these checks, ERP measurements were taken according to the time windows for P1 and N170 at the electrodes from the regions of interest (occipital, occipitotemporal, and frontal-central areas in left and right hemispheres).

5.6.2.2 Clustering

Clustering involves averaging across multiple electrode sites to improve signal-noise ratio, obtain clearer and cleaner results, and increase statistical power (Dien, 2017). However, there is no clear consensus for which electrodes should be used when clustering at any given area. Due to the differences of both naming convention and specific location of electrodes between EEG systems, the sites suggested by the literature will be a guide and surrounding sites will also be considered for analysis. Therefore, in order to better define which electrodes should be grouped together to form clusters to represent the key areas of interest, a combination of visual inspection, physical proximity and location of electrodes, descriptive statistics, and reliability analysis were taken into account for both P1 and N170 timeframes, using amplitudes as a gauge. The principal channel for each cluster, however, was selected a priori (Dien, 2017), based on being representative of distinct component activity (e.g., O1 for left occipital, PO8 for right occipitotemporal measures, respectively).

Cluster analysis used area amplitude data per channel averaged across responses to the real word (RW) stimuli in Study 1 and the real word primes in Study 2, as the real word stimulus type was part of both studies and intrinsic to the aims of the research. Additionally, responses to the Study 2 stimuli were recorded semi-passively (requiring attention, but not a behavioural responses), so less influenced by experimental factors. Performing such analyses on each variable and collating the outputs in any meaningful way to find suitable and reliable patterns for clustering would be impractical and prone to many statistical and experimentwise errors, but the collapsing of related variables (i.e., all measuring responses to real words) allows a simpler, more direct, and more powerful approach to give a relatively stable and more objective basis for clustering.

Measures for electrodes in left and right occipital, occipitotemporal, and frontal-central areas were assessed for their suitability to create clusters representing each region. For each area, a combination of 95% confidence intervals, correlations, and t-tests was employed (see Appendix I for full statistical output), using only electrodes eligible for the area e.g., right frontal-central electrodes were not analysed with left occipital electrodes. Sites with similar ERP waveform shapes in the same area as one another were considered for clustering with statistical analysis used to support these conclusions. If an electrode was supported by analysis, but not in direct proximity, however, it would not be accepted as part of a cluster. Clusters must include only neighbouring electrodes within a reasonable distance from the maximal electrode.

Visual inspection of grand average ERP waveforms was also considered to check for any anomalies or other potential activity of interest. However, care should be taken not to make deep inferences from this alone as investigations into ERP/neuroscience methodologies suggest that basing comparisons and measurements on the data too directly can lead to Type I/II errors that statistics cannot account for (Luck, 2014). Instead, the statistical approaches provided guidance and a more objective, statistical basis for supporting what visual inspection suggested to be the optimal electrode clusters. However, visual inspection and, more broadly, examining the waveforms too deeply before making decisions has been strongly advised against by some ERP researchers (e.g., Luck & Gaspelin, 2016). Due to a lack of conformity and therefore possibility for a priori clusters, some cluster analysis was deemed appropriate and more objective than simple visual inspection.

Descriptive statistics (including means, standard deviations, correlations, and 95% confidence intervals) were used to highlight the site of maximal amplitude for each component of interest within each area of interest. Similar means and SDs, positive correlations, and largely overlapping 95% CIs between electrodes per area indicated that they were similar enough to be clustered. The electrode combinations submitted for statistical cluster analysis are presented in Table 2 and marked in Figure 8.

Table 2: Submitted and accepted channels/electrodes of the ERP cluster analysis

Region	Hemisphere	Principal	Neighbours	Rejected	Components
Occipital	Left	O1	POO9h, OI1h	POO3h	**P1-O(L)**
	Right	O2	OI2h, POO4h, POO10h		**P1-O(R)**
Occipito-temporal	Left	PO7	PO9, PPO9h, PO5, PPO5h, P5		**P1-OT(L)**, **N170-OT(L)**
	Right	PO8	PO10, PPO10h, PO6, PPO6h, P6		**P1-OT(R)**, **N170-OT(R)**
Frontal-central	Left	FC3	FC5, FFC3h, FCC5h FCC3h	FFC5h	**N100-FC(L)**,
	Right	FC4	FC6, FFC4h, FFC6h, FCC4h, FCC6h		**N100-FC(R)**

Component naming convention adopted from Dien (2009).

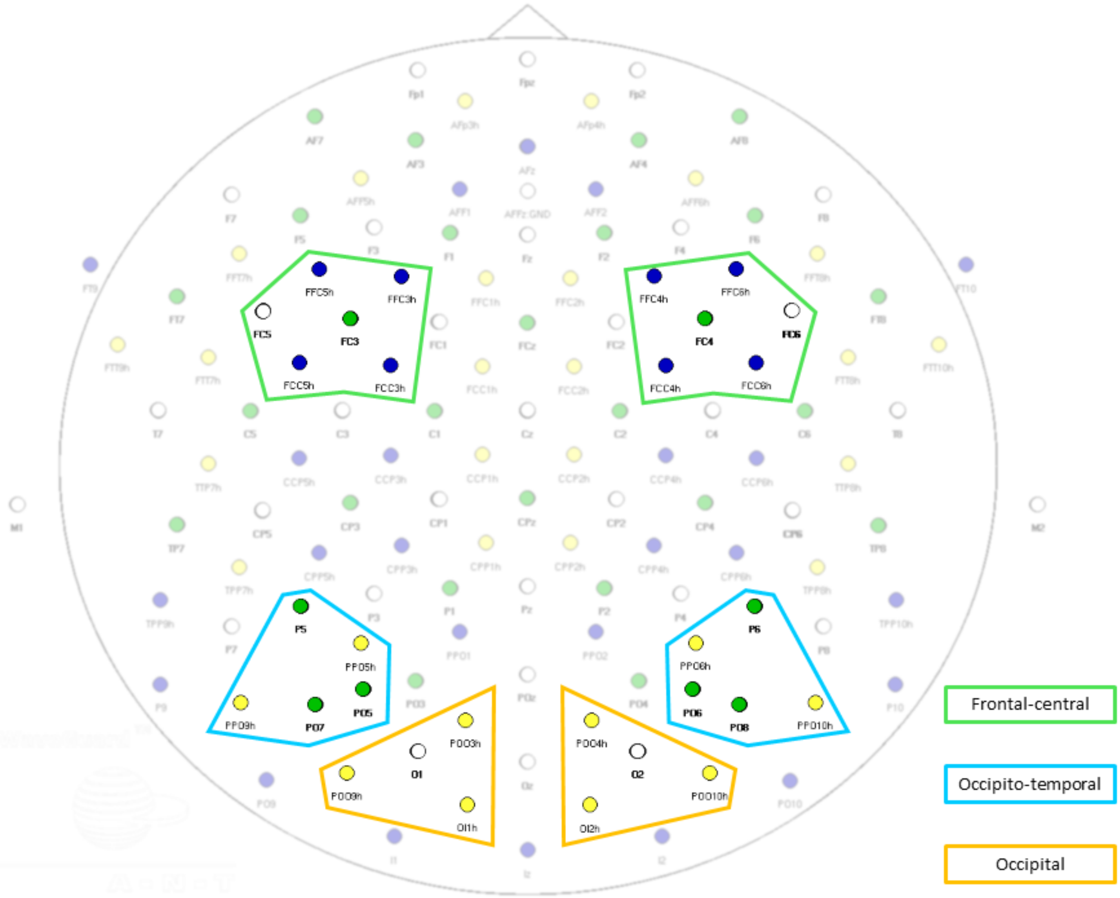

Figure 8: Electrode clusters on 2D electrode layout

Table 2 also shows how cluster analysis defined the clusters, including which channels were accepted and rejected for inclusion in averaging per cluster and which components they created for two distinct timeframes (~100ms and ~170ms) per experimental condition for each of the six discrete regions, using area amplitude, mean amplitude, and 50% area latency measurements. These definitions were also used to create the prime control in Study 2.

5.7 Analysis

Behavioural data from E-Prime 2 Pro (PST, 2017) and, following necessary pre-processing (see §5.6.2), ERP data from asa pro (v4.9.3; ANT Neuro, 2015) were prepared in Microsoft Excel (Microsoft Corporation, 2018) for statistical analysis using SPSS (v26; IBM, 2020), JASP (v0.14.1; JASP Team, 2020), and jamovi (v1.6.23; The jamovi project, 2021). All behavioural and ERP measurements are based on data from correct responses only. ESL groups did not significantly differ in English language test scores (i.e., IELTS or equivalent) or Language History Questionnaire scores, but there was a significant difference in phonology test scores with Spanish-English being more accurate than Chinese-English (see §5.2.1.3).

5.7.1 Statistical tests

Multivariate and/or multi-factorial statistics are sometimes favoured for such rich data sets as collected in ERP studies (Dien, 2017). Considering the independent and dependent variables available for analysis in the current research, a factorial multivariate approach could be given a strong argument. However, it has been suggested that using simpler, more direct and powerful univariate statistical analyses is a better approach, based on inherent issues with multivariate statistics (including potentially over-complicated interpretations) and considerations concerned with the nature of EEG/ERP data and sample size (Luck, 2014). Large multivariate factorial analyses may provide additional information and support for interactions between factors, but specific answers might not necessarily be provided and could even be missed (e.g., non-significant interaction stemming from the complexity of the design).

While the widely-used factorial ANOVA approach is also not perfect e.g., when applied to response time data (Whelan, 2008), its frequent usage in VWR and ERP studies provides a familiar framework to present findings.

5.7.2 Statistical analysis

All analyses were planned, method-based, and literature-driven: the main comparisons for each study were determined a priori, guided by the ERP literature, following the respective designs, and ultimately defined by the aims of the research. Differences between experimental conditions and groups, respectively, will form the foundation of all discussion and interpretation, along with the use of any further analyses (e.g., between-conditions correlations, behavioural-ERP correlations, P1-N170 correlations, hemisphere contrasts).

Behavioural measures of accuracy (% correct) and response times (in milliseconds) as dependent variables will be analyzed with separate univariate ANOVAs. The Condition x Group interactions are the main focus in order to examine pairwise contrasts of Condition within and between groups. ERP measures of (positive/negative) area amplitude and fractional 50% (positive/negative) area latency (positive/negative depending on component and deflection) as dependent variables will also be analyzed with separate univariate ANOVAs per timeframe/cluster/measure (e.g., P1-O, N170-OT). While ERP dependent variables can be considered per individual timeframe/cluster/hemisphere/measure (e.g., P1-O(R) latency, N170-OT(L) amplitude) and so running univariate statistical analysis for each of these dependent variables was an option, current research questions concern lateralization of ERP effects, so hemisphere will also be entered as a factor of the ANOVA. The Condition x Hemisphere x Group interactions from the ANOVA will be the main focus in order to examine pairwise contrasts of Condition within and between groups for each hemisphere. Main effects of Condition and Hemisphere and interactions not involving the Group factor, however, are arguably meaningless in this context because they ignore the critical Group factor. Averaging across groups would create functionally redundant measures for comparison and introduce confounds due to the experimental groups varying so fundamentally on a psycholinguistic level. Interpreting such effects that effectively ignore the Group factor, would, therefore, not be useful or advantageous, so will not be considered. This helps with the familywise error rate and issues of implicit multiple comparisons. More directly concerned with controlling error rates and for multiple comparisons, post-hoc group comparisons will use the Tukey correction, as it is more suited to independent-samples tests, and comparisons within groups will use the Holm correction, which is more appropriate for repeated-samples tests.

It is important to also note that statistical tests were only conducted on viable measures e.g., negative ERP measures were not tested for the posterior P1 components and factors were not entered into ANOVA analyses if they were deemed to obfuscate or potentially confound results. Following Dien (2009) in treating activity at different timeframes and different scalp locations/clusters as essentially (or, at least, potentially) different ERP components e.g., P1-OT(L) akin to the P1-PO7, cluster and hemisphere were not entered as factors in the ANOVAs and each cluster was analyzed separately with a Condition (2) x Group (3) ANOVA mixed factorial ANOVA only for relevant measures e.g., positive area amplitude and 50% positive area latency for occipital and occipitotemporal P1 measures, while negative amplitude and latency measures were not tested for the positive posterior P1 deflections. This helps to reduce implicit multiple comparisons that would otherwise be present in larger ANOVAs (especially with redundant factors/levels) and increase statistical power. Implicit multiple comparisons were reduced further still by the a priori selection of main effects and interactions of interest in that even though main effects and interactions for all entered factors would be computed for each ANOVA, only a sub-set were used as the others are practically of less or no importance to the research e.g., a main effect of Condition might suggest a difference in the respective ERP measure (e.g., P1-LeftOP mean amplitude), but this difference is based on averaging across the groups, essentially ignoring a vital element of the research, confounding the finding and its meaning.

5.7.3 Data preparation

The Z-score approach was used to identify outliers, which were defined as absolute values ≥1.96 SD from the mean (~2.5% from each end of the distribution) for accuracy and response times (which were additionally trimmed if <200ms or >2000ms). The same Z-score approach was used for ERP amplitude and latency data with a difference being that the EEG data has already been cleaned with potential outliers removed through ERP pre-processing procedures. Values identified as outliers were replaced with the 5% trimmed mean of the data without the outliers (so the replacement mean is not biased by the outliers), as used by robust analysis of variance techniques (Dien, 2017). Replacing outliers with a meaningful and potentially legitimate value that is a robust descriptor of central tendency helps to maintain statistical power without significantly impacting the overall mean or other statistics (Cousineau & Chartier, 2015; Field, 2013,

p. 199; Tabachnick & Fidell, 2007). In other words, the use of 5% trimmed means helps to protect the analysis from persisting outliers (Dien, 2010). This is preferable to replacement with artificial upper/lower values (as in Winsorizing), which can result in outliers not being fixed, instead creating different extreme values depending on the unit. For example, adding 1 unit to the next highest non-outlier for an ERP amplitude measure means that the legitimate highest value is then an entire microvolt lower than the new highest value, which could still be an outlier. Lastly, Shapiro-Wilk, along with Z-skewness and Z-kurtosis, will be used to check each variable for normality of distribution. Unless stated otherwise, parametric assumptions were met without significant violation, allowing data to be accepted as parametric.

5.8 Ethics

All aspects of this research were approved by the Sheffield Hallam University institutional ethics board prior to being carried out (please see Appendix C for the letter of confirmation). The research was supported by Sheffield Hallam University with funding being provided through a bursary to the lead researcher. The research was conducted without any commercial or financial relationships and there was no conflict of interest.

5.9 Summary of rationale

Considering the visual requirements of reading and the universal ontogenetic primacy of phonology in natural languages, orthographic and phonological processes are the underpinnings of VWR and fundamental to understanding how written language comprehension is achieved by L1 and L2 readers. The current research, therefore, is multithreaded, focused on orthographic and phonological processing both within and between the English, Spanish-English, and Chinese-English groups. Overarching the research is the broad question of whether orthographic and/or phonological effects are observed on behavioural and/or ERP measures across these groups. Aside from investigating orthographic and phonological processing of English within the groups, the main premise is whether the timing and nature of processing in one group are also found in another, including how they compare in terms of timing and/or characteristics (behavioural or ERP).

Across groups, the focus is on how L2 processing in L2 readers corresponds to L1 profiles and whether any orthographic or phonological effects can be at all attributed to language profile. This also involves the extent that theories of monolingual (e.g., BIAM, DRC) and bilingual (e.g., BIA+, Multilink) VWR apply to L2 reading. For comparisons between groups, the key question concerns the similarities and differences of behavioural and ERP measures in the context of orthographic and phonological VWR processing. This is especially related to the extent of difference between native and target languages i.e., how different Spanish is from English relative to how different Chinese is from English. This also includes how processing in L2 reading of English deviates from native processing and allows using the native English group for comparisons between L1 and L2 reading in order to investigate how L2 processing of English might work.

Considering the depth of difference between language profiles of the Spanish-English bilingual (non-native, alphabetic L1) and Chinese-English bilingual (non-native, non-alphabetic L1) groups, including the possible benefits and/or difficulties for reading English, as well as directly relevant pre-200ms ERP comparisons of these specific groups being sparse in the literature, it is difficult to say how processing of the same non-native, second, alphabetic language (English) will differ between groups. Nevertheless, behavioural measures will show the potential for native-like performance through comparable patterns to native readers, even though both accuracy and response times will reflect poorer performance in ESL readers e.g., slower response times and more errors in the ESL groups. In particular, Chinese-English will likely show poorer performance for phonologically-oriented stimuli/tasks e.g., pseudohomophones, orthographically incongruent rhymes, while Spanish-English will have less issue with the alphabetic stimuli. While overall processing, responses, and behaviour during reading can appear similar between native and proficient ESL readers, lower-level cognition and electrophysiological responses are potentially very different due to the foundations of L1 that are at least partially if not mostly relied on for L2 processing.

The focus across groups also includes how and when late bilinguals with different types of L1, including alphabetic or non-alphabetic (e.g., logographic), process L2 orthography and phonology and how processing might differ in bilinguals compared with monolingual native readers of that same language (in this case, English). There is no question that L1, in terms of its linguistic components, psycholinguistic processes, and any language-specific strategies, can influence L2 processing in bilinguals or that bilinguals can achieve “native-like” proficiency in their L2 (Dijkstra et al., 2019; van Heuven & Dijkstra, 2010). More typically these contrasts are between L1 and L2 within a population e.g., Dutch (L1) and English (L2) in Dutch-English speakers, which can be very informative, but this area has not been sufficiently investigated in terms of L1 language profiles (e.g., L1 language type; outlined in §1.1) and their relevance to L2 processing. In other words, an important aspect of the current research is its perspective of comparing English as the L2 across bilingual groups (as opposed to L1 and L2 within a group), examining contrasts in L2 behaviour and brain activity between groups instead, which is necessary to directly contrast L2 VWR processes in readers with different L1 profiles. Furthermore, this is a significant extension to the typical perspective of the bilingualism ERP literature, investigating how similar/different L2 processing is between bilinguals of different language profiles, such as ESL populations with native languages from different writing systems e.g., alphabetic vs logographic.

Importantly, the relative novelty of this approach also extends to contributing evidence of orthographic and phonological processing in terms of behavioural and early ERP measures to the bilingualism literature. These aspects in the pre-200ms timeframe are not sufficiently addressed in the literature, as work in this area is largely concerned with word-naming latencies, interlingual processing conflicts, or sentence-level processing of semantic and syntactic variables. While these approaches are vital to understanding the bilingual brain, they are typically associated with later processing and brain activation when considering the timeframe of single word recognition (<300ms per word) and, therefore, the timeframe of early brain activity (i.e., sub-300ms) from a neuroscience perspective (e.g., ERP).

Based on a wealth of VWR research using electrophysiological measures, the first ~200ms after seeing a word-like stimulus is a vital foundation for the neural processing necessary for language comprehension and associated behavioural responses. Mainly concerned with orthographic and phonological (as opposed to e.g., semantic and syntactic), the ~200ms timeframe in terms of electrophysiological activity is occupied by occipital and occipitotemporal P1 components at ~100ms and occipitotemporal N170 components at ~170ms (Dien, 2009), which form the main ERP focus of the current research. Serial accounts of VWR might broadly attach initial visual/orthographic evaluation to the posterior activity at ~100ms (Dien, 2009) with deeper orthographic and orthographic-phonological processing occurring at ~170ms (Grainger & Holcomb, 2009). Other accounts have, however, cast doubt on whether VWR processing is so strictly serial, whether parallel processing plays a larger part, or simply that VWR processing is more dynamic than many current theories allow. For instance, such early phonological activity as observed at ~100ms in left frontal-central regions (Ashby et al., 2009; Klein et al., 2015; Wheat et al., 2010), alongside the continuing debate around the aforementioned activity up to ~200ms, question the very nature of VWR processing, from the point of seeing a stimulus to understanding it as a linguistic form with specific sound and meaning.

The focus on the initial ~200ms of brain activity and observing similarities and differences in processing across groups necessarily funnels down into the more specific enquiries of the research: whether evidence of orthographic processing is observed in the first ~200ms , especially on the P1-O, P1-OT, and N170-OT; and whether evidence of phonological processing is observed in the first ~200ms of seeing a word (without masked priming), especially on the N100-FC and N170-OT. The first inquiry of orthographic and phonological stimulus-level processing, therefore, involves the initial (~100ms) occipital activation represented by the P1 component, including whether its nature is more than a visual response and is sensitive to linguistic features. The second involves the stimulus-based and task-related natures of the occipitotemporal N170 in terms of sensitivity to orthography and/or a role in phonological processing. Lastly, early (~100ms) frontal-central activity will be investigated with respect to it representing early phonological processing and its involvement in language processing, such as any indications of parallel processing. More study-specific rationale for the focus on these components and timeframes will be discussed further in respective sections.

Considering the broad support for a feedforward relationship between orthographic and phonological processing (e.g., Dien, 2009; Holcomb & Grainger, 2006; Hauk et al., 2012), taken together with the compelling evidence of ~100ms phonological activation (e.g., Wheat et al., 2010; Ashby, 2010) and the possibility of initial visual processing followed in parallel by orthographic and phonological processing (e.g., Cornelissen et al., 2010; Pammer et al., 2004), orthographic effects are expected to occur earlier than phonological effects in some scenarios e.g., grapheme-phoneme conversion, but there will be an equivalency in others e.g., real words in the English group, suggesting context-dependent serial/parallel processing. However, while effects within the early <200ms ERP timeframe (on P1 and N170 components) have been observed in previous studies and might be expected in the English group, the lack of prior focus on this early timeframe in the bilingualism literature makes it difficult to hypothesize for the ESL groups. However, it is reasonable to accept that the association of orthographic expertise of the N170 could allow effects on it to be observed in the Chinese-English due to the strongly orthographic focus of their L1. Meanwhile, both P1 and N170 effects could be observed in the Spanish-English group due the relevance of the group's alphabetic expertise to the L2 target. Nonetheless, <200ms could also be too early to observe direct electrophysiological effects from VWR processes in late-bilingual non-proficient ESL readers, though the observed responses will still be of interest due again to the lack of prior focus on this early timeframe in ESL readers and to help inform how bilingual readers perceive and process their L2 stimuli, especially with an eye to their different language profiles (alphabetic vs non-alphabetic). However, the distinct contrast between groups does provide a window for behavioural and early ERP responses to be investigated and provide some insight into L2 processing. Furthermore, finding commonalities within and between groups of L1 and L2 readers of a particular language (e.g., English) can provide insight into which psycholinguistic elements are focused on most and therefore which elements of the language are most important to its processing.

Chapter 6: Orthographic and phonological lexicalities in readers of English with different L1 profiles (Study 1)

6.1 Overview

This chapter will document the background, rationale, method, and results of Study 1, ending with a discussion of findings. This first study examines behavioural and early pre-200ms ERP responses to orthography- and phonology-oriented lexicality decisions with contrasts between native (English monolingual), non-native alphabetic L1 (Spanish-English bilingual), and non-native non-alphabetic L1 (Chinese-English bilingual) groups of English readers. In line with the aims of the thesis, these distinct ESL populations are the focus to investigate how L1 profiles (as discussed in §1.1) contribute to reading English as a second language, while also contrasting with native-level reading of English. Alongside comparisons of behavioural performance (via accuracy and response times), the main aims of the study are to provide evidence concerning whether the occipital/occipitotemporal P1 reflects lexicality, the possibility of early ~100ms frontal-central phonological activity, and the orthographic and/or phonological nature of the occipitotemporal N170 during VWR, each with emphasis on how observations differ across groups. An orthographic lexical decision task (oLDT) using real words and pseudohomophones and a phonological lexical decision task (pLDT) using pseudohomophones and pseudowords will separately provide the means to observe the necessary orthographic and phonological processing, respectively, within and between groups.

6.2 Background

Lexicality, in the context of VWR, concerns the linguistic status of a stimulus and the recognition of it as language (Coch & Mitra, 2010). The classic lexicality effect is observed when real words are processed more efficiently (i.e., behaviourally faster and more accurately) compared with pseudowords (Balota & Spieler, 1999; Fiez et al., 1999; Hauk, Patterson, et al., 2006). Investigating such lexicality effects typically employs a form of lexical decision task (LDT), where participants indicate whether visually-presented stimuli (traditionally, real words and pseudowords) are words or not. While the LDT paradigm is well-established, the contrast between real words and pseudowords is relatively dense, involving orthographic, phonological, and semantic factors, and is much more complex than simply different stimulus types; direct comparison between them, therefore, leaves a substantial gap in interpretation. Real words have the full array of legal and legitimate[3] orthography, phonology, and semantics. Pseudowords such as *tuss* and *fode*, however, lack legitimate orthography, phonology, and semantic associations, making them stark contrasts to real words, having only legal orthography and phonology to look word-like and act as fake words. The aforementioned examples (*tuss*, *fode*), for instance, are both pronounceable and phonology can be extrapolated from their legal orthography, even though neither the overall orthography nor the phonology exist as such in English. Pseudohomophones

[3] Legal linguistic elements are those found in the language, whereas legitimate linguistic elements follow the rules of the language but are not necessarily found in the language. Please see §1.2 for more details.

(e.g., *kave*, *gote*), meanwhile, lack only legitimate orthography, retaining legal and legitimate phonology/semantics that allows them to also look word-like and indirectly imitate the phonology of real words. The contrasts between recognizing stimuli as being real (real words), sounding real (pseudohomophones), or not being real at all (pseudowords) reflect the automatic retrieval of phonological/semantic information (Twomey et al., 2011; Ziegler et al., 2001). The use of these three stimulus types (as opposed to just two e.g., real words and pseudowords), therefore, operationalizes the lexicality decision for examining both orthographic and phonological processing more discretely. Importantly, the ability to read pseudowords demonstrates a route to word-sound (phonology) identification using grapheme-phoneme conversion or, at least, an alternative indirect mechanism to decode visual representations into internal phonology (Coltheart et al., 2001; Ellis et al., 2004). Using pseudowords and pseudohomophones as stimuli essentially forces the indirect, grapheme-to-phoneme route. Therefore, research using pseudowords and pseudohomophones can be used to demonstrate this route by requiring readers to use it, while research using both real words and pseudowords can provide some indication of any differences between this indirect grapheme-to-phoneme route and a more direct lexical route.

Lexicality decisions during VWR fundamentally concern identifying whether a stimulus is associated with an item in lexical memory (Balota et al., 2004), essentially being orthographic (Grainger & Holcomb, 2009; Grainger & Jacobs, 1996). However, as the orthographic lexicon theoretically only contains entries for letter-strings with semantic associations (McNorgan et al., 2015), lexicality can be considered to be fundamentally semantic. This follows the notion of a direct route from orthography to semantics in line with the DRC model (Coltheart et al., 2001) and that lexicality decisions automatically activate semantic networks (McNorgan et al., 2015), which act in a top-down fashion to help process the orthographic input (Twomey et al., 2011). However, this is not to say that semantic processing is necessary for lexicality decisions, as VWR processing can proceed without the semantic system (Coltheart, 2005; Grainger & Jacobs, 1996). Nonetheless, semantics is integral to lexicality, even if not strictly necessary for the decision, with initial semantic activation suggested to occur simultaneously with lexical activation at ~200ms (Hauk et al., 2012). Therefore, the pre-200ms and arguably pre-semantic timeframe, which includes the P1 and N170 ERP components (introduced and discussed in §4.4.1.1.1 and §4.4.1.2, respectively), is vital for investigating how the orthographic and phonological processes that precede lexical/semantic activation contribute to lexicality effects (and VWR more generally).

6.2.1 ERP discrepancies in lexicality effects

Lexicality effects have been observed as early as ~100ms (e.g., Sereno et al., 1998) and as late as ~400ms (e.g., Lehtonen et al., 2012), being reported for a variety of ERP components that have associations with psycholinguistic processes, such as the P1 (Hauk, Patterson, et al., 2006; Segalowitz & Zheng, 2009), P150 (Proverbio et al., 2004), N170 (Hauk, Davis, et al., 2006; C. D. Martin et al., 2006), N200 (Coch & Mitra, 2010), and even N400 (Lehtonen et al., 2012). This demonstrates an especially broad timeframe for ERP activity associated with a particular psycholinguistic property (P1 at ~100ms to N400 at ~350-500ms), hence findings of lexicality effects on both relatively early components (e.g., P1, P150) and the later N400 have been taken to suggest that both low-level, sublexical processing and higher-level lexical processing can reflect orthographic fluency (Coch & Mitra, 2010). Amplitudes within ~200ms in particular (e.g., P1, N170) have shown varying differences between words and pseudowords compared to meaningless letter strings (e.g., *jwpk*), for example, as well as being larger for high-frequency than low-frequency words (Proverbio et al., 2004), which relates to both lexicality and the word superiority effect (Coch & Mitra, 2010; C. D. Martin et al., 2006). Combined with evidence of unfamiliar and irregular orthography eliciting larger P150 ERP amplitudes within ~200ms (cf. orthographic familiarity and regularity effects; Coch & Mitra, 2010), such lexicality and frequency effects and their interaction indicate brain activity in this timeframe playing a significant role in sub-lexical orthographic processing (Grainger & Holcomb, 2009). Essentially, different aspects of lexicality could be processed at different times, raising questions about lexicality effects and specifically which element of lexicality is the root (Coch & Mitra, 2010). Therefore, following the linguistic distinctions between real words, pseudohomophones, and pseudowords (as outlined in §1.2), and the discrepancies of ERP amplitude and latency between findings discussed here, lexicality is better considered in different forms, such as orthographic and phonological lexicalities, as investigated in the current study.

In the ~100ms timeframe, lexicality effects on ERP amplitudes are inconsistent and the way the P1 component in particular is sensitive to lexicality is unclear (Hauk, Davis, et al., 2006). For instance, some studies show larger amplitudes to pseudowords and strings of consonants than words (e.g., Sereno et al., 1998), while others report no difference between words and pseudowords (Hauk, Davis, et al., 2006), words and letter-strings (C. D. Martin et al., 2006), or words and symbols (Emmorey et al., 2017). Such discrepancies could be due to different task methodologies, though even in comparable LDTs, the amplitude of the P1 at occipital sites has been reported to be larger to real words (Segalowitz & Zheng, 2009), smaller to real words (Taroyan & Nicolson, 2009), and "virtually identical" to real words compared with pseudowords (Hauk, Davis, et al., 2006, p. 1389). Furthermore, in a study directly comparable to the current research due to using orthographic and phonological LDTs with native English participants, there were no reported differences in P1 amplitudes between real words and pseudohomophones or between pseudohomophones and pseudowords at occipital or other areas (Taroyan, 2015).

The variability between reported lexicality effects could reflect different types of lexicality and underlying processes, possibly stemming from effects of different, albeit related psycholinguistic properties being attributed to "lexicality". Considering the effects of both orthography and phonology at ~100ms (e.g., Wheat et al., 2010), it is not unreasonable to suggest that the P1 is not responding to lexicality per se, but to a visual or sublexical property, such as word shape or n-gram frequency. For instance, as well as to lexicality, ERP activity at ~100-150ms (e.g., P1, P150) has shown sensitivity to sublexical legal orthography, being of greater magnitude to both words and pseudowords than to non-word letter-strings (Proverbio et al., 2004).

Akin to the P1 timeframe, there is also controversy about the direction of effects on N170 amplitude in terms of various lexicality effects (Pattamadilok et al., 2015). While lexicality effects in the N170 timeframe are not uncommon (Coch & Mitra, 2010), the nature of such effects is still unclear. Within VWR research and in terms of reading English, the N170 difference in the classic lexicality effect between real words and pseudowords is inconsistent; some studies have reported a larger N170 to words (Coch & Meade, 2016; Mahé et al., 2012; Maurer, Brandeis, et al., 2005), while it has also been reported as being smaller to real words than pseudowords (Compton et al., 1991) and others report no significant difference (Bentin et al., 1999; Maurer, Brem, et al., 2005). Despite the discrepancies, there still seems to be a strong implication that the N170 is sensitive to linguistic forms that distinct orthographic and phonological types of lexicality, as examined in the current study, could help explain. For instance, the N170 has been found to be larger to real words and pseudowords than letter strings, non-words, and symbols (Emmorey et al., 2017), suggesting that it is sensitive to pronounceable linguistic stimuli or related to them looking linguistically legitimate (e.g., Coch & Mitra, 2010; C. D. Martin et al., 2006). However, others report minimal or no difference between different types of legally orthographic stimuli, such as real words, pseudowords, and consonant strings (Simon et al., 2004), while it is also smaller to consonant strings than pseudowords (McCandliss et al., 1997). The N170 is larger to legal orthographic stimuli than false fonts and typically larger to letter strings than strings of other non-linguistic symbols (Appelbaum et al., 2009; Bentin et al., 1999; Mahé et al., 2012; Maurer, Brandeis, et al., 2005), while also distinguishing between letter strings and other letter-like visual stimuli, such as symbol strings (Helenius et al., 1999; Maurer, Brem, et al., 2005), "forms" and alphanumeric symbols (Bentin et al., 1999), shapes and dots (Eulitz et al., 2000).

While its specificity to language is debated, the N170 appears to be sensitive to known and meaningful orthographic representations with an element of word superiority being present and possible connections to phonology (Maurer et al., 2008). Furthermore, lexicality-type effects where real words elicit larger N170 amplitudes than non-word linguistic stimuli have been found in a number of other languages and of various language types besides English and alphabetic, including French (Bentin et al., 1999), and German (Maurer, Brem, et al., 2005), as well as Korean and both alphabetic

Japanese Kana and non-alphabetic Japanese Kanji (Fu et al., 2012), syllabic Hiragana, logographic Kanji, and moraic Katakana scripts of Japanese (Maurer et al., 2008), providing more support for this potentially universal response to visually presented words. However, the classic lexicality comparison (real words vs pseudowords) may not be sufficient to observe the necessary detail, as more specific orthography- and phonology-oriented psycholinguistic factors seem to be involved, calling again for investigation into separate orthographic and phonological lexicalities.

6.2.2 Orthographic lexicality

As outlined in the previous section, the classic lexicality effect (real words vs pseudowords) can be considered a form of orthographic lexicality decision, as it can theoretically be achieved based on visual/orthographic familiarity (van der Mark et al., 2009). However, orthographic lexicality requires a more precise definition that constrains the broad definition of lexicality to primarily reflect an orthographic distinction. The lexical decision between real words and pseudohomophones requires a level of orthographic processing beyond visual evaluation that reflects orthographic lexicality, as only legitimate orthography separates the two stimulus types: both have access to meaning via the phonological lexicon. Real words and pseudohomophones both have legal orthography and legal and legitimate phonology, but pseudohomophones lack legitimate orthography. Therefore, these stimuli likely recruit different processing routes in that real words can be recognised directly using visual/orthographic processes, while pseudohomophones cannot, requiring an indirect and non-lexical pathway that activates the real-word phonology and allows lexical access (van der Mark et al., 2009). Orthographic legitimacy could also be managed with visual/orthographic processing in the sense of feature detectors of the interactive-activation model (McClelland & Rumelhart, 1981), which would also account for the distinct responses to pseudohomophones.

Many properties of words can also be applied to sublexical units (Grainger & Holcomb, 2009), so the familiarity, regularity, and frequency of graphemes, phonemes, and n-grams could contribute to the overall lexicality effect, leading to differences between conditions that could drive misinterpretations as lexicality effects. Familiarity, for instance, is not limited to whole words either, extending to sublexical units, as shown by effects of bigram frequency (Hauk et al., 2012). Lexical and sublexical frequency and familiarity are rooted in the overtly visual/orthographic nature of reading (Mechelli et al., 2005), presenting key differences between linguistic stimuli and supporting the importance of such visual/orthographic factors as the foundation of processing during reading (Hauk et al., 2012; Holcomb & Grainger, 2006; Proverbio et al., 2004). The legitimate orthography that differentiates pseudohomophones from the real words with which they share phonology entails differences of familiarity and frequency, among other orthography-related psycholinguistic factors, on lexical and sublexical levels (Ziegler et al., 2001). Although the letters themselves are familiar, groups of letters do not necessarily provide visual cues for lexical status (Twomey et al., 2011), but it is possible that visual – not even orthographic – features allow words to be recognised, at least as words (Grainger & Holcomb, 2009). Visual familiarity (in terms of word length and shape relative to font size and context, for example) is, therefore, a key factor in orthographic lexicality decisions and familiarity may largely stem from frequent exposure to a stimulus in that exposure and familiarity are correlated (Tanaka-Ishii &

Terada, 2011). Furthermore, effects of sublexical familiarity in terms of orthographic typicality (referring to spelling regularity based on n-gram frequency) have been found to interact with lexicality at ~158ms at temporoparietal sites following an ~100ms effect of typicality at occipitoparietal sites (Hauk, Patterson, et al., 2006). Likely generators of these components may not be specialized to language per se, but in visual processing that include such processes as letter identification and word shape recognition (e.g., Price & Devlin, 2003; Vogel et al., 2014). Visual familiarity could explain its apparent specialization to words: it is not specialization to words, but specialization to forms that have been seen and used hundreds or thousands of times (Cornelissen et al., 2009). However, the observation of visual familiarity effects when lexical frequency is controlled shows a clear distinction between these properties (Twomey et al., 2013), posing questions of whether lexicality can influence the P1 and N170 when orthographic familiarity, lexical frequency, and visual familiarity are controlled (Hauk, Davis, et al., 2006), as in the current study.

6.2.3 Routes to lexicality decisions

It is important to remember that the stimuli involved in the orthographic lexical decision task (oLDT; real words vs pseudohomophones) used in the current study (described in §6.3.3) can also reflect different theoretical routes to word recognition with a direct lexical pathway being used for real words and an indirect route involving a form of sounding out (e.g., GPC) for pseudohomophones. Due to this difference in processing and in legitimate orthography between real words and pseudohomophones, only orthographic evaluation is necessary for accurate responses in the oLDT and not further phonological or semantic processing (even if occurring automatically), which would not help with the task due to both stimulus types having legal and legitimate phonology and semantics that propose all oLDT stimuli to be real words. Slower and less accurate performance to pseudohomophones compared with real words (e.g., Braun et al., 2015; Briesemeister et al., 2009; Twomey et al., 2011) would, therefore, illustrate the less direct and more cognitively demanding route required to identify them as stimuli that do not precisely represent real English words (at least in the context of the oLDT). However, this implies that different processing routes would be used between real words and pseudohomophones, even though only initial orthographic evaluation to indicate orthographic familiarity was required (Twomey et al., 2011; van der Mark et al., 2009).

Considering the initial stage of psycholinguistic processing as orthographic evaluation, stimulus type (real words, pseudohomophones, pseudowords) would typically define which processing route and what further processing is required (i.e., direct/indirect). Though simplistic and assuming matched stimuli on various psycholinguistic variables including relatively high frequency n-grams and only familiar irregular spellings (if any), this describes a serial process from orthographic evaluation to direct/lexical recognition (real words) or indirect GPC processing (pseudohomophones, pseudowords) based on the outcome of the evaluation, resulting in lexicality/pseudohomophone effects (Twomey et al., 2011). Alternatively, the same orthographic evaluation and output could be made alongside automatic indirect processing, where the orthographic evaluation and the orthographic-phonological conversion work in parallel, resulting in agreement or disagreement and consequently the pseudohomophone effect (Briesemeister et al., 2009). Such parallel processing could be considered unnecessary if one or the other route can successfully process the stimulus, though evidence of how orthographic incongruence can interfere with phonological decisions (and vice versa) suggest that automatic routes working in parallel describes early orthographic-phonological processing (e.g., Twomey et al., 2011). Therefore, it is not unreasonable to conclude that processing continues based on the initial orthographic evaluation when seeing a linguistic stimulus – real words directly, pseudohomophones (and pseudowords) indirectly – and it is this difference in the extent of processing that accounts for differences in performance and possibly early ERP activity (Braun et al., 2009, 2015; Briesemeister et al., 2009). Furthermore, the possibly automatic processing beyond orthographic evaluation makes sense in terms of checking negative responses, which would mainly be associated with pseudohomophones and, therefore, accounts for additional time and effort spent on pseudohomophones in the oLDT. Essentially, if orthographic evaluation was the end of the line due to it being all that is required for the oLDT, there would not be any difference in processing between real words and pseudohomophones that are balanced on all controllable orthographic (and phonological) psycholinguistic properties (length, n-gram frequency etc.). However, this is not what studies report in contrasts between real words and pseudohomophones (e.g., Braun et al., 2009, 2015; Briesemeister et al., 2009), though the specific root of the processing difference or conflict is not clear enough, especially considering early ERPs associated with orthographic processing, such as P1 and N170. As mentioned earlier, there are various differing and even opposing contentions concerning these components and the ~200ms post-stimulus timeframe, all of which feed into either orthographic lexicality, phonological lexicality, or both, and require deeper investigation using time-sensitive neuroscience methods as in the current research.

6.2.4 Phonological lexicality

Processing pseudowords and pseudohomophones requires differential activation and processing to “real” language (Coltheart et al., 2001; McNorgan et al., 2015), making them markedly more complex than reading known real words, especially high frequency ones, being more like reading low/no-frequency words. Due to their lack of legitimate orthography, neither are true lexical items and so both theoretically use the same indirect orthography-phonology pathway (Coltheart et al., 2001). This differential processing is demonstrated through poorer accuracy, longer response times, and modulated neural activity to such stimuli that constitute orthographic (real words vs pseudohomophones) and phonological (pseudohomophones vs pseudowords) lexicality effects (Twomey et al., 2011). Akin to the pseudohomophone effect, phonological lexicality is observed through differences in responses between pseudohomophones and pseudowords (Braun et al., 2009). However, there is some overlap within the comparison, as neither stimulus type is orthographically legitimate nor allows direct access to semantics; while lexicality can be considered fundamentally semantic (McNorgan et al., 2015), the psycholinguistic features of pseudowords and pseudohomophones do not obligate semantic processing. Deciding that a pseudoword is not a real word is arguably an orthographic decision based on visual/orthographic familiarity and is not semantic at all, requiring only orthographic evaluation. Likewise, deciding that a pseudohomophone is not a real word is also arguably a visual/orthographic decision, though deciding whether a pseudohomophone has the legitimate phonology of a real word cannot be made with visual/orthographic processing alone and is clearly phonological. Accurate phonological lexicality judgements such as correct responses in a phonological LDT, therefore, require phonological processing and access to the phonological lexicon, even if semantic activation occurs too.

Based on theories of reading development and reading disorders, phonological access from orthographic input depends on the word being familiar to the reader (McNorgan et al., 2015). However, neither pseudowords or pseudohomophones are orthographically familiar, having never (or very rarely) been seen before, not having a place in the orthographic lexicon (Coltheart et al., 2001), and allowing only sublexical orthography to aid their recognition. Although pseudowords can activate semantic processes through similarity to real words with large neighbourhoods or if they have large orthographic or phonological neighbourhoods (Holcomb et al., 2002; Nation & Cocksey, 2009), pseudowords are made of sublexical units in a novel configuration that does not provide a direct path to legitimate meaning because they have no legitimate meaning. Pseudowords, therefore, likely require more serial processing than real words in that a process, such as letter identification, must at least begin, if not be completed, before other processing, such as grapheme-phoneme conversion, can occur (Coltheart et al., 2001; McNorgan et al., 2015).

Such an indirect route is also required to process pseudohomophones because the orthographic representations of these stimuli are not in lexical memory either (Coltheart et al., 2001). Pseudohomophones, however, can access legitimate phonology in the phonological lexicon (and, by extension, semantics) that pseudowords cannot, giving pseudohomophones a phonological benefit, but an orthographic obstacle (Ziegler et al., 2001). Both pseudohomophones and pseudowords, therefore, involve an element of incongruence. Orthographic incongruence in pseudowords is on a simpler lexicon-level, where the orthography of the stimulus does not match any of the orthographic entries in the lexicon, while orthographic incongruence in pseudohomophones is on both lexicon- and stimulus-level, where the orthography of the stimulus does not match the real word target. The legitimate phonology in pseudohomophones, therefore, results in a kind of double-edged sword of orthographic-phonological conflict.

Translating external visual stimuli (e.g., words) into internal phonological representations involves top-down processes influencing the way bottom-up visual/orthographic input is managed through word- and language-specific knowledge and statistical information about sublexical characteristics, combinations, and correspondences (Mechelli et al., 2005). Some psycholinguistic information, such as lexical and grammatical rules, psycholinguistic properties, such as lexical status, familiarity, and the sublexical phonology of letter combinations, can be considered top-down information, conceptualized to be in the mind of the reader and not part of the written word directly (Twomey et al., 2011). Indeed, following the principle of the word superiority effect (WSE), lexical status (i.e., whether the stimulus is a real word or not) influences sublexical orthographic processing, demonstrating top-down feedback during VWR (Twomey et al., 2011). For instance, phonological lexicality and pseudohomophone effects demonstrate top-down processing in VWR in that the known and familiar (top-down) sublexical phonology of pseudohomophones results in an orthography-phonology conflict and impacts judgement of the (bottom-up) linguistic stimulus, showing that top-down phonological information can influence VWR, even when it is essentially irrelevant to the choice (Twomey et al., 2011). Although it involves this conflict between the partially bottom-up orthographic input and top-down phonological processing, such effects are accepted as evidence of automatic phonological recoding during reading and tantamount to a phonological variant of lexicality (Ziegler et al., 2001). Furthermore, this top-down phonological information appears to be activated rapidly and can conflict with mismatching orthographic form (Twomey et al., 2011), the resolution of which could be the root of the decreased efficiency and increased difficulty in processing pseudohomophones compared with relatively conflict-free words and pseudowords (Briesemeister et al., 2009). It is such automatic and rapid top-down influence of internal phonology and of internal orthography-phonology relationships on VWR processes that the current study aims to investigate through behavioural and ERP responses to discrete orthographic and phonological lexical decision tasks.

In terms of timing and ERPs, pseudohomophone effects have been reported at ~150ms (e.g., Braun et al., 2009), while other studies reported them on the N400 (~350-500ms) but not earlier timeframes (e.g., Briesemeister et al., 2009) demonstrating a discrepancy between findings that requires further investigation. Furthermore, other findings of phonological processing observed within 100ms post-stimulus suggest that the orthography-phonology conflict in pseudohomophones is resolved markedly earlier than the N400 timeframe (e.g., Ashby et al., 2009; Wheat et al., 2010). Such findings propose especially early phonological processing that involves parallel processing of orthography and phonology, while also suggesting phonological effects at 100-200ms and 400ms are distinct. In the earlier timeframe of ~150-200ms (e.g., P150, N170), larger ERP amplitudes to pseudowords (spelling controls) compared with pseudohomophones have been suggested to reflect orthography-phonology conflicts (Braun et al., 2009). However, lexical memory holds no full orthographic representations for pseudohomophones and no full orthographic or phonological representations for pseudowords (Coltheart et al., 2001). Therefore, activity in this timeframe appears to reflect sublexical orthography-phonology processing, including any conflicts (e.g., pseudohomophones), proposing it as an index of n-gram frequency. This is supported by reports of bigram analysis and n-gram frequency effects, which represent a large part of sublexical processing, being strongly associated with the N170 (Hauk, Davis, et al., 2006; Holcomb & Grainger, 2006), hence it being imperative that such psycholinguistic variables be strictly controlled between stimuli/conditions. While low-level effects of graphemes, phonemes, or n-grams alone do not constitute lexicality (Coch & Mitra, 2010), they underpin and can influence the processing required for lexicality decisions (Hauk et al., 2008), requiring such sublexical units to be tightly controlled for effects of whole-word orthography and phonology to be observed.

Neuroimaging studies have shown a relationship between occipitotemporal activity and phonological lexicality with greater activation to pseudohomophones than pseudowords as well as real words (Twomey et al., 2011; van der Mark et al., 2009). Based on pseudohomophones requiring conflict resolution that real words and pseudowords do not, increased occipitotemporal activation to pseudohomophones could represent the integration of orthographic and phonological information, providing evidence for the role of the occipitotemporal region as an interface between bottom-up visual properties and top-down “nonvisual” (phonological and semantic) processes (Twomey et al., 2011). It is such findings that this study aims to replicate using ERPs (in particular, the occipitotemporal N170 in this case), also providing insight into the timing of this processing, which is a primary focus of the current work.

As with pseudowords, orthographic evaluation of pseudohomophones concludes that the stimulus does not exist in orthographic memory (Coltheart et al., 2001). Unlike for pseudowords that do not have legitimate whole-word phonology in memory, however, reading pseudohomophones results in a processing conflict when their legitimate phonology is activated, as demonstrated by responses to pseudowords being typically more efficient than to pseudohomophones in LDT contexts (Braun et al., 2009). For orthographic evaluation to have such an effect, it must draw its conclusions before phonological activation, otherwise the phonological lexicon would be accessed, the phonology of the pseudohomophone would be recognized as sounding real, and the decision would be complete. Furthermore, for the orthography-phonology conflict to exist, it follows that phonological activation begins before orthographic evaluation is complete or, at least, it follows automatically (Ziegler et al., 2001). Conceptually, phonological activation based on orthographic evaluation implies a serial nature to early word recognition processes, following a more traditional dual-route model (e.g., Coltheart et al., 2001), while phonological activation based on visual evaluation does not and permits parallel processing of orthography and phonology, as posited by the BIAM (Grainger & Holcomb, 2009).

6.2.5 Orthographic and phonological lexicality in bilinguals

The perspective on bilingualism in the current research, as stated in the introduction to this thesis, is different from the usual perspectives on bilingualism in the literature. This difference is due to the focus on L2 processing in two very different ESL bilingual populations (Spanish-English and Chinese-English late bilinguals) and, importantly, not contrasting bilinguals' L2 and L1 processing within groups, instead examining L2 processing between groups taking the bilinguals' L1 language profile into account. Therefore, previous bilingualism research relevant to the current study (in terms of participants/samples, conditions/stimuli/tasks and/or methodology i.e., behavioural, EEG/ERP) has generally used significantly different perspectives (Sun-Alperin & Wang, 2011). These have either been comparing L1 and L2 within a single population (e.g., Spanish and English stimuli in Spanish-English participants), comparing groups with similar stimuli/tasks but using native English participants with a second language (e.g., English participants learning Spanish and a native Spanish group), using comparable stimuli/tasks but with ESL bilinguals very different from the current work (e.g., Dutch-English, German-English or Arabic-English), using early/native-like bilinguals (as opposed to late-bilinguals), investigating interlingual factors using the L1 of the ESL bilingual sample (e.g., Spanish-English participants but Spanish stimuli), or using similar tasks and stimuli to the current study but only focused on L1 processing (e.g., Chinese-English and Chinese monolingual participants with only Chinese stimuli).

Even the larger and more extensive behavioural literature (i.e., without a neuroscience component) appears to offer very few direct like-for-like comparisons regarding condition/task (using English word/pseudohomophone and pseudohomophone/pseudoword comparisons) and group (Spanish-English and/or Chinese-English bilinguals) factors for results, findings, or insights of the ESL groups' performance. For instance, there are studies using samples of French-English e.g., Haigh & Jared (2007), Dutch-English e.g., Duyck (2005), Portuguese-English e.g., Vale (2011), and Arabic-English e.g., Taha & Khateb (2013), but not Spanish-English or Chinese-English in a study sufficiently comparable on other factors e.g., stimuli, task, and methodological approach. For example, studies such as Boukrina et al. (2014) and Carrasco-Ortiz et al. (2012b) had likenesses to the current work, but their participant inclusion and/or conditions/tasks were still much too distinct to be directly helpful, while studies such as Sun-Alperin and Wang (2011) only used real words and pseudowords (no pseudohomophones) and the Spanish-English bilinguals were young children. In any case, there is a shortage of information about the behavioural VWR performance and, especially, early brain activity during L2 VWR associated with orthographic and phonological processing in late bilingual ESL readers with an alphabetic or non-alphabetic L1, such as Spanish-English and Chinese-English, respectively. Indeed, this relative novelty of the current research precluded finding previous studies so similar as to make the current one a true replication. However, this does strongly highlight that novelty and the gap in knowledge about late-bilingual ESL processing in populations with distinct types of L1 (Sun-Alperin & Wang, 2011), even more so considering the early timeframe and EEG/ERP aspects too. The main reference for behavioural hypotheses and results will, therefore, be Twomey et al. (2011) due to the overlapping task methodology and Ota et al. (2010b) due to the similar groups and perspective. ERP hypotheses and discussion of results, meanwhile, will be based largely on these behavioural results alongside current knowledge of P1 and N170 activity in English monolinguals, linking where possible to theories of bilingualism.

Although different stimulus conditions and tasks were used and without a neuroscience component, Ota et al. (2010b) had a similar perspective on L2 VWR and involved L1 orthography and phonology in relevant late ESL bilinguals with different language profiles (Spanish-English and Japanese-English). The visual semantic categorization task employed used homophones (e.g., *see* for *sea*) or minimal pairs not present in the ESL participants' L1 (e.g., *fun* for *fan*) for the category exemplars to create a contrast between orthographic and phonological processing. More false positive errors and slower processing were observed for homophones than orthographic controls across all groups, suggesting, at least on the surface, that native and ESL readers employ similar strategies, including L2 readers with an alphabetic or non-alphabetic L1. For the minimal pairs condition, phonological "near-homophone" effects were reported on accuracy but not response times in Japanese-English (non-alphabetic L1), but not Spanish-English (alphabetic L1) participants (Ota et al., 2010b). More specifically, certain phonemic contrasts from English not found in Japanese (/æ/–/ʌ/, /b/–/v/, and /l/–/r/) were observed to result in more false positive errors in the Japanese-English, while the Spanish-English group was unaffected. As the Spanish-English participants would be familiar with these phonemic contrasts (as they are also present in Spanish), these results show that phonological mediation as in native (L1) processing is also apparent in bilingual processing of non-native words or, at least, sub-lexical sounds (phonemes). It was concluded that L1 phonology influences L2 VWR, but the homophone and near-homophone effects reported depend somewhat on the L1 orthographic system (Ota et al., 2010b). This feeds directly into the L2 processing perspective of the current research, which is firmly focused on how English is read by L2 readers and explicitly not how ESL and L1 processing contrast within bilinguals.

6.2.6 The current study

Central to both orthographic and phonological lexicalities and, therefore, the orthographic and phonological experiments of the current study, is the pseudohomophone effect, which is observed when pseudohomophones are processed less efficiently (e.g., more slowly and less accurately) than real words or pseudowords (Braun et al., 2009). To operationalize this, the task methodology for the current study largely follows that of Twomey et al. (2011), which employed orthographic and phonological lexical decision tasks with English monolinguals. The orthographic lexical decision task (oLDT) was used to focus on orthographic processing and involves real words (RW) and pseudohomophones (PH1) that respectively require different processing routes and reflect the distinction of orthographic lexicality. In contrast, the phonology focused task is a phonological lexical decision task (pLDT), which uses pseudohomophones (PH2 – matched but different stimuli from PH1) and pseudowords (PW) that entail the same initial processing path as one another but diverge in terms of phonological lexicality. These experiments represent a two-task (oLDT and pLDT) approach to studying orthographic and phonological processing (respectively), examining how different psycholinguistic strategies for word recognition, such as a direct lexical pathway or indirect grapheme-phoneme conversion, work in L1 and L2 readers with different language profiles. Due to the fMRI approach used previously with this task methodology (Twomey et al., 2011) and the use of an EEG/ERP methodology in the current work, some minor amendments were required (as outlined in §6.3.2), but the current study effectively replicates this task methodology using EEG/ERP instead of fMRI.

The specific neuroimaging findings of Twomey et al. (2011) cannot be replicated or directly extrapolated upon in the current work (due to the divergence in the neuroscience methods of fMRI and EEG/ERP). However, the behavioural results and interpretations of findings in the context of VWR processing are all invaluable to the current research, which aims to complement the aforementioned findings of Twomey (2011) and other relevant studies (e.g., Braun et al., 2009, 2015; Briesemeister et al., 2009; van der Mark et al., 2009) as discussed earlier in the chapter. As such, evidence of orthographic/phonological processing of the physical visual/orthographic and internal phonological representations occurring in series or parallel (or in a cascaded (DRC; Coltheart et al., 2010) or overlapping (BIAM; Grainger & Holcomb, 2009) fashion) is an important aspect of this study. Such evidence will be observed through task effects with time-focused electrophysiological measures to complement previous fMRI findings (e.g., Twomey et al., 2011), as well as extending across bilingual ESL readers.

As a primary facet of the thesis, Study 1 will contrast all measures within and between the English, Spanish-English (non-native alphabetic L1), and Chinese-English (non-native non-alphabetic L1) groups. Specific to Study 1, these measures are concerned with orthographic and phonological processing through orthographic and phonological lexicality (cf. the pseudohomophone effect), as well as any impact of internal phonology as observed through contrasts between the orthographic and phonological LDTs. The main aims are to examine the sensitivity of ERP responses around ~100ms (the P1 timeframe) to lexicality manipulations (especially orthographic, as per the typically more orthography-related response of the occipital P1) and to investigate the orthographic and/or phonological nature of activity during the N170 timeframe (including visual familiarity and orthography-phonology integration).

In terms of ERPs, the focus on P1- and N170-related activity here is mainly because these early timeframes have been shown to be important for VWR in native readers of English, especially concerning orthographic and phonological processing, but such early ERP activity is often overlooked in bilingualism studies in favour of later ERPs (e.g., N400, P600) and, therefore, is not well understood in the context of bilingual VWR. Therefore, the current work aims to examine these timeframes in ESL readers with different language profiles to see how similar or different reading-related brain activity and behavioural performance are between L1 and L2 readers, as well as between different ESL groups (Spanish-English and Chinese-English).

6.2.6.1 Hypotheses

As discussed earlier in this chapter, relatively little is known about how ESL bilinguals (especially late bilinguals) process lexicality decisions in English (i.e., their L2) and how different L1 language profiles (e.g., alphabetic or logographic) might impact such processing in relation to native readers and between ESL groups. Behaviourally, it can be expected that ESL bilinguals will generally perform as well (if not better) than native readers in VWR tasks (such as lexical decision) when they are early bilinguals with a native-like proficiency (Clahsen & Felser, 2006). How this behaviour and, indeed, brain activity looks in late bilinguals with good but not necessarily native-like levels of English proficiency (e.g., levels acceptable at UK universities), however, is not yet sufficiently documented. Furthermore, how differences in terms of initial orthographic and phonological processing and associated early <200ms ERP activity potentially stemming from language profiles of different ESL groups has also not yet been sufficiently studied, but some contrasts between groups are expected.

Based on the same tasks as the current study with native English monolinguals (Twomey et al., 2011), responses to pseudohomophones (PH1) will be more accurate but slower than real words (RW) in the oLDT and more accurate but slower than a different set of pseudohomophones (PH2) between tasks, while pLDT responses will be faster to pseudohomophones (PH2) than pseudowords (PW) with no observed difference in accuracy. Although no directly comparable results for these tasks and Spanish-English or Chinese-English participants were found, it will be of significant interest to contrast the results of the current study and these ESL groups with results of Twomey (2011).

Considering the late bilingual Spanish-English and Chinese-English language profiles involved in the current research alongside the evidence primarily concerning reading alphabetic scripts (discussed earlier), some differences in behavioural performance are expected. Behavioural measures are expected to show a similar pattern of orthographic/phonological processing between native and ESL readers, albeit the latter being slower and less accurate in general. However, some behavioural differences are expected between ESL groups: Spanish-English being faster to pseudowords and pseudohomophones based on their phonology-oriented L1, while Chinese-English being more accurate with orthographically-oriented stimuli that can be decoded via whole word routes e.g., real words due to their visual/orthographic logographic L1, which may also be a hindrance for the more phonological manipulations e.g., pseudohomophones.

While the effects and their operationalization reported in Ota et al. (2010) are distinct from the pseudohomophone or lexicality effects investigated in the current study, they still suggest a similar pattern in the Spanish-English and Chinese-English groups will be found in the current study. Specifically, Chinese-English will exhibit increased phonological errors shown through accuracy for pseudohomophones in the oLDT and an equivalent error rate between conditions in the pLDT, while the Spanish-English will not show an equivalent phonological difficulty in the oLDT and will fare better with pseudohomophones in the pLDT.

The Spanish-English group, due to the phonological nature of their alphabetic L1, is likely to show a performance advantage over the Chinese-English group in terms of response times for accessing the phonology of pseudowords (pLDT) and pseudohomophones (both oLDT and pLDT). However, the objective difference between English and Chinese (or logographic scripts in general) could also serve as an advantage in terms of avoiding linguistic conflicts between language types, especially when visual /orthographic processing is all that is theoretically required for the task, as in the oLDT. The lack of an equivalent grapheme-phoneme conversion process in the Chinese-English bilinguals' L1 could also make processing the pseudohomophones and pseudowords difficult though. It is, therefore, expected for the Spanish-English group to perform better overall than the Chinese-English group with the English group approaching ceiling effects in most conditions. In terms of ERP response, however, some overlap is expected between each possible pair with the most similar being between English and Spanish-English, due to their shared experience of alphabetic orthography. Meanwhile, Chinese-English brain responses could show a different pattern and strategy of VWR processing, which would likely involve differences in P1 and/or N170 lateralization and condition effects compared with English and Spanish-English groups.

Considering prior associations between VWR processes and ERP responses, behavioural patterns are anticipated to be reflected in the ERP responses to some extent, though inconsistencies from previous studies (as discussed earlier in this chapter and in the introductory chapters) make it unclear when, where, and how (i.e., in which component/timeframe or measure), especially in the ESL groups. However, it is vital to recall that one of the main rationales of this research is to examine the early pre-200ms neurophysiological timeframes in distinctive language groups in the context of such orthography- and phonology-oriented experimental conditions precisely because this information is missing from the literature and is essential to inform related and subsequent/consequent processes.

Despite the lack of previous findings related to early pre-200ms ERP activity during such specifically orthographic and/or phonological VWR contexts, it is reasonable to expect that, along with the native English group, the Spanish-English group shows ERP effects related to phonological processing at 100ms at left frontal-central sites (cf. Ashby, 2010; Wheat et al., 2010), while effects most likely in the Chinese-English group are on right-lateralized P1-OT and N170-OT. Finding how orthographic and phonological lexicality decisions can influence early ERP activity will inform the broader question of what psycholinguistic processing occurs or can occur as early as ~100ms post-stimulus. If lexicality effects are observed at ~100ms on ERP measures (e.g., P1-O, P1-OT, or N100-FC amplitude or latency), they will reflect more specific orthographic and/or phonological processing in contrast with general RW/PW lexicality effects in this timeframe, which are so far not consistently reported in the literature. This unexplained inconsistency of findings (as discussed earlier in the chapter), however, reduces confidence in a hypothesis one way or another, despite various control measures, such as matching bigram frequency and other psycholinguistic variables (see 5.4.1) and avoiding peak ERP measures, being strengthened or added. Ultimately, effects at ~100ms could require the broader RW/PW distinction and the narrower comparisons of RW/PH and PH/PW, respectively, may not be sufficient.

Considering the association of P1-OT responses during VWR and orthographic processing (Dien, 2009), a difference in its amplitude or latency between ESL groups would at least support the notion that processing alphabetic scripts varies between bilinguals with an orthographically shallow L1 (e.g., Spanish-English) compared with bilinguals with a more deeply orthographic L1 (e.g., Chinese-English). More specifically concerning the conditions of the orthographic task (RW,PH1), a distinction in right hemisphere P1-OT response is expected at least in the native English group (Dien, 2009), showing sensitivity to the distinction between the legitimate orthography of real words and the legal but not legitimate orthography of the pseudohomophones. The same or similar such effect might reasonably be anticipated in the Spanish-English group too, considering their alphabetic background and English proficiency alongside the relative high frequency of the real word stimuli. However, prior testing and analyses of this component and timeframe in an L2 VWR context have not been available to refine the hypothesis.

Taking into account the reports of lateralized N170 activity in Chinese readers (Maurer, Brandeis, et al., 2005; Yum & Law, 2021), along with the orthographic nature of both Chinese and the right-lateralized P1-OT, it is not unreasonable to anticipate some effect in the Chinese-English group between RW and PH1 or between tasks on P1-

OT amplitude or latency that reflects orthographic processing. Regarding the pLDT, the matched orthographic legitimacy between pseudohomophones and pseudowords (i.e., neither have legitimate orthography) proposes that no effects or differences are to be expected between pLDT conditions (PH2,PW) on P1-OT activity due to its typically orthographic nature. Nevertheless, it should not be dismissed that alternative aspects of the P1-OT could be reflected in bilinguals if VWR processing or strategies for L2 reading is sufficiently distinct from native readers, which would also open up the possibility of group differences. Again, however, this is conjecture due to the lack of sufficient relevant information from appropriate previous research on the occipitotemporal P1 of ESL bilinguals in orthographic or phonological contexts.

Regarding phonology and N100-FC activity, the controversy of phonological activation during VWR occurring as early as ~100ms, as discussed earlier, presents further difficulty for hypothesizing. Activity at left frontal-central sites in response to phonological manipulations has been documented and is not disputed as a possibility per se, but the nature of such activity is in question and the context in which it occurs is still unclear. This is at least the case for native readers and it could indeed be an effect specific to masked priming (as opposed to phonological effects, priming or otherwise), but there is not yet any evidence of an equivalent effect during L2 VWR in bilinguals. While this may be due to the lack of focus on this early timeframe in bilingualism research, there is still no precedent for L2 phonology being activated either at ~100ms post-stimulus or at these frontal-central areas in bilinguals, so it should not necessarily be expected to be observed in the current study.

Based on the largely visual/orthographic nature of the N170 and that many effects on it have been reported in the VWR literature, it is expected that N170 responses in the native English group will differ between real words and pseudohomophones. Orthographic/phonological lexicality effects are expected on the N170 in native readers, which will help determine the origin of the effects: orthographic lexicality will imply familiarity and/or the requirement for grapheme-phoneme conversion (GPC), while phonological lexicality will imply validity/quality of the GPC outcome i.e., whether it is recognized as a real word's phonology. Taken together, orthographic and phonological lexicality effects will inform whether the N170 is associated with GPC at all or is more likely to be a visual/orthographic component that has been associated with phonology previously through epiphenomenon.

N170 lateralization is expected to differ between the English group and the Chinese-English group for real word stimuli. It could be expected that pseudowords (and perhaps pseudohomophones) might be processed more bilaterally in the Chinese-English due to their orthographic novelty (or lack of whole word orthographic familiarity), which would mirror the N170 lateralization typically found during native reading of Chinese. In contrast, it is expected that N170 lateralization in the Spanish-English group will more closely resemble that of the English group due to the familiarity of the native-like alphabetic stimuli in which they share deeper expertise that the Chinese-English group do not.

Lastly, at the overall task level (i.e., oLDT vs pLDT, PH1 vs PH2), orthographic effects are expected to occur earlier than phonological effects in some scenarios e.g., grapheme-phoneme conversion as required for reading pseudohomophones and pseudowords, but there will be an equivalency in others e.g., direct lexical access as in real words in the English group, suggesting serial or parallel processing of orthography and phonology to be context-dependent.

6.3 Method

6.3.1 Participants

Full details of the participants can be found in section 5.2 and the following provides an overview of the basic participant information. The English group (n=20) included monolingual native readers of English, while the native Spanish group (n=20) and Chinese group (n=20) included late-bilinguals with English as their second language (IELTS of 6.5 or equivalent). All participants had normal or corrected-to-normal vision and no participants reported specific learning or language disabilities or neurological disorders.

6.3.2 Design

Study 1 used orthographic and phonological variants of lexical decision task (LDT), experiments that were designed as a combined two-task approach. The orthographic LDT (oLDT) and phonological LDT (pLDT) distinguish "visual" and "non-visual" elements of word recognition, respectively, and separate visual familiarity from phonology/semantics between tasks (van der Mark et al., 2009). The oLDT requires participants to indicate whether real words (e.g., *tool*) and pseudohomophones (e.g., *fome*) are real English words or not, which highlights orthographic processing, as it can be completed by visual information alone. In contrast, the pLDT requires participants to indicate whether pseudowords (e.g., *lish*) and a different set of pseudohomophones (e.g., *gole*) have the phonology (i.e., sound) of real English words, therefore focusing on phonological activation. Following the notion of different forms of lexicality based on different psycholinguistic properties being processed, orthographic and phonological variants of LDT could show orthographic and phonological lexicality, respectively, which can be used to isolate the processing involved.

The task methodology of the current study is a broad replication of the task methodology used in an fMRI context with native English readers (Twomey et al., 2011), implementing several necessary methodological changes. In place of fMRI, an EEG/ERP approach was used and the design optimised accordingly, mainly through the stimuli and the timings of the trials. The EEG/ERP approach suits the focus on timing of this research and the stimuli were more tightly constrained in terms of psycholinguistic variables (e.g., semantic association). These changes could be considered a significant departure from the original Twomey et al. (2011) study, but their purpose is to increase statistical power and the chance that any findings are not due to aspects of the stimuli not controlled adequately, but to relevant aspects of the stimuli. The current study also extends it to bilingual groups as well as the previously used native English population. The oLDT and pLDT have not previously been conducted using non-native readers of the target language, which is an important novel aspect of Study 1.

6.3.3 Stimuli

The oLDT used 60 real words (RW; e.g., *tool*) and 60 pseudohomophones (PH1; e.g., *fome*), while the pLDT used a different 60 pseudohomophones (PH2; e.g., *gole*) and 60 pseudowords (PW; e.g., *lish*) as stimuli (see Appendix F for the full stimulus sets). General information and discussion of stimulus properties can be found in section 5.4.2 with details pertaining specifically to Study 1 below.

6.3.3.1 Stimulus creation

Using the pool of real words as base words (see §5.4.3 for details), each pseudohomophone and pseudoword was created by substituting a single phonotactically appropriate phoneme in a base word. Pseudohomophones are here defined as linguistic stimuli with legal but not legitimate orthography that will reasonably be pronounced with the same phonology as its real word counterpart. In other words, pseudohomophones are linguistic stimuli that look fake but spelt in a way that would still reasonably lead to the expected base word pronunciation (and therefore the correct meaning through the phonology). Pseudowords, on the other hand, also exhibit legal orthography, but do not have the legitimate phonology or semantics of real words and pseudohomophones. Essentially, pseudowords are truly fake words, but should still adhere to the phonotactic rules of the respective language (here, English) and therefore be pronounceable. Vitally, pseudowords and pseudohomophones all conformed to the same orthographic, phonotactic, morphosyntactic, and syntagmatic rules of English that real words do (Hauk et al., 2012).

Every effort was made in creating the stimulus sets to ensure pseudowords and pseudohomophones would only be reasonably expected to be pronounced in the designed way. To this end and to reduce the inclusion of phonologically ambiguous stimuli as much as possible, the values for frequency (CELEX), biphone frequency, phonological neighbourhood, and imageability were taken from their base words (and therefore their intended and expected phonological form) in order to compute comparisons and ensure matching across conditions. Values true to the pseudohomophones' orthographic forms were used for controlling number of letters (which was matched to the base word), bigram frequency, and orthographic neighbourhood. Stimuli were also trialled in a brief pilot study (conducted to check the stimuli and task design), with native English, Spanish-English, and Chinese-English associates, and via scale-based ratings of the orthographic and phonological efficacy of the stimuli (see §5.4.4).

6.3.3.2 Stimulus control

Descriptive statistics for each condition across core psycholinguistic variables are presented in Table 3. The necessary control of psycholinguistic variables was checked and supported by no significant differences (α=.05) across all appropriate statistical comparison tests on all viable pairwise permutations of the conditions: RW/PH1 (oLDT), PH2/PW (pLDT), PH1/PH2 (between-task), as well as RW/PH2, RW/PW, and PH1/PW (for completeness).

Table 3: Means (standard deviations) for oLDT and pLDT stimulus sets (n=60 per condition[6])

	RW	PH1	PH2	PW
Letters	4.27 (0.45)	4.22 (0.42)	4.23 (0.43)	4.32 (0.47)
CELEX frequency[1] *	74.8 (103.82)	51.22 (84.22)	61.49 (87.65)	-
Subjective frequency[2] *	455.43 (112.62)	429.95 (122.04)	428.13 (121.73)	-
Bigram frequency[3]	1948.65 (1174.74)	2081.34 (1437.64)	1918.99 (1425.27)	1796.13 (1047.78)

Biphone frequency[3]	601 (524.49)	654.72 (618.85)	568.87 (419.51)	-
Neighbourhood[4]	8.87 (4.83)	7.65 (4.63)	7.83 (4.75)	8.25 (4.31)
Phonological neighbourhood	20.35 (7.7)	20.15 (7.96)	20.32 (8.72)	-
Imageability[5] *	514.53 (79.8)	476.8 (166.93)	489.18 (156.94)	-
Age of Acquisition[6] *	296.16 (75.58)	320.85 (79.31)	309.96 (77.41)	-

[1] British English, units per million [2] Balota, Pilotti, & Cortese (2001) [3] Token measure, units per million [4] Coltheart's N [5] Retrieved from the MRC database (Coltheart, 1981) [6] Not all stimuli have AoA values in the Bristol norms (Stadthagen-Gonzalez & Davis, 2006), hence the entered number of stimuli/values per analysis were lower than the number of stimuli in the respective condition: RW (n=44), PH1 (n=48), PH2 (n=45)

* Using data from base words for pseudohomophone and pseudoword statistics.

Note: Unless otherwise specified, all stimulus data was retrieved using N-Watch (Davis, 2005).

6.3.3.2.1 Stimulus survey testing

Survey data for each stimulus type was collected online from 12 random, unknown, and anonymous ESL readers who did not participate in the main research. These tests were conducted using Qualtrics (Qualtrics, 2005) on the participants' own personal devices at their convenience. Participants were shown the real English words (RW), pseudohomophones (PH1 & PH2), and pseudowords (PW) used in Study 1. They were asked to rate each stimulus on a scale of 0-5, where 0 denoted the stimulus sounded nothing like a real English word and 5 denoted the stimulus sounded exactly the same as a real English word. Stimuli were displayed one at a time in black Arial (24pt) on a white background with no time limit to respond. Table 4 shows the descriptive statistics of each stimulus type.

Table 4: Descriptive statistics of oLDT and pLDT stimulus checking tasks

	Mean (SD)	Range
Real words (RW)	4.99 (1.67)	4.91 - 5.00
Pseudohomophones (PH1 & PH2)	4.72 (0.14)	4.40 - 4.91
Pseudowords (PW)	0.98 (0.23)	0.40 - 1.44

The pseudowords *holl* and *nish* were rated slightly less consistently than others, but all stimuli were rated within acceptable ranges for their respective stimulus types in that the means of stimuli that had legal and legitimate phonology (RW, PH1, and PH2) resided within the first quartile and the mean for the stimuli without legitimate phonology (PW) were in the fourth quartile. This also allowed a reasonable margin of error due to the potential effect of pseudohomophones' lack of legitimate orthography on their ratings compared with real words.

6.3.4 Apparatus

Full details and discussion of all used materials can be found in section 5.5.

6.3.5 Procedure

For the oLDT, participants were asked to indicate as quickly and accurately as possible whether real words (RW, e.g., *tool*) and pseudohomophones (PH1, e.g., *fome*) were real English words or not. For the pLDT, participants were asked to indicate whether pseudohomophones (PH2, e.g., *gole*) and pseudowords (PW, e.g., *lish*) have the exact phonology (i.e., sound) of real English words. Both tasks shared the same trial format, as shown in Figure 9.

Time was taken to explain the distinctions between choices in each task and the same instructions, examples, and practice trials were given to all participants to attempt controlling demand characteristics. The difference between oLDT conditions was relatively straightforward, but clarification was especially important for the pLDT. Through instructions, discussion, and practice, it was ensured that participants understood that stimuli responded to with the pseudohomophone button must sound "exactly the same" as real English words and not just sound "like" real English words in the way that the pseudoword *yarm* could be said to sound like real words *farm* and *balm*, which was not the object of the exercise and might have dissolved any effects of it.

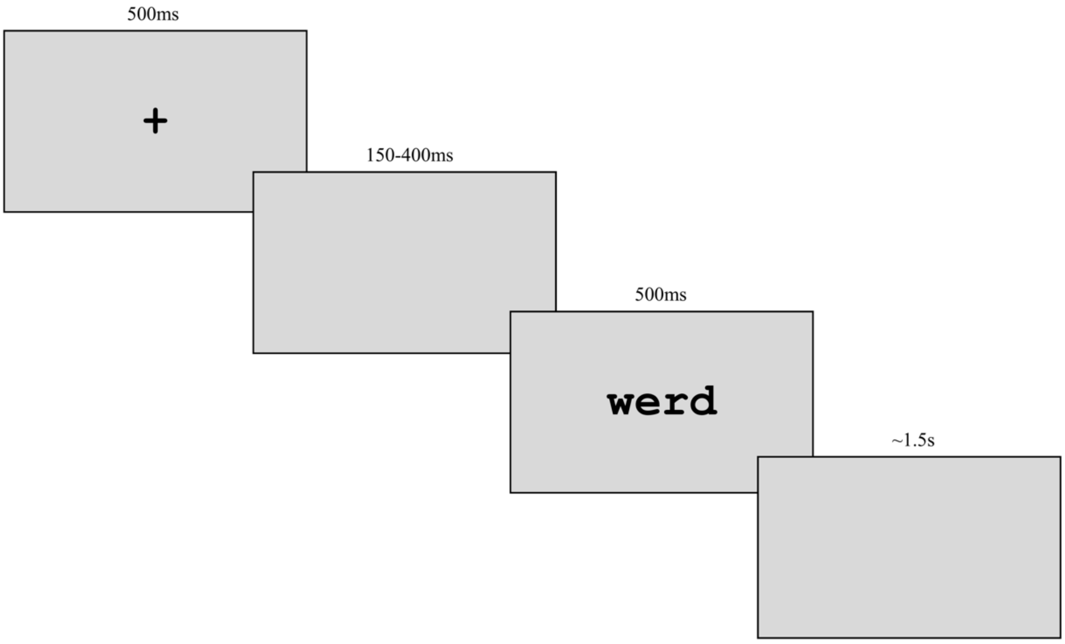

Figure 9: Stimulus presentation format of a single trial in orthographic and phonological LDTs

The interstimulus interval (ISI) between fixation and stimulus was pseudorandomly jittered, using a rectangular distribution of times (150, 200, 250, 300, 350, 400ms) that follow the 60Hz refresh rate (16.67ms refresh duration) of the monitor and allow for the recommended minimum SOA of 640ms (Perfetti & Liu, 2005), providing an average SOA of 2775ms with 2650ms minimum and 2900ms maximum durations. Each task, including the half-time break, was expected to last 6-8 minutes.

6.3.6 Analysis

Across all behavioural and ERP measures, comparison between RW and PH1 (oLDT) within each group provides an indication of orthographic lexicality, while comparison between PW and PH2 (pLDT) demonstrates phonological lexicality. Comparison of the two pseudohomophone conditions, PH1 (oLDT) and PH2 (pLDT), reflects any difference in top-down processing in terms of a task effect (Twomey et al., 2011).

In terms of inferential statistics, oLDT, pLDT, and Task analysis of the behavioural data each used a univariate 3 x (2) mixed factorial ANOVA with one between-participants factor of language group (English, Spanish-English, Chinese-English) and one within-subject factor of condition (analysis-dependent). For ERP data, a univariate mixed factorial 2 (Condition: analysis-dependent, stated per analysis) x 2 (Hemisphere: left/right) x 3 (Group: English/Spanish-English/Chinese-English) ANOVA was conducted for each combination of component/cluster (P1-O, P1-OT, N100-FC, N170-OT) and measure (positive/negative area amplitude, 50% positive/negative area latency). Where appropriate, the ANOVA outputs will be followed by relevant pairwise post-hoc results (using Tukey corrections for group comparisons and Holm corrections for comparisons within groups). Only statistically significant results will be reported and any main effects, interactions, and post-hocs not reported should be understood as non-significant. Full statistical outputs, including 95% confidence intervals, are included in Appendices J (behavioural), K (ERP amplitudes), and L (ERP latencies).

6.4 Results

For the measures of accuracy, response times, and each ERP component/region measure, the following sections will include descriptive statistics and inferential statistical test results from the contrasts of groups (English, Spanish-English, Chinese-English) and experimental conditions (RW vs PH1 from oLDT, PW vs PH2 from pLDT, and overall oLDT and pLDT measures (averaged across conditions within tasks) as well as PH1 vs PH2 for task-level effects).

6.4.1 Data preparation

Data were prepared according to the description and explanation in section 5.7.3.

6.4.1.1 Pre-analysis item evaluation

While cross-linguistic relationships in the stimuli were avoided as much as possible through the stimulus creation methods described in sections 5.4.3 and 6.3.3.1, as well as manual checks by language experts and volunteer native speakers of Spanish and Chinese, the Study 1 stimulus sets included Spanish nouns *bote* and *doce* as pseudohomophones, neither being pronounced identically between languages. Neither behavioural nor ERP responses to these stimuli were outliers in the data. With only these items, interlingual elements in the stimulus sets were minimal (0.83% of Study 1 stimuli), effectively random throughout the experimental procedure, and without the power to influence averaged overall results. While such a small minority are highly unlikely to affect any behavioural or ERP analysis, especially after rigorous data cleaning and processing (e.g., artefact rejection, filtering, averaging, outlier removal), all data associated with them was removed from analysis. This, combined with the comprehensive pseudo-randomization and counterbalancing, minimizes any detriment to the analysis.

6.4.2 Orthographic lexical decision task (oLDT)

Accuracy

Based on the descriptive statistics for oLDT accuracy as shown in Table 5, the same pattern was observed between conditions and between groups. Accuracy was higher in the English group than in the Spanish-English group and higher in the Spanish-English group than the Chinese-English group for both oLDT conditions (RW,PH1), for which accuracy was consistently higher for RW than PH1 across all groups. The notably high standard deviation for PH1 in the Chinese-English group is due to a wider spread of accuracy scores within the group.

Table 5: Means and standard deviations for oLDT accuracy (% correct)

Condition	Group	Mean (% correct)	SD
	English	97.97	3.04
Real Word	Spanish-English	95.51	4.23
	Chinese-English	93.22	5.04
	English	94.08	3.88
Pseudohomophone1	Spanish-English	87.67	7.81
	Chinese-English	80.92	14.72

The ANOVA on accuracy showed a significant main effect of Group, $F(2,57)=14.12, p<.001, \eta^2_p=.33$, but no significant Condition x Group interaction, $F(2,57)=3.04, p=.056, \eta^2_p=.1$. Averaging across conditions, accuracy was higher in English than both Spanish-English ($p=.032$) and Chinese-English ($p<.001$) and higher in Spanish-English than in Chinese-English ($p=.025$). There was also a significant main effect of Condition, $F(1,57)=33.00, p<.001, \eta^2_p=.37$, where accuracy to RW was higher than to PH1 when averaging across groups, but this collapses the Group factor and so is not relevant to the analysis.

6.4.2.1 Response times

Across all groups, fewer than 5% of RT scores were removed due to outliers (4.38% of English, 4.04% of Spanish-English, and 4.87% of Chinese-English). Based on the descriptive statistics for response times in the oLDT conditions (RW, PH1) per group as shown in Table 6, RTs to RW were shortest in the English group and shortest in the Chinese-English group to PH1, while RTs were longest in the Spanish-English group for both conditions.

Table 6: Means and standard deviations for oLDT response times (ms)

Condition	Group	Mean (ms)	SD
	English	581.12	96.38
Real Word	Spanish-English	672.52	111.92
	Chinese-English	609.9	133.75
	English	671.33	116.75
Pseudohomophone1	Spanish-English	781.08	130.75
	Chinese-English	632.76	178.03

The ANOVA on RTs showed significant main effects of Condition, $F(1,57)=54.92, p < .001, \eta^2_p=.49$, and Group, $F(2,57)=4.564, p=.014, \eta^2_p=.14$, as well as a significant Condition x Group interaction, $F(2,57)=6.83, p=.002, \eta^2_p=.19$. Pairwise post-hoc comparisons showed response times to RW were faster than to PH1 in both English, ($p<.001$), and Spanish-English, ($p<.001$), but no significant difference in Chinese-English. Between groups, post-hocs showed response times were only faster in Chinese-English than Spanish-English to PH1 ($p=.009$).

6.4.2.2 ERPs

The following sections document left and right occipital P1 (P1-O), occipitotemporal P1 (P1-OT) and N170 (N170-OT), and frontal-central N100 (N100-FC) ERPs from the English, Spanish-English, and Chinese-English groups to the oLDT conditions (RW,PH1). See Figure 8 (5.6.2.2) for the electrode cluster layout.

6.4.2.2.1 Occipital P1 (P1-O)

Figure 10 depicts occipital ERP data at 100ms to oLDT conditions.

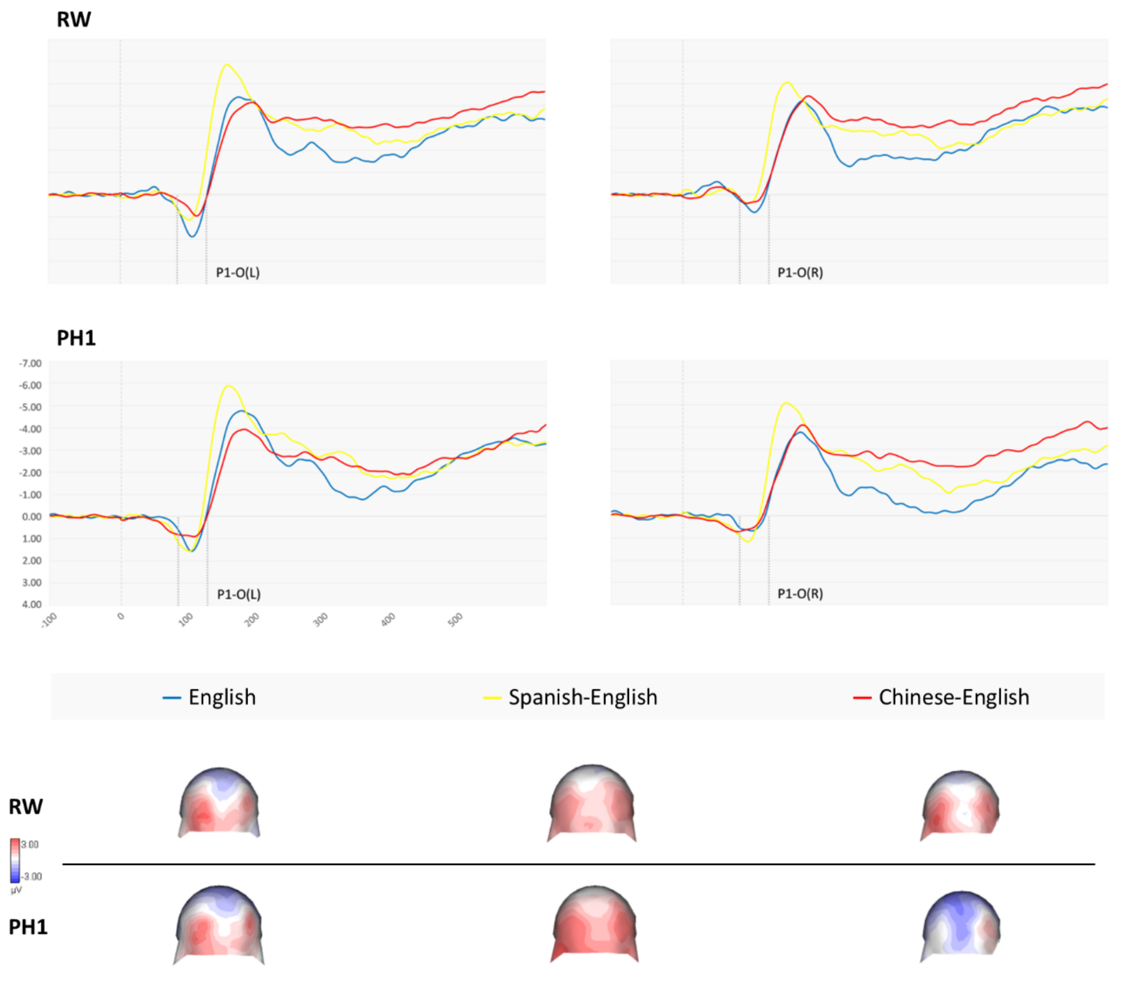

Figure 10: Event-related potentials and topographic voltage maps of left and right occipital clusters at 100ms for oLDT conditions (RW,PH1) across groups

6.4.2.2.1.1 P1-O amplitude

Table 7 shows descriptive statistics for P1-O amplitude across groups for the oLDT conditions.

Table 7: Means (SDs) of P1-O amplitude (µV) in oLDT

		English	Spanish-English	Chinese-English
Real Word	Left	0.86 (0.55)	0.76 (0.66)	0.96 (0.60)
Pseudohomophone1		0.69 (0.48)	1.18 (0.87)	0.85 (0.62)
Real Word	Right	0.66 (0.41)	0.54 (0.52)	0.71 (0.55)

Pseudohomophone1		0.61 (0.35)	0.83 (0.49)	0.55 (0.31)

The oLDT Condition x Hemisphere x Group ANOVA on P1-O amplitude showed no significant main effects, but a significant interaction of oLDT Condition x Group, $F(2,57)=3.74, p=.03$, $\eta^2_p=0.12$. However, no post-hoc tests showed significance after Holm-Bonferroni corrections were applied to control FWER.

6.4.2.2.1.2 P1-O latency

Table 8 shows descriptive statistics for P1-O latency across groups for the oLDT conditions.

Table 8: Means (SDs) of P1-O latency (ms) in oLDT

		English	Spanish-English	Chinese-English
Real Word	Left	100.06 (7.12)	96.16 (8.08)	103.12 (5.23)
Pseudohomophone1		99.31 (2.88)	94.79 (6.67)	99.02 (7.77)
Real Word	Right	97.79 (6.31)	96.42 (7.80)	98.75 (8.75)
Pseudohomophone1		97.6 (6.06)	92.61 (4.72)	97.65 (6.38)

The oLDT Condition x Hemisphere x Group ANOVA on P1-O latency revealed significant main effects of oLDT Condition (PH1<RW), $F(1,57)=6.35, p=.015, \eta^2_p=0.1$, Hemisphere (Right<Left), $F(1,57)=8.00, p=.006, \eta^2_p=0.12$, and Group, $F(2,57)=4.83, p=.012, \eta^2_p=0.14$, but no statistically significant interactions. Post-hocs between groups showed P1-O latency (averaging across hemisphere and condition) to be earlier in Spanish-English than in Chinese-English ($p=.013$).

6.4.2.2.2 Occipitotemporal P1 (P1-OT)

Figure 11 depicts occipitotemporal ERP data at 100ms to oLDT conditions.

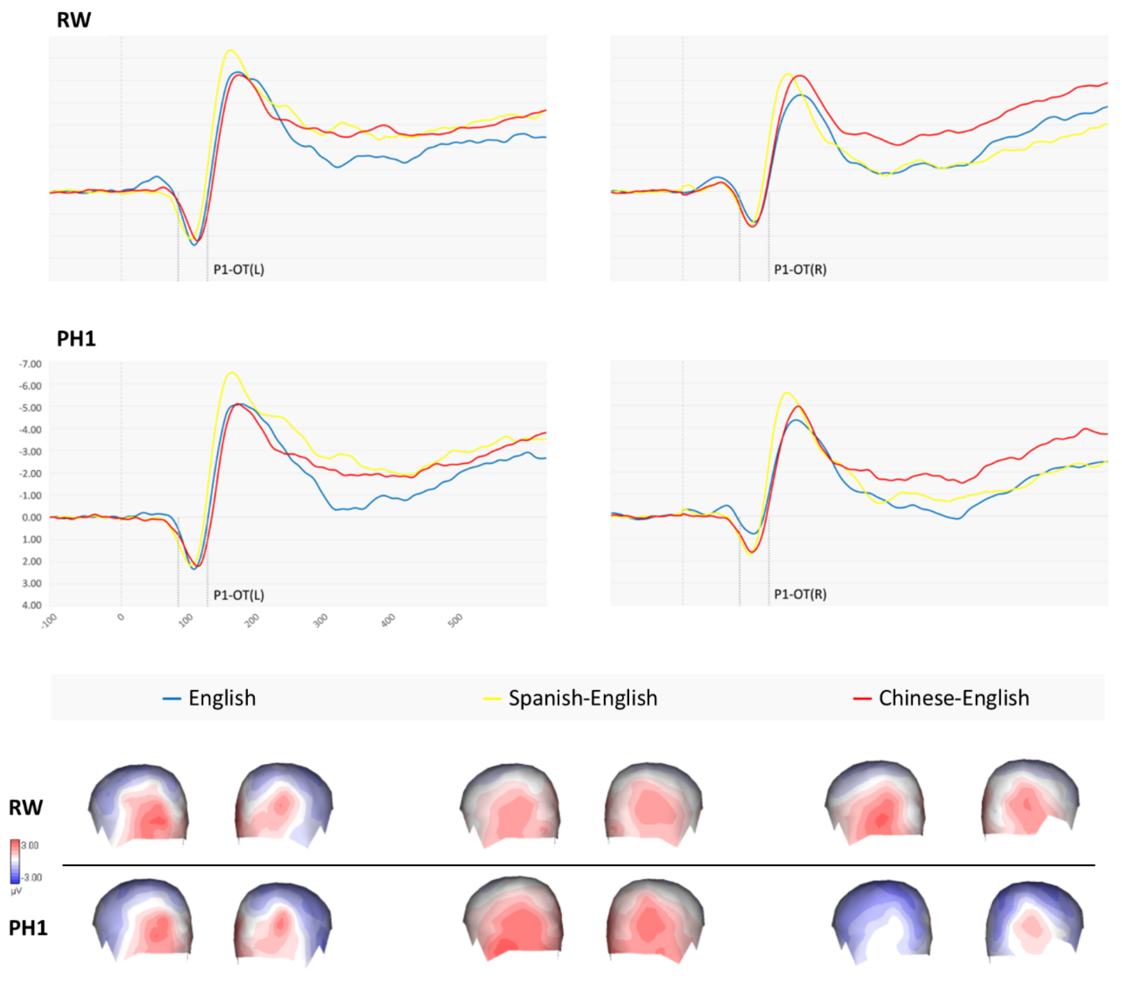

Figure 11: Event-related potentials and topographic voltage maps of left and right occipitotemporal clusters at 100ms oLDT conditions (RW,PH1) across group

6.4.2.2.2.1 P1-OT amplitude

Table 9 shows descriptive statistics for P1-OT amplitude across groups for the oLDT conditions.

Table 9: Means (SDs) of P1-OT amplitude in oLDT (µV)

		English	Spanish-English	Chinese-English
Real Word	Left	0.84 (0.47)	0.78 (0.59)	0.98 (0.60)
Pseudohomophone1		0.64 (0.41)	1.06 (0.76)	0.88 (0.54)
Real Word	Right	0.71 (0.38)	0.85 (0.51)	1 (0.77)
Pseudohomophone1		0.72 (0.38)	0.93 (0.70)	1.01 (0.59)

No significant main effects or interactions were shown by the oLDT Condition x Hemisphere x Group ANOVA on P1-OT amplitude.

6.4.2.2.2.2 P1-OT latency

Table 10 shows descriptive statistics for P1-OT latency across groups for the oLDT conditions.

Table 10: Means (SDs) of P1-OT latency in oLDT (ms)

		English	Spanish-English	Chinese-English
Real Word	Left	101.61 (5.93)	98.08 (6.27)	103.42 (4.82)
Pseudohomophone1		102.36 (3.60)	98.53 (4.72)	102.75 (5.84)
Real Word	Right	98.62 (5.11)	96.02 (5.19)	97.93 (5.61)
Pseudohomophone1		101.34 (5.52)	94.12 (6.12)	99.07 (4.37)

The oLDT Condition x Hemisphere x Group ANOVA on P1-OT latency revealed significant main effects of Hemisphere (Right<Left), $F(1,57)=15.22, p<.001, \eta^2_p=0.21$, and Group, $F(2,57)=10.2, p<.001, \eta^2_p=0.26$, but no significant interactions. Post-hocs between groups showed P1-OT latency (averaging across hemisphere and condition) to be earlier in Spanish-English than in English (*p*<.001) and Chinese-English (*p*<.001).

6.4.2.2.3 Frontal-central N100 (N100-FC)

Figure 12 depicts frontal-central ERP data at 100ms to oLDT conditions.

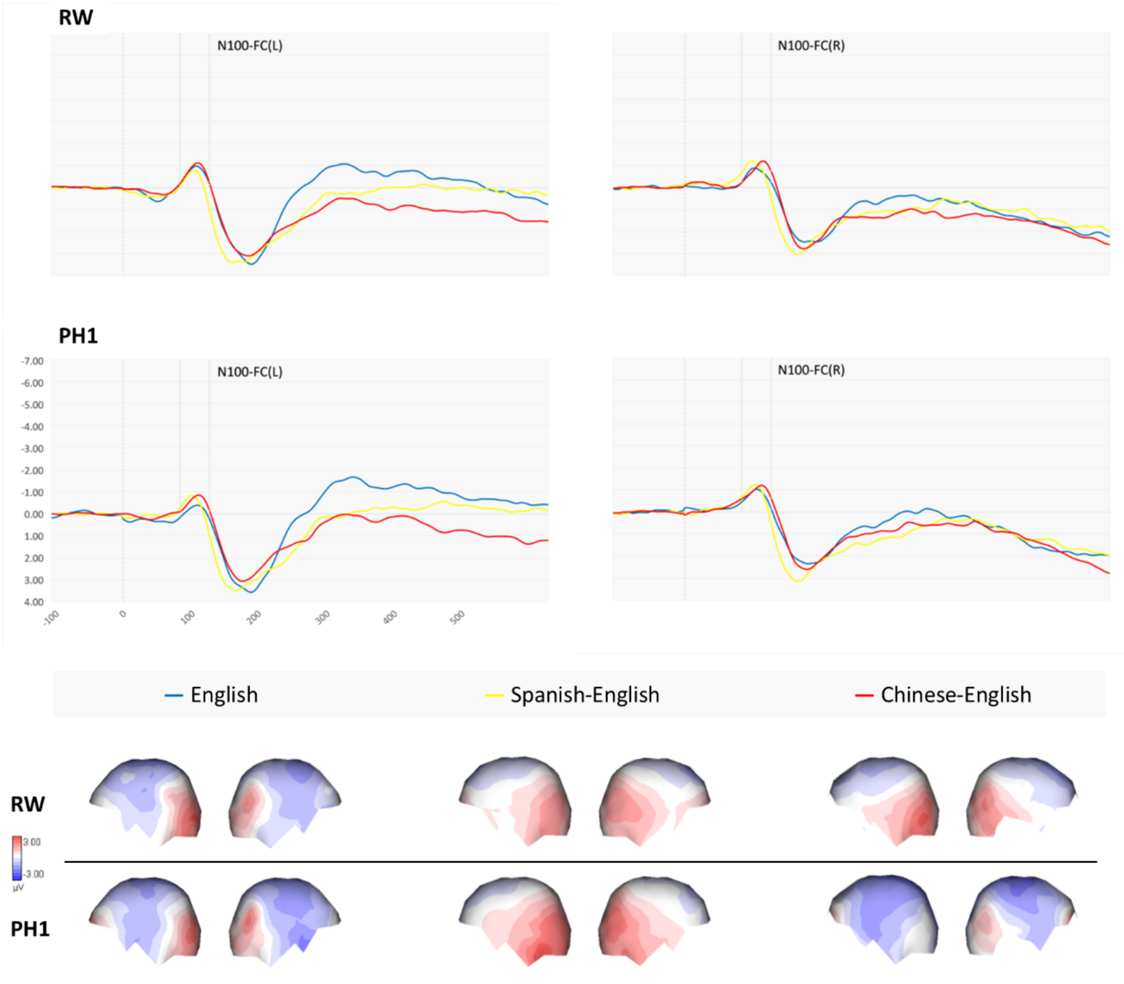

Figure 12: Event-related potentials and topographic voltage maps of left and right frontal-central clusters at 100ms for oLDT conditions (RW,PH1) across groups

6.4.2.2.3.1 N100-FC amplitude

Table 11 shows descriptive statistics for N100-FC amplitude across groups for the oLDT conditions.

Table 11: Means (SDs) of N100-FC amplitude in oLDT (µV)

		English	Spanish-English	Chinese-English
Real Word	Left	-0.54 (0.31)	-0.56 (0.28)	-0.61 (0.41)
Pseudohomophone1		-0.45 (0.28)	-0.61 (0.38)	-0.67 (0.44)
Real Word	Right	-0.55 (0.30)	-0.49 (0.25)	-0.66 (0.37)
Pseudohomophone1		-0.57 (0.32)	-0.66 (0.39)	-0.76 (0.51)

No significant main effects or interactions were shown by the oLDT Condition x Hemisphere x Group ANOVA on N100-FC amplitude.

6.4.2.2.3.2 N100-FC latency

Table 12 shows descriptive statistics for N100-FC latency across groups for the oLDT conditions.

Table 12: Means (SDs) of N100-FC latency in oLDT (ms)

		English	Spanish-English	Chinese-English
Real Word	Left	101.35 (5.97)	96.39 (7.57)	102.18 (4.40)
Pseudohomophone1		103.75 (6.93)	96.44 (7.05)	101.03 (4.11)
Real Word	Right	98.89 (7.23)	95.64 (5.95)	103.11 (4.02)
Pseudohomophone1		100.38 (5.33)	94.01 (7.02)	100.43 (6.08)

The oLDT Condition x Hemisphere x Group ANOVA on N100-FC latency showed main effects of Hemisphere (Right<Left), $F(1,57)=6.1, p=.017, \eta^2_p=0.1$, and Group, $F(2,57)=9.79, p<.001, \eta^2_p=0.26$, as well as a significant interaction of oLDT Condition x Group, $F(2,57)=4.78, p=.012, \eta^2_p=0.14$. Averaging across hemispheres, N100-FC latency to RW was earlier in Spanish-English than in Chinese-English ($p=.002$), while latency to

PH1 was earlier in Spanish-English than both English (p=.001) and Chinese-English (p=.015).

6.4.2.2.4 Occipitotemporal N170 (N170-OT)

Figure 13 depicts occipitotemporal ERP data at 100ms to oLDT conditions.

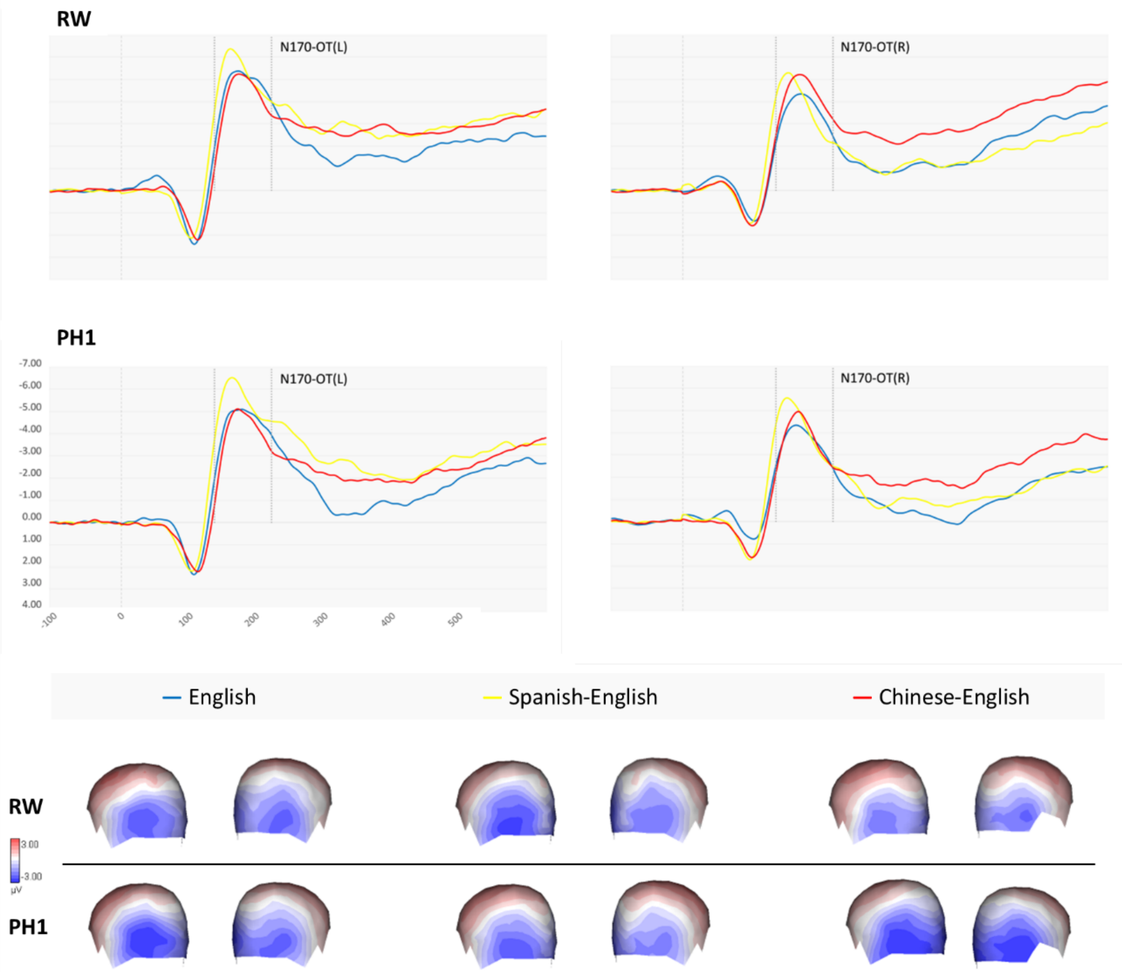

Figure 13: Event-related potentials and topographic voltage maps of left and right occipitotemporal clusters at 170ms for oLDT conditions (RW,PH1) across groups

6.4.2.2.4.1 N170-OT amplitude

Table 13 shows descriptive statistics for N170-OT amplitude across groups for the oLDT conditions.

Table 13: Means (SDs) of N170-OT amplitude in oLDT (µV)

		English	Spanish-English	Chinese-English
Real Word	Left	-1.91 (0.94)	-2.28 (1.42)	-1.7 (0.72)
Pseudohomophone1		-2.43 (1.01)	-2.79 (1.48)	-2.24 (1.49)
Real Word	Right	-1.78 (0.69)	-2.04 (1.35)	-2.2 (1.1)
Pseudohomophone1		-1.83 (0.95)	-2.1 (1.18)	-2.31 (1.53)

The oLDT Condition x Hemisphere x Group ANOVA on N170-OT amplitude showed a significant main effect of oLDT Condition (PH1>RW), $F(1,57)=9.19, p=.004, \eta^2_p=0.14$, and a significant interaction of Hemisphere x oLDT Condition, $F(1,2)=9.76, p=.003, \eta^2_p=0.15$, neither of which involve the Group factor so will not be investigated further.

6.4.2.2.4.2 N170-OT latency

Table 14 shows descriptive statistics for N170-OT latency across groups for the oLDT conditions.

Table 14: Means (SDs) of N170-OT latency in oLDT (ms)

		English	Spanish-English	Chinese-English
Real Word	Left	173.1 (7.15)	164.49 (7.63)	170.39 (7.14)
Pseudohomophone1		169.95 (6.09)	165.83 (6.40)	170.21 (7.54)
Real Word	Right	170.07 (8.34)	157.69 (8.69)	168.50 (6.96)
Pseudohomophone1		167.46 (8.81)	160.41 (8.03)	166.83 (5.55)

The oLDT Condition x Hemisphere x Group ANOVA on N170-OT latency revealed significant main effects of Hemisphere (Right<Left), $F(1,57)=15.53, p<.001, \eta^2_p=0.21$, and Group, $F(2,57)=11.88, p<.001, \eta^2_p=0.29$, as well as a significant interaction of oLDT Condition x Group, $F(2,57)=6.12, p=.004, \eta^2_p=0.18$. Averaging across hemispheres, post-hocs between groups showed N170-OT latency to RW to be earlier in Spanish-English than in both English ($p<.001$) and Chinese-English ($p<.001$). N170-OT latency to RW (irrespective of hemisphere) was also earlier than to PH1 in English ($p=.044$).

6.4.2.3 Summary

Faster response times to RW than to PH1, reflecting effects of orthographic lexicality, were observed in both English and Spanish-English (but not in Chinese-English), which links with the bilateral N170-OT being earlier to RW than to PH1 in English. Between groups, overall accuracy was higher in English than both Spanish-English and Chinese-English and also higher in Spanish-English than in Chinese-English, while response times were only faster in Chinese-English than Spanish-English (to PH1). Also potentially linked with behavioural results, bilateral N170-OT was earlier in Spanish-English than in both English and Chinese-English to RW. Lastly, bilateral N100-FC was earlier in Spanish-English than in Chinese-English (but not English) to RW, while earlier in Spanish-English than both English and Chinese-English to PH1, indicating differences in processing between groups.

6.4.3 Phonological lexical decision task (pLDT)

6.4.3.1 Accuracy

Based on the descriptive statistics for pLDT accuracy shown in Table 15, the same pattern was observed between groups as in the oLDT with accuracy being higher in the English group than in the Spanish-English group and higher in the Spanish-English group than the Chinese-English group for both pLDT conditions (PW, PH2). However, differences in accuracy between conditions were not consistent, as accuracy was higher for PW than PH2 in the English group, but higher for PH2 than PW in both Spanish-English and Chinese-English groups. The notably high standard deviations, such as for both PH2 and PW in the Chinese-English group and for PW in the Spanish-English group, are simply due to a wider spread of accuracy scores within the groups.

Table 15: Means and standard deviations for pLDT accuracy

Condition	Group	Mean (% correct)	SD
Pseudohomophone2	English	84.12	8.12
	Spanish-English	71.23	10
	Chinese-English	66.75	14.7
Pseudoword	English	92.08	6.44
	Spanish-English	64.08	12.62
	Chinese-English	62.25	17.20

The Condition x Group ANOVA on pLDT accuracy showed a significant main effect of Group, $F(2,57)=41.83, p<.001, \eta^2_p=.60$, and a significant Condition x Group interaction, $F(2,57)=4.8, p=.012, \eta^2_p=.14$. Pairwise post-hoc comparisons showed English accuracy to PH2 was significantly higher than in both Spanish-English, ($p=.011$), and Chinese-English, ($p<.001$). Likewise, English accuracy to PW was also higher than in both Spanish-English,

(p<.001), and Chinese-English, (p<.001). However, post-hocs showed no significant differences in accuracy between Spanish-English and Chinese-English or between conditions within any group.

6.4.3.2 Response times

Across all groups, around 5% of RT scores were removed due to outliers (5.08% of English, 3.26% of Spanish-English, and 4.86% of Chinese-English). Based on the descriptive statistics for pLDT response times shown in Table 16, RTs to PH2 were shorter in the English group than in the Spanish-English group, but only marginally shorter than in the Chinese-English group, while RTs to PW were clearly shorter in the English group than in the Spanish-English group and shorter in the Spanish-English group than in the Chinese-English group. Between conditions, RTs were shorter to PH2 than PW in both English and Spanish-English groups, but shorter to PW than PH2 in the Chinese-English group.

Table 16: Means and standard deviations for pLDT response times (ms)

Condition	Group	Mean (ms)	SD
Pseudohomophone2	English	702.36	126.64
	Spanish-English	787.51	152.01
	Chinese-English	703.94	219.33
Pseudoword	English	886.16	112.65
	Spanish-English	812.36	242.53
	Chinese-English	689.69	234.01

The Condition x Group ANOVA on pLDT RTs showed a significant main effect of Condition (PH2<PW), $F(1,57)=5.62, p=.021, \eta^2_p=.09$, and a significant Condition x Group interaction,

$F(2,57)=4.91, p=.011, \eta^2_p=.15$. Pairwise post-hoc comparisons showed responses to PW were faster in Chinese-English than in English, ($p=.014$), while the only difference between conditions was that English responses to PH2 were faster than to PW ($p=.003$).

6.4.3.3 ERPs

The following sections document left and right occipital P1 (P1-O), occipitotemporal P1 (P1-OT) and N170 (N170-OT), and frontal-central N100 (N100-FC) ERPs from the English, Spanish-English, and Chinese-English groups to the pLDT conditions (PW,PH2). See Figure 8 (5.6.2.2) for the electrode cluster layout.

6.4.3.3.1 Occipital P1 (P1-O)

Figure 14 depicts occipital ERP data at 100ms to pLDT conditions.

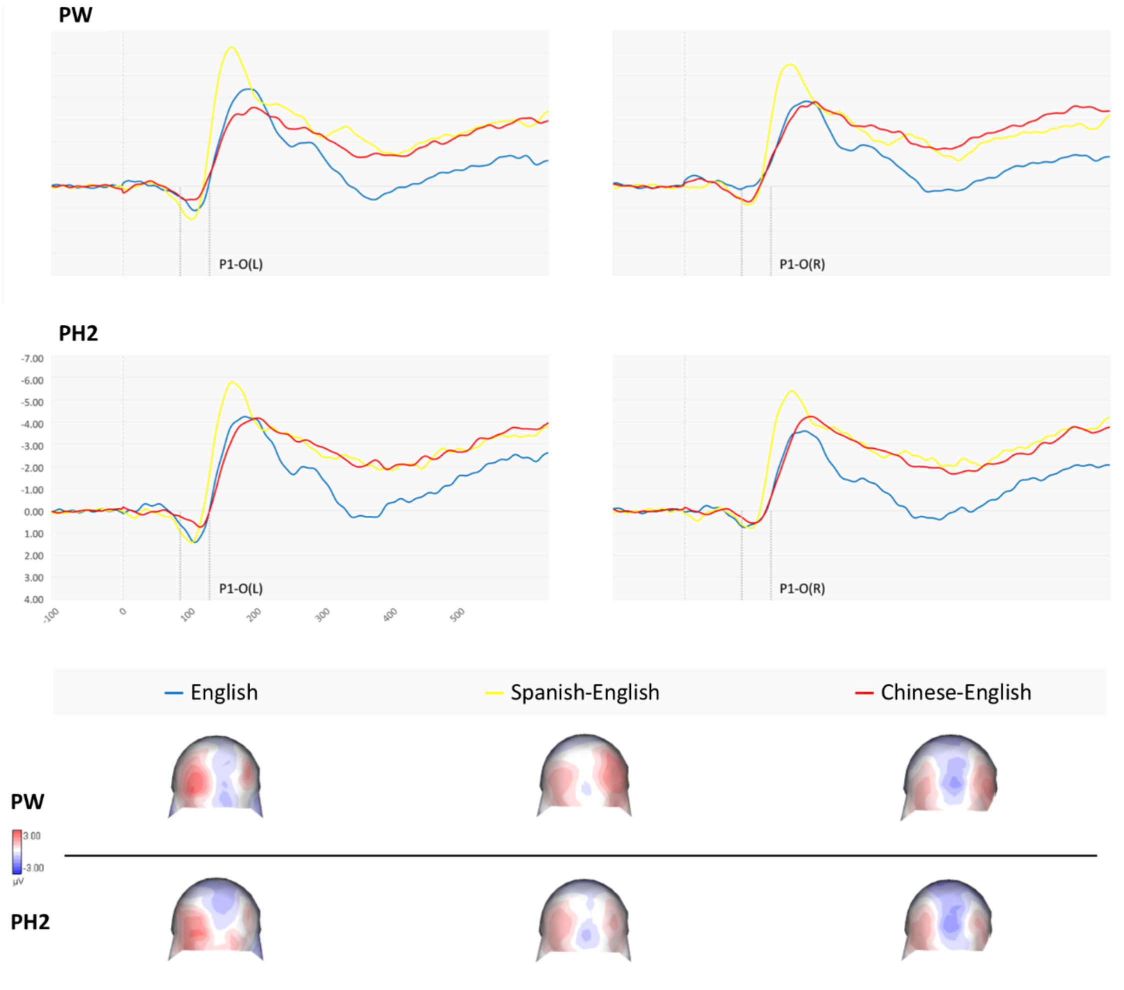

Figure 14: Event-related potentials and topographic voltage maps of left and right occipital clusters at 100ms for pLDT conditions (PW,PH2) across groups

6.4.3.3.1.1 P1-O amplitude

Table 17 shows descriptive statistics for P1-O amplitude across groups for the pLDT conditions.

Table 17: Means (SDs) of P1-O amplitude in pLDT (µV)

		English	Spanish-English	Chinese-English
Pseudoword	Left	0.74 (0.47)	0.79 (0.57)	0.76 (0.54)
Pseudohomophone2		0.69 (0.56)	0.61 (0.46)	0.71 (0.4)
Pseudoword	Right	0.45 (0.35)	0.76 (0.47)	0.62 (0.46)

Pseudohomophone2		0.62 (0.39)	0.67 (0.35)	0.66 (0.35)

The pLDT Condition x Hemisphere x Group ANOVA on P1-O amplitude showed no significant main effects or interactions.

6.4.3.3.1.2 P1-O latency

Table 18 shows descriptive statistics for P1-O latency across groups for the pLDT conditions.

Table 18: Means (SDs) of P1-O latency in pLDT (ms)

		English	Spanish-English	Chinese-English
Pseudoword	Left	102.13 (8.14)	94.04 (6.73)	98.41 (7.30)
Pseudohomophone2		98.77 (6.72)	94.94 (9.07)	99.82 (7.21)
Pseudoword	Right	97.75 (8.67)	90.75 (6.76)	93.04 (7.40)
Pseudohomophone2		98.63 (7.04)	93.25 (5.71)	95.23 (8.16)

The pLDT Condition x Hemisphere x Group ANOVA on P1-O latency showed significant main effects of Hemisphere (Right<Left), $F(1,57)=26.72, p<.001, \eta^2_p=0.32$, and Group, $F(2,57)=5.59, p=.006, \eta^2_p=0.16$, as well as a significant interaction of pLDT Condition x Hemisphere, $F(1,2)=6.57, p=.013, \eta^2_p=0.1$. Averaging across conditions (PW,PH2) and hemispheres, the only significant difference was that P1-O latency was earlier in Spanish-English than English in the pLDT overall ($p=.004$).

6.4.3.3.2 Occipitotemporal P1 (P1-OT)

Figure 15 depicts occipitotemporal ERP data at 100ms to pLDT conditions.

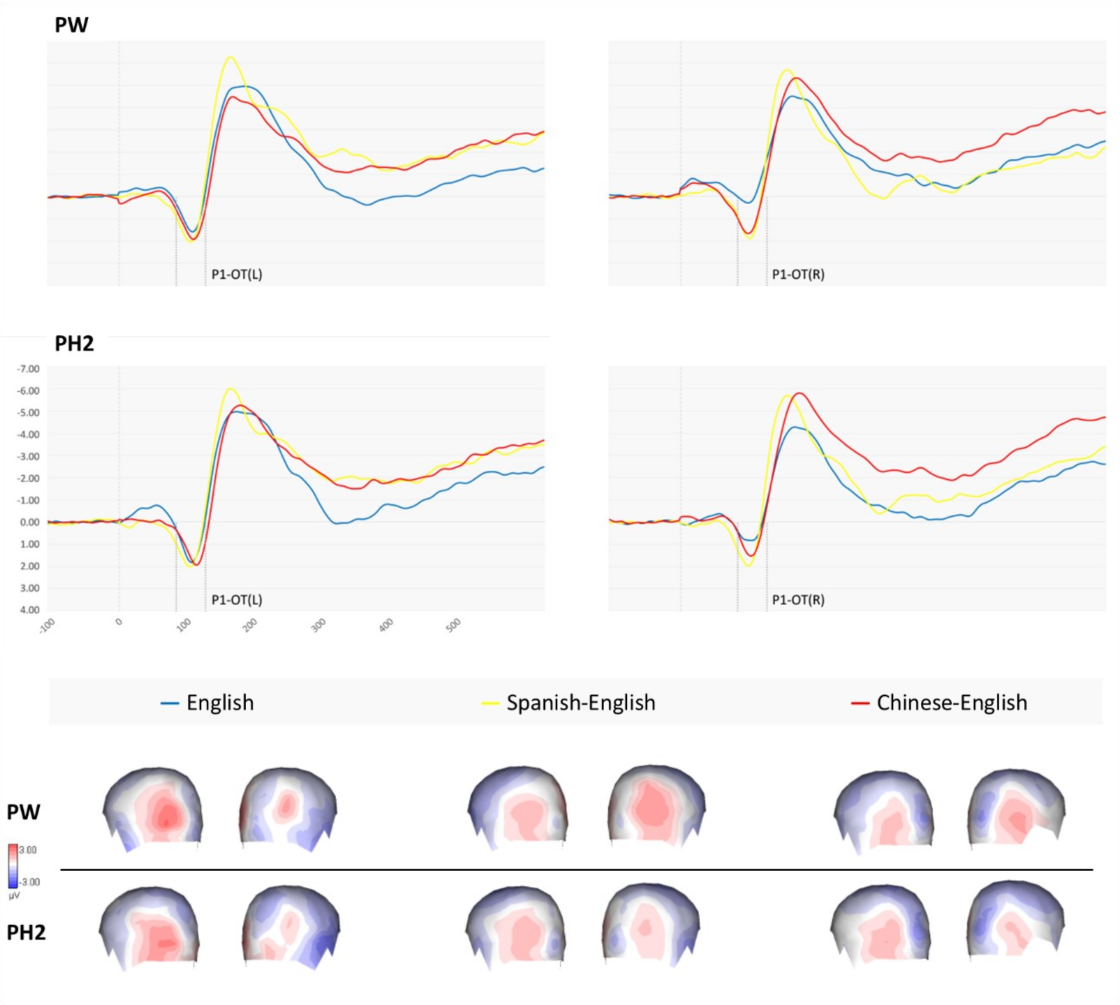

Figure 15: Event-related potentials and topographic voltage maps of left and right occipitotemporal clusters at 100ms for pLDT conditions (PW,PH2) across groups

6.4.3.3.2.1 P1-OT amplitude

Table 19 shows descriptive statistics for P1-OT amplitude across groups for the pLDT conditions.

Table 19: Means (SDs) of P1-OT amplitude in pLDT (µV)

		English	Spanish-English	Chinese-English
Pseudoword	Left	0.76 (0.51)	0.82 (0.47)	0.68 (0.4)
Pseudohomophone2		0.74 (0.6)	0.71 (0.54)	0.86 (0.45)
Pseudoword	Right	0.52 (0.2)	1.36 (0.9)	0.91 (0.45)
Pseudohomophone2		0.41 (0.28)	0.81 (0.38)	0.82 (0.43)

The pLDT Condition x Hemisphere x Group ANOVA on P1-OT amplitude showed significant main effects of pLDT Condition (PW>PH2), $F(1,57)=6.08, p=.017$, $\eta^2_p=0.1$, and Group, $F(2,57)=4.34, p=.018, \eta^2_p=0.13$, as well as significant interactions of pLDT Condition x Group, $F(2,57)=5.61, p=.006, \eta^2_p=0.16$, Hemisphere x pLDT Condition, $F(1,2)=8.63, p=.005$, $\eta^2_p=0.13$, and Hemisphere x Group, $F(2,57)=7.9, p<.001, \eta^2_p=0.22$. Post-hocs for the pLDT x Group interaction (averaging across hemispheres) showed that P1-OT amplitude was larger to PW in English than Spanish-English ($p<.001$), while larger to PW than PH2 in Spanish-English ($p=.002$). Averaging across conditions (PW,PH2), post-hocs for the Hemisphere x Group interaction showed P1-OT(R) was smaller in English than in both Spanish-English right ($p<.001$), and Chinese-English ($p=.004$), overall in the pLDT.

6.4.3.3.2.2 P1-OT latency

Table 20 shows descriptive statistics for P1-OT latency across groups for the pLDT conditions.

Table 20: Means (SDs) of P1-OT latency in pLDT (ms)

		English	Spanish-English	Chinese-English

Pseudoword	Left	103.51 (5.14)	99.94 (6.13)	100.09 (6.08)
Pseudohomophone2		102.14 (5.7)	98.7 (6.82)	101.75 (5.95)
Pseudoword	Right	98.63 (5.5)	94.57 (6.01)	96.34 (5.6)
Pseudohomophone2		99.89 (5.65)	95.22 (6.07)	99.13 (7.16)

The pLDT Condition x Hemisphere x Group ANOVA on P1-OT latency showed significant main effects of Hemisphere (Right<Left), $F(1,57)=21.79, p<.001, \eta^2_p=0.28$, and Group, $F(2,57)=4.92, p=.011, \eta^2_p=0.15$. Averaging across conditions (PW and PH2) and hemispheres, post-hocs showed P1-OT was earlier in Spanish-English than English ($p=.008$).

6.4.3.3.3 Frontal-central N100 (N100-FC)

Figure 16 depicts frontal-central ERP data at 100ms to pLDT conditions.

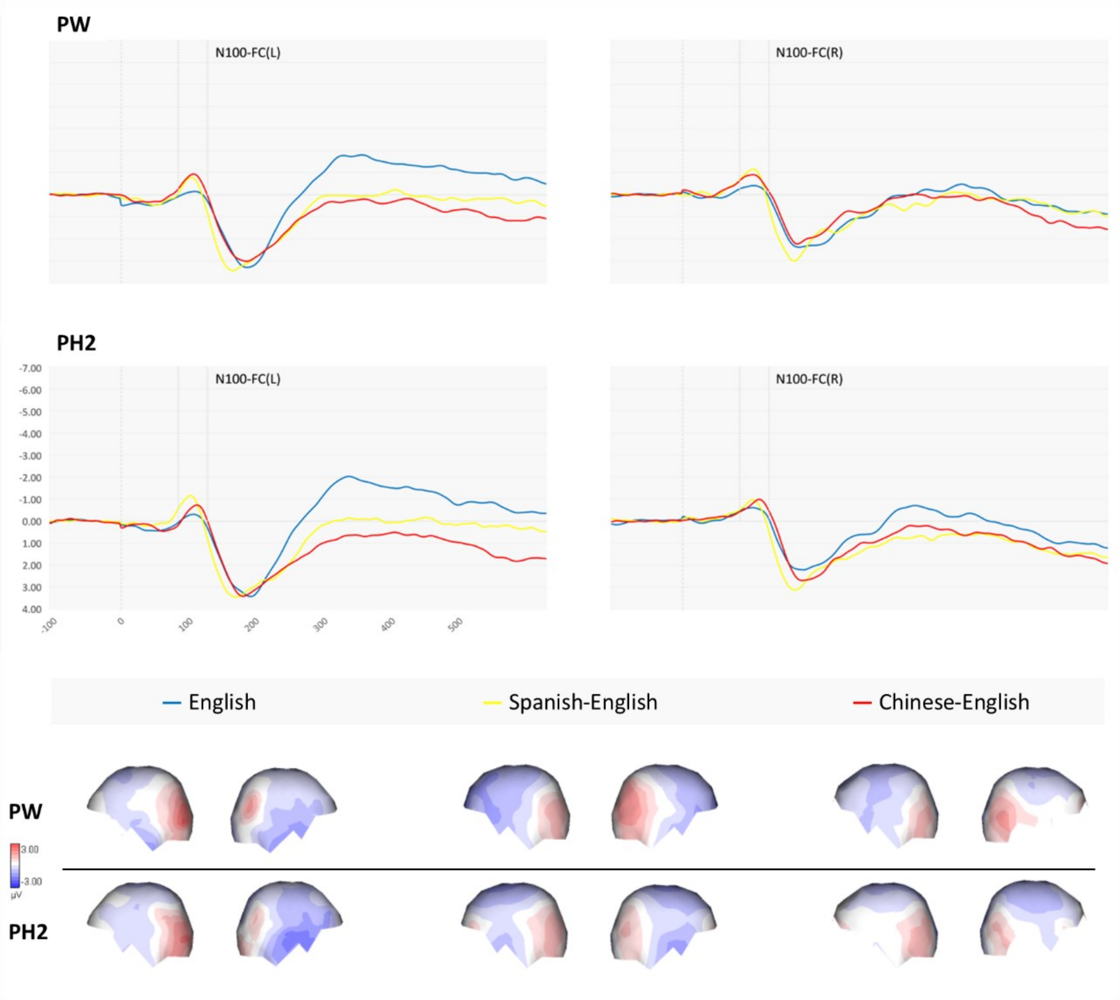

Figure 16: Event-related potentials and topographic voltage maps of left and right frontal-central clusters at 100ms for pLDT conditions (PW,PH2) across groups

6.4.3.3.3.1 N100-FC amplitude

Table 21 shows descriptive statistics for N100-FC amplitude across groups for the pLDT conditions.

Table 21: Means (SDs) of N100-FC amplitude in pLDT (µV)

		English	Spanish-English	Chinese-English
Pseudoword	Left	-0.45 (0.3)	-0.94 (0.62)	-0.63 (0.45)
Pseudohomophone2		-0.34 (0.17)	-0.57 (0.34)	-0.56 (0.33)
Pseudoword	Right	-0.34 (0.17)	-0.8 (0.49)	-0.54 (0.33)

Pseudohomophone2		-0.52 (0.43)	-0.71 (0.43)	-0.68 (0.37)

The pLDT Condition x Hemisphere x Group ANOVA on N100-FC amplitude showed a significant main effect of Group, $F(2,57)=6.86, p=.002, \eta^2_p=0.19$, pLDT Condition x Hemisphere interaction, $F(1,2)=12.21, p<.001, \eta^2_p=0.18$, and pLDT Condition x Group interaction, $F(2,57)=5.42, p=.007, \eta^2_p=0.16$. Averaging across hemispheres, post-hocs showed N100-FC amplitude was larger in Spanish-English than English ($p<.001$), while larger to PW than PH2 in Spanish-English ($p=.013$).

6.4.3.3.3.2 N100-FC latency

Table 22 shows descriptive statistics for N100-FC latency across groups for the pLDT conditions.

Table 22: Means (SDs) of N100-FC latency in pLDT (ms)

		English	Spanish-English	Chinese-English
Pseudoword	Left	97.51 (7.88)	96.54 (6.1)	101.47 (5.15)
Pseudohomophone2		101.26 (6.01)	97.14 (7.18)	100.35 (7.38)
Pseudoword	Right	100.87 (6.89)	95.27 (6.05)	95.36 (6.07)
Pseudohomophone2		99.38 (8.69)	95.35 (5.3)	99.82 (6.17)

The pLDT Condition x Hemisphere x Group ANOVA on N100-FC latency showed a significant main effect of Group, $F(2,57)=3.85, p=.027, \eta^2_p=0.12$, and a significant interaction of pLDT Condition x Hemisphere x Group, $F(2,57)=5.24, p=.008, \eta^2_p=0.16$. Pairwise post-hoc comparisons revealed only that N100-FC(R) was significantly earlier than N100-FC(L) in Chinese-English to PW, ($p=.037$).

6.4.3.3.4 Occipitotemporal N170 (N170-OT)

Figure 17 depicts occipitotemporal ERP data at 170ms to pLDT conditions.

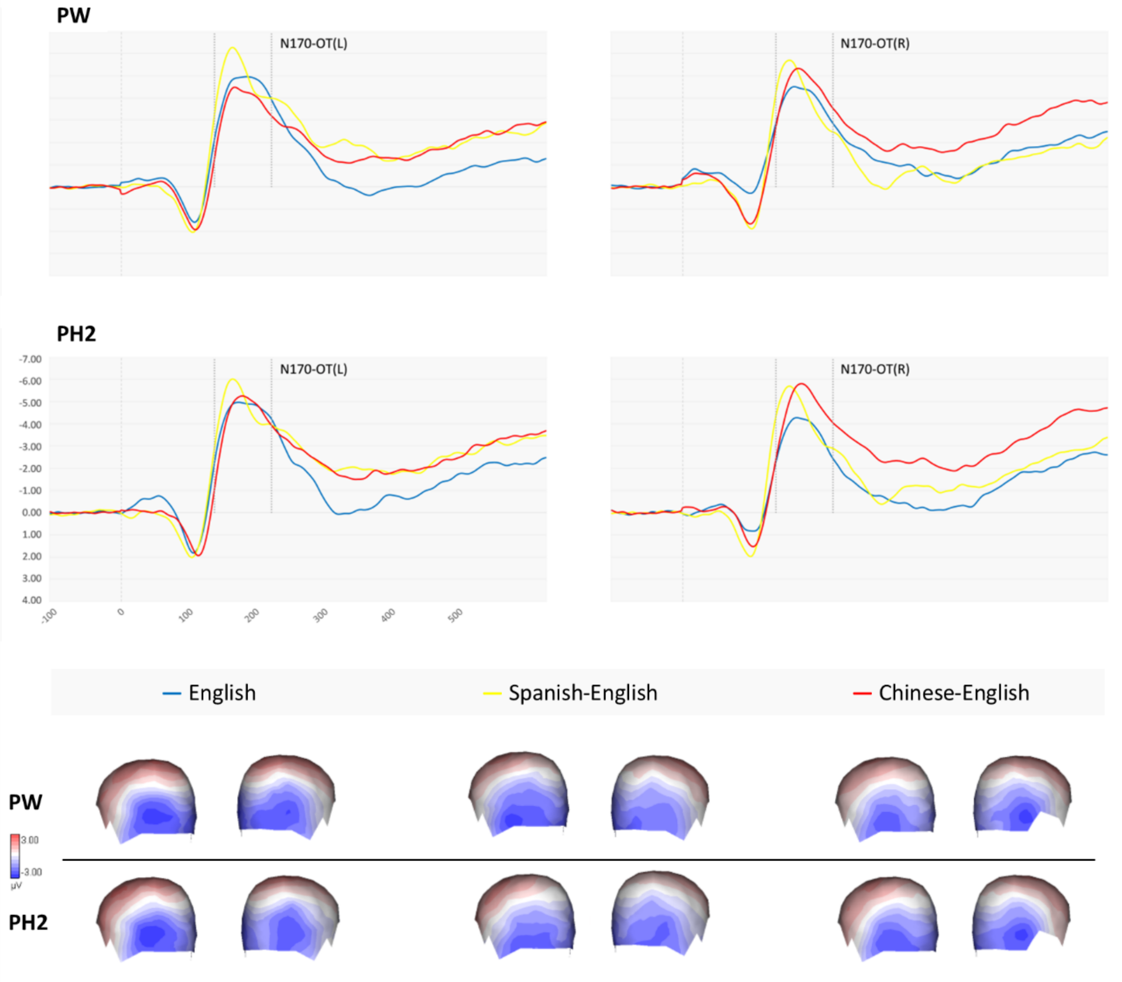

Figure 17: Event-related potentials and topographic voltage maps of left and right occipitotemporal clusters at 170ms for pLDT conditions (PW,PH2) across groups

6.4.3.3.4.1 N170-OT amplitude

Table 23 shows descriptive statistics for N170-OT amplitude across groups for the pLDT conditions.

Table 23: Means (SDs) of N170-OT amplitude in pLDT (µV)

		English	Spanish-English	Chinese-English
Pseudoword	Left	-1.95 (0.79)	-2.13 (1.26)	-1.75 (0.64)
Pseudohomophone2		-2.05 (0.83)	-1.93 (1.17)	-1.8 (0.53)
Pseudoword	Right	-2.12 (0.91)	-2.01 (1.45)	-2.09 (0.94)
Pseudohomophone2		-2.01 (0.86)	-1.94 (1.38)	-2.16 (1.02)

The pLDT Condition x Hemisphere x Group ANOVA on N170-OT amplitude showed no significant main effects or interactions.

6.4.3.3.4.2 N170-OT latency

Table 24 shows descriptive statistics for N170-OT amplitude across groups for the pLDT conditions.

Table 24: Means (SDs) of N170-OT latency in pLDT (ms)

		English	Spanish-English	Chinese-English
Pseudoword	Left	172.73 (7.25)	165.62 (4.7)	168.57 (9.42)
Pseudohomophone2		171.63 (8.7)	167.07 (7.78)	171.32 (8.52)
Pseudoword	Right	166.8 (6.97)	158.4 (8.92)	167.12 (10.5)
Pseudohomophone2		166.53 (6.89)	161.63 (9.79)	167.97 (5.65)

The pLDT Condition x Hemisphere x Group ANOVA on N170-OT latency showed only significant main effects of Hemisphere (Right<Left), $F(1,57)=21.99, p<.001, \eta^2_p=0.28$, and Group, $F(2,57)=6.26, p=.003, \eta^2_p=0.18$. Averaging across conditions and hemispheres, post-hocs between groups showed N170-OT was earlier in Spanish-English than both English ($p=.006$) and Chinese-English ($p=.015$).

6.4.3.4 Summary

Faster response times to PH2 than to PW showed a phonological lexicality effect in English, but in neither ESL group. Also showing a phonological lexicality effect, bilateral P1-OT and N100-FC were both larger to PW than PH2 in Spanish-English. Furthermore, bilateral N100-FC was larger in Spanish-English than English, reflecting a group distinction between alphabetic L1 readers. As expected, accuracy was higher in English than both Spanish-English and Chinese-English to both PH2 and PW, but not significantly different between Spanish-English and Chinese-English. However, response times were faster in Chinese-English than in English, while bilateral P1-OT was larger in English than Spanish-English to PW.

6.4.4 Task effects: oLDT (PH1) vs pLDT (PH2)

Overall oLDT and pLDT measures (averaged across conditions within tasks) and the two pseudohomophone conditions from each task (PH1 and PH2 from oLDT and pLDT, respectively) were further analyzed to observe any task-level effects. The levels of the Condition factor in the following Task analyses are oLDT (RW and PH1 collapsed) and pLDT (PW and PH2 collapsed), while the additional pairwise comparisons for the Pseudohomophone factor used PH1 and PH2.

6.4.4.1 Accuracy

Table 25 shows the descriptive statistics for accuracy across tasks per group.

Table 25: Means and standard deviations for accuracy (% correct) across tasks

Condition	Group	Mean	SD
	English	95.941	2.67
oLDT	Spanish-English	91.588	4.707
	Chinese-English	87.069	7.47
	English	88.232	5.975
pLDT	Spanish-English	67.656	6.926
	Chinese-English	64.502	12.324

The Task x Group ANOVA for accuracy showed main effects of Task (oLDT>pLDT), $F(1,57)=575.60, p<.001, \eta^2_p=0.91$, and Group, $F(2,57)= 32.33, p<.001, \eta^2_p=0.53$, as well as a significant Task x Group interaction, $F(2,57)=47.58, p<.001, \eta^2_p=0.63$. Pairwise post-hoc comparisons showed accuracy to oLDT was higher than to pLDT in English, ($p<.001$), Spanish-English, ($p<.001$), and Chinese-English, ($p<.001$). Between groups, accuracy to oLDT was significantly higher in English than Chinese-English ($p=.003$), while accuracy to pLDT was significantly higher in English than both Spanish-English ($p<.001$) and Chinese-English ($p<.001$). The Pseudohomophone (PH1,PH2) x Group ANOVA, however, did not show a significant interaction.

6.4.4.2 Response times

Table 26 shows the descriptive statistics for response times across tasks per group. Across all groups, around 5% of RT scores were removed due to outliers (4.76% of English, 4.41% of Spanish-English, and 5.36% of Chinese-English).

Table 26: Means and standard deviations for response times (ms) across tasks

Condition	Group	Mean	SD
oLDT	English	629.161	109.64
	Spanish-English	726.799	113.822
	Chinese-English	621.326	150.815
pLDT	English	797.688	118.622
	Spanish-English	799.935	154.318
	Chinese-English	696.813	189.84

The Task x Group ANOVA for response times showed only a significant main effect of Group, $F(2,57)=3.927, p=.025, \eta^2_p=.12$. Averaging across tasks, post-hocs showed Chinese-English were faster than Spanish-English overall (p=.029).

6.4.4.3 ERPs

The following sections document left and right occipital P1 (P1-O), occipitotemporal P1 (P1-OT) and N170 (N170-OT), and frontal-central N100 (N100-FC) ERPs from the English, Spanish-English, and Chinese-English groups to the pseudohomophone conditions (PH1,PH2) and tasks overall (oLDT,pLDT). See Figure 8 (5.6.2.2) for the electrode cluster layout.

6.4.4.3.1 Occipital P1 (P1-O)

Figure 18 depicts occipital ERP data at 100ms to pseudohomophone conditions.

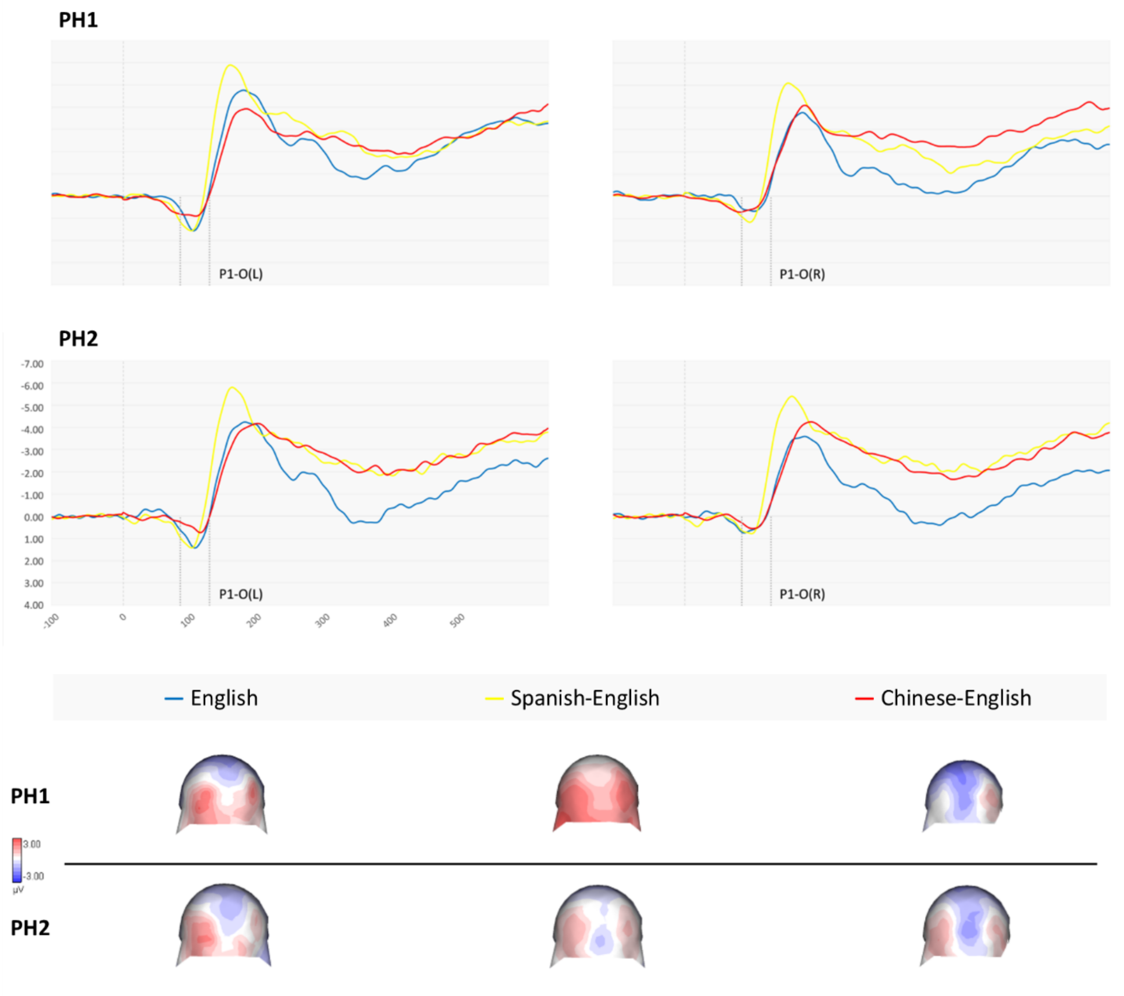

Figure 18: Event-related potentials and topographic voltage maps of left and right occipital clusters at 100ms for task/pseudohomophone comparisons (PH1,PH2) across groups

6.4.4.3.1.1 P1-O amplitude

Table 27 shows descriptive statistics for P1-O amplitude across groups and tasks.

Table 27: Means (SDs) of P1-O amplitude across tasks (µV)

		English	Spanish-English	Chinese-English
oLDT	Left	0.77 (0.46)	0.97 (0.42)	0.91 (0.54)

pLDT		0.71 (0.47)	0.7 (0.45)	0.74 (0.39)
oLDT	Right	0.63 (0.3)	0.68 (0.36)	0.63 (0.4)
pLDT		0.53 (0.33)	0.71 (0.31)	0.64 (0.35)

The Task x Hemisphere x Group ANOVA on P1-O amplitude showed significant main effects of Task (oLDT>pLDT), $F(1,57)=4.67, p=.035, \eta^2_p=0.08$, and Hemisphere (Left>Right), $F(1,57)=14.26, p<.001, \eta^2_p=0.2$, as well as a significant Hemisphere x Task interaction, $F(1,2)=7.54, p=.008, \eta^2_p=0.12$, and the full three-way Task x Hemisphere x Group interaction, $F(2,57)=3.48, p=.038, \eta^2_p=0.11$. However, post-hocs showed no significant pairwise comparisons after the Holm-Bonferroni corrections were applied to control FWER.

6.4.4.3.1.2 P1-O latency

Table 28 shows descriptive statistics for P1-O latency across groups and tasks.

Table 28: Means (SDs) of P1-O latency across tasks (ms)

		English	Spanish-English	Chinese-English
oLDT	Left	99.68 (4.57)	95.48 (6.14)	101.07 (5.76)
pLDT		100.45 (6.65)	94.49 (5.33)	99.12 (5.99)
oLDT	Right	97.69 (5.67)	94.52 (5.41)	98.2 (6.17)
pLDT		98.19 (7.42)	92 (4.62)	94.14 (7.02)

The Task x Hemisphere x Group ANOVA on P1-O latency showed significant main effects of Task (pLDT<oLDT), $F(1,57)=5.94, p=.018, \eta^2_p=0.09$, Hemisphere (Right<Left), $F(1,57)=21.03, p<.001, \eta^2_p=0.27$, and Group, $F(2,57)=5.6, p=.006, \eta^2_p=0.16$, as well as a significant interaction of Task x Group, $F(2,57)=3.57, p=.034, \eta^2_p=0.11$. Averaging across hemispheres, post-hocs showed P1-O latency was earlier in Spanish-English than English to pLDT (*p*=.009) and earlier to pLDT than to oLDT in Chinese-English (*p*=.039).

6.4.4.3.2 Occipitotemporal P1 (P1-OT)

Figure 19 depicts occipitotemporal ERP data at 100ms to pseudohomophone conditions.

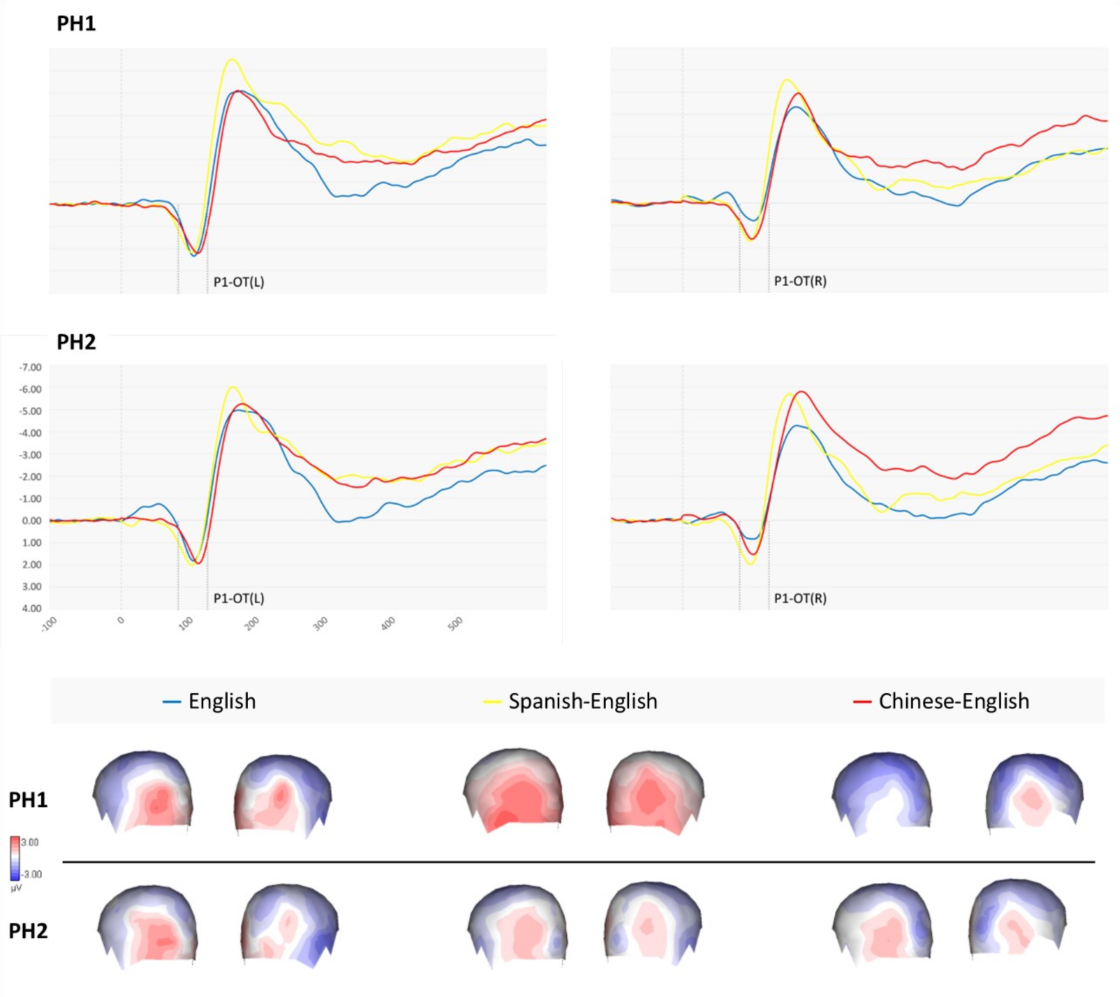

Figure 19: Event-related potentials and topographic voltage maps of left and right occipitotemporal clusters at 100ms for task/pseudohomophone comparisons (PH1,PH2) across groups

6.4.4.3.2.1 P1-OT amplitude

Table 29 shows descriptive statistics for P1-OT amplitude across groups and tasks.

Table 29: Means (SDs) of P1-OT amplitude across tasks (µV)

		English	Spanish-English	Chinese-English
oLDT	Left	0.74 (0.38)	0.92 (0.51)	0.93 (0.48)
pLDT		0.75 (0.51)	0.77 (0.44)	0.77 (0.35)
oLDT	Right	0.72 (0.29)	0.89 (0.56)	1.01 (0.62)
pLDT		0.46 (0.21)	1.09 (0.57)	0.87 (0.38)

The Task x Hemisphere x Group ANOVA on P1-OT amplitude showed the main effect of Group, $F(2,57)=3.44, p=.039, \eta^2_p=0.11$, and the full three-way Task x Hemisphere x Group interaction to be significant, $F(2,57)=6.62, p=.003, \eta^2_p=0.19$. Pairwise post-hoc comparisons at task-level showed that P1-OT(R) to pLDT was larger in Spanish-English than English ($p=.002$). Pairwise post-hoc comparisons between pseudohomophone conditions (PH1,PH2), meanwhile, showed P1-OT(L) amplitude in Spanish-English was significantly larger to PH1 than to PH2 ($p=.008$), and P1-OT(R) amplitude in English was significantly larger to PH1 than to PH2 ($p<.001$).

6.4.4.3.2.2 P1-OT latency

Table 30 shows descriptive statistics for P1-OT latency across groups and tasks.

Table 30: Means (SDs) of P1-OT latency across tasks (ms)

		English	Spanish-English	Chinese-English
oLDT	Left	101.99 (4.43)	98.3 (4.86)	103.09 (4.4)

pLDT		102.83 (4.93)	99.32 (4.75)	100.92 (5.34)
oLDT	Right	99.98 (5.06)	95.07 (4.98)	98.5 (4.47)
pLDT		99.26 (5.15)	94.9 (4.22)	97.73 (5.71)

The Task x Hemisphere x Group ANOVA on P1-OT latency showed significant main effects of Hemisphere (Right<Left), $F(1,57)=23.66, p<.001, \eta^2_p=0.29$, and Group, $F(2,57)=8.5, p<.001, \eta^2_p=0.23$. Averaging across tasks and hemispheres, post-hocs showed P1-OT was significantly earlier in Spanish-English than both English (p<.001) and Chinese-English ($p=.01$).

6.4.4.3.3 Frontal-central N100 (N100-FC)

Figure 20 depicts frontal-central ERP data at 100ms to pseudohomophone conditions.

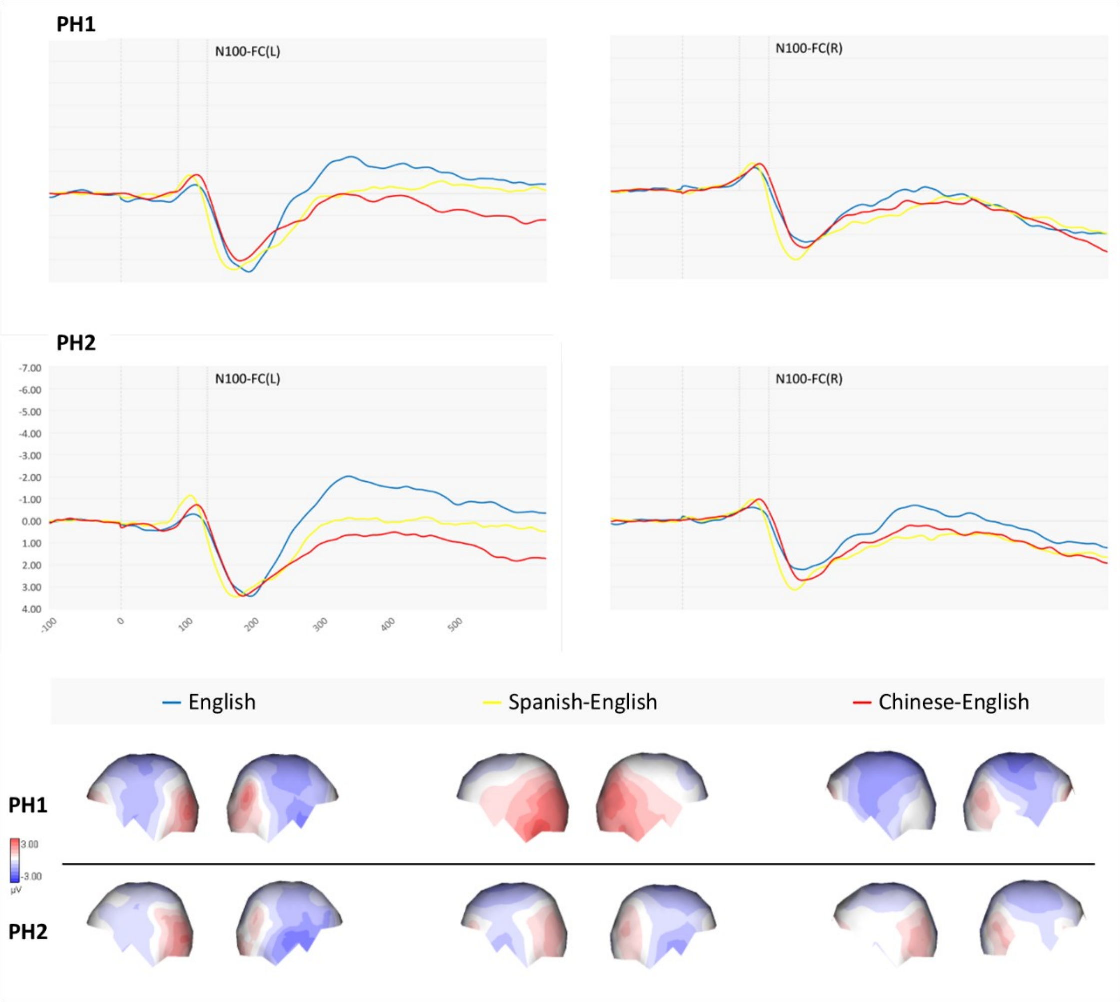

Figure 20: Event-related potentials and topographic voltage maps of left and right frontal-central clusters at 100ms for task/pseudohomophone comparisons (PH1,PH2) across groups

6.4.4.3.3.1 N100-FC amplitude

Table 31 shows descriptive statistics for N100-FC amplitude across groups and tasks.

Table 31: Means (SDs) of N100-FC amplitude across tasks (µV)

		English	Spanish-English	Chinese-English
oLDT	Left	-0.49 (0.22)	-0.58 (0.29)	-0.64 (0.33)
pLDT		-0.4 (0.21)	-0.76 (0.44)	-0.6 (0.31)
oLDT	Right	-0.56 (0.28)	-0.57 (0.24)	-0.71 (0.36)
pLDT		-0.43 (0.25)	-0.75 (0.42)	-0.61 (0.28)

The Task x Hemisphere x Group ANOVA on N100-FC amplitude showed a significant main effect of Group, $F(2,57)=3.88, p=.026, \eta^2_p=0.12$, and a significant interaction of Task x Group, $F(2,57)=8.31, p<.001, \eta^2_p=0.23$. Averaging across hemispheres, post-hocs showed N100-FC was larger to pLDT than oLDT in Spanish-English ($p=.029$) and larger to pLDT in Spanish-English than English ($p=.002$).

6.4.4.3.3.2 N100-FC latency

Table 32 shows descriptive statistics for N100-FC latency across groups and tasks.

Table 32: Means (SDs) of N100-FC latency across tasks (ms)

		English	Spanish-English	Chinese-English
oLDT	Left	102.55 (5.82)	96.41 (6.45)	101.61 (3.52)
pLDT		99.38 (5.89)	96.84 (5.05)	100.91 (5.15)
oLDT	Right	99.63 (5.06)	94.82 (5.88)	101.77 (4.46)
pLDT		100.13 (6.49)	95.31 (3.54)	97.59 (5.41)

The Task x Hemisphere x Group ANOVA on N100-FC latency showed significant main effects of Hemisphere (Right<Left), $F(1,57)=8.21, p=.006, \eta^2_p=0.13$, and Group, $F(2,57)=8.58, p<.001, \eta^2_p=0.23$, as well as the full three-way Task x Hemisphere x Group interaction, $F(2,57)=5.63, p=.006, \eta^2_p=0.17$. Post-hoc comparisons showed N100-FC(L) to oLDT was earlier in Spanish-English than English ($p=.018$), while N100-FC(R) to oLDT was also earlier in Spanish-English than English ($p=.003$).

6.4.4.3.4 Occipitotemporal N170 (N170-OT)

Figure 21 depicts occipitotemporal ERP data at 170ms to pseudohomophone conditions.

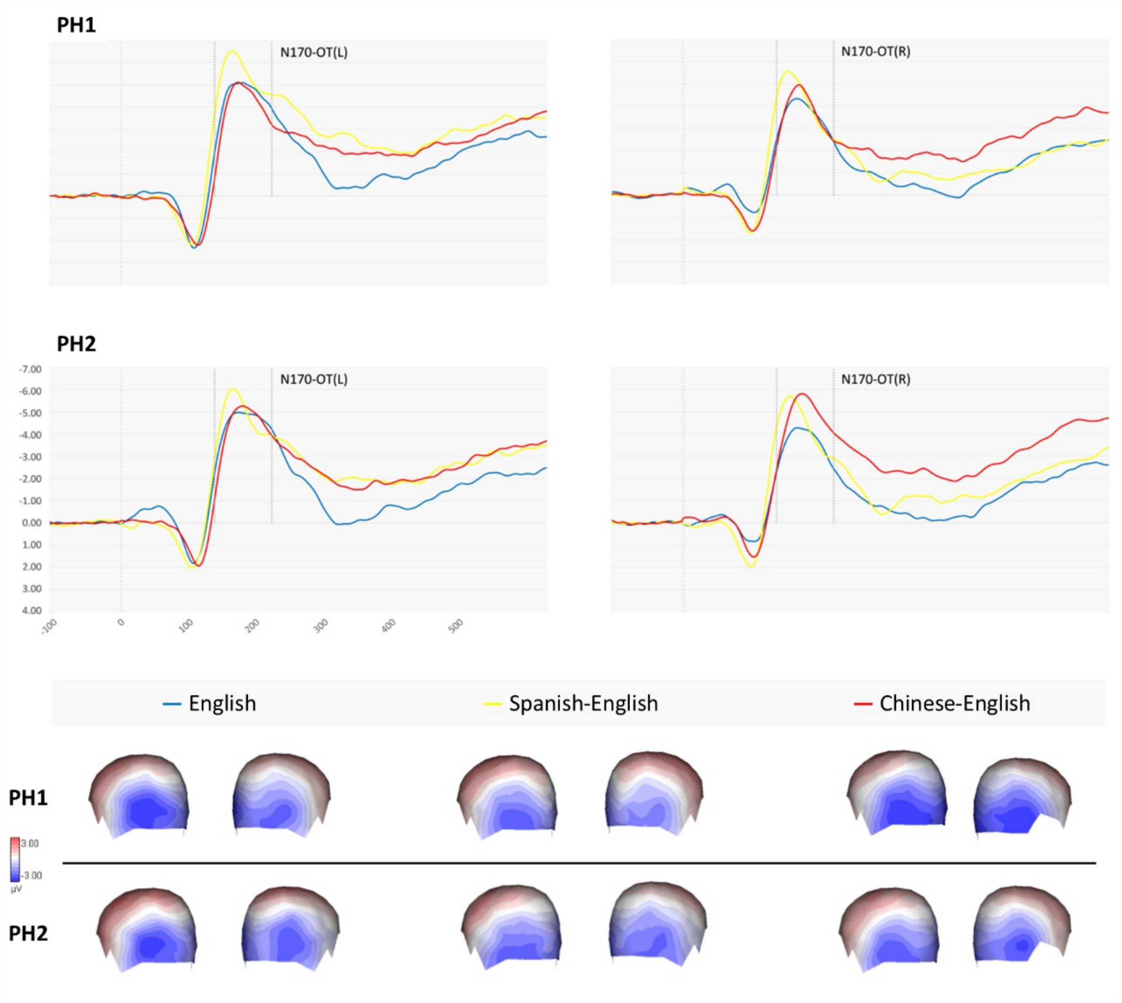

Figure 21: Event-related potentials and topographic voltage maps of left and right occipitotemporal clusters at 170ms for task/pseudohomophone comparisons (PH1,PH2) across groups

6.4.4.3.4.1 N170-OT amplitude

Table 33 shows descriptive statistics for N170-OT amplitude across groups and tasks.

Table 33: Means (SDs) of N170-OT amplitude across tasks (µV)

		English	Spanish-English	Chinese-English

oLDT	Left	-2.17 (0.85)	-2.53 (1.35)	-1.97 (1.02)
pLDT		-2 (0.77)	-2.03 (1.08)	-1.78 (0.53)
oLDT	Right	-1.81 (0.71)	-2.07 (1.21)	-2.26 (1.24)
pLDT		-2.07 (0.85)	-1.98 (1.34)	-2.12 (0.91)

The Task x Hemisphere x Group ANOVA on N170-OT amplitude showed only a significant interaction of Hemisphere x Task, $F(1,2)=4.46, p=.039, \eta^2_p=0.07$, which essentially disregards the Group factor so will not be investigated further.

6.4.4.3.4.2 N170-OT latency

Table 34 shows descriptive statistics for N170-OT latency across groups and tasks.

Table 34: Means (SDs) of N170-OT latency across tasks (ms)

		English	Spanish-English	Chinese-English
oLDT	Left	171.52 (6.12)	165.16 (6.05)	170.3 (6.79)
pLDT		172.18 (7.35)	166.34 (5.48)	169.95 (7.96)
oLDT	Right	168.77 (8.1)	159.05 (7.46)	167.67 (5.87)
pLDT		166.66 (6.58)	160.02 (8.39)	167.55 (7.49)

The Task x Hemisphere x Group ANOVA on N170-OT latency showed significant main effects of Hemisphere (Right<Left), $F(1,57)=26.32, p<.001, \eta^2_p=0.32$, and Group, $F(2,57)=9.85, p<.001, \eta^2_p=0.26$. Averaging across tasks and hemispheres, post-hocs showed N170-OT was significantly earlier in Spanish-English than both English ($p<.001$) and Chinese-English ($p<.001$).

6.4.4.4 Summary

As expected due to the relative difficulty of the tasks, accuracy in the oLDT was higher than in the pLDT in all groups, while only Chinese-English were faster than Spanish-English overall. Attributable as phonological effects, bilateral N100-FC in Spanish-English was larger in the pLDT than in the oLDT, while P1-OT(L) was significantly larger to PH1 than to PH2 in Spanish-English and P1-OT(R) was significantly larger to PH1 than to PH2 in English. These findings show clear differences between the orthography- and phonology-oriented tasks as well as between groups linked to language profiles.

6.5 Discussion

The overarching contrast between the processing of real words and linguistic stimuli that are not real words (e.g., pseudohomophones or pseudowords) has traditionally been the backbone of lexicality effects, but the more specific task methodology of the current study aimed to dissect it further by using separate orthographic and phonological tasks (oLDT and pLDT, respectively). Behavioural and ERP measures to RW and PH1 from the oLDT were analyzed to investigate effects of orthographic lexicality, which can also be seen as orthographic familiarity through the legitimate orthography of real words found in RW that PH1 does not possess. Responses to pseudowords (PW) and pseudohomophones (PH2) from the pLDT, meanwhile, were analyzed in order to investigate effects of phonological lexicality, which also reflects a pseudohomophone effect through the phonological legitimacy of pseudohomophones.

The main avenues of enquiry with this approach concerned how behavioural and electrophysiological measures of orthography- and phonology-oriented lexicality differ within and between native English and bilingual ESL groups with an alphabetic L1 (Spanish-English) or non-alphabetic L1 (Chinese-English). Moreover, the study investigates whether lexicality is reflected by the occipital P1 (P1-O) and if there is any indication of early phonological activation in left frontal-central N100-FC responses. Additionally, it examines the orthographic and/or phonological nature of occipitotemporal ERP responses at ~100 and ~170ms (P1-OT and N170-OT, respectively). Regarding group-level interpretations, it is important to remember and reasonable to accept that any differences between the ESL groups and the English group are largely due to the difference in English proficiency (as highlighted by the ceiling effects to RW in the English group). Therefore, the main contrasts of interest are between the ESL groups (of equivalent proficiency), using the English group as a control.

6.5.1 Behavioural performance

The behavioural facet of the study involved any lexicality effects within and between the ESL groups (using the native English group as a control), while providing a behavioural performance reference for the ERP results where applicable. Before ERP results, therefore, this section will outline and discuss the behavioural findings.

Starting with the oLDT, accuracy overall was higher in English than in both ESL groups, but also higher in the alphabetic L1 ESL group (Spanish-English) than in the non-alphabetic L1 ESL group (Chinese-English), indicating an advantage for the groups with an alphabetic L1, as expected due to the lifelong familiarity of an alphabetic script in English and Spanish-English. However, no differences in accuracy were observed between RW and PH1 conditions within groups or per condition between groups, showing that the oLDT conditions were interpreted with similar success in each group and no group made more errors to one condition or the other. This result for English differs substantially from Twomey et al. (2011) where the native English monolingual participants were less accurate to real words than pseudohomophones in the same orthographic lexical decision task context. Briesemeister et al. (2009) also reported opposing results to Twomey et al. (2011) with accuracy being higher for real words than pseudohomophones, though that study did use a linguistically different pool of stimuli and participants (German), which could account for the difference. Sometimes, this kind of discrepancy is due to a difference in condition-balancing between studies where, for example, the real word condition included more irregular and/or ambiguous terms than did the base words for the pseudohomophone condition, providing an effect in favour of the pseudohomophones. However, Twomey et al. (2011) used many of the stimulus control parameters and methods used in the current study and suggesting any other psycholinguistic factor being the culprit would be unsupported conjecture. Importantly, the lack of difference in accuracy between oLDT conditions was not only observed in the native English group, but also in both ESL groups, who arguably have more reason to be more or less accurate to real words due to less time learning and using English. The reason for this discrepancy between studies is, therefore, not clear.

Fully in line with Twomey et al. (2011), Briesemeister et al. (2009), and Braun et al. (2015), however, response times within groups were faster to RW than to PH1 in English, as well as in Spanish-English, but not in Chinese-English. This combination of results strongly suggests an advantage for the groups with an alphabetic script for their L1 and the resulting familiarity from everyday usage. The faster responses to RW than to PH1 demonstrate a clear effect of orthographic lexicality alongside response inhibition from the orthographic-phonological conflict of pseudohomophones (legal and legitimate phonology, but only legal orthography). These results suggest that traditional lexicality effects contrasting real words with pseudowords have an orthographic root and do not involve phonological activation, at least concerning behavioural responses.

Response times between groups, however, were only faster in Chinese-English than Spanish-English to PH1. As response times to RW were similar across groups, this difference in response times to pseudohomophones between ESL groups suggests that Chinese-English participants were inhibited by the orthographic-phonological incongruence inherent to pseudohomophones less than Spanish-English participants.

This difference between ESL groups proposes a difference in VWR processing when reading pseudohomophones. Considering the language profiles of the ESL groups and the inherently different ways of learning English as a second language i.e., with Spanish, an alphabetic language as a basis or with the non-alphabetic logographic Chinese/Mandarin as a basis, the use of different cognitive strategies for deciphering novel L2 stimuli makes sense. Furthermore, Chinese-English showed no difference in RTs between oLDT conditions, supporting the interpretation that the Chinese-English participants responded similarly regardless of whether the stimuli were real English words (RW) or pseudohomophones (PH1), also suggesting a similar cognitive process for both stimulus types.

As observed in the oLDT, but this time supporting Twomey et al. (2011), there were no differences in accuracy between pLDT conditions within any group, suggesting that PH2 and PW stimuli were responded to similarly within all groups. Between groups in the pLDT, accuracy was higher in English than in both Spanish-English and Chinese-English to both PH2 and PW. This was not unexpected due to the contrast between the monolingual native English group and the late-bilingual ESL groups, for whom English is not only their second language but with which significantly less time in their lives has been spent learning and using it. Furthermore, there were also no differences in accuracy between Spanish-English and Chinese-English, suggesting the ESL groups responded with similar efficacy to the pLDT conditions.

Response time results also support the suggestion that efficacy was similar between ESL groups as no differences were observed between them either. The only RT difference between groups was that Chinese-English were faster than English to PW, which is also part of a speed-accuracy trade-off in the PW condition, where Chinese-

English were faster, but English were more accurate. While English and Spanish-English processed PW similarly (especially following the lack of difference in accuracy from PH2), perhaps due to their respective alphabetic experience, the faster but less accurate responses in the Chinese-English group suggest there was a difference in VWR processing when reading and responding to pseudowords. The group RT difference to PW is perhaps better explained as English being slower and taking additional time to check the pseudowords (perhaps phonologically) compared with Chinese-English who processed the pseudowords orthographically as whole words and did not perform additional checks, hence the poorer accuracy (M. Wang et al., 2003). This PW RT difference with English not also being observed for Spanish-English indirectly suggests a distinction between the ESL groups and processing of pseudowords at least. As mentioned above, the Spanish-English group may also have made additional checks on pseudowords as with the English group, considering their shared alphabetic background, but just did not do so significantly faster than the English or slower than the Chinese-English groups.

The only behavioural difference between pLDT conditions was that responses to PH2 were faster than to PW in the English group, replicating Twomey et al. (2011) and demonstrating a phonological lexicality effect. Considering that the stimuli of PH2 and PW both used equally legal orthography and were matched on various psycholinguistic factors (see §5.4.1), differing only in phonological legitimacy, this is clear evidence of facilitation from the legitimate phonology of pseudohomophones over the legal but not legitimate phonology of pseudowords. This facilitation of processing times by PH2 compared with PW and a resulting pseudohomophone effect, however, was not present in the ESL groups. On the surface, this suggests pseudohomophones and pseudowords were processed similarly in the ESL groups despite the potentially facilitative legitimate phonology of the PH2 pseudohomophones, indicating the ESL groups were more focused on orthography, which follows the equivalent orthographic (un)familiarity of both pLDT conditions and lessens the potential effect of phonology. This different pattern of results between native and ESL groups could also involve additional orthographic-phonological checking of PW for legitimate phonology (in case they were pseudohomophones) by the ESL groups, which takes additional time, negating any relative facilitation from the pseudohomophones of PH2.

In addition to the condition-specific comparisons discussed already, the behavioural measures can be considered overall per task as a task-level contrast. Between groups, accuracy to oLDT (RW,PH1 averaged) was only higher in English than Chinese-English, while accuracy to pLDT (PH2,PW averaged) was higher in English than both Spanish-English and Chinese-English. While not surprising that the English group fared better in both tasks, it is notable that there was no difference in accuracy between Spanish-English and Chinese-English groups themselves, suggesting some similarity of approach for both tasks (oLDT and pLDT).

Within groups, overall accuracy to oLDT was higher than to pLDT within English, Spanish-English, and Chinese-English, as was expected due to oLDT being a simpler task in general while also involving (high frequency) real words that are easier to read and recognize. Somewhat unexpectedly, however, there were no overall differences in RTs between tasks within groups, which is likely explained by the more difficult and incongruent conditions per task (PH1 and PW in the oLDT and pLDT, respectively) evoking longer RTs than the easier, more congruent conditions (RW and PH2 in the oLDT and pLDT, respectively), but to an extent within each task that averaged out similarly.

6.5.2 Lexicality and occipital P1 (P1-O)

The aim was to explore how early ~100ms brain activity behaves in late-bilingual ESL readers in contrast with one another as well as with native English monolinguals. No statistically significant differences were observed on P1-O amplitude between oLDT condition (RW,PH1) within group or per condition between groups. This follows several reports from Hauk and colleagues (e.g., Hauk, Davis, et al., 2006; Hauk et al., 2012; Hauk, Patterson, et al., 2006) and others (e.g., Emmorey et al., 2017), supporting the notion that P1-O amplitude does not reflect lexicality or word recognition per se (Gibbons et al., 2022; D. Zhang et al., 2014), refuting claims of larger P1 amplitudes to real words (e.g., Segalowitz & Zheng, 2009) and to pseudowords (e.g., Taroyan, 2015). Not only did the results of the current study suggest that P1-O amplitude did not differ based on orthographic lexicality, but, importantly, did not meaningfully differ in any group. This alludes to this early ~100ms stage of processing potentially being similar across groups, though deeper and more specific testing is needed.

The pattern of results from the oLDT followed into the pLDT, where there were also no effects of condition (PH2,PW) or group found on P1-O amplitude, which indirectly supports the idea of P1-O activity being visual/orthographic in nature and not phonological as tested with the pLDT. It also suggests that P1-O amplitude in a phonological lexical decision task is similar between L1 and L2 readers of English. Overall, though, no evidence was found that suggests differences in lexicality, orthographic or phonological, influence P1-O amplitude. Therefore, in contrast to some previous findings (e.g., Segalowitz & Zheng, 2009; Taroyan, 2015), P1-O amplitude does not appear to reflect either orthographic or phonological lexicality when stimulus conditions are closely controlled.

P1-O latency, on the other hand, was observed to be earlier in Spanish-English than English and Chinese-English groups, dependent on task. Averaging across oLDT conditions (RW,PH1), bilateral P1-O latency was earlier in Spanish-English than in Chinese-English, while bilateral P1-O latency when averaging across pLDT conditions (PH2,PW) was earlier in Spanish-English than English. Latency being earlier in Spanish-English than Chinese-English in the oLDT could relate to the Spanish-English readers' stronger and inherent familiarity with the alphabetic script that is largely shared between English and Spanish (hence a lack of latency difference between Spanish-English and English) that the late-bilingual Chinese-English participants would only have encountered more formally when learning English. Latency being earlier in Spanish-English than English in the pLDT could also relate to the familiarity of alphabetic script for Spanish-English participants, but with the suggestion of a cognitive strategy different from native English readers that leans toward the phonological L1 profile of the Spanish-English participants. For instance, the phonological context of the pLDT paired with the phonology-oriented processing that native Spanish readers are accustomed to in their L1 possibly inhibited the visual/orthographic processing associated with the P1-O, effectively moving that stage of processing along more quickly than in the English group, favouring a phonological processing route not reflected by the P1-O. Similarly, the oLDT results could reflect the same urgency in the Spanish-English to process phonologically, but factoring in the inherently orthographic processing of the native Chinese readers (due to their logographic language profile) for which visual/orthographic processing associated with the P1-O is naturally more pertinent.

The earlier P1-O latencies in Spanish-English, at least compared with the English group in the pLDT, echo the observation of Italian-English bilinguals (i.e., another ESL group with a similarly orthographically shallow L1 to the Spanish-English) reading pseudowords more quickly and accurately than English monolinguals (Paulesu et al., 2000). One conclusion drawn from this behavioural finding was that the ESL readers, due to their phonology-oriented L1, were predisposed to an orthography-phonology conversion approach to reading. It is not unreasonable to suggest that the earlier P1-O latencies observed here in Spanish-English could also be explained by this same readiness for phonological conversion from the occipitally processed visual/orthographic input. This readiness combined with the natural expectancy from the orthographically shallow native Spanish could account for the earlier latency of an occipital component more closely associated with visual/orthographic processing.

6.5.3 Nature of occipitotemporal P1 (P1-OT)

The main aim of investigating the association of P1-OT activity with visual word recognition was to observe how it presents in ESL readers. In particular, right occipitotemporal P1 responses (peaking at ~100ms around electrode PO8) have often been a feature of research into early visual word recognition processes (Dien, 2009). Despite this precedent from the literature, no effects of orthographic lexicality were found on P1-OT amplitude, suggesting that neither left, right, nor bilateral P1-OT amplitude was sensitive to orthographic lexicality effects of real words (RW) compared with pseudohomophones (PH1) and that the P1-OT was not lateralized to either stimulus type in any group.

This relative similarity of results in all groups suggests that fundamental orthographic processing associated with the P1 over occipitotemporal areas is not dissimilar between native and ESL readers or, indeed, between readers of English with different L1 profiles. However, a right-lateralized occipitotemporal P1 was expected, at least in the native English group if not the ESL groups too. The most likely reasons for not finding this previously observed effect concern the control of frequency effects through the relative high frequency of the stimuli, control of more visual/physical characteristics, such as the matched orthographic word lengths, as well as control of orthographic neighbourhood sizes between conditions (Hauk & Pulvermüller, 2004), and that this effect has most commonly been observed through the traditional lexicality contrast of real words and pseudowords i.e., not real words and pseudohomophones as used here in the oLDT (Segalowitz & Zheng, 2009).

In the pLDT, bilateral P1-OT amplitude was larger to PW than PH2 in Spanish-English, showing an effect of phonological legitimacy, where processing the pseudowords (that lack the legitimate phonology of the pseudohomophones) appears more cognitively taxing, taking more effort. This phonological lexicality effect being observed in Spanish-English and not Chinese-English is likely connected to the experience and expertise with the alphabetic script of the Spanish-English participants that the late-bilingual Chinese-English participants could not match. This same effect not being present in English participants, meanwhile, could be due to the experience and expertise of their own native and only language. Following that bilateral P1-OT amplitude to PW was larger in English than Spanish-English, suggesting increased orthographic-phonological processing effort in English participants, this could also propose a difference in processing between English and Spanish-English for reading written language with legal orthography but not phonology (i.e., pseudowords).

Averaging across pLDT conditions (PH2,PW), right hemisphere P1-OT amplitude was larger in both Spanish-English and Chinese-English than in English. Considering such an averaged measure to represent novel orthographic processing with a requirement for phonological output i.e., to complete the task, this tentatively supports the importance of right-hemisphere P1-OT activity for orthography-phonology mapping, as required when reading pseudohomophones and pseudowords (and unfamiliar real words), suggesting it to also be important in ESL readers as in native English readers (Dien, 2009). This finding also suggests that processing effort for pLDT responses was more intense in Spanish-English and Chinese-English, which follows the contrast in levels of proficiency and experience between native and ESL groups.

Bilateral P1-OT latency averaged across oLDT conditions (RW,PH1) was earlier in Spanish-English than in both English and Chinese-English. As it is fair to accept that P1-OT latencies were not dissimilar in the oLDT between English and Chinese-English (groups with a much more orthographic nature of L1 – English with the deep orthography and Chinese-English with the logography of Chinese), one interpretation is that the earlier P1-OT responses in Spanish-English relate to their orthographically shallow and phonology-based L1 language profiles. This finding, therefore, follows one of the overarching motivations for the current research in terms of examining the role of L1 in L2 reading and provides rationale for looking further into how the nature of L1 profiles influences visual word recognition (in both L1 and L2 readers).

Bilateral P1-OT latency averaged across pLDT conditions (PH2,PW) was also earlier in Spanish-English than English, but not different between the ESL groups, pointing to a difference in processing between readers with an alphabetic L1. As with the similar P1-O result, this P1-OT observation could also relate to the script familiarity for Spanish-English participants, but with a strategy that leans toward the Spanish-English phonological L1 profile. The natural phonology-oriented processing of native Spanish readers and the phonological context of the pLDT possibly favoured phonological processing not reflected by the P1-OT and/or inhibited the orthographic processing associated with the P1-OT, resulting in less time spent at this level of processing than in the English group. This result also echoes the faster and more accurate reading of pseudowords by Italian-English bilinguals (Paulesu et al., 2000) and that the phonology-oriented L1 of the ESL readers were predisposed to an orthography-phonology conversion route to VWR, as posited in the DRC (Coltheart et al., 2001) and BIAM (Grainger & Holcomb, 2009).

Overlapping results are not surprising between P1-O and P1-OT as their measurement timeframes are identical (i.e., 80-120ms) and the scalp locations are in relatively close proximity. Indeed, the pattern of earlier P1-OT latencies in Spanish-English echoes the P1-O latency results, but with a key difference of P1-OT latency being earlier in Spanish-English than English to oLDT conditions (RW,PH1), which was not found for P1-O. Considering the aforementioned association of the P1-OT to orthographic processing (Dien, 2009), at least in native readers, it might instead be better to think of this result as P1-OT latencies being later in English than Spanish-English. Therefore, this finding suggests quicker or perhaps lesser orthographic processing in Spanish-English potentially due to less reliance and emphasis on orthographic processing based on the shallow orthographic depth and phonological nature of their Spanish L1. This implies a deviation of VWR processing strategy between English and Spanish-English where Spanish-English bilinguals might appear to move onto phonological processing more quickly and readily than English native readers.

6.5.4 Early phonological activation and N100-FC

The focus on early pre-200ms and, especially, ~100ms frontal-central ERP responses to phonological stimuli follows the compelling but somewhat controversial evidence for phonological activation as early as ~100ms (e.g., Ashby, 2010; Klein et al., 2015; Pammer et al., 2004; Wheat et al., 2010). Due to the emphasis on phonology, findings from within the pLDT and between tasks will take priority in this discussion.

Between conditions in the pLDT, bilateral N100-FC amplitude was larger to PW than PH2 in Spanish-English, showing a clear pseudohomophone effect where activity was lesser, implying easier processing, for pseudohomophones. Alternatively, this can be seen as an effect of phonological lexicality through processing pseudowords and their lack of phonological legitimacy appearing to require more cognitive effort. Though this was not the left-lateralized activity reported previously (e.g., Ashby, 2010; Wheat et al., 2010) and found only in the Spanish-English ESL group (as opposed to the native English monolingual group), this provides important support for the potential for frontal-central EEG activity that is specific to legitimate phonology in nature.

Notably, this effect was not observed in English or Chinese-English, indicating this sensitivity of N100-FC amplitude to phonological processing to be somewhat receptive and more easily observed in Spanish-English readers, following their L1-based sensibility for phonology. Taken further, it is important to consider why the effect was not observed in English or Chinese-English. While Spanish-English were similarly experienced with the alphabetic script (if not more so due to knowing two languages that use it), English participants being naturally more experienced with its implementation as English could be why processing between PW and PH2 in English evoked no difference in frontal-central effort at ~100ms, especially in combination with the relatively high frequency base words of the stimuli. If this is the case, it suggests the PW>PH2 effect in Spanish-English to reflect the additional phonological effort required to accurately determine that the pseudowords did not have the legitimate phonology of real English words that the pseudohomophones did. It would, therefore, be worthwhile to test native English monolinguals with less straightforward e.g., lower frequency phonological stimuli, as well as to investigate this phonological effect further in the Spanish-English group.

Participants in the late bilingual Chinese-English group, meanwhile, were only familiar with English phonology more recently and do not have a language profile that either foregrounds phonology (as in Spanish) or that would necessarily provide the skills for grapheme-phoneme conversion (as in all languages with an alphabetic script). The lack of phonological effect (on N100-FC or elsewhere) could, therefore, indicate a different cognitive approach to deciphering English-based non-real linguistic stimuli. Further phonology-focused investigation, however, is required to support this notion and define what such an alternative processing strategy would be.

Between groups in the pLDT, bilateral N100-FC amplitude to PW was larger in Spanish-English than English, mirroring P1-OT amplitude that was larger in English. These contrasting results suggest that processing associated with N100-FC activity takes precedence in Spanish-English participants (or perhaps ESL readers with an orthographically shallow alphabetic L1) while equivalent processing is more pertinently associated with P1-OT activity in native English readers. However, as this finding repeats the pattern on P1-OT amplitude, it is also conceivable that this mirrored effect is the dipolar opposite of the P1-OT finding and, therefore, further specific research is needed to dissociate these observations.

Despite the full three-way interaction between pLDT Condition, Hemisphere, and Group on N100-FC latency, the only pairwise finding was that N100-FC was earlier in the right than left hemisphere in Chinese-English to PW, showing right-lateralized frontal-central activity to pseudowords in Chinese-English participants. Considering this effect was found only to pseudowords (but not pseudohomophones) and only in the group with a non-alphabetic L1 suggests that it could reflect a difference in strategy for decoding unknown and unfamiliar written forms in L2 or, at least, that the processing reflected by this effect is more pertinent in the ESL readers with a non-alphabetic L1. Furthermore, this was not observed in English or Spanish-English participants, both groups of which have more experience with the alphabetic script and have an alphabetic L1.

6.5.5 The nature of the N170 and its hemispheric laterality in VWR

Studying N170-OT activity in orthographic and phonological contexts between groups with different L1 profiles was to explore whether its nature lies in orthographic processing, orthography-phonology mapping, or phonological activation, as well as to observe its hemispheric lateralization, which has been shown as an index for visual language familiarity. From an example incorporating both condition and hemisphere factors, the N170-OT has previously been reported to be larger in the left but smaller in the right for orthographic stimuli compared with “nonorthographic” stimuli (Bentin et al., 1999). However, no such observations were found in the current study: no group effects were observed on N170-OT amplitude for either oLDT conditions (RW,PH1) or pLDT conditions (PW,PH2). Therefore, there was no evidence for hemispheric laterality differences between groups and no evidence for sensitivity of N170 amplitude to real words, orthography, phonology, or any language-specific processing in any group, as effects between oLDT conditions (RW,PH1) or pLDT conditions (PW,PH2) would show. These results follow the findings of studies where no amplitude differences were observed between such stimuli as words, pseudowords, and consonant strings (e.g., Hauk, Davis, et al., 2006; Maurer, Brem, et al., 2005).

Latencies of the bilateral N170-OT observed in the current study, meanwhile, were earlier to RW than to PH1 in English, showing a clear effect of orthographic familiarity in the native group, echoing the response times being faster to RW than to PH1. While no effects of oLDT condition were found on bilateral N170-OT latency in the ESL groups, it was earlier in Spanish-English than in both English and Chinese-English to RW. On the surface, this suggests that the processing associated with the N170 was completed more quickly in Spanish-English. As the N170-OT latency in Spanish-English was earlier even than in the native English readers, it is fair to consider this an effect of group properties in the sense of being based on something intrinsic to the Spanish-English readers. As emphasized when contrasted with the native languages of English and Chinese-English, Spanish-English represent ESL readers with an orthographically shallow, phonology-based written language. On the other hand, this could be associated with the same finding for P1-OT latency, following on in time and cognitive process, potentially only being earlier because the prior processing was earlier, thus the processing reflected by the bilateral N170-OT in Spanish-English may not be distinct from English and Chinese-English.

The only observed N170-OT effect in the pLDT was on bilateral N170-OT latency, which was again earlier in Spanish-English than both English and Chinese-English, this time when averaging across pLDT conditions (PW,PH2). This effect not being found for each or either condition separately, is especially relevant because the two pLDT conditions are both orthographically unfamiliar, so would need some form of grapheme-phoneme conversion to be read, and using them as an overall measure averages out the phonological legitimacy that would otherwise distinguish them. While this requirement of processing would also be true for English and Chinese-English participants (and anyone else), the distinction of earlier N170-OT latency in Spanish-English once again ties into the phonological nature of their L1 and indicates a potential difference of processing strategy between groups.

Bilateral N170-OT latency being earlier in Spanish-English than both English and Chinese-English to pLDT conditions (PW,PH2) overall also shines a light back on the equivalent result to RW in the oLDT and, specifically, why it was only observed to RW and not PH1 (or to both overall). In contrast with the pLDT, the oLDT conditions require different strategies to be read with a direct lexical approach for real words (RW) and an indirect GPC approach for pseudohomophones (PH1), which underlies why the earlier latencies were only observed to RW in the oLDT. These N170-OT effects from the oLDT and pLDT, as well as the similar findings on P1-OT latency, therefore, tie in together as further evidence of the group with a phonological and orthographically shallow L1 processes alphabetic scripts differently from readers with more orthographic focus in their L1 (i.e., English, Chinese).

6.5.6 Task effects and top-down processing

Responses to tasks overall (oLDT,pLDT) and the pseudohomophone conditions per task (PH1,PH2) were analyzed to observe top-down processing between orthography- and phonology-dominant scenarios (oLDT and pLDT, respectively) and how they compare between groups. For the oLDT measure, processing related to oLDT conditions theoretically averages out (i.e., direct lexical for RW, GPC for PH1) and so does not directly represent either pathway, while the processing pathway of the pLDT conditions (GPC for both PH2 and PW) is compounded in the pLDT measure, providing an overall measure of grapheme-phoneme conversion.

Starting with the occipital P1, bilateral P1-O latency was earlier to pLDT than to oLDT in Chinese-English, showing an effect of task and, by extension, top-down processing. The later latency in the orthographic task compared with the phonological suggests processing associated with the P1-O took longer in the oLDT for the Chine-English group, which follows the notion of the P1-O (in VWR contexts) being associated with initial visual/orthographic processing. Indeed, the non-alphabetic L1 and orthography-centred language profile of the Chinese-English group might reasonably expect differential occipital activation, especially for illegitimate linguistic stimuli (pseudowords, pseudohomophones) and lesser-known words. However, the main difference between tasks was the inclusion of legitimate real words in the oLDT where the pLDT had none and it is not unreasonable to accept this as the reason for this effect on P1-O latency in the Chinese-English group. Lastly, as this pattern of processing was specific to the Chinese-English (non-alphabetic L1) group and not found in either alphabetic L1 group, it could also generalize to ESL readers with a non-alphabetic L1, though further testing is required to support this.

Right-lateralized P1-OT to pLDT was larger in Spanish-English than English, following the idea that P1-OT and especially right-lateralized P1-OT activity is closely associated with orthographic processing and is likely fuelled by the equivalent pattern to PW within the pLDT. Both PW (within the pLDT) and pLDT itself (in contrast with oLDT) are conditions that reflect phonology-related processing through an indirect grapheme-phoneme conversion pathway. In this case, therefore, right-lateralized P1-OT activity in the Spanish-English group appears to reflect the orthography-phonology mapping necessary for legal orthographic input that has no record in the lexicon e.g., pseudowords. The larger amplitude of the right-lateralized P1-OT (relative to the native English group) suggests increased effort of processing in Spanish-English and that this component/timeframe is even more important for bilingual VWR than already documented for native readers (Dien, 2009), at least in late ESL bilinguals with an alphabetic L1.

In the native English group, the right-lateralized P1-OT amplitude was also significantly larger to PH1 than to PH2, which indicates a task-level difference due to the matched sets of pseudohomophones used for PH1 and PH2, though there was no overall difference between oLDT and pLDT. This dissociation suggests that this finding highlights the top-down influence of the task on the grapheme-phoneme conversion required to read pseudohomophones. The larger right-lateralized P1-OT amplitude to PH1 reflects increased effort for GPC when the task (oLDT) did not explicitly require it and less relative effort when the task (pLDT) did require it, especially considering that both conditions of the pLDT required it and no switching between direct lexical and GPC methods was needed, unlike in the oLDT, the conditions of which explicitly demand such switching (i.e., direct lexical for RW, GPC for PH1). Mirroring the English finding, albeit with opposite hemispheric lateralization, larger left-lateralized P1-OT amplitude to PH1 than to PH2 in the Spanish-English group also indicates a task-level difference (as the pseudohomophone conditions were matched), providing evidence of both a similarity and difference in Spanish-English with the native English group.

Moving to frontal-central activity, bilateral N100-FC was larger to pLDT than oLDT in Spanish-English, which follows the notion that frontal-central activity is associated with phonological processing with amplitudes being larger in the phonological task than the orthographic, even this early in the VWR timeframe. Further, this was only observed in the arguably more phonology-sensitive Spanish-English group, which could be due to the familiar alphabetic script and an increased sensitivity to phonology from the Spanish-English participants' L1 that the Chinese-English participants do not share. It not being observed in the Chinese-English group could also be attributed to the non-alphabetic orthographic L1 and an alternate strategy being used by such ESL readers. As for this effect not being found in the English group, it is likely another case of the stimuli not being taxing enough for native readers or, indeed, that Spanish-English participants do have a different processing strategy. As previously discussed, however, the two tasks (oLDT and pLDT) are not equally difficult and, indeed, the pLDT is significantly more cognitively taxing (as shown in behavioural measures). Therefore, the increased N100-FC activity could instead reflect this increased difficulty of task.

Bilateral N100-FC was also larger to pLDT in Spanish-English than English, showing increased activity during a phonological task in a group with an orthographically shallow phonological L1 compared with native readers. Furthermore, this was observed at a cluster associated with phonological processing and in the same early timeframe as the controversial early phonological activation found in previous studies. In terms of latency, bilateral N100-FC to oLDT was earlier in Spanish-English than English. This could be indicative of the Spanish-English group being more sensitive to orthography-phonology processing due to the shallow orthographic and phonological nature of their L1.

Later in the ERP timeline, there were no significant task-level results involving Group for N170-OT amplitude or latency, meaning that processing associated with this timeframe/cluster combination was not observed within any group in this study. More specifically, no overall distinction between orthographic processing in the oLDT and phonological processing in the pLDT was found for N170-OT, suggesting that these operationalizations of psycholinguistic processes are not sufficient to affect N170-OT amplitudes or latencies during VWR.

6.6 Study summary

Several findings – behavioural and electrophysiological – speak to the distinction between native monolingual (English), alphabetic L1 late bilingual (Spanish-English), and non-alphabetic L1 late bilingual (Chinese-English) groups in terms of VWR processing. Results suggest that ESL groups employ different cognitive strategies for VWR, particularly when reading pseudohomophones, which appears to relate to the language profiles of the ESL groups.

Following previous conclusions from the literature (e.g., McNorgan et al., 2015), results suggest that traditional lexicality effects (real words vs pseudowords) have an orthographic basis and do not require phonological processing. Despite varying reports of larger and smaller amplitudes to different linguistic stimuli, P1-O amplitude was not observed to differ between oLDT conditions (RW,PH1; within groups) or between groups and was not observed to reflect early orthographic or, indeed, psycholinguistic processing more generally. There was also a stark lack of differences in N170-OT amplitude within or between tasks despite the various reports in the literature of the psycholinguistic processing sensitivities of the occipitotemporal N170. N170-OT amplitude measures from the current study, therefore, do not lend further support to the psycholinguistic associations of the component. N170-OT latency, however, did show an effect of orthographic lexicality, being earlier to RW than to PH1 in the English group and echoing the faster response times to RW than to PH1. Orthographic lexicality effect was only observed in the alphabetic L1 groups, relating to language profile and the lifelong everyday usage of an alphabetic script in English and Spanish-English groups. Meanwhile, effects of phonological lexicality reflecting an increased effort for pseudoword processing relative to pseudohomophones were highlighted in the pLDT by larger bilateral P1-OT and N100-FC amplitudes to PW than PH2 in the Spanish-English group. Furthermore, this effect distinguished between groups, as it was not observed in English or Chinese-English, only in the Spanish-English ESL group, likely connected to their alphabetic and orthographically shallow L1 background. Considering top-down effects as per Twomey et al. (2010) and comparing across orthographic and phonological tasks, different patterns of P1-OT lateralization and effects between pseudohomophones (PH1,PH2) once again distinguished between groups.

This concludes Study 1, which investigated orthographic and phonological processing in native and ESL readers with a two-task approach involving separate orthographic and phonological lexical decision tasks. The following chapter will document Study 2, which approached orthographic and phonological processing in native and ESL readers with a single behavioural task (rhyme recognition) as a complement and contrast to Study 1 and an alternate perspective.

Chapter 7: Orthographic and phonological processing of rhyme recognition in ESL readers with different L1 profiles (Study 2)

7.1 Overview

This chapter will document the background, rationale, method, and results of the second study of the thesis, ending with a discussion of findings. Study 2 was designed to directly complement Study 1 (5.9) in terms of the overall research focus through an additional perspective on orthographic and phonological processing during VWR. As discussed in section 5.3.1, Study 2 complements Study 1 by integrating both orthographic and phonological experimental manipulations in the same behavioural task (visual rhyme judgement), contrasted with using separate tasks as in Study 1, and uses real English words only (no pseudohomophones or pseudowords as in Study 1). These factors provide alternative perspectives more akin to natural language processing that are critical for the outcomes of Study 2 to be considered alongside findings from Study 1 for a clearer overall understanding.

Visual rhyme judgement tasks require participants to decide whether pairs of visually presented words rhyme or not. As will be outlined in more detail later in this chapter, such tasks centre on a form of phonological priming, relying on memory processes to rehearse or recall the first word of the pair (the prime) for phonological comparison with the second word of the pair (the target). The specific rhyme judgement task used in Study 2 takes a step further to manipulate orthography and phonology separately and interactively, employing an orthogonal design operationalized through orthographically and phonologically congruent and incongruent rhymes and non-rhymes. Effectively, word pairs can look like they rhyme and do rhyme (P+O+[4] e.g., file-MILE), look like they rhyme but do not rhyme (P-O+ e.g., worm-FORM), rhyme but do not look like they rhyme (P+O- pairs e.g., goal-BOWL), and do not look like they rhyme and do not rhyme (P-O- e.g., joke-GATE). These combinations of orthographic/phonological relationship between word pairs provide the means to observe the necessary orthographic and phonological processing, respectively, within and between groups.

More specifically, Study 2 examines behavioural and early pre-200ms ERP responses to orthographically and phonologically congruent and incongruent word pairs with contrasts between native (English monolingual), non-native alphabetic L1 (Spanish-English bilingual), and non-native non-alphabetic L1 (Chinese-English bilingual) groups of English readers. In line with the aims of the thesis, these distinct ESL populations are the focus to investigate how L1 profiles (as discussed in §1.1) can contribute to reading English as a second language, as well as contrasting with native-level reading of English. Alongside contrasts of behavioural performance (via accuracy and response times), the main aims of the study are to provide evidence pertaining to whether processing associated with the occipital/occipitotemporal P1 contributes to orthographic (or phonological) processing, the possibility of early ~100ms frontal-central phonological activity (as in Study 1), and the orthographic and/or phonological nature of the

[4] **P** = Phonological, **O** = Orthographic; **+** denotes congruence, **-** denotes incongruence; e.g., P+O+ is orthographically and phonologically congruent e.g., file-MILE, while P-O+ is orthographically congruent, but not phonologically congruent e.g., worm-FORM.

occipitotemporal N170 during VWR (also as in Study 1), each with emphasis on how observations differ across groups.

7.2 Background

When manipulations of orthography and phonology are orthogonal, phonological similarity is typically more influential to word processing than orthographic similarity (Alario et al., 2007), suggesting that, all else being equal, phonological activation carries more weight than orthographic activation. What this entails for both L1 and L2 readers, especially when orthographic and/or phonological aspects of a bilingual language profile deviate from the target language (e.g., English from Spanish or Mandarin), however, is not yet fully understood (K. I. Martin, 2017). For instance, an alphabetic but orthographically shallow L1 as in Spanish-English bilinguals and a logographic L1 as in Chinese-English bilinguals are likely to have distinct influences on L2 reading of English. Furthermore, phonological effects enduring regardless of ostensible orthographic, morphological, and syntactic/phonotactic confounds, as found between heterographic homophones (e.g., *court-caught*) and pseudohomophones (e.g., *cawt*), suggests that phonological similarity must be more salient and influential than orthographic (Alario et al., 2007). However, for fairer estimations of the relative strengths of orthographic and phonological effects, such orthographic differences, including potential differences in n-gram frequency as well as visual familiarity and complexity need to be acknowledged (as in the current research), as they could influence orthographic and phonological processing differently (Hauk et al., 2006). Nonetheless, this circumstance of enduring phonology can be likened to orthographically incongruent but phonologically congruent (e.g., rhyming) pairs, such as *bass-face* in that phonology is shared between words, while orthography is largely not. However, this contrasts with the effect that occurs with orthographically congruent and phonologically incongruent word pairs as used in the current study. For example, *couch* primes target *touch*, which works as a distractor and the orthographic similarity does not facilitate but inhibits recognition due to the combination with phonological incongruence (Alario et al., 2007; Ferrand & Grainger, 1994). Inhibitive effects in such orthographically-similar non-rhymes (*couch-touch*) suggests that the orthographic code can be as strong and sometimes stronger than the phonological, even when phonology is required for the task. Overall, this provides support for the role of phonology in VWR and that the underlying orthographic and phonological mechanisms could be investigated further by orthogonally manipulating orthographic and phonological congruence between word pairs, as in visual rhyme judgement.

7.2.1 VWR and visual rhyme judgement

Phonological processing is sufficient for recognition of rhyming words when reading, but, critically, both orthographic evaluation and phonological recoding are necessary: the orthography provides the question in the form of the stimuli asking whether they rhyme, while the internal phonological representation is required for the answer of whether they rhyme or not (Alario et al., 2007; Bitan et al., 2007). While it is, therefore, not disputed that rhyme is fundamentally phonological and that phonological processing will occur when making rhyme judgements (Coch et al., 2008), it must also be acknowledged that visual rhyme recognition also requires orthographic processing of both onset and rime (i.e., the whole word) for successful completion (Bitan et al., 2007). Rhyme is, therefore, a valuable psycholinguistic vehicle for observing explicit processing of both orthography and phonology with any interactions between them during VWR, which would not be the case for a task without overt phonological involvement.

With the phonological basis of rhyme (through the necessary phonological equivalency of rimes between word pairs), effects of rhyme on behavioural responses are often discussed in terms of phonological priming i.e., phonologically matched primes/targets requiring less processing than mismatched pairs (Coch et al., 2008; Khateb et al., 2007). However, this is not to be confused with traditional definitions of priming as in semantic priming, for example, or in other priming paradigms e.g., masked priming. In the context of cognitive psychology, the phenomenon of priming "refers to instances in which prior experience with a specific stimulus influences subsequent behavior [sic] in the absence of intentional remembrance a form of indirect, nonconscious, non-declarative, or implicit memory" (Wagner & Koutstaal, 2002, p28). Based on this definition, cognitive effects requiring more than one related stimulus presented in series might be better considered as working memory effects because the first stimulus of the pair, the prime, is required to be actively held in working memory until the subsequent target stimulus is perceived and a comparison between them made, reflecting "intentional remembrance". In terms of processing, then, it appears the key conceptual difference between traditional priming and priming effects related to working memory concerns the aspects of implicit memory and conscious rehearsal. The key practical difference, therefore, appears to be the task and what the task encourages to happen in the inter-stimulus interval between prime and target in terms of that working memory: if the task is somewhat independent of the prime-TARGET association, such as a lexical decision to the target stimuli, this can be traditional ("nonconscious") priming but if the task is linked to the relationship between "prime" and "target", such as responding whether the prime and target rhyme, it is not traditional priming per se. This can also be conceptualized as whether cognition is working "backwards" (traditional priming, not explicitly aware of needing to retain the prime) or "forwards" (explicit awareness that there is potentially an association between prime and target and, therefore, the need to retain the prime).

The rhyme effect from a rhyme judgement task is not priming in a traditional sense and requires explicit rehearsal and use of the phonological loop (for the correct responses and the effect itself). Citing Brown (1987), Coch et al. (2008, p231) explains that “rhyming judgments [sic] require that the ending sound (rime) of the prime be held in working memory for later comparison to the rime of the target” and that the phonological loop of working memory is used to store and manipulate phonological information. Traditional priming, meanwhile, does not impact accuracy directly because the task for which accuracy is being measured (e.g., LDT for semantic priming) is not directly related to the prime-TARGET association and is an implicit retroactive effect. Whether through rehearsal or recall, the orthography and phonology of the prime is (re)processed alongside that of the TARGET during a rhyme judgement task, resulting in the prime-TARGET congruence effects (in terms of orthographic and/or phonological interference) that are the focus of the current study. Therefore, the term *priming* will still be used in describing the orthographic and phonological relationship between primes and targets but referring to a process of conscious rehearsal and intentional remembrance.

Considering visual rhyme recognition as the retrieval of phonological representations of orthographic inputs and using internal phonological rehearsal or recall mechanisms to compare them, a visual-phonology processing pathway might seem most efficient. However, VWR tasks rooted in phonology, such as rhyme recognition, need implicit orthographic activation for the inherently required phonological access (Alario et al., 2007), so a visual-phonology pathway would not suffice. Essentially, the route to phonology requires initial orthographic processing through an initial visual-orthographic process, as postulated in weak phonological theories of VWR (Rastle & Brysbaert, 2006). Reading the individual (high frequency, familiar, real) words of a rhyming pair, therefore, theoretically follows a direct orthographic-semantic route that activates the corresponding phonology (Coltheart et al., 2001). Such an account might be acceptable if rhyme recognition was truly polarized and did not involve orthography-phonology interactions that have been shown to inhibit behavioural and influence ERP responses. For instance, behavioural and electrophysiological effects have been reported from a range of phonology-oriented stimuli, such as homophones (e.g., Carrasco-Ortiz et al., 2012a), pseudohomophones (e.g., Wheat et al., 2010), and matched word onsets (e.g., Ashby, 2010), as well as rhyming and non-rhyming pairs (e.g., Alario et al., 2007; Grossi et al., 2001; Weber-Fox et al., 2003). The occurrence of such inhibitive phonological effects (e.g., longer RTs, different ERP amplitudes than neutral control conditions) in any paradigm, however, suggests that attempts to access stimulus phonology are made, but processing conflicts, whether between phonology and orthography or phonology and another factor, lead to inferior performance (Van Wijnendaele & Brysbaert, 2002). Phonological activation when reading phonology-oriented stimuli, such as rhymes and pseudohomophones, should supersede any complications for processing, such as incongruent orthography in a rhyming pair. However, phonological activation does not always resolve conflicts with orthographic processing as with orthographically dissimilar rhymes e.g., *gown-own*, *cone-own* (Grossi et al., 2001; Weber-Fox et al., 2003), which speaks to the relative strength of orthography in VWR. Furthermore, if phonology was accessed directly from the visual input, before or without orthographic processing, the result would be sufficient for such phonological decisions as rhyme judgement to be successfully completed without any interference from orthography. However, orthography-phonology conflicts do occur (Alario et al., 2007; Weber-Fox et al., 2003), implying that responses are based on a preliminary level of orthographic processing that precedes full phonological activation.

7.2.2 Visual rhyme judgement effects

Effects of rhyme on behavioural and electrophysiological measures during VWR are well-documented, manifesting as faster behavioural responses to rhyming pairs and ERP differences between rhyming and non-rhyming pairs (Coch et al., 2008; Khateb et al., 2007; Weber-Fox et al., 2003). Observing an overall rhyme effect, such as faster responses to rhymes than non-rhymes (Coch et al., 2008; Khateb et al., 2007; Weber-Fox et al., 2003), requires only straightforward rhyming and non-rhyming stimulus pairs. For instance, a simple task could just include orthographically-similar rhymes (P+O+, e.g., file-MILE), which are fully congruent and involve cumulative orthographic and phonological priming, and orthographically-dissimilar non-rhymes (P-O-, e.g., joke-GATE), which do not involve any elements of explicit orthographic or phonological priming. Neither P+O+ or P-O- rhyme pairs, however, evoke the processing conflicts of pairs that are overall incongruent in the sense of either their orthography or phonology does not agree with the rhyme/non-rhyme status. For instance, in an orthogonal design of rhyme judgement task, there would also be the orthographically-dissimilar rhyme pair (P+O-, e.g., goal-BOWL), with its combination of phonological priming and orthographical incongruence, and the orthographically-similar non-rhyme (P-O+, e.g., worm-FORM) with its inhibitive orthographic priming in a non-rhyming pair (Weber-Fox et al., 2013), both of which provide direct contrast with respective congruent controls (P+O+, P-O-). The incongruent conditions (P+O-, P-O+) are particularly important due to the inherent emphasis on phonology in rhyme judgement that can neither be mediated nor fulfilled based on the degree of orthographic correspondence between primes and targets (Bitan et al., 2007). Orthographically-dissimilar rhyming pairs (P+O-) have little or no orthographic correspondence, potentially suggesting an incorrect non-rhyme response, while orthographically-similar non-rhyming pairs (P-O+) have almost complete orthographic correspondence, leading toward an incorrect response that they rhyme. Furthermore, this cognitive conflict from incongruent stimulus pairs, such as *bass-face* (P+O-) and *couch-touch* (P-O+), illustrates an intrinsic link between the orthography and phonology of words in terms of the way they are read, showing how orthographic priming can be facilitative or inhibitive to phonological processing (Coch et al., 2008).

All conditions, congruent or not, rhyming or not, require internal phonological evaluation, which is emphasized by the orthogonal design. Theoretically, however, only the P+O- pairs have an orthographic reason to be checked, while P-O+ pairs, due to the orthographic similarity that lulls readers into a false sense of security, do not. The increased processing complexity of orthographically-similar non-rhymes (P-O+) in particular could be described in terms of inhibitive priming due to the initial perception that the words are phonologically similar based on orthographic similarity and pre-phonological visual/orthographic processing, which then conflicts with phonological evaluation (Weber-Fox et al., 2013). This implies serial processing, following the concept of extracting phonology from the orthographic input and assuming orthographic processing to occur before phonological (see §4.4). While orthography is the more immediately salient stimulus property (and whether orthographic processing precedes or occurs alongside phonological processing), it is likely to have the initial influence on processing, even if phonological processing has a stronger influence in later timeframes (Grossi et al., 2001; Khateb et al., 2007; Weber-Fox et al., 2003). Incorrect responses are, therefore, likely made based on the orthographic relationship between stimuli that misleads the reader into thinking that the words sound similar (as in P-O+) or that they do not sound similar (as in P+O-), which points to orthography having the strength to influence processing in a phonological task (Weber-Fox et al., 2003). Furthermore, as phonological congruence typically has a stronger influence on word processing than orthographic congruence when orthographic and phonological factors are manipulated orthogonally (Alario et al., 2007), the incongruent conditions are likely to have differential effects in VWR. For instance, P+O- is likely to have a stronger and facilitative effect, while P-O+ is likely to be the most inhibitive and disruptive to VWR (Grossi et al., 2001; Khateb et al., 2007; Weber-Fox et al., 2003).

When orthographic evaluation does not match the phonological conclusion, as in rhyme judgement of orthographically-dissimilar rhymes (P+O-) and orthographically-similar non-rhymes (P-O+), behavioural performance, as observed through significantly longer response times and higher error rates compared with other conditions (Weber-Fox et al., 2003), is worse than for congruent conditions (P+O+, P-O-) due to the orthography-phonology conflict that needs to be resolved (Rugg & Barrett, 1987). Not only are responses to P-O+ pairs notably more difficult than to P+O- in phonology-based VWR tasks (e.g., rhyme judgement, LDT with pseudohomophones), such incongruence has also been found to increase activation of both orthographic and phonological systems, initiating a "battle" between them (Bitan et al., 2007; Pas et al., 2016), resulting in modulations of later ERPs (early N450, Grossi et al., 2001; N450, Khateb et al., 2007; N350, Weber-Fox et al., 2003). Therefore, as orthography and phonology do interact during VWR cognition, questions concern how and when conflicts occur. These questions not only extend to bilingual readers for the same reasons as native monolingual readers, but for the additional perspective of how orthographic and/or phonological aspects of bilinguals' L1 profiles (e.g., phonological L1 as in Spanish-English or logographic L1 as in Chinese-English) might influence L2 processing in terms of behavioural performance and brain activity during VWR.

As visual rhyme processing must logically be based on orthographic evaluation to some extent, incongruent orthography (in terms of P+O- and P-O+, where orthography does not "agree" with the phonology and, thus, the correct rhyme decision) could interfere with a direct orthographic-semantic or orthographic-phonology route. It is, therefore, not unreasonable that complementary parallel processes, here referring to the potential for the orthography and phonology of the target to be processed in parallel as opposed to in series (i.e., orthography→phonology) could ease such cognitive conflicts. For instance, a visual/orthographic pathway to phonology (Grainger & Holcomb, 2009), parallel orthographic/phonological processing (McClelland & Rogers, 2003), or grapheme-phoneme conversion (Coltheart et al., 2001) in some cases could be used to resolve the orthography-phonology conflict through additional evaluation of the input stimulus and consequent provision of more evidence to influence VWR and the cognitive/behavioural response. The orthogonal design of the rhyme judgement task with all four permutations of orthographic and phonological congruence (P+O+, P+O-, P-O+, P-O-) allows observation of which orthographic and/or phonological element(s) of the prime-TARGET associations influence responses to the target. Effects of incongruent conditions in visual rhyme recognition (e.g., P+O-, P-O+) could be the result of a lack of cooperation in parallel processing stemming from the disagreement between orthographic activation and the phonology-based correct rhyme decision for the target stimulus (Cornelissen et al., 2009; Grainger & Holcomb, 2009). Accounts involving parallel processing e.g., BIAM (Grainger & Holcomb, 2009) and/or orthography-semantic routes that also activate phonology e.g., DRC (Coltheart et al., 2001) might both predict the processing conflicts of the incongruent conditions to occur within a similar timeframe. However, this assumes orthographic and phonological processing to require similar cognitive resources in terms of importance to the task at hand or extent of cognitive workload, which may not be the case and is, therefore, an important question for the current research.

7.2.3 ERPs and effects of rhyme

Previous studies using the rhyme judgement paradigm (comparing participant responses to rhyming and non-rhyming word pairs) have, understandably, used it almost exclusively for investigating rhyme recognition and, in terms of ERP studies, the later

N350/N450 components. Principally, non-rhyming word pairs have consistently been reported to elicit larger amplitudes when compared with rhyming pairs (Coch, Hart, & Mitra, 2008; Khateb et al., 2007), sometimes regardless of orthographic similarity (Botezatu et al., 2015) in the 300-500ms timeframe associated with N350, N400, and N450 components. Language-related brain activity underlying visual rhyme recognition, however, is not rhyme-specific and the language-related brain activity associated with rhyme recognition, such as the N450, is attuned to (and arguably dependent on) earlier orthographic and phonological processing mechanisms (and interactions between them) that underpin the aforementioned facilitative or inhibitive effects of rhyme (Bitan et al., 2007; Rugg & Barrett, 1987; Weber-Fox et al., 2003). It is the combination of underlying and/or preceding processing and brain activity that culminates in any differential behavioural effects and sufficient work has not been done on ERP activity that potentially culminates in explicit phonological processing outputs, such as rhyme recognition. In other words, visual rhyme recognition is based on the processing of interactive visual, orthographic, and phonological inputs that must converge for the resulting recognition of a rhyme. Therefore, it is not unreasonable to accept that later ERP components reflecting rhyme recognition (e.g., N450) would be influenced by earlier orthographic and/or phonological processes not specifically related to rhyme recognition. Moreover, orthographic and phonological processing, including the resolution of conflicts between them, have been posited to occur much earlier than the N350/N450 timeframe (e.g., Klein et al., 2015; Pammer et al., 2004; Wheat et al., 2010) with phonological effects prior to 200ms influencing real-time word recognition processes, while effects at such later ERP timeframes as N350 and N450 correspond with phonological integration at a sentence level (Ashby, 2010).

ERP studies using visual rhyme recognition consistently report a selection of later (post-300ms) components associated with rhyme processing (e.g., N350, N450), where non-rhyming pairs are found to elicit smaller N350 and larger N450 amplitudes than rhyming pairs (Grossi et al., 2001). The rhyming effect on the N450 has also been found to be less pronounced to orthographically similar alphabetic word pairs (Chen et al., 2010), suggesting it to be affected by orthography as well as being an index of phonological processing (Botezatu et al., 2015). Modulations of central and parietal N450 and left frontal-central N350 amplitudes have also been attributed to interactions between orthography and phonology in both native (Khateb et al., 2007) and L2 readers e.g., Chinese-English bilinguals (Chen et al., 2010), suggesting potential similarities of rhyme processing in L1 and L2 reading of English (e.g., Weber-Fox et al., 2003). This provides important rationale for extending the investigation of ERP components preceding these N350/N450 effects (e.g., P1, N170) to L2 readers of English and from populations with significantly different language profiles (Spanish-English and Chinese-English, respectively). In particular, pre-200ms ERP activity can be examined for evidence of how similar (or different) such early orthographic and/or phonological processing is within and between native English and ESL groups during VWR. For instance, similar ERP activity between groups would signify a similar VWR approach to understanding written English, whether as L1 or L2, while distinct ERP patterns could show a tendency (or not) for L1 processing strategies (i.e., accommodation) over learned L2 methods (i.e., assimilation). More specifically, occipital/occipitotemporal and frontal-central activity (i.e., P1-O/P1-OT and N100-FC) will be examined between groups to see how such early initial processing of orthography and phonology, respectively, might manifest in readers with different language profiles.

The task methodology of Botezatu et al. (2015), though concentrated on the N450 rhyme effect (and, therefore, with a very different focus in terms of ERPs and analysis), was almost identical and tested a very similar set of groups to the current study (English monolinguals, Spanish-English, and Chinese-English, but early not late bilinguals), so provides a usable benchmark. No rhyme effect was observed between conditions that used orthographically similar words (e.g., *right-fight*, *dough-cough*; i.e., P+O+ and P-O+) on ERPs between English, Spanish-English, and Chinese-English groups (as in the current research, albeit with more fluent early bilinguals), which was concluded to mean that the phonological processing underlying rhyme judgement is similar across L1 and L2 readers of English (Botezatu et al., 2015). In the same study (Botezatu et al., 2015), the rhyme effect on N450 amplitude from orthographically dissimilar pairs (P+O-) was shown to be modulated by the factor of group. This N450 rhyme effect was larger in English than Chinese-English, but no different between Spanish-English and either other group, which was taken as evidence for the N450 being a phonology-based index of language proficiency due to the bilingual groups in particular not being matched in English proficiency and the Spanish-English participants having the advantage of alphabetic experience (Botezatu et al., 2015). However, which aspects of L2 proficiency, such as the extent that L1 orthography and phonology can influence L2 phonological processing, and more broadly how an alphabetic or non-alphabetic (e.g., logographic) L1 profile can influence L2 VWR are still unclear. These findings highlight the importance and potential influence of orthography on phonological processing in phonological tasks, but the necessary brain activity preceding rhyme effects remains to be sufficiently investigated, especially concerning its similarity in L1 and L2 readers.

Other previous studies also used very similar task methodologies, using all the same orthographic and phonological manipulations for the same conditions, as in the current study. Weber-Fox et al. (2013) was one such study, using English words with native English participants, so the behavioural results that showed most impact from orthographically similar non-rhymes (e.g., gown-own) are at least directly relevant to the current study. However, the ERP focus was on the later N350, amplitudes of which reflecting orthographic and phonological congruence as well as rhyme while latency being left-lateralized across conditions, and, most significantly, analysis used peak measures of amplitude and latency, which should be avoided without very good reason (Luck, 2014; see also §5.3.3.2). Classon et al. (2013) was another study using a very similar task methodology (in its short ISI iteration of the rhyme task), showing higher error rates in incongruent than congruent pairs and in P+O- than **P−O+** with no mention of response time differences. However, the study used Swedish stimuli and Swedish participants (i.e., very different language profiles), and the ERP focus was also mainly on later activity (N400, N450, P600) and, again, using peak measures. That said, a 100-300ms ("N2") window was considered (Classon et al., 2013), though was only examined at centro-parietal sites not involved in the current research and no effects were found in the comparable short ISI study, so the stretch of inference required would be unwise.

In a potentially comparable study with a directly relevant sample of Chinese-English late bilinguals, Chen et al. (2010) observed shorter response times for rhyming than non-rhyming English word pairs as well as ceiling effects for accuracy in a basic rhyme judgement task using English stimuli. However, there was no manipulation of orthography, just phonology via rhyme and non-rhyme conditions using only orthographically incongruent conditions (P+O- and P-O-) and the sole ERP focus was on the later N450. Similarly, Coch, Hart, and Mitra (2008) also used a more limited rhyme judgement paradigm without orthographic manipulations and observed larger N400/N450 amplitudes for rhyming than non-rhyming word pairs in native English participants. Unfortunately, response times were not recorded and, furthermore, the ERP focus was strictly on the N400/450 and CNV (contingent negative variation) components and, therefore, only on frontal and central sites, respectively, and at later timeframes. Lastly, the participants and stimuli in Khateb et al. (2007) were French, but a comparable rhyme paradigm was used and there is some linguistic overlap between both English and Spanish with French, respectively, so the observation of larger left fronto-temporal N350 to non-rhyming than rhyming pairs could still be worth considering, especially in relation to the early left frontal-central activity under scrutiny in the current research. However, in addition to the ERP focus again being on the N350/N450, only phonologically related (i.e., rhyming) and phonologically unrelated (i.e., non-rhyming) conditions were used and it was unclear about orthographic control or manipulations of conditions, leaving meaningful inference about the current study difficult.

It is important to reiterate that these studies and their conclusions were based on activity of the later largely rhyme-specific N450 component (~450ms post-stimulus) and the destination does not denote the route taken. As discussed earlier, it is still not known what happens in terms of orthographic and phonological processing underlying the associated rhyme effects, providing a valid timeframe before this N450 activity to explore that coincides with the first ~200ms post-stimulus known to be vital in VWR. The nature of rhyme-specific activation (e.g., N450 at ~450ms) is relatively late in terms of typical VWR timing (see §1.2 and §4.4.1), thus allowing the study of earlier processes without impact of the later rhyme effect and focusing instead on the stimulus-level orthographic and phonological foundations of VWR through earlier ERP activity e.g., P1-OT, N100-FC, N170-OT. Activity up to ~200ms has been linked with both orthographic and phonological processing (Holcomb & Grainger, 2006), effects of which have been shown to be dependent on the type of priming, where modulation from orthographic priming through orthographically similar primes and targets was found to occur earlier than from phonological priming from phonologically similar primes and targets (Grainger et al., 2006). Broadly coinciding with either P1 or N170 timeframes, occipital activity at 50-175ms ("P120") and 150-250ms ("N180") and temporal activity at 75-200ms ("N120") have been reported to both primes and targets in visual rhyme tasks (Grossi et al., 2001). Furthermore, ERP activity leading up to ~200ms in silent reading activities has been suggested to reflect the detection of conflict between sublexical orthography and phonology based on early analysis of the visual input and violations of expectations of orthographic form (Classon et al., 2013), which fits with the kind of processing required for incongruent rhyme/non-rhyme pairs. Importantly, the activity that underlies such an index of phonological processing as the rhyme effect is the activity that is not only critical to understanding how native readers process written language, but how bilinguals with different L1 profiles read their L2. However, the details of how this earlier activity contributes to the recognition of orthography and phonology before rhyme processing is not yet clear and whether such earlier electrophysiological brain activity shows similar patterns to the documented later rhyme-related components is germane to the current research. Further investigation into the early orthographic and phonological processing that precedes the N350/N450 rhyme effects in both L1 and L2 readers is, therefore, required.

The early timeframe focus of the current research follows the argument that there must be processing of and, possibly, interactions between orthography and phonology earlier than the aforementioned N350/N450 rhyme effects to allow the overall rhyme effect to exist. Considering rhyme effects in terms of cognitive mismatch (either just between phonological codes of prime/target or also involving orthographic congruency), internal orthographic/phonological processing during visual rhyme judgement could be related to mismatch negativity observed on earlier components (e.g., Froyen et al., 2010; Frances et al., 2021), where an oddball stimulus interrupts the continuous processing of a repeating stimulus or stimulus type (e.g., the letter *n* presented repeatedly, interspersed with a pseudorandom letter *x*). However, there are distinct differences in the underlying processes involved in typical mismatch paradigms (e.g., visual and auditory are not orthographic and phonological, respectively). In particular, phonology-related effects from tasks involving the auditory modality have been shown to influence such early <200ms EEG/ERP processing, potentially suggesting that comparable influence could be observed in VWR. Furthermore, it should be emphasized that orthographic, phonological, and rhyme effects do also occur in L2/ESL readers e.g., Botezatu et al. (2015), so such early activity also requires investigation in L2 readers.

7.2.4 The current study

The theoretical rationale for Study 2 follows the main aims of the thesis in terms of observing the nature of early pre-200ms orthography- and phonology-related processing and how it might differ based on orthographic and phonological aspects of distinct language profiles (English, Spanish-English, and Chinese-English). In terms of methodology, the current study is very similar to the aforementioned Botezatu et al. (2015) study, which used an ERP approach to observe rhyme effects on N450 amplitude in English monolinguals alongside Spanish-English, and Chinese-English bilinguals via a rhyme judgement task that also manipulated orthographic congruence between word pairs. The key departures from prior research of the current study are the focus on amplitudes and latencies of much earlier pre-200ms ERP activity, following the objectives of the thesis, in addition to the novel methodological perspective of incorporating all orthographic and phonological manipulations into one single orthogonal paradigm and analysis instead of analyzing orthographically similar and dissimilar word pairs separately (where all pairwise contrasts could not be examined). The use of a rhyme judgement paradigm to investigate underlying and early orthographic and phonological processing is a major part of the original contribution to knowledge for the current research. It is also important to note that the hypothesized priming involved only real English words in both prime and target positions and whole-word conscious online processing, as opposed to masked or partial priming as in Holcomb & Grainger (2006), Wheat et al. (2010), and others.

While rhyme itself may not be the crux of VWR or reading, the orthographic and phonological activation required for its recognition form the backbone of written language comprehension. It is, therefore, pertinent to examine how these processes fare in ESL readers with different language profiles in contrast to the native group, allowing insight into all groups' VWR processes. Consequently, whether orthographic and phonological priming manifest similarly in native and ESL readers is of particular interest.

7.2.4.1 Hypotheses

As in Study 1, such novelty of method and specific aims has a side-effect that no prior research precisely fits the bill for hypothesis creation or direct comparisons of results. For instance, prior rhyme-based ERP studies in VWR (e.g., Botezatu et al., 2015; Coch et al., 2008; Khateb et al., 2007; Weber-Fox et al., 2003) used similar task methodologies, but different analysis approaches with different research objectives and none with the focus on early ERP measures as in the current study. However, elements of some previous studies can still be called upon for insights if not direct comparisons to form broad hypotheses.

In terms of behaviour, more errors and longer response times are expected from orthographic priming than phonological priming that will be stronger and more influential on processing e.g., larger effect sizes. Non-rhymes with congruent orthography (P-O+) will pose most difficulty for all groups, followed by rhymes with incongruent orthography (P+O-), resulting in slower responses with more errors than the relative controls (P+O+,P-O-). Based on bilinguals being more reliant on word-level processing in L2 as opposed to sublexical processing (Diependaele, Duñabeitia, Morris, and Keuleers, 2011), it is reasonable to consider the ESL groups having fewer issues with the orthographically congruent non-rhymes (P-O+) and orthographically incongruent rhymes (P+O-). However, the mismatch between orthographic congruence and the correct answer for the task is anticipated to be sufficient to make these conditions difficult for both native and ESL groups. Evidence for L1-oriented processing in the ESL groups is expected to be observed through orthographically similar non-rhyming pairs being more cognitively demanding for the Spanish-English group and orthographically dissimilar rhymes will disadvantage the Chinese-English group.

Behavioural responses are expected to be faster and more accurate for orthographically congruent rhymes (compared with non-rhymes) and orthographically incongruent non-rhymes (compared with rhymes) with the latter being the least cognitively taxing overall (Botezatu et al., 2015). Taking behavioural performance further, it is expected that the P-O+ condition will cause the greatest orthography-phonology conflict, which will be clearly observed in the behavioural results, followed in terms of difficulty by the P+O- condition. The congruent conditions (P+O+ and P-O-) are not expected to present difficulties in responses and there is little reason to suggest that patterns of behavioural results will differ between groups for these relative control conditions.

Regarding differences between groups, the relatively shallow orthographic depth and phonological nature of Spanish and the related greater reliance on spelling-sound associations (Botezatu et al., 2015), orthographic congruence in non-rhymes (as in P-O+) is expected to be most problematic to Spanish-English readers of English. By extension, it is reasonable to consider that effects associated with VWR found in Spanish-English readers of English reflect orthography-phonology conversion processes. Furthermore, readers with an orthographically deep or complex L1 (e.g., English, Chinese-English) are less likely to use such an assembled strategy, instead using a whole-word addressed approach (Botezatu et al., 2015). It should be noted, however, that such a whole-word strategy requires sufficient proficiency and experience, meaning it is imperative for groups to be matched on these factors as in the current research if credible comparisons of orthographic and/or phonological processing are to be made.

The nature of the relationship between behavioural and ERP effects in such conditions is unknown, but it is expected that the N170 at least in the native group will reflect stimulus-related orthographic/phonological processing required for the task. Phonological priming (through the contrast between P-O- and P+O-) will be reflected in ERPs at ~100ms in the native group. This could also be observed in the Spanish-English group if a phonological approach to ESL reading is not used and in the Chinese-English group if an orthographic/whole-word approach to ESL reading is used. Incongruent orthography (O- conditions) will be reflected in the N170, potentially in all groups due to its importance in orthographic processing and the necessity of phonology for the task, as well as the stimuli all being real words that are familiar to all groups (cf. expertise).

In terms of ERPs, the well-documented effects of rhyme on the later N450 component, where amplitudes are higher for non-rhymes than rhymes (e.g., Botezatu et al. 2015), along with the postulation that brain activity preceding this timeframe (~300-500ms) is involved with orthography and/or phonology make it reasonable to consider that earlier ERP responses (e.g., at ~100ms and/or ~170ms) will also reflect differences based on the orthographic and/or phonological processing required (dependent on condition). Due to more directly comparable rhyme-based studies not investigating the same groups and/or, vitally, the same ERP timeframes/clusters as the current study, hypotheses of ERP differences will be kept broad and two-tailed. They will be based on the associations of pre-200ms ERPs (e.g., P1-O, P1-OT, N100-FC, and N170-OT) and orthographic and/or phonological processing using different task methodologies. For instance, the aforementioned masked priming studies that highlighted the early ~100ms phonological activation and serve as rationale for the N100-FC investigation in the current research all allow relevant inference (e.g., Klein et al., 2015; Wheat et al., 2010). It is, therefore, hypothesized that phonological effects will be observed on N100-FC in English and Spanish-English groups (due to lifelong experience of an alphabetic script and grapheme-phoneme conversion), following Wheat et al. (2010), Pammer et al. (2004), and Klein et al. (2015). Based on its involvement in orthographic-phonological mapping when reading alphabetic scripts (Amora et al., 2022; Dien, 2009; Yum & Law, 2021), N170-OT responses in the native group are expected to show differential responses to orthographic and phonological congruence and phonological effects are also anticipated across groups unless VWR strategies are similar between groups. Orthographic priming, meanwhile, is hypothesized to influence P1-O and P1-OT measures across groups, while orthographic incongruence is also expected to be pertinent to these and the N170-OT. Between groups, differences in ERP measures are expected due to the distinctions in language profile between the groups, reflecting greater sensitivity to phonology in the Spanish-English and an orthographic focus in the Chinese-English groups. It is hypothesized that N100-FC and N170-OT will show phonology-related effects in the Spanish-English group and that orthography-related effects on P1-O and P1-OT will be found in the Chinese-English group. Difference in amplitude or latency would support the notion that the group with an orthographically shallow L1 (e.g., Spanish) processes alphabetic scripts differently than readers with a more deeply orthographic L1 (e.g., English, Chinese).

7.3 Method

7.3.1 Participants

Participants for Study 2 were the same as for Study 1 (see §5.2 for details), data from each participant being collected for both studies in the same session.

7.3.2 Design

As outlined in section 7.1 and in more detail throughout 7.2, Study 2 used a rhyme recognition task (RRT) based on phonological priming in which the participant must indicate whether each presented stimulus pair rhymes or not. The variant of RRT used in this research builds on the basic polarized rhyme design (rhyme vs non-rhyme) by using an orthogonal design, the framework for which has been established in the literature (e.g., Classon et al., 2013; Grossi et al., 2001; Weber-Fox et al., 2003). This task methodology is used to focus on orthographic and phonological priming, as the main psycholinguistic phenomena, along with the potential interaction and conflict between orthography and phonology. Table 35 illustrates the four conditions, which represent the within-participants factors of phonological (P) and orthographic (O) similarity (+) and dissimilarity (-), operationalized through orthographic and phonological priming between forty stimulus pairs of real English words per condition (see Appendix M for full stimulus list). Effectively, P+O+ stimulus pairs look like they rhyme and do rhyme (e.g., file-MILE), P-O+ pairs look like they rhyme but do not rhyme (e.g., worm-FORM), P+O- pairs do not look like they rhyme but do rhyme (e.g., goal-BOWL), and P-O- pairs do not look like they rhyme and they do not rhyme (e.g., joke-GATE).

Table 35: Characteristics of RRT conditions with stimulus examples

	Phonologically congruent (**P+**)	Phonologically incongruent (**P-**)
Orthographically congruent (**O+**)	**file MILE**	**worm FORM**
Orthographically incongruent (**O-**)	**goal BOWL**	**joke GATE**

Forming an orthogonal design by employing orthographically-dissimilar rhymes (P+O-), orthographically-similar non-rhymes (P-O+) in addition to the standard rhymes (P+O+) and non-rhymes (P-O-) allows the relative strengths of orthographic and phonological processes to be explored concurrently to provide insight into the extent to which they are used in VWR. Such a design allows direct focus on orthographic and phonological processes through the facilitative orthographic or phonological priming or inhibitive orthographic effects they incur (Rastle & Brysbaert, 2006; Riele et al., 1996; Seidenberg & Tanenhaus, 1979). These are reflected in any rhyme-related facilitative or inhibitive effects that can be compared directly (Alario et al., 2007; Ashby, 2010; Van Orden, 1987).

As in Table 35 and Figure 22, primes were presented in lower case and the targets in upper case to reduce potential effects physical familiarity, visual repetition, and visual coding (Baddeley et al., 2002; Landi & Perfetti, 2007), which could and likely would confound results (Holcomb & Grainger, 2006). For instance, orthographically and phonologically congruent pairs in lower case (e.g., file, mile) would also be visually congruent, which the mismatch of letter case helps to prevent, leading to all pairs being visually incongruent.

Additionally, the task was implemented across two separate variants (RRT-A and RRT-B), the order of which was fully counterbalanced within each group and participants completed the Study 1 tasks in between, which acted as a palate-cleanser for the iterations of RRT (though all participants had sufficient breaks between as well as within tasks). The RRT-A and RRT-B variants differed only in that the stimulus pairs were reversed in one compared to the other in order to counterbalance position and (lower/upper) case per stimulus (Vergara-Martíneza, Perea, & Leone-Fernandez, 2020). For example, *file-MILE* in one variant became *mile-FILE* in the other, therefore using the same stimuli in both possible prime and target positions. Although repetition of stimuli might appear methodologically ill-advised due to potential priming effects conflicting with performance and results (indeed, repetition of phonological forms was removed from the Study 1 stimuli), carefully controlled and purposeful repetition of stimuli across trial blocks can be a methodological boon. Wheat et al. (2010), for instance, employed the same target stimuli three times per participant within a study. This approach can radically improve signal-to-noise ratio due to doubling (or tripling in the case of Wheat et al. (2010) the number of trials and therefore increasing internal reliability and validity of the findings (Luck, 2004). Furthermore, using stimuli in both prime and target positions (in different trials) can also control any effects of stimulus order within trials (Luck, 2004). As mentioned, however, the order of the variants was counterbalanced within each group, which also controls for any potential priming effects between repetitions of stimulus pairs (i.e., between variants). This way, any priming effects occurring between trial blocks and/or visual effects occurring between primes and

TARGETS would be averaged out. Any priming effects between trial blocks would likely be minimal, though, due to being effectively masked by the tasks of Study 1 in between and through the use of common high frequency words (Rugg, 1990). Furthermore, it is expected (due to the relative ease of the task overall, the aforementioned high frequency of the stimuli, and the English language proficiency of the participants) that ceiling effects of accuracy will be observed in most conditions, suggesting that participants already know whether or not the pairs rhyme and therefore that repeating them would not have a significant impact on performance or response. For transparency and to reduce any effects of participants realising the stimuli are repeated, participants were informed that their second RRT trial block would involve the same stimuli. Overall, using the two variants doubles the potential number of trials for the analyses and, therefore, improves the quality of data per participant (increasing the number of trials is important to improve signal-to-noise ratio), while also counterbalancing any adverse effects from using just one of the prime-TARGET orders per stimulus pair.

Figure 22 shows the sequence of each of the 160 pseudorandomly presented RRT trials (40 per condition). The ISI between fixation and prime as well as between prime and target stimuli was pseudorandomly jittered, using a rectangular distribution of times (150, 200, 250, 300, 350, 400ms) that follow the 60Hz refresh rate (16.67ms refresh duration) of the monitor and allow for the recommended minimum SOA of 640ms (Perfetti & Liu, 2005), creating a variable SOA. The mean ITI was 2000ms and the mean total trial duration was 4660ms with a minimum of 4300ms and maximum of 5000ms. Each variant of the RRT was expected to last approximately 14 minutes including two ~1-minute and one ~3-minute breaks.

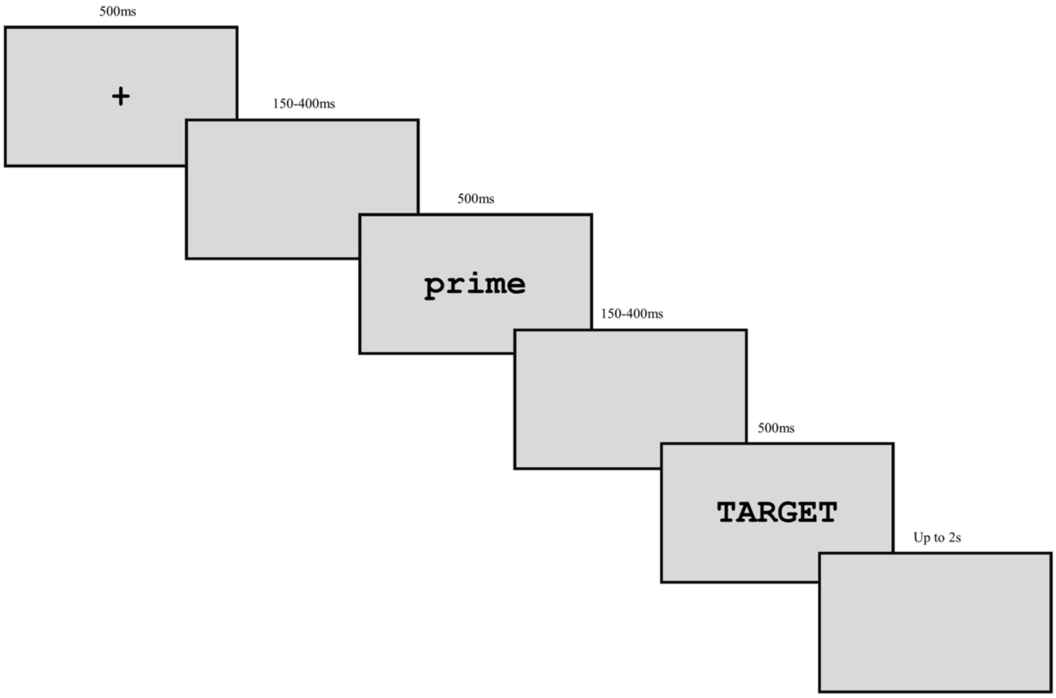

Figure 22: Stimulus presentation format of a single trial in the rhyme recognition task

Contamination and potential confounds can be minimized by matching primes on psycholinguistic factors (see §5.4.1) across conditions. In some paradigms used to examine orthographic and phonological processing, primes are part of the experimental manipulation and have different properties between conditions, as in different stimulus types as primes (e.g., pseudohomophones, Wheat et al., 2010), and onset-based partial orthographic/phonological primes (e.g., Ashby, 2010). This is an important methodological point for the rhyme task in the current study, as both primes and targets are always real words and matched on the aforementioned psycholinguistic factors between all conditions with conditions only differing based on the orthographic and/or phonological relationship between primes and targets within each condition.

As with Study 1, measures of behavioural performance and EEG were recorded to all behavioural targets. Electrophysiological responses to primes were also recorded in order to compute prime-target ERP difference waves for each condition per group as well as a collapsed measure of semi-passive responses to real words. While maintaining important design characteristics concerned with early processes and the ERP methodology (e.g., timing constraints), all efforts were made to make the tasks as straightforward and relatively easy as possible for participants. This was largely because two of the three groups are not native English speakers and the task being unnecessarily difficult could confound results. However, it should be noted that the P-O+ stimulus pairs (e.g., bead-HEAD) are typically more challenging for any participants, native readers or otherwise (Coch, Hart, & Mitra, 2008; Weber-Fox et al., 2013). As well as completing the orthogonal design of the RRT, this condition was important for maintaining attention, helping keep participants focused and avoiding strategy formation. Therefore, while results and responses to this condition are of significant interest to the research, poorer performance is expected and neither analysis nor conclusions rely on it.

7.3.3 Stimuli

The principles of stimulus creation (see §5.4.3) and stimulus matching (see §5.4) were also followed for the stimuli in Study 2, though, due to the design, more details and stipulations are required. Four equal conditions (P+O+, P+O-, P-O+, P-O-) of 40 stimuli each formed the main stimulus sets for RRT-A and RRT-B from 320 different real words that adhere to the stimulus criteria outlined in section 5.4.2 (see Appendix M for the full sets).

7.3.3.1 Stimulus control

Appropriate statistical tests verified that there were no significant differences (α=.05) on a stimulus level between conditions for any controlled psycholinguistic factors (see §5.4.1), as well as no significant differences between primes and targets within each condition. As irregular words are integral to the experimental manipulations of the current study, particularly the incongruent conditions (P+O-, P-O+), it is important to note that controlling the relative frequency and familiarity of stimuli was especially essential to reduce or remove effects of regularity. Descriptive statistics for RRT stimulus properties are shown in Table 36 and Table 37.

Table 36: Means (standard deviations) of RRT-A stimulus properties (RRT-A primes, RRT-B targets)

	P+O+ (n=40)	P+O- (n=40)	P-O+ (n=40)	P-O- (n=40)
Letters	4.43 (0.5)	4.43 (0.5)	4.28 (0.45)	4.33 (0.47)
CELEX frequency[1]	60.22 (74.67)	72.52 (92.35)	63.63 (94.41)	50.57 (57.07)
Subjective frequency[2]	444.08 (79.8)	433.03 (98)	422.7 (126.34)	423.6 (135.69)
Bigram frequency[3]	1914.08 (1279.43)	1901.49 (1381.78)	1933.06 (807.56)	1851.35 (1507.57)

Biphone frequency[3]	757.46 (951.75)	639.52 (523.01)	962.18 (1368.95)	603.02 (694.73)
Neighbourhood[4]	7.63 (3.89)	7.43 (4.61)	7.85 (4.63)	8 (4.6)
Phonological neighbourhood	15.35 (7.56)	17.75 (6.67)	15.53 (7.84)	17.68 (9.24)
Imageability[5]	476.33 (182.79)	489.13 (145.47)	419.7 (218.95)	479.13 (181.94)
Age of Acquisition[6]	294.10 (80.36)	320.10 (78.85)	346.12 (88.47)	336.57 (82.93)

[1] British English [2] Balota, Pilotti, & Cortese (2001) [3] Token measure [4] Coltheart's N [5] Retrieved from the MRC database (Coltheart, 1981) [6] Not all stimuli have AoA values in the Bristol norms (Stadthagen-Gonzalez & Davis, 2006), so the entered number of stimuli/values per analysis were lower than the number of stimuli in the respective condition.

Note: Unless otherwise specified, all stimulus data was retrieved using N-Watch (Davis, 2005).

Table 37: Means (standard deviations) of RRT-B stimulus properties (RRT-B primes, RRT-A targets)

	P+O+ (n=40)	P+O- (n=40)	P-O+ (n=40)	P-O- (n=40)
Letters	4.43 (0.5)	4.43 (0.5)	4.28 (0.45)	4.35 (0.48)
CELEX frequency[1]	115.22 (307.05)*	49.7 (86.12)	75.91 (157.77)	43.39 (59.04)
Subjective frequency[2]	437.73 (121.65)	423.65 (134.08)	442.93 (99.59)	443.8 (91.34)

Bigram frequency[3]	2016.83 (1335.9)	1676.54 (943.36)	1990.25 (877.7)	1521.42 (1222.64)
Biphone frequency[3]	1261.8 (2010.25)	645.81 (58423)	981.11 (1445.55)	807.37 (1502.51)
Neighbourhood[4]	8.13 (4.24)	5.63 (5.04)	7.35 (3.66)	6.9 (4.79)
Phonological neighbourhood	17.45 (7.02)	16.6 (7.78)	15.1 (8.13)	16.9 (7.45)
Imageability[5]	492.73 (163.47)	452.3 (185.98)	449.25 (177.9)	485.63 (197.43)
Age of Acquisition[6]	305.04 (67.97)	320.85 (78.31)	303.11 (78.59)	288.04 (80.21)

[1] British English [2] Balota, Pilotti, & Cortese (2001) [3] Token measure [4] Coltheart's N [5] Retrieved from the MRC database (Coltheart, 1981) [6] Not all stimuli have AoA values in the Bristol norms (Stadthagen-Gonzalez & Davis, 2006), so the entered number of stimuli/values per analysis are lower than the number of stimuli in the respective condition.

[*] The mean and SD for P+O+ frequency are notably high due to the inclusion of *time* (CELEX=1792.91).

Note: Unless otherwise specified, all stimulus data was retrieved using N-Watch (Davis, 2005).

7.3.3.1.1 Scale-based support

The validity of the RRT stimulus pairs (in terms of the extent they rhymed or did not rhyme) were also checked independently and separately from the main experiment. Scale-based survey data for each stimulus pair were collected online from 12 anonymous individuals who did not participate in the main research. These surveys were conducted using Qualtrics on the participants' own personal devices at their leisure. Participants were shown the rhyming and orthographically congruent (P+O+), rhyming but orthographically incongruent (P+O), orthographically congruent but non-rhyming (P-O+), and the phonologically and orthographically incongruent (P-O-) stimulus pairs used in Study 2. While the stimulus pair was still on-screen, participants were asked to rate each stimulus, in their own time, on a scale of 0-5, where 0 denoted the stimulus pair did not rhyme at all and 5 denoted the stimulus pair rhymed perfectly. Stimulus pairs were displayed one at a time in **Arial** (24pt) with both stimuli on-screen together, positioning to the left of centre and to the right of centre, respectively, which clearly distinguished them as separate stimuli. As response time was not important or being recorded, this was sufficient for the purpose of the survey. Table 38 shows the means, standard deviations, and range of each stimulus type.

Table 38: Descriptive statistics of RRT stimulus checking surveys

	Mean	SD	Range
P+O+	4.73	0.33	3.27 – 5.00
P+O-	4.53	0.36	3.27 – 5.00
P-O+	0.63	0.30	0.00 – 1.33
P-O-	0.00	0.01	0.00 – 0.08

Means for all stimuli were within expected thresholds of the first quartile for rhyming pairs (P+O+, P+O-) and the fourth quartile for non-rhyming pairs (P-O+, P-O-), strongly supporting the stimuli for use in their respective conditions. The results also showed the expected pattern according to rhyme likeness: P+O+ > P+O- > P-O+ > P-O-.

7.3.3.2 Semantic association testing

As outlined in section 5.4.1.9, phonological and semantic processing are directly linked (Brunswick, 2010) and semantic relatedness between word pairs in a priming paradigm (such as the one used in the current study) has been widely documented to facilitate processing of target stimuli (e.g., Franklin et al., 2007; Grossi et al., 2001; Moreno & Kutas, 2005), semantic activation must also be balanced across conditions to minimize any involvement of semantic processing that could confound observations of phonological (and orthographic processing of interest to the current research. Therefore, to check that no pairs of stimuli (whether they were to be presented as stimulus pairs or not) were significantly related semantically, all 102,080 possible combinations of real word pairs from the 320 RRT stimuli were submitted for matrix comparison using Latent Semantic Analysis (LSA; Landauer et al., 1998). Due to the vast amount of data that is obtained through such analysis for so many combinations of word pairs, only a summary will be provided here.

As an indication of the semantic relatedness within the entire stimulus set (0:unrelated to 1:related), the mean for the 149 pairs (0.29% of all pairs) that achieved greater than 0.5 on the LSA output was 0.6 (SD=0.08) with a range of 0.51-0.89. However, only 19 of these achieved greater than 0.7 and only 2 above 0.8, none of which were presented together as experimental stimulus pairs or in adjacent trials.

7.3.4 Apparatus

The materials used for Study 2 were identical to those used for Study 1, details of which can be found in section 5.5.

7.3.5 Procedure

Participants were asked to indicate as quickly and accurately as possible whether or not stimulus pairs rhymed. Responses were made to target stimuli to denote rhyming pairs (P+O+ and P+O-) and non-rhyming pairs (P-O+ and P-O-) by pressing the appropriate button of the two allocated buttons, which were counterbalanced within participant groups. It was clearly explained what constitutes a rhyme and that only ear/sound/phonological rhymes (e.g., *love-dove*, as found in songs and poems) were acceptable and that so-called eye rhymes (e.g., *lemon-demon*) were not admissible as rhymes in this task and should be deemed non-rhyming. The same instructions, examples and practice were given to all participants in an attempt to control demand characteristics. As practice, four trials for each condition, which were not part of the main test stimuli, were pseudorandomly presented to participants exactly as in the real experiments. Practice trials were completed before the task and evaluated by both researcher and participant to ensure the participant understood the requirements of the task. Following successful practices, half of each group completed RRT-A before RRT-B and the other half completed RRT-B before RRT-A, all of whom completed the tasks of Study 1 in between RRT variants. Participants were informed that the second variant was the same as the first rhyme task in terms of what they had to do and the words they would see but were still reminded of the instructions and prompted to complete practice trials again to ensure their understanding. Each variant of the RRT took approximately 14 minutes including two ~1-minute and one ~3-minute breaks.

7.3.6 Analysis

First, a brief overview of the conditions: the P+O+ condition represents both orthographic and phonological repetition, P+O- reflects phonological repetition with orthographic conflict, P-O+ involves orthographic repetition with phonological conflict, and P-O- provides a control condition with no explicit orthographic or phonological repetition, only stimulus-type repetition (i.e., another word). The fully congruent conditions (P+O+ and P-O-) were used as relative controls and baselines to compare with the incongruent conditions (P+O- and P-O+). Congruent and incongruent conditions will be compared directly to focus on orthographic processing (P+O+ vs P+O- and P-O- vs P-O+) and phonological processing (P+O+ vs P-O+ and P-O- vs P+O-), respectively, while also allowing subtractive methods (e.g., differences between P+O- and P-O- to represent phonology; differences between P-O+ and P-O- to represent orthography). The P-O- condition also acts as an overall control, as it involves no grapheme- or phoneme-level orthographic or phonological priming.

Comparisons between conditions (i.e., responses to targets) highlight the relative strengths of the associated orthographic and phonological processing. The extent of difference between conditions highlights which components/areas are sensitive to stimulus-type repetition, orthographic repetition, phonological repetition, orthographic conflict, and phonological conflict. Significant differences between conditions can be attributed to different contrasts of orthographic and phonological priming dependent on the control condition used. The contrast between P-O- and P-O+ provides a measure of orthographic priming, while a difference between P-O- and P+O- reflects phonological priming. Alternatively, comparing P+O+ with P+O- indicates phonological repetition or conflict of orthography, while the contrast between P+O+ and

P-O+ demonstrates orthographic repetition or conflict of phonology. Taken a step further, effect sizes of each contrast can be used for insight into which is most influential on VWR in terms of their respective effects.

Apart from the additional considerations for ERP analysis due to the nature and complexities of priming and counterbalancing across two separate instances of the task (described earlier in this chapter), Study 2 will follow many of the analysis principles as Study 1 (as described in §5.7), focusing on the a priori contrasts between groups and experimental conditions. A mixed factorial ANOVA was conducted per behavioural measure (accuracy, RTs to TARGET stimuli) and ERP component/cluster/measure combination (P1-O amplitude/latency, P1-OT amplitude/latency, N100-FC amplitude/latency, N170-OT amplitude/latency). For the behavioural data, a univariate two-way (4) x 3 mixed factorial ANOVA with one within-participants factor of Condition (P+O+, P+O-, P-O+, P-O-) and one between-participants factor of language group (English, Spanish-English, Chinese-English) was performed per analysis. For ERP data, a univariate mixed factorial 4 (Condition: P+O+, P+O-, P-O+, P-O-) x 2 (Hemisphere: left/right) x 3 (Group: English, Spanish-English, Chinese-English) ANOVA was conducted for each combination of component/cluster (e.g., P1-O, N170-OT) and measure (positive/negative amplitude, 50% positive/negative area latency). Corrections for violations sphericity (according to Mauchly's test) were applied to calculations of effects and interactions involving the 4-level within-participants factor of Condition, using Greenhouse-Geisser (when $\varepsilon<0.75$) or Huynh-Feldt (when $\varepsilon>0.75$). Where appropriate, the ANOVA outputs will be followed by relevant pairwise post-hoc results (using the Tukey correction for group comparisons and the Holm correction for within-group comparisons). Only statistically significant results will be reported and any main effects, interactions, and post-hocs not reported can be understood as non-significant. Full statistical outputs, including 95% confidence intervals, are included in Appendices N (behavioural), O (ERP amplitudes) and P (ERP latencies).

7.4 Results

For the measures of accuracy, response times, and each ERP component/region measure (see Figure 8 for the layout of electrode clusters), the following sections will include descriptive statistics and inferential statistical test results from the contrasts of groups (English, Spanish-English, Chinese-English) and experimental conditions (P+O+, P+O-, P-O+, P-O-).

7.4.1 Data preparation

Across both RRT-A and RRT-B, 4.60% of English, 5.65% of Spanish-English, and 6.42% of Chinese-English RT scores were removed due to outliers (absolute values <200ms, >2000ms, or ≥1.96 SD from the mean). As RT and ERP analyses only involves correct responses, 6.22% of English, 35.42% of Spanish-English, and 28.73% of Chinese-English data was not included in RT or ERP analyses (see Table 39 for overall accuracy data). Apart from accuracy in the P-O- condition showing a non-normal distribution in English and Spanish-English across RRT-A and RRT-B (based on *Zskew*<-1.96 and *Zkurtosis*>1.96), representing a ceiling effect that is expected for native alphabetic readers in such a relatively easy and congruent control condition, all variables were within reasonable parametric boundaries in terms of normality and outliers. Table 39 shows that each group performed very similarly across RRT-A and RRT-B, validating the plan to collapse across task iterations, using RRT-A and RRT-B for counterbalancing, and to analyze RRT-A and RRT-B as one. Supporting this statistically, all paired-samples *t*-tests between RRT-A and RRT-B, conducted per condition within each group, showed non-significance ($p > \alpha$) for accuracy and RTs ($\alpha = .0125$, Bonferroni-corrected; see Appendix Q for full output).

Table 39: Overall accuracy (%) means (SDs) across all conditions for RRT-A and RRT-B in all groups

	English	Spanish-English	Chinese-English
RRT-A	95 (3.52)	64.28 (8.28)	71.28 (10.53)
RRT-B	93.38 (4.7)	64.88 (8.77)	71.25 (7.98)
Overall	93.78 (4.17)	64.58 (7.83)	71.27 (9.0)

7.4.1.1 Pre-analysis item evaluation

While interlingual relationships in the stimuli were avoided wherever possible, the Study 2 stimulus sets included Spanish noun and cognate *base*, which has distinctly different pronunciation between English and Spanish. Neither behavioural nor ERP responses to these stimuli were outliers in the data. With just this one (0.31% of stimuli), true interlingual stimulus conflicts were, therefore, minimal, effectively random throughout the experimental procedure, and without the power to influence averaged overall results. While such a small minority is highly unlikely to affect any behavioural or ERP analysis, especially after sufficient and rigorous data cleaning and processing (e.g., artefact rejection, filtering, averaging, outlier removal), all data associated with them was removed from analysis. This, combined with the comprehensive pseudo-randomization and counterbalancing, minimizes any detriment to the analysis.

7.4.2 Accuracy

Figure 23 shows the means and standard deviations for accuracy for each condition of the RRT for all groups.

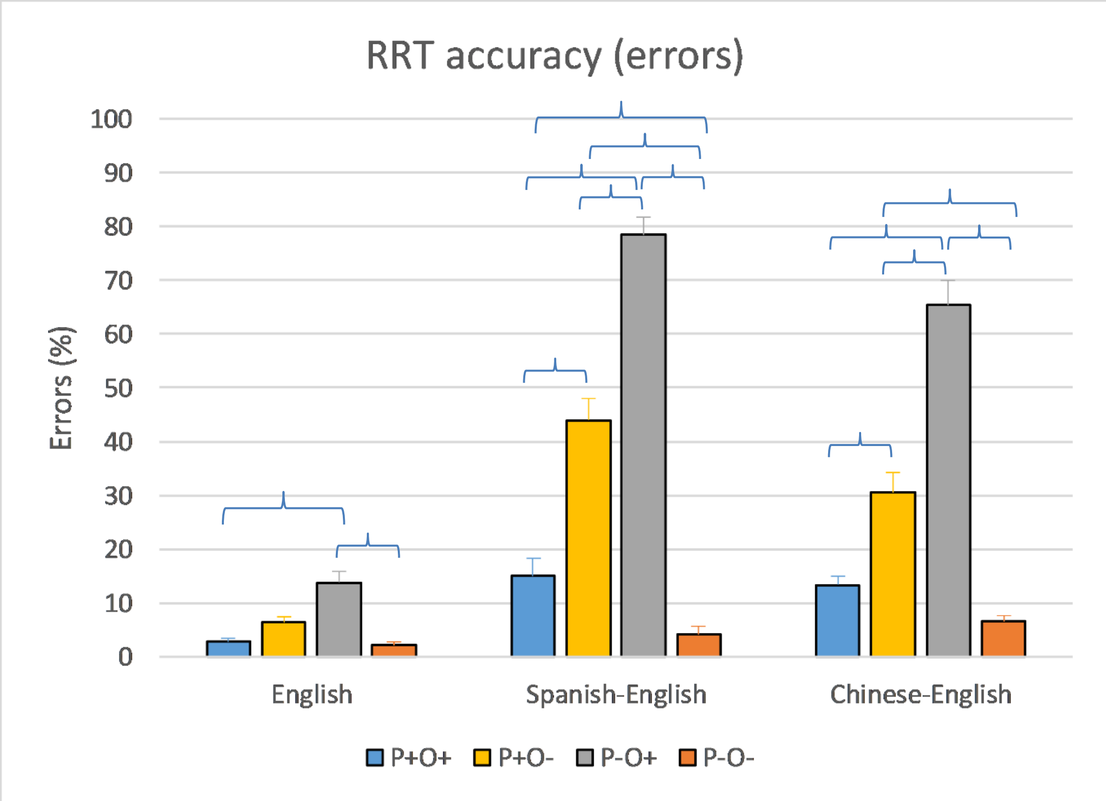

Figure 23: Descriptive statistics for RRT accuracy (% errors) for all conditions across all groups (error bars represent SDs)

The (4) x 3 ANOVA showed significant main effects of Condition, $F(1.93,109.91)=237.53, p<.001, \eta^2_p=0.81$, and Group, $F(2,57)=86.95, p<.001, \eta^2_p=0.75$), as well as a significant interaction of Condition x Group on accuracy, $F(3.86,109.91)=36.59, p<.001, \eta^2_p=0.56$. Pairwise post-hoc comparisons between conditions (within groups) and between groups (within conditions), respectively, are reported below.

7.4.2.1 Between conditions, within groups

Accuracy to P-O- was higher than to P+O- in Spanish-English, $p<.001$, MD=39.78, $d=2.1$, and Chinese-English, $p<.001$, MD=23.99, $d=1.27$, but not in English. Likewise, accuracy to

P+O+ was also higher than to P+O- in Spanish-English, $p<.001$,MD=28.88,$d=2.64$, and Chinese-English, $p<.001$,MD=17.28,$d=1.46$, but not in English. These results reflect the inhibitive orthography/phonology incongruence in P+O- negatively impacting accuracy in the ESL groups.

Accuracy to P-O- was higher than to P-O+ in English, $p=.023$,MD=11.63,$d=1.39$, in Spanish-English, $p<.001$, MD=74.33,$d=5.45$, and in Chinese-English, $p<.001$, MD=58.77,$d=2.97$, as expected due to the relative difficulty of the incongruent orthography/phonology in P-O+ and the ease of P-O-. Similarly, accuracy to P+O+ was higher than to P-O+ in all groups: English $p=.042$,MD=10.94,$d=1.2$, Spanish-English, $p<.001$,MD=63.44,$d=2.64$, and Chinese-English, $p<.001$,MD=52.06,$d=2.55$. These results reflect the inhibitive orthography/phonology incongruence in P-O+ negatively impacting accuracy across all groups.

Accuracy to P-O- was higher than to P+O+ in Spanish-English, $p=.042$,MD=10.9,$d=0.68$, but not in English or Chinese-English, indicating that the lack of any orthographic or phonological similarity in P-O- facilitated and/or the presence of both orthographic and phonological priming in P+O+ did not facilitate accuracy in the Spanish-English group, while these control conditions were responded to similarly in English and Chinese-English groups. Lastly, accuracy to P+O- was also higher than to P-O+ in Spanish-English, $p<.001$,MD=34.55,$d=1.37$, and in Chinese-English, $p<.001$,MD=34.78,$d=1.67$, but not in English, as expected for all groups (including the native English) due to the relative difficulty of the orthographically similar non-rhyme condition (P-O+).

7.4.2.2 Between groups, within conditions

Between groups, accuracy to P+O- was higher in English than in both Spanish-English, $p<.001$,MD=37.57,d=2.83, and Chinese-English, $p<.001$,MD=24.21,d=1.96, while also higher in Chinese-English than Spanish-English, p=.024,MD=13.36,d=-0.77. As expected due to the relative difficulty of orthographically similar non-rhymes, accuracy to P-O+ was also higher in English than in both Spanish-English, $p<.001$,MD=64.75,d=5.19, and Chinese-English, $p<.001$,MD=51.63,d=3.21, while also higher in Chinese-English than in Spanish-English, p=.029,MD=-13.13,d=-0.74. Accuracy was not significantly different between groups to either P+O+ or to P-O- (the relative control conditions).

7.4.3 Response times

Figure 24 shows means and standard deviations of response times across all groups and conditions of the RRT. Across all groups, around 5% of RT scores were removed due to outliers (4.60% of English, 6.42% of Spanish-English, and 5.65% of Chinese-English).

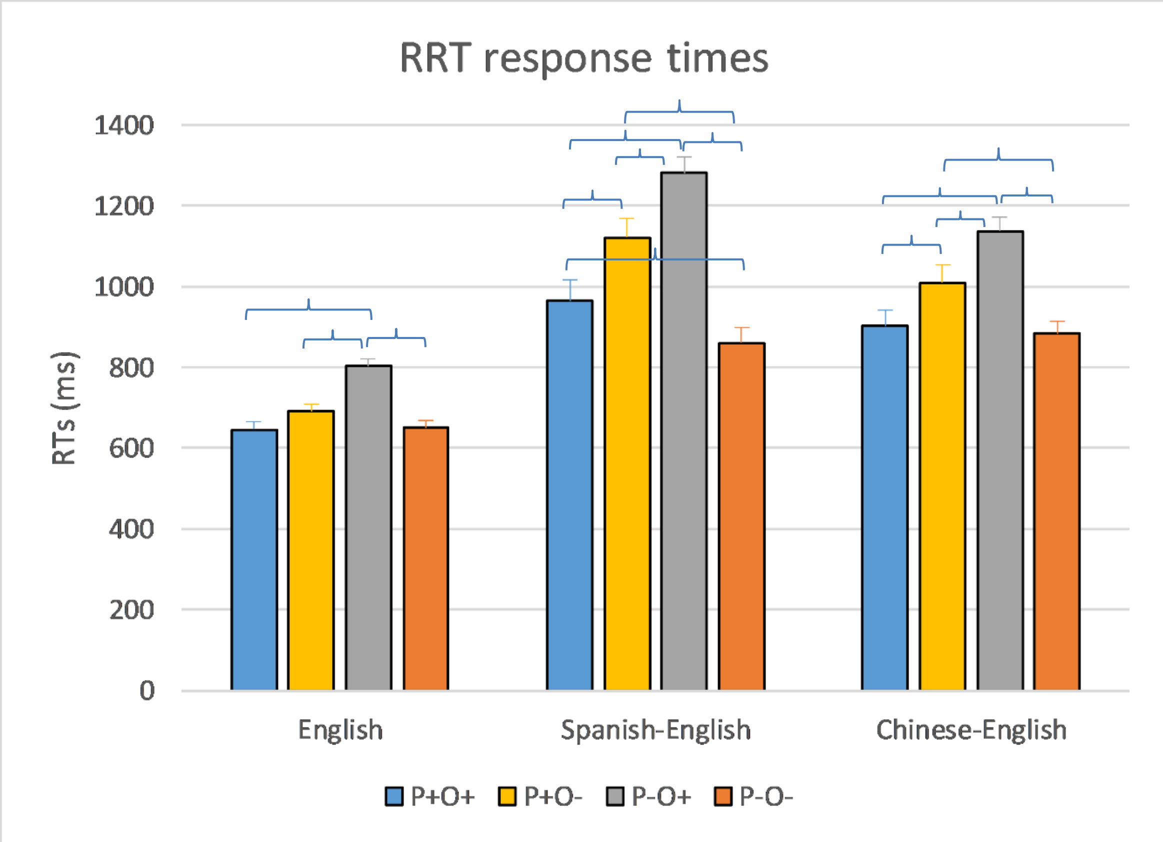

Figure 24: Response times (ms) for all conditions across all groups (error bars represent SDs)

The (4) x 3 ANOVA showed significant main effects of Condition, $F(2.08,118.51)=155.7, p<.001, \eta^2_p=0.73$, and Group, $F(2,57)=33.83, p<.001, \eta^2_p=0.54$, as well as a significant interaction of Condition x Group on response times, $F(4.16,118.51)=12.73, p<.001, \eta^2_p=0.31$. Post-hocs between conditions (within groups) and between groups (within conditions), respectively, are reported below and in Tables 46-51.

7.4.3.1 Between conditions, within groups

Between conditions, response times to P+O- were slower than to P-O- in Spanish-English, $p<.001$,MD=-260.4,d=-1.54, and Chinese-English, $p=.001$,MD=-126.35,d=-0.87, but not in English, indicating an effect of the incongruent orthography/phonology of P+O- and the phonological dissimilarity in P-O-. Likewise, response times to P+O- were slower than to P+O+ in Spanish-English, *p*<.001,MD=-157.3,*d*=-1.63, and in Chinese-English, *p*<.001,MD=-106.14,*d*=-1.6, but not in English, this time showing the impact of the congruent orthography/phonology and phonological priming in P+O+ and the incongruent orthography/phonology of P+O-.

As expected, response times to P-O+ were slower than to P-O- in English, *p*<.001, MD=-151.07, *d*=-3.42, Spanish-English, *p*<.001,MD=-421.28,*d*=-3.22, and Chinese-English, *p*<.001,MD=-253.48,*d*=-3.22. Response times to P-O+ were also slower than to P+O+ in English, *p*<.001,MD=-155.98,*d*=-3.21, in Spanish-English, *p*<.001,MD=-318.18,*d*=-2.04, and in Chinese-English, *p*<.001,MD=-233.27,*d*=-2.58. Both results support the notion of orthographically similar non-rhymes (P-O+) and their orthography/phonology incongruence negatively impacting behavioural responses.

Response times between the relative controls were faster to P-O- than to P+O+ in Spanish-English only, *p*<.001,MD=-103.1,*d*=-0.58, showing that the presence of both orthographic and phonological priming in P+O+ did not facilitate response times and/or the lack of any orthographic or phonological similarity in P-O- resulted in faster responses in the Spanish-English group, while these control conditions were responded to similarly in English and Chinese-English groups. Lastly, response times between the incongruent conditions, were slower to P-O+ than to P+O- in English, *p*<.001, MD=-110.84, *d*=-2.28, in Spanish-English, *p*<.001,MD=-160.88,*d*=-1.07, and in Chinese-English, *p*<.001,MD=-127.13,*d*=-1.10, as expected due to the relative difficulty of the orthographically similar non-rhyme condition (P-O+).

7.4.3.2 Between groups, within conditions

Between groups, response times were faster in English than in Spanish-English to P+O+, *p*<.001,MD=-317.27,*d*=-1.83, to P+O-, *p*<.001,MD=-429.42,*d*=-2.59, to P-O+, *p*<.001,MD=-479.47,*d*=-3.42, and to P-O-, *p*=.005,MD=-209.25,*d*=-1.56, as expected considering the relative differences in language profiles between groups i.e., English as L1 for English, L2 for Spanish-English. Expected also and for the equivalent reasons, responses in English were also faster than in Chinese-English to P+O+, *p*<.001,MD=-256.59,*d*=-1.85, to P+O-, *p*<.001,MD=-317.58,*d*=-2.11, to P-O+, *p*<.001,MD=-333.87,*d*=-2.59, and to P-O-, *p*<.001,MD=-231.46,*d*=-1.99. Response times between Spanish-English and Chinese-English, however, did not differ significantly to any condition.

7.4.4 ERPs

The following sections document left and right occipital P1 (P1-O), occipitotemporal P1 (P1-OT) and N170 (N170-OT), and frontal-central N100 (N100-FC) ERPs from the English, Spanish-English, and Chinese-English groups to the RRT conditions (P+O+,P+O-,P-O+,P-O-). See Figure 8 (5.6.2.2) for the electrode cluster layout.

7.4.4.1 Occipital P1 (P1-O)

Figure 25 depicts occipital ERP data to RRT conditions at 100ms.

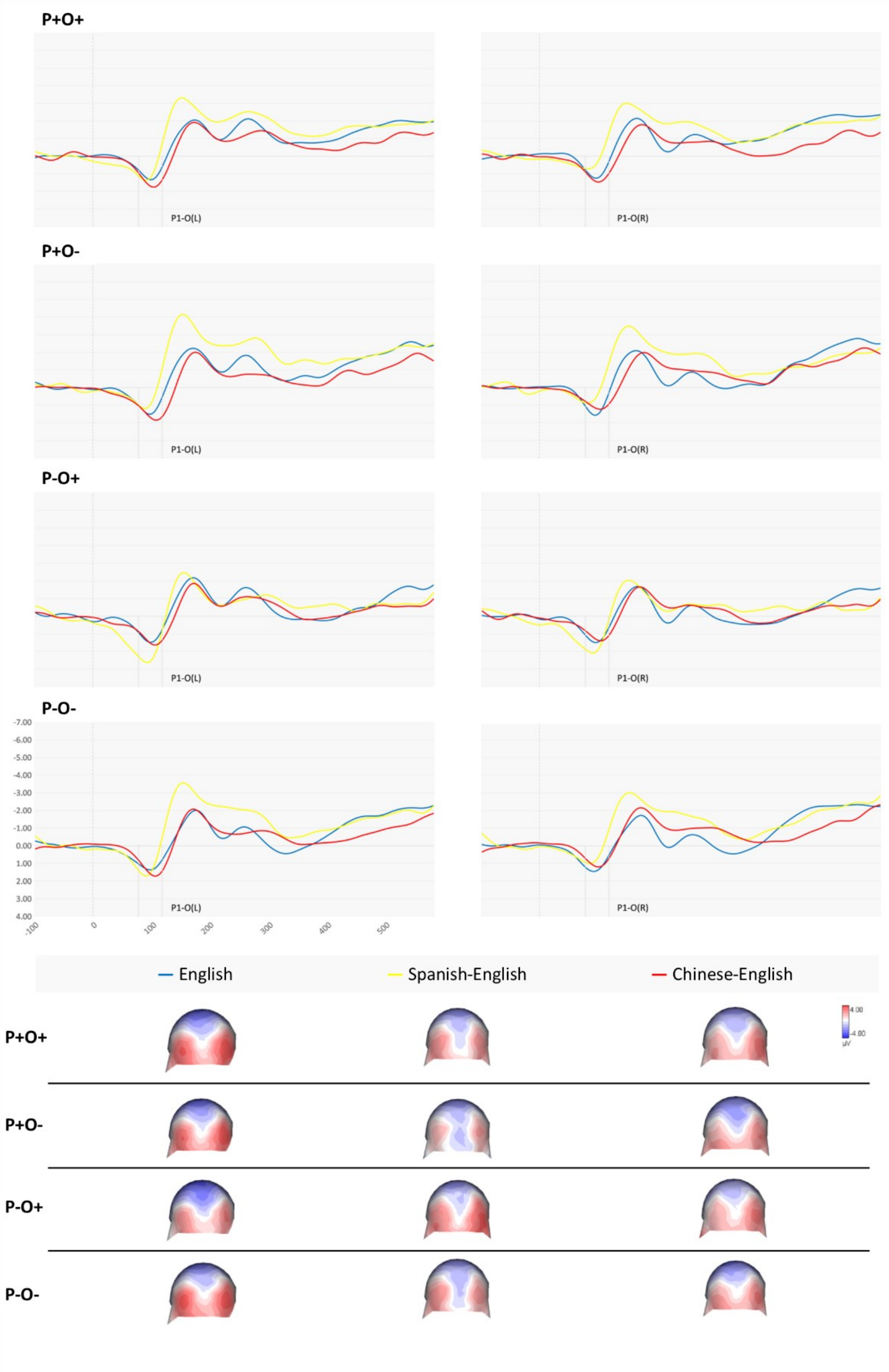

Figure 25: Event-related potentials and topographic voltage maps of left and right occipital clusters at 100ms for RRT conditions (P+O+,P+O-,P-O+,P-O-) across groups

7.4.4.1.1 P1-O amplitude

Table 40 shows means and standard deviations for prime-TARGET differences in positive area amplitude measures of P1-O across conditions and groups.

Table 40: Means (SDs) of P1-O amplitude prime-TARGET differences (μV)

Condition	Hemisphere	English	Spanish-English	Chinese-English
P+O+	Left	-0.33 (0.58)	-0.32 (0.54)	-0.44 (0.51)
P+O-		-0.46 (0.56)	-0.21 (0.41)	-0.35 (0.52)
P-O+		-0.38 (0.55)	-0.94 (0.86)	-0.31 (0.72)
P-O-		-0.28 (0.56)	-0.17 (0.32)	-0.47 (0.43)
P+O+	Right	-0.4 (0.51)	-0.19 (0.4)	-0.38 (0.61)
P+O-		-0.57 (0.56)	-0.25 (0.45)	-0.25 (0.54)
P-O+		-0.4 (0.57)	-0.76 (0.72)	-0.39 (0.63)
P-O-		-0.3 (0.49)	-0.05 (0.39)	-0.35 (0.57)

Means below zero reflect a larger (more positive) amplitude to TARGET stimuli than primes.

The (4) x (2) x 3 ANOVA showed a significant main effect of Condition, $F(3,171)=7.15, p<.001, \eta^2_p=0.11$, and a significant interaction of Condition x Group, $F(4.12,117.29)=8.46, p<.001, \eta^2_p=0.23$, on P1-O amplitude. Averaging across hemispheres, post-hocs showed prime-TARGET differences of P1-O amplitude to P-O+ were significantly larger than to P+O+, $p<.001$, MD=0.59, $d=0.74$ in Spanish-English.

7.4.4.1.2 P1-O latency

Table 41 shows means and standard deviations for prime-TARGET differences in 50% positive area latency measures of P1-O across conditions and groups.

Table 41: Means (SDs) of P1-O latency prime-TARGET differences (ms)

Condition	Hemisphere	English	Spanish-English	Chinese-English
P+O+	Left	3.21 (3.55)	0.63 (3.58)	0.01 (3.7)
P+O-		2.77 (2.72)	-0.2 (4.68)	-0.09 (4.67)
P-O+		1.98 (3.39)	0.11 (4.39)	1.93 (4.34)
P-O-		3.44 (3.68)	0.13 (2.92)	-1.32 (3.55)
P+O+	Right	1.36 (4.81)	-1.32 (3.3)	-1.07 (4.39)
P+O-		1.46 (5)	-0.67 (3.44)	-1.51 (5.32)
P-O+		0.69 (4.2)	-1.97 (4.13)	1.2 (6.09)
P-O-		2.8 (4.55)	-1.36 (4.45)	-1.16 (4.84)

Means above zero reflect a shorter latency to TARGET stimuli than to primes; means below zero reflect a longer latency to TARGET stimuli than to primes.

The (4) x (2) x 3 ANOVA showed main effects of Hemisphere, $F(1,57)=6.66, p=.012, \eta^2_p=0.1$, and Group on P1-O latency, $F(2,57)=6.04, p=.004, \eta^2_p=0.17$, as well as a significant interaction of Condition x Group, $F(6,171)=2.86, p=.011, \eta^2_p=0.09$, on P1-O latency. Averaging across hemispheres, post-hocs showed no significant effects on prime-TARGET differences between conditions (within groups). Between groups, however, prime-TARGET differences in English P1-O latency to P-O- (averaged across hemispheres) were significantly larger than in Chinese-English, $p=.014$, MD=4.37, $d=1.17$. The prime-TARGET differences reflected shorter latencies in English and longer latencies in Chinese-English to P-O- TARGETs.

7.4.4.2 Occipitotemporal P1 (P1-OT)

Figure 26 depicts occipitotemporal ERP data at 100ms to RRT conditions.

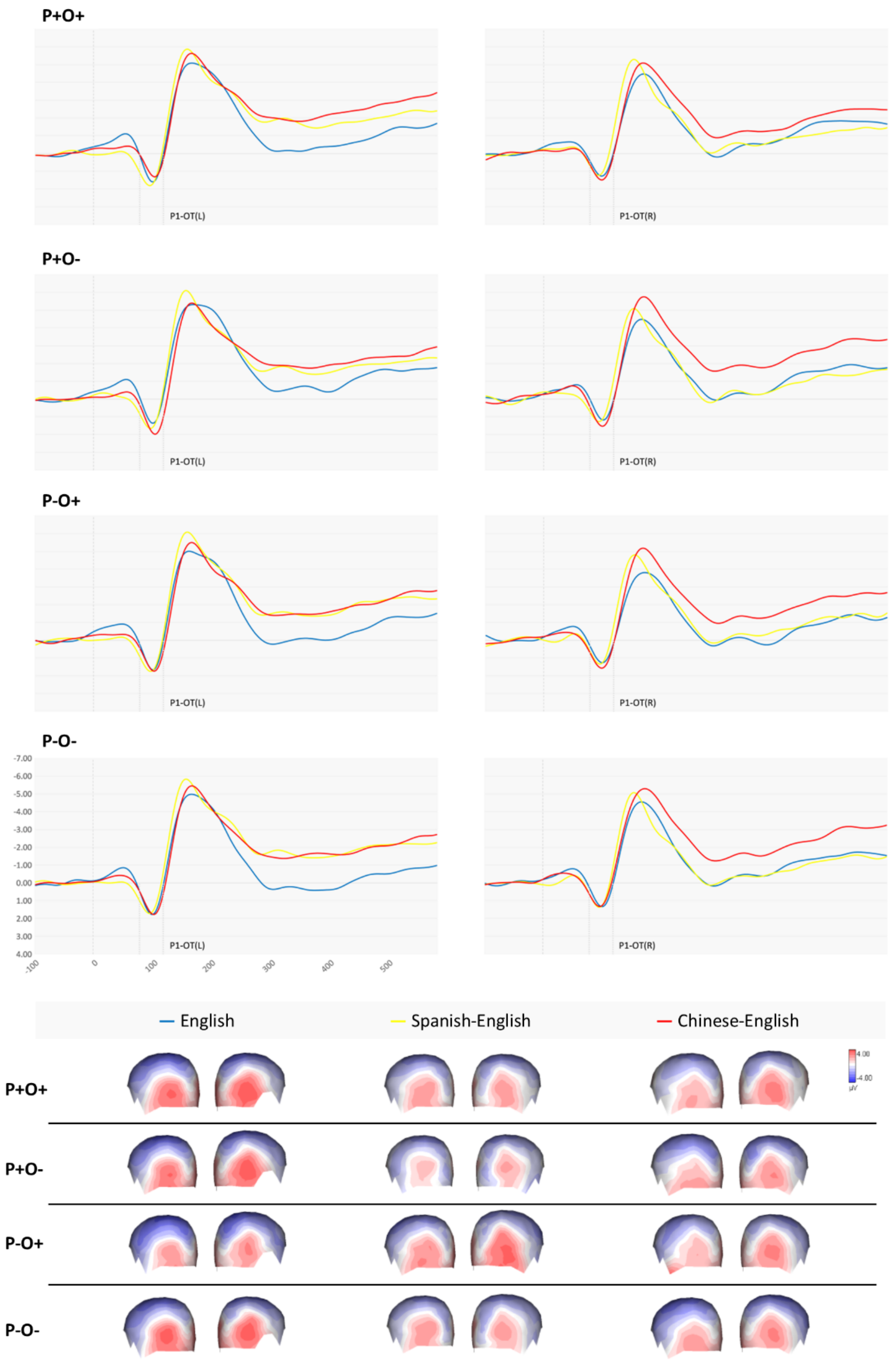

Figure 26: Event-related potentials and topographic voltage maps of left and right occipitotemporal clusters at 100ms for RRT conditions (P+O+,P+O-,P-O+,P-O-) across groups

7.4.4.2.1 P1-OT amplitude

Table 42 shows means and standard deviations for prime-TARGET differences in positive area amplitude measures of P1-OT across conditions and groups.

Table 42: Means (SDs) of P1-OT amplitude prime-TARGET differences (uV)

Condition	Hemisphere	English	Spanish-English	Chinese-English
P+O+	Left	-0.49 (0.53)	-0.31 (0.39)	-0.44 (0.53)
P+O-		-0.52 (0.47)	-0.18 (0.31)	-0.3 (0.4)
P-O+		-0.47 (0.5)	-0.59 (0.69)	-0.22 (0.55)
P-O-		-0.5 (0.61)	-0.22 (0.37)	-0.47 (0.48)
P+O+	Right	-0.38 (0.61)	-0.26 (0.36)	-0.54 (0.63)
P+O-		-0.63 (0.47)	-0.39 (0.44)	-0.39 (0.48)
P-O+		-0.6 (0.61)	-0.75 (0.58)	-0.49 (0.48)
P-O-		-0.63 (0.64)	-0.27 (0.47)	-0.55 (0.55)

Means below zero reflect a larger (more positive) amplitude to TARGET stimuli than primes.

The (4) x (2) x 3 ANOVA showed a main effect of Condition, $F(3,171)=2.75, p=.045, \eta^2_p=0.05$, and a significant interaction of Condition x Group, $F(6,171)=6.27, p<.001, \eta^2_p=0.18$, on P1-OT amplitude. Averaging across hemispheres, post-hocs showed that prime-TARGET differences in Spanish-English P1-OT amplitude to P-O+ were significantly larger than to P+O+, $p<.001$,MD=0.39,$d=0.75$, larger than to P+O-, $p<.001$,MD=0.38,$d=0.74$, and larger than to P-O-, $p<.001$,MD=0.42,$d=0.78$. There were no significant post-hocs between groups for any condition.

7.4.4.2.2 P1-OT latency

Table 43 shows means and standard deviations for prime-TARGET differences in 50% positive area latency measures of P1-OT across conditions and groups.

Table 43: Means (SDs) of P1-OT latency prime-TARGET differences (ms)

Condition	Hemisphere	English	Spanish-English	Chinese-English
P+O+	Left	2.32 (2.77)	0.89 (2.88)	1.03 (2.21)
P+O-		2.24 (1.96)	1.54 (2.66)	0.75 (2.15)
P-O+		1.89 (3.17)	0.63 (4.71)	1.68 (3.72)
P-O-		1.81 (2.1)	1.95 (2.54)	-0.2 (2.53)
P+O+	Right	1.93 (2.58)	0.75 (3.12)	-0.24 (2.2)
P+O-		2.16 (2.35)	-0.17 (3.66)	0.39 (2.31)
P-O+		2.17 (2.27)	-0.91 (4.58)	0.82 (3.52)
P-O-		1 (2.53)	1 (2.92)	0.48 (2.96)

Means above zero reflect a shorter latency to TARGET stimuli than primes; means below zero reflect a shorter latency to primes than TARGET stimuli.

The (4) x (2) x 3 ANOVA showed a main effect of Group, $F(2,57)=4.28, p=.018, \eta^2_p=0.13$, and a significant interaction of Condition x Group, $F(6,171)=2.97, p=.009, \eta^2_p=0.09$, on P1-OT latency. Averaging across hemispheres, however, post-hocs showed no significant differences in prime-TARGET differences between conditions or groups.

7.4.4.3 Frontal-central N100 (N100-FC)

Figure 27 depicts frontal-central ERP data at 100ms to RRT conditions.

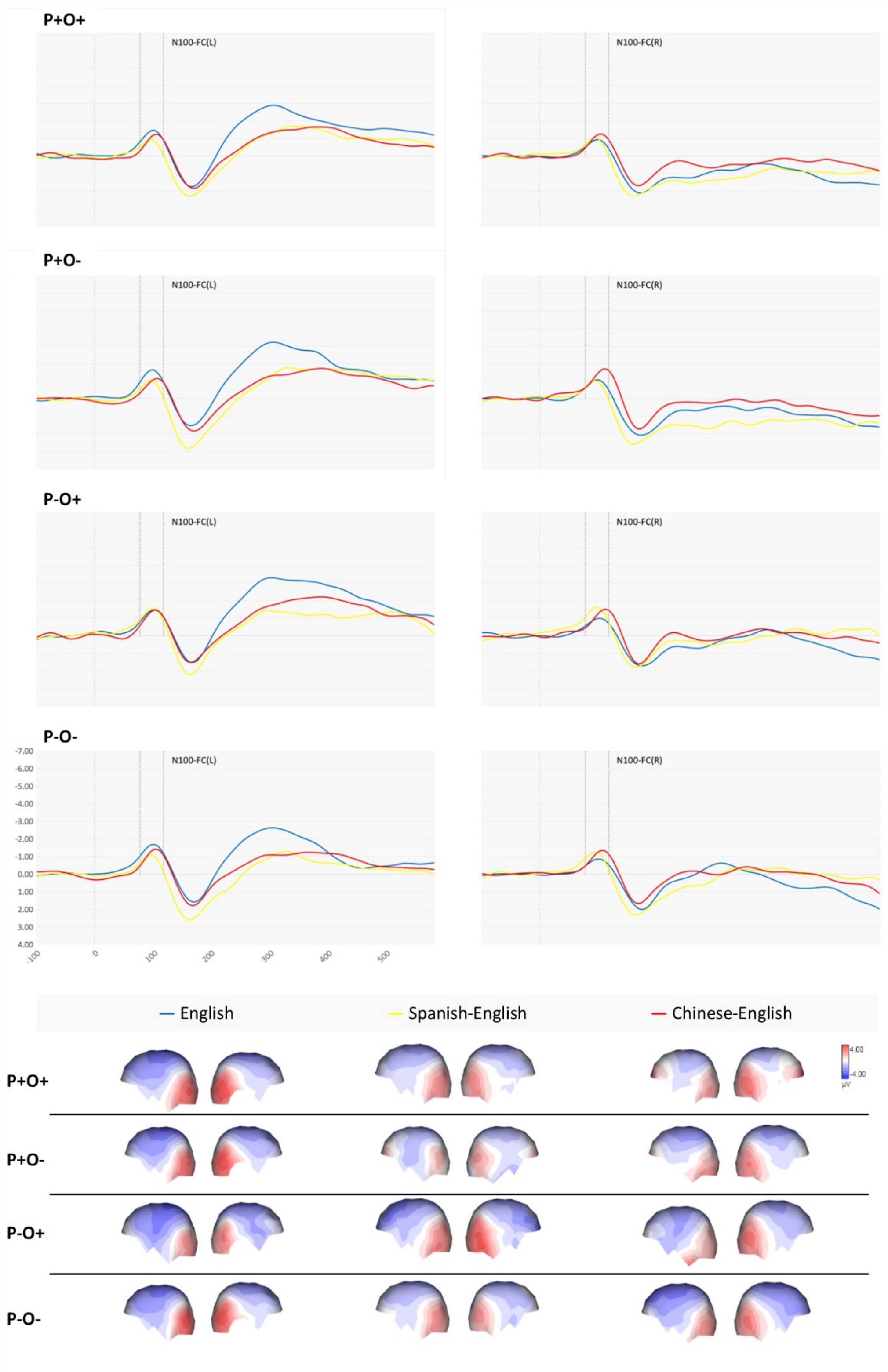

Figure 27: Event-related potentials and topographic voltage maps of left and right frontal-central clusters at 100ms for RRT conditions (P+O+,P+O-,P-O+,P-O-) across groups

7.4.4.3.1 N100-FC amplitude

Table 44 shows means and standard deviations for prime-TARGET differences in negative area amplitude measures of N100-FC across conditions and groups.

Table 44: Means (SDs) of N100-FC amplitude prime-TARGET differences (µV)

Condition	Hemisphere	English	Spanish-English	Chinese-English
P+O+	Left	0.35 (0.33)	0.19 (0.3)	0.24 (0.32)
P+O-		0.48 (0.37)	0.19 (0.31)	0.32 (0.33)
P-O+		0.42 (0.38)	0.34 (0.35)	0.4 (0.41)
P-O-		0.48 (0.39)	0.09 (0.3)	0.29 (0.34)
P+O+	Right	0.26 (0.26)	0.08 (0.24)	0.25 (0.3)
P+O-		0.27 (0.24)	0.13 (0.3)	0.36 (0.32)
P-O+		0.27 (0.29)	0.24 (0.4)	0.32 (0.35)
P-O-		0.19 (0.33)	0.16 (0.27)	0.21 (0.31)

Means above zero reflect a larger (more negative) amplitude to TARGET stimuli than primes.

The (4) x (2) x 3 ANOVA showed main effects of Condition, $F(3,171)=5.18, p=.002, \eta^2_p=0.08$, and Hemisphere, $F(1,57)=5.99, p=.017, \eta^2_p=0.1$, as well as a significant full three-way interaction of Condition x Hemisphere x Group, $F(5.18,147.52)=2.29, p=.047, \eta^2_p=0.07$, on N100-FC amplitude. Pairwise post-hocs showed prime-TARGET differences in N100-FC(L) amplitudes to P-O- were larger in English than Spanish-English, $p=.037$, MD=0.39, $d=0.75$.

7.4.4.3.2 N100-FC latency

Table 45 shows means and standard deviations for prime-TARGET differences in 50% negative area latency measures of N100-FC across conditions and groups.

Table 45: Means (SDs) of N100-FC latency prime-TARGET differences (ms)

Condition	Hemisphere	English	Spanish-English	Chinese-English
P+O+	Left	0.47 (2.78)	-1.11 (3.34)	-0.63 (3.49)
P+O-		1.31 (2.48)	-1.6 (3.87)	0.57 (3.5)
P-O+		0.73 (3.58)	-2.42 (3.67)	1.2 (3.38)
P-O-		1.26 (2.6)	-1.28 (3.07)	0.78 (3.63)
P+O+	Right	1.22 (2.95)	-0.18 (2.78)	0.77 (2.52)
P+O-		2.51 (4.31)	0.34 (2.59)	-0.8 (3.32)
P-O+		1.29 (3.57)	1.8 (4.13)	-1.51 (3.69)
P-O-		3.21 (3.39)	-0.42 (2.55)	-1.53 (2.68)

Means above zero reflect a shorter latency to TARGET stimuli than primes; means below zero reflect a shorter latency to primes than TARGET stimuli.

The (4) x (2) x 3 ANOVA showed a significant main effect of Group, $F(2,57)=5.81, p=.005, \eta^2_p=0.17$, as well as a significant interaction of Hemisphere x Group, $F(2,57)=6.75, p=.002, \eta^2_p=0.19$, and a significant three-way interaction of Condition x Hemisphere x Group, $F(6,171)=6.35, p<.001, \eta^2_p=0.18$, on N100-FC latency. Post-hocs revealed that prime-TARGET differences in Spanish-English N100-FC(L) latencies to P-O+ were significantly smaller than N100-FC(R), $p<.001$, MD=-4.22, $d=-0.8$, where N100-FC(R) to P-O+ reflected shorter latencies to TARGETs, while N100-FC(L) exhibited longer latencies to TARGETs. The prime-TARGET differences in N100-FC(R) latency to P-O- were significantly larger in English than in Chinese-English, $p<.001$, MD=4.75, $d=1.56$. The prime-TARGET differences reflected shorter latencies to P-O- TARGETs than to primes in English, but longer latencies to P-O- TARGETs than primes in Chinese-English.

7.4.4.4 Occipitotemporal N170 (N170-OT)

Figure 28 depicts occipitotemporal ERP data at 170ms to RRT conditions.

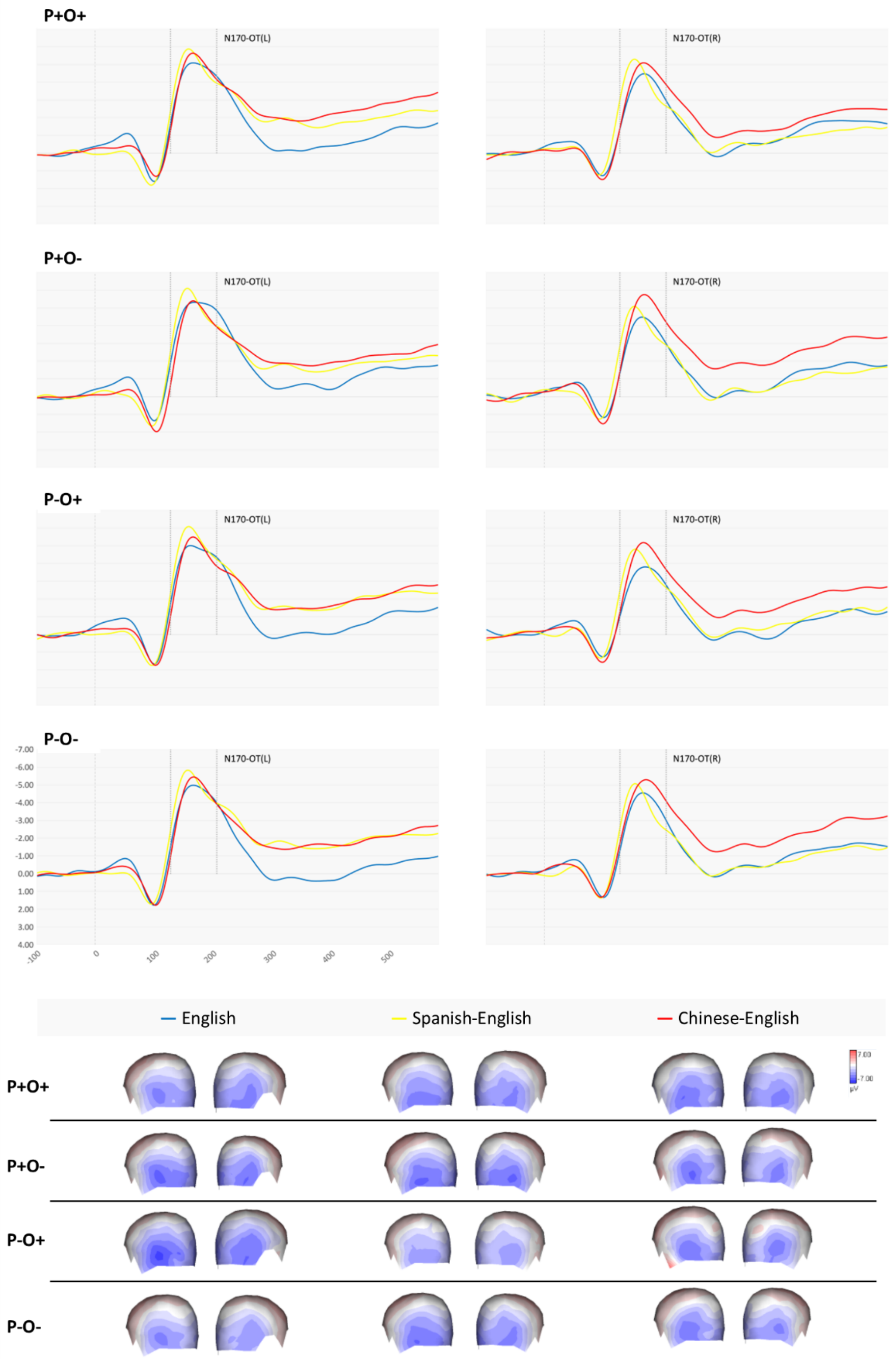

Figure 28: Event-related potentials and topographic voltage maps of left and right occipitotemporal clusters at 170ms for RRT conditions (P+O+,P+O-,P-O+,P-O-) across groups

7.4.4.4.1 N170-OT amplitude

Table 46 shows means and standard deviations for prime-TARGET differences in negative area amplitude measures of N170-OT across conditions and groups.

Table 46: Means (SDs) of N170-OT amplitude prime-TARGET differences (µV)

Condition	Hemisphere	English	Spanish-English	Chinese-English
P+O+	Left	-0.95 (0.75)	-0.57 (0.56)	-0.67 (0.5)
P+O-		-0.8 (0.77)	-0.44 (0.52)	-0.56 (0.44)
P-O+		-0.92 (0.92)	-0.87 (0.79)	-0.74 (0.47)
P-O-		-0.84 (0.89)	-0.54 (0.5)	-0.72 (0.5)
P+O+	Right	-0.61 (0.85)	-0.45 (0.77)	-0.86 (0.78)
P+O-		-0.48 (0.88)	-0.39 (0.72)	-0.68 (0.66)
P-O+		-0.68 (0.93)	-0.62 (1.04)	-0.88 (0.77)
P-O-		-0.57 (0.66)	-0.41 (0.82)	-0.99 (0.77)

Means above zero reflect a larger (more negative) amplitude to TARGET stimuli than primes.

The (4) x (2) x 3 ANOVA showed only a significant main effect of Condition on N170-OT amplitude, $F(3,171)=7.27, p<.001, \eta^2_p=0.11$, and there were no significant pairwise post-hocs.

7.4.4.4.2 N170-OT latency

Table 47 shows means and standard deviations for prime-TARGET differences in 50% negative area latency measures of N170-OT across conditions and groups.

Table 47: Means (SDs) of N170-OT latency prime-TARGET differences (ms)

Condition	Hemisphere	English	Spanish-English	Chinese-English
P+O+	Left	2.54 (5.54)	0.34 (4.85)	0.17 (4.22)

P+O-		2.34 (4.98)	1.1 (6.26)	-1.83 (3.61)
P-O+		1.42 (5.35)	-0.4 (9.77)	-3.61 (4.46)
P-O-		0.48 (5.37)	-0.04 (3.59)	-0.7 (3.91)
P+O+	Right	5.12 (6.5)	0 (7.41)	2.33 (5.84)
P+O-		2.59 (4.93)	-1.01 (4.18)	1.37 (6)
P-O+		5.15 (5.68)	0.27 (8.85)	1.14 (7.77)
P-O-		3.07 (4)	-1.22 (4.1)	2.02 (5.97)

Means above zero reflect a shorter latency to TARGET stimuli than primes; means below zero reflect a shorter latency to primes than TARGET stimuli.

The (4) x (2) x 3 ANOVA showed significant main effects of Hemisphere, $F(1,57)=4.35, p=.042, \eta^2_p=0.07$, and Group on N170-OT latency, $F(2,57)=3.85, p=.027, \eta^2_p=0.12$. There were no significant pairwise post-hoc comparisons.

7.5 Discussion

Following the main objective of the thesis, Study 2 investigated how native (English monolingual), non-native alphabetic L1 (Spanish-English bilingual), and non-native non-alphabetic L1 (Chinese-English bilingual) L1 profiles (as discussed in §1.1) might impact orthographic and phonological processing of English. To this end and to complement the approach of Study 1 with a novel perspective, behavioural performance (via accuracy and response times) and both amplitude and latency measures of occipital/occipitotemporal P1, occipitotemporal N170, and frontal-central activity at ~100ms were observed during a visual rhyme judgement paradigm, orthogonally matched using orthographically and phonologically congruent and incongruent word pairs. All peaking pre-200ms and theoretical precursors to later behavioural and electrophysiological activity, these ERP components in particular were analyzed to examine the contribution of the occipital/occipitotemporal P1 to initial ~100ms orthographic evaluation, the potential for early ~100ms phonological activity at frontal-central sites, and the nature of the occipitotemporal N170 in terms of orthographic and/or phonological processing. The current study combined these factors in order to observe both behavioural and ERP indicators of orthographic and phonological processing based on a single behavioural task for evidence of how measures can be similar or different between different groups of English readers, especially two distinct types of late bilingual (Spanish-English and Chinese-English).

Behavioural performance and later ERP responses in rhyme judgement have been shown to be similar but not identical across native English and ESL readers, such as Spanish-English and Chinese-English bilinguals (Botezatu et al., 2015). Rhyme judgement combined with the flexible linguistic rules of English, therefore, offers a method of investigating VWR processing across L1 and L2 readers. Importantly, orthographic congruence within the word pairs is an understandably significant factor in such visual rhyme judgements and so manipulating both orthography and phonology of real word pairs was used to highlight the orthographic and phonological processes used in VWR more generally, the underpinnings of rhyme judgement.

Behavioural and ERP measures to the orthogonal conditions of the rhyme judgement task were analyzed to investigate the orthographic and phonological processes underlying the necessarily orthographic and phonological task of visual rhyme recognition in native English and ESL readers. The contrast between the fully congruent relative control P+O+ (e.g., file-MILE) and P+O- (e.g., goal-BOWL) reflects orthographic effects, as both phonological and orthographic priming and congruence are present in P+O+, but orthographic incongruence and only phonological priming in P+O-, leaving any difference between them to orthographic congruence. The contrast between P+O+ and P-O+ (e.g., worm-FORM), meanwhile, reflects phonological effects with only orthographic priming present in P-O+, leaving any difference between them to phonological congruence. The contrast between P-O- (e.g., joke-GATE), the other relative control due to involving neither phonological nor orthographic priming, and P+O- also reflects phonological effects due to P+O- involving orthographic incongruence and only phonological priming, leaving any difference between them to phonological priming. Lastly, the contrast between P-O- and P-O+ reflects orthographic effects with P-O+ involving phonological incongruence and only orthographic priming, leaving any difference between conditions to orthographic priming.

Moreover for ERP analyses, the prime-TARGET difference measures that were subjected to statistical analyses are essentially the differences in ERP measures per ERP timeframe/cluster between the overall prime and the target from each condition, which reflect whether ERP amplitudes/latencies to the target were facilitated (larger amplitudes, earlier latencies) or inhibited (smaller amplitudes, later latencies) relative to the prime. It is also important to note that the primes provided no indication of the identity or condition of their subsequent targets, but the context of the rhyme judgement task increased awareness of phonological congruence (i.e., rhyming) and of potentially problematic orthography.

7.5.1 Behavioural performance

As with Study 1, the purpose of the behavioural analysis was twofold: to examine effects of orthographic and phonological manipulations within a single task based on real word rhymes/non-rhymes within and between the ESL groups with different L1 profiles (using the native English group as a control), while providing a behavioural performance reference point for the ERP results where applicable. Before discussing ERP results, this section will outline and discuss the various findings concerning the behavioural measures (accuracy and response times).

Between groups, response times in English were faster than in Spanish-English and Chinese-English to all conditions P+O+, P+O-, P-O+, and P-O-, while Spanish-English and Chinese-English did not differ in RTs to any condition P+O+, P+O-, P-O+, and P-O-. In terms of accuracy, there were no significant differences between any groups to either of the relative controls that involved either orthographic and phonological priming

(P+O+) or neither (P-O-), providing some validation of these conditions as controls. Meanwhile, accuracy was higher in English than in both Spanish-English and Chinese-English and higher in Chinese-English than Spanish-English for both P+O- and P-O+ conditions, showing a clear distinction for the incongruent conditions. Similarly, Botezatu et al. (2015) reported English monolinguals to be more accurate than both Spanish-English and Chinese-English groups to all conditions, but only more accurate and faster to P+O+ and P-O+ than both ESL groups. Botezatu et al. (2015) also reported no differences in accuracy between (early) bilingual Spanish-English and Chinese-English groups to P-O-, but no differences to P+O- as well, and Spanish-English were observed to be faster than Chinese-English to P+O+ and P-O+ (Botezatu et al., 2015), showing a different pattern of results across groups than in the current study. Following the conclusion of Botezatu et al. (2015) that the observed differences between Spanish-English and Chinese-English early bilinguals were due to English proficiency, these discrepancies between studies could also be down to proficiency and differences between early and late bilinguals (as in Botezatu et al. (2015) and the current study, respectively).

Despite divergences from a study with a very similar methodology (Botezatu et al., 2015), the accuracy and RT results from the current study are relatively unambiguous, clearly showing the expected behavioural performance advantage of the native monolingual English group over both ESL late bilingual groups and the relative equivalency of the control conditions across groups. Effect sizes were consistently higher in comparisons between English and Spanish-English than between English and Chinese-English for accuracy, but were relatively similar for RTs per condition. This suggests the Chinese-English group were closer to the native English group in proficiency for rhyme judgement than Spanish-English. Effect sizes in accuracy were also relatively similar between Spanish-English and Chinese-English and there were no RT differences between these groups, further showing similar behavioural performance between ESL groups. In contrast, Botezatu et al. (2015) observed the relative effect of the orthographically congruent non-rhyme (P-O+) to be larger in Spanish-English than in Chinese-English on both RTs and accuracy and larger in Spanish-English than English on accuracy. This different pattern of results can, again, be attributed to the proficiency differences of groups between studies (i.e., early vs late bilinguals) and between groups in Botezatu et al. (2015).

While there were no direct accuracy or RT differences between ESL groups to the relative controls (P+O+, P-O-), accuracy was higher and RTs were faster to P-O- than to P+O+ in Spanish-English, but not in English or Chinese-English. Considering the theoretically cumulative orthographic and phonological priming in P+O+ to be facilitative (by matching congruent orthography and phonology), this finding likely shows that the lack of any orthographic or phonological priming in P-O- was also facilitative for Spanish-English but even more so compared with P+O+. Overall, it shows that P-O- was an easier condition than P+O+ and, therefore the easiest condition, at least for the Spanish-English group. Being observed only in the Spanish-English group points to the shallower more straightforward orthography-phonology processing properties of the Spanish language in contrast with the more orthographically complex English and Chinese languages. The lack of similarity within stimulus pairs in P-O- could, therefore, have been more readily accepted by the Spanish-English group and scrutinized more by the English and Chinese-English groups, accounting for this performance difference.

Responses were slower and accuracy lower to P+O- than to both relative controls (P+O+,P-O-) in Spanish-English and Chinese-English, but not in English. Coch, Hart, and Mitra (2008) also found no difference in accuracy between P+O- and P-O- in native English participants, though Chen et al. (2010) found no differences in accuracy between rhymes and non-rhymes (assumed to be equivalent to P+O- and P-O-, respectively) in Chinese-English late bilinguals, which could be due to differences in stimulus control or participant criteria between studies. However, RTs were also observed to be faster to P+O- (rhymes) than P-O- (non-rhymes) in Chinese-English bilinguals (Chen et al., 2010), which is opposite from the results of the current study, but again could be due to differences in stimuli, as the extent of orthographic control or (inadvertent) manipulation was not clearly reported. Botezatu et al. (2015) also reported Chinese-English bilinguals to be slower and less accurate to P+O- than to P-O-, but in the same comparison Spanish-English bilinguals and English monolinguals were only less accurate (not slower too). In any case, further replication attempts using this task methodology and conditions is required to root out these inconsistencies between studies.

The phonological priming in P+O-, despite the orthographic incongruence, sufficed for the English to be as accurate to P+O- as the fully congruent controls (P+O+, P-O-). However, the incongruence of the condition overall (phonologically but not orthographically congruent) was enough to negatively impact responses in terms of both accuracy and RTs in both ESL groups. Effect sizes were generally large for all pairwise comparisons between P+O- and relative controls (P+O+, P-O-) in Spanish-English and Chinese-English, mostly similar for RTs (d=1.54-1.63) except being smaller between P+O- and P-O- in Chinese-English (d=0.87), but distinctly larger for accuracy in Spanish-English (d=2.1-2.64) compared to Chinese-English (d=1.27-1.46). The differences in effect sizes for accuracy between ESL groups likely just reflect the relatively poor accuracy to P+O- in Spanish-English.

Finally, accuracy was worse to P-O+ than to both relative controls (P+O+, P-O-) in all groups. This was expected and follows previous studies, such as Weber-Fox et al. (2013) also reporting responses being slower and less accurate to P-O+ compared with all other conditions in native English participants and Botezatu et al. (2015) reporting all groups being slower and less accurate to P-O+ than P+O+. As also expected, response times to P-O+ were slower than to P-O-, P+O+, and P+O- in all groups, confirming that this orthographically congruent non-rhyme condition was the most difficult condition across groups. While P-O+ was fully anticipated to be the most difficult, it was not expected to this extent and such relatively low accuracy was not observed in previous studies using a similar task methodology (e.g., Botezatu et al., 2015). Considering accuracy to P-O+ was below chance levels (i.e., <50%) for both Spanish-English and Chinese-English, though, tests of other measures involving P-O+ should be interpreted with increased caution, though, as stated in the Method section (7.3.2), poorer performance was expected and conclusions should not rely on it.

7.5.2 The orthographic nature of P1-O and P1-OT

The main purpose of investigating occipital and occipitotemporal P1 measures was to observe the orthographic nature of P1-O and P1-OT responses to orthographic manipulations and when only real words are used as stimuli (as opposed to pseudowords or pseudohomophones as in Study 1). More specifically, the aim was to investigate any similarities and differences between native monolingual and ESL bilingual readers in terms of early ~100ms P1-O and P1-OT responses, brain activity not well covered in the VWR literature, especially considering the contrasting L1 profiles. While phonological processing has not been readily associated with P1-O or P1-OT activity (and was not hypothesized to be associated with it in this study either), potential phonological leanings were not deemed a priority for the analysis. However, the task used was designed to allow observation of both orthographic and phonological processing, so both orthographic and phonological effects will be considered.

Between groups, prime-TARGET differences of bilateral P1-O latencies to P-O- were larger in English than in Chinese-English, but not different between English and Spanish-English, nor between Spanish-English and Chinese-English, the two ESL groups. This result shows a larger impact of the target in Chinese-English compared with English in P-O-, when the targets shared neither onset nor rime with the prime e.g., joke-GATE. This result was echoed in the behavioural analysis, reflecting longer response times in Chinese-English than English to P-O-. The alphabetic L1 profiles underlying VWR processing in both English and Spanish-English groups go some way to explain them not being different in this regard, suggesting that processing at this ~100ms point in the timeline is similar between them, at least in a VWR context where orthography and phonology are underscored. However, if Spanish-English L2 processing in this context was so similar to native English monolinguals, a difference might reasonably have been expected between Spanish-English and Chinese-English (as observed between English and Chinese-English), but none was found.

Between groups, there were no significant pairwise comparisons for any condition on P1-OT amplitude, suggesting similar P1-OT amplitudes across groups in the context of rhyme judgement in VWR and, furthermore, that this context has no impact on processing associated with P1-OT. Regarding P1-OT latency too, there were no significant pairwise comparisons of prime-TARGET differences between conditions or groups, suggesting this early occipitotemporal ERP response was not dissimilar in timing across groups, regardless of condition. Considering the associations of P1-OT (Dien, 2009), this lack of statistically significant results between groups tentatively supports the hypothesis for P1-OT activity reflecting low-level perceptual processing of orthography (Dien, 2009), which is perceptually lower than the orthographic manipulations in the RRT, and that this holds across monolinguals and bilinguals, at least when reading English.

Between conditions, the prime-TARGET differences in bilateral P1-O and P1-OT amplitudes to P-O+ were larger than to P+O+ in Spanish-English, which echoed the response time finding. The contrast of P-O+ and P+O+ reflects phonological congruence in that the incongruent phonology of the orthographically congruent non-rhymes (P-O+) influenced processing associated with the bilateral P1-OT more than the orthographically and phonologically congruent rhymes (P+O+). However, this is at odds with the usual visual/orthographic associations of P1-O and P1-OT activity during VWR (Dien, 2009). This considered, any effects observed in this timeframe (80-120ms, ~100ms) over occipital or occipitotemporal areas would reasonably be expected to be visual/orthographic in nature also. Such assumptions, however, are largely based on research involving native monolingual VWR processing (and, further, typically in English participants), not L2 VWR processing and not in bilinguals with such a fundamentally phonological language profile as Spanish-English participants. Therefore, while a tentative proposition without further evidence or replication, this finding suggests a different VWR pathway in Spanish-English bilinguals reading English.

Comparing P-O+ and P-O-, meanwhile, reflects an orthographic effect through orthographic congruence and bilateral P1-OT amplitude was also larger to P-O+ than to P-O- in Spanish-English, thus showing both phonological and orthographic effects in P1-OT in the Spanish-English group only. This P-O+ and P-O- contrast shows influence of the repeated orthography in P-O+ compared with the relative control that shared neither orthography nor phonology between prime and target (P-O-). If the influence of P-O+ was just orthographic, an effect would not be expected between it and P+O+. Likewise, an effect would not be expected between it and P-O- if the influence was just phonological. However, both were observed in the Spanish-English group and, importantly, there were no P1-OT amplitude differences between relative controls (P+O+,P-O-), showing that one was no more facilitative or inhibitive than the other. Therefore, it is reasonable to attribute this effect to the stark orthographic/phonological manipulation and overall incongruence in P-O+ itself, created from the combination of orthographic similarity, phonological difference, and the context of visual rhyme recognition.

The larger prime-TARGET differences in bilateral P1-OT amplitude to P-O+ than to P+O- in Spanish-English provides further support that the orthographically congruent non-rhyme condition (P-O+) required more cognitive effort from processes associated with the P1-OT and proved the more difficult incongruent condition compared with the orthographically incongruent rhyme condition (P+O-). The behavioural findings in terms of RTs in every comparison with P-O+ concur with this interpretation, though evidence from ERPs was only found in Spanish-English.

On that point, it should be noted that larger amplitudes to P-O+ were not only observed in Spanish-English bilinguals, but observed in the group with an L1 that has a shallow orthography, more direct orthography-phonology mappings, and arguably a greater dependence on phonology during VWR, while not in the English or Chinese-English groups with a more orthographically complex language profile. This finding only being observed in Spanish-English, the group with the phonology-oriented orthographically shallow L1, is one of the most important factors. It lends weight to it as a phonological effect, suggesting that processing may work differently in Spanish-English bilinguals compared with bilinguals with a more orthographically complex language profile. Furthermore, this phonological effect not being found in English or Chinese-English, the groups with orthographically complex language profiles is noteworthy, as effects on P1-O and/or P1-OT would more likely require an orthographic manipulation in these groups.

As noted in the discussion of behavioural performance, however, the accuracy of the Spanish-English group in particular for the P-O+ condition require that these P1-OT amplitude results be considered with caution. The poorer behavioural accuracy for P-O+ in Spanish-English could be interpreted as a link with the larger P1-OT amplitude differences, associated with the additional processing effort required for the P-O+ condition. However, the extent of behavioural errors to P-O+ in Spanish-English make it more likely that the significant statistical differences between P-O+ and other conditions are due to smaller included trial counts in the ERP averaging and, therefore, poorer signal-to-noise ratio. The P1-OT amplitude findings presented here could, indeed, be legitimate, but a replication with increased trial counts is required before they can be deemed reliable.

7.5.3 Early phonological activation and N100-FC

As in Study 1, the focus on early pre-200ms frontal-central ERP responses to phonological stimuli follows the compelling but somewhat controversial evidence for phonological activation as early as ~100ms (e.g., Ashby, 2010; Pammer et al., 2004; Wheat et al., 2010). Due to the emphasis on phonology in a VWR context, the rhyme judgement task of the current study was designed and is primed for evaluating such early phonological activation with the added benefit of using real words within a single task that manipulates orthography equivalently alongside the phonology.

Following the full three-way interaction of Condition x Hemisphere x Group on N100-FC amplitude, pairwise comparisons showed prime-TARGET differences in N100-FC(L) amplitudes to P-O- were larger in English than Spanish-English, which is mirrored by response times that were faster in English. As P-O- involves neither phonological nor orthographic priming, designed as the true baseline measure for the other conditions, prime-TARGET differences in it essentially reflect reading a pair of orthographically and phonologically (as well as semantically) unrelated English words in relatively quick succession with only the instructions of the (rhyme judgement) task to influence processing i.e., the same instructions as for all other conditions. Therefore, this finding shows that the presentation of the second orthographically and phonologically unrelated word resulted in a stronger N100-FC(L) response in English compared with the same response in Spanish-English. Though not orthographic nor phonological priming as discussed previously and in other parts of this analysis, it appears that the orthographic and/or phonological novelty of the second unrelated word i.e., P-O-TARGET was sufficient to facilitate brain activity in left frontal-central scalp sites at ~100ms post-stimulus, at least relative to the Spanish-English group. While this finding is not directly in line with those of Wheat et al. (2010) and Pammer et al. (2004), not being the observation of a phonological processing difference between conditions per se, it does support the notion of left frontal-central ERP activity at ~100ms being pertinent to processing during VWR. The very early ~100ms timeframe is what was contentious about previous reports of such early phonological activation and is what is at least partially supported here.

The full three-way interaction of Condition x Hemisphere x Group was also observed for N100-FC latency. For latency, however, N100-FC(R) to P-O- was significantly larger in English than in Chinese-English, which also mirrored faster response times in English. However, this is likely due to the nature of processing between prime and target in the P-O- condition, as the prime-TARGET differences reflected shorter latencies to P-O- TARGETs than primes in English, but shorter latencies to primes than P-O- TARGETs in Chinese-English. In other words, targets of the P-O- condition resulted in an earlier N100-FC in English, facilitating the processing associated with the N100-FC, while the same P-O- targets inhibited N100-FC activity in Chinese-English (such inhibition was also apparent in Spanish-English for P-O-, but statistical significance was not reached).

Between hemispheres, prime-TARGET differences of N100-FC latency to P-O+ were smaller in the left than right hemisphere in Spanish-English, where N100-FC(R) to P-O+ reflected shorter latencies to TARGETs, while N100-FC(L) exhibited longer latencies to TARGETs. While the same caution should be taken as with the P1-OT findings regarding trial counts to P-O+, the longer N100-FC(L) latency can be associated with additional phonological processing, based on the overall phonological incongruence (in the context of rhyme judgement) of P-O+. Again, while not directly following the findings of Wheat et al. (2010), it does support the notion of phonological activity at ~100ms over left frontal-central sites.

7.5.4 The nature of the N170 and its hemispheric laterality in VWR

As with Study 1, the attention to occipitotemporal N170 activity here was to investigate its nature during VWR in terms of orthographic processing, orthography-phonology mapping, and phonological activation, as well as to observe its hemispheric lateralization according to orthographic/phonological manipulations and differences between groups with different L1 profiles. No relevant main effects, interactions, or pairwise comparisons were observed for N170-OT amplitude within or between any condition, hemisphere, or group. Accepting that such little statistical significance from so many calculations does not reflect Type II errors, this widespread lack of differences in N170-OT amplitude suggests that the brain activity in this timeframe (130-200ms, ~170ms) at occipitotemporal scalp sites and associated VWR processing are similar across English, Spanish-English, and Chinese-English groups, at least in the context of rhyme judgement. More specifically, this lack of effects shows that neither orthographic nor phonological manipulations of the RRT impacted N170-OT amplitude and, importantly, that there was no clear lateralization of the N170-OT response in any group.

The only relevant statistically significant finding on N170-OT latency was a main effect of Group, where larger prime-TARGET differences in English than both Spanish-English and Chinese-English were observed (averaging across conditions and hemispheres). As this result does not involve the condition or hemisphere factors, it has little bearing on the questions this study was designed to explore. However, the difference in English reflects a facilitatory priming effect and that appears to be of greater magnitude in the native monolingual English group than in either Spanish-English or Chinese-English ESL groups, which is of general note to VWR research and, based on overall N170-OT latencies between native monolingual and bilingual groups (e.g., Maurer et al., 2008; Tong et al., 2016), it could be worth revisiting in future research.

Based on the variety of prior associations with N170-OT activity, especially concerning orthography and script familiarity (Maurer et al., 2008), as well as orthographic-phonological mapping, the relative lack of statistically significant results from N170-OT analyses is somewhat surprising. However, also lacking are comparable investigations in the literature using more corresponding samples, stimuli, and tasks. From the results of the current study, therefore, it appears that N170-OT responses do not distinguish between English monolinguals, late bilingual Spanish-English, and late bilingual Chinese-English ESL readers, and are not affected by overt orthographic or phonological priming.

7.6 Study summary

Several distinctions in VWR processing were observed between groups based on measures from the rhyme judgement task. Behavioural responses (accuracy and response times) were similar overall to previous reports (e.g., Botezatu et al, 2014; Chen et al., 2010; Coch, Hart, & Mitra, 2008), but not identical, which is suspected to concern English proficiency in terms of the ESL participants being early (previous research) or late (current research) bilinguals. Response times in English were faster than in Spanish-

English and Chinese-English to all conditions P+O+, P+O-, P-O+, and P-O-, while Spanish-English and Chinese-English did not differ in RTs to any condition P+O+, P+O-, P-O+, and P-O-. In terms of accuracy, there were no differences between any groups to either of the relative controls (P+O+, P-O-), though accuracy was higher in English than in both Spanish-English and Chinese-English, as was expected for a native monolingual group compared with late bilingual ESL groups. Accuracy was also higher in Chinese-English than Spanish-English for both P+O- and P-O+, suggesting the orthographic-phonology incongruence of these conditions was more problematic for the Spanish-English group used to a shallow orthography in their L1. The overall incongruence of the P+O- condition (phonologically but not orthographically congruent) resulted in lower accuracy and slower responses to P+O- than to both relative controls (P+O+,P-O-) in both ESL groups, but not in the English group who were as accurate to P+O- as the fully congruent controls (P+O+, P-O-).

Much of the behavioural performance was as expected: native English participants were more accurate and quicker to respond overall and the incongruent conditions resulted in processing deficits, especially P-O+ in all groups. This orthographically similar non-rhyme condition (P-O+) was confirmed to be the most difficult across all groups, the extent to which was not expected: accuracy to this condition was below chance (i.e., <50%) for both ESL groups. The overall similarity of behavioural performance between ESL groups, however, was also not expected. That said, differences in the Spanish-English group but not in English or Chinese-English groups (e.g., lower errors and RTs to P-O-) suggest language profiles being pertinent to behavioural performance.

In terms of ERPs at ~100ms, there were no differences in P1-OT amplitude or latency between groups, but the prime-TARGET differences of bilateral P1-O latencies to P-O- were smaller in Chinese-English than in the native English group, while not different between English and Spanish-English or between the two ESL groups, showing a distinction in the orthographic processing of English in the Chinese-English group.

Mirroring the faster response times in the English group (than both Spanish-English and Chinese-English groups), prime-TARGET differences in N100-FC(L) amplitudes were larger in English than Spanish-English to P-O-. Following the especially phonological nature of the L1 of the Spanish-English group and the previously reported sensitivity to phonological processing at left frontal-central sites in the ~100ms timeframe (e.g., Wheat et al., 2010), this finding highlights the potential involvement of N100-FC activity in orthographic/phonological processing of English, at least in native readers. Meanwhile, prime-TARGET differences of N100-FC(R) latencies to P-O- were significantly larger in English than in Chinese-English, showing orthographically and phonologically dissimilar word targets to impact processing associated with right-lateralized frontal-central activity at ~100ms in the English group more than the Chinese-English group, suggesting a distinct VWR strategy between these groups.

Lastly (and somewhat unexpectedly considering the previous reports of N170-related VWR effects), there were no directly relevant findings from the analyses of N170-OT amplitude or latency within or between conditions of the rhyme judgement task. The only slight exception was the overall larger prime-TARGET differences of N170-OT latency observed in the native English group compared with both Spanish-English and Chinese-English groups, suggesting a relative delay of processing associated with bilateral N170-OT activation in the ESL groups. Findings suggest that different cognitive strategies for VWR are employed by the different ESL groups. As anticipated (e.g., Weber-Fox et al., 2013), this is especially the case when dealing with incongruent orthography/phonology and can be seen to follow the language profiles of the ESL groups.

Chapter 8: General discussion

This thesis has documented original empirical research into how initial orthographic and phonological processing of English manifests in terms of brain and behaviour in skilled readers with fundamentally different L1 profiles. The overall objective of the research was twofold: examine the timeframe of orthographic and phonological processing during English VWR and investigate the potential influence of L1 language profile on such processing of L2 (English) orthography and phonology in bilingual ESL readers. The research, therefore, examined contrasts of native monolingual (English), non-native alphabetic L1 late bilingual (Spanish-English), and non-native non-alphabetic L1 late bilingual (Chinese-English) readers of English in terms of behavioural accuracy and response times beside amplitude and latency measures of precursory (pre-200ms) ERP activity. The specific motivations of the ERP approach were to examine pre-lexical and pre-attentive orthographic and phonological processing, focusing on ERP indices previously implicated in early VWR processing: the occipital P1, occipitotemporal P1, frontal-central N100, and occipitotemporal N170. The occipital P1, as opposed to its other associations (e.g., lower-level visual attention processes), was used to examine initial orthographic processing, while the occipitotemporal P1, also examined for its association with orthographic processing, was used to explore early orthography-phonology mapping. Additionally, the investigation of ~100ms frontal-central activity followed reports of controversially early phonological activation and its likelihood to be linguistic. Lastly, the nature of the occipitotemporal N170 was inspected due to its orthographic and/or phonological role in VWR and the association of its lateralization with script expertise. Vitally, how such antecedent ERP activity is associated with VWR behavioural responses within and between groups rounded the aims of the research. Consequently, the approach used a combined behavioural/ERP methodology involving different linguistic stimuli (words, pseudohomophones, pseudowords, sight and sound rhymes) in a novel complement of lexical decision task and rhyme judgement task variants to identify distinctions in orthographic and phonological processing within and between groups. The two central themes of processing timeframe and bilingual distinctions will be discussed in the following two sections, respectively, albeit with some overlap in each due to the inextricable relationship between themes in the context of the current work.

8.1 Early orthographic and phonological processing

Considering the notion that the foundation of VWR involves phonological representations being rapidly if not automatically extrapolated from orthographic input (Rastle & Brysbaert, 2006), initial orthographic and phonological processing is vital for understanding later processing and how written language is read and understood. More specifically, the questions relate to the timing and interactive nature of orthography- and phonology-oriented responses to linguistic stimuli and the extent that phonological activation is required during VWR cf. strong/weak phonological theories (e.g., Frost, 2003; Rastle & Brysbaert, 2006). All of this also relates to whether such processes are consistent across readers with different language profiles that differ in orthographic depth and thus their reliance on phonology, such as ESL readers with an alphabetic L1 e.g., Spanish-English or logographic L1 e.g., Chinese-English. This section will discuss the main overall findings in terms of the timing of orthographic and phonological processing.

Based on the lack of differences found between groups or between conditions using an orthography-focused task (the oLDT), the occipital P1 did not reflect early orthographic processing in any group. While this was not unexpected and is in line with some previous work (e.g., Hauk et al., 2006), it does oppose other claims of such early VWR-related associations with the occipital P1 (e.g., Segalowitz & Zheng, 2009). Following its connections to the visual areas of the (extrastriate) cortex as a visually evoked potential (VEP), the occipital P1 is more typically associated with visual processing as well as being an index of attention (Dien, 2009; Luck et al., 1994; Mangun & Hillyard, 1991). Considering clear cognitive links between vision, attention, and VWR, early orthographic processing being reflected by the occipital P1 is not unreasonable. However, a more likely explanation is that the sources of previously reported P1 effects are prelinguistic or early visual/orthographic factors that were not manipulated or investigated in the current research e.g., n-gram frequency, word shape (Hauk et al., 2006; Hauk et al., 2008; Dien, 2009).

While the occipital P1 appeared largely neutral to orthographic and phonological manipulations in the current research, the occipitotemporal P1 was further evidenced as being associated with orthographic processing and orthography-phonology mapping in the context of VWR (Dien, 2009). Findings highlighted the dynamic context-dependence of VWR processing (cf. Huettig & Ferreira, 2022), using the route to VWR that appears optimal based on bottom-up aspects of stimuli and task as well as top-down factors of language profile, suggesting that the relative exclusion of top-down processing (i.e., the task/decision subsystem from affecting the word identification subsystem) be reconsidered in the BIA+ (van Heuven & Dijkstra, 2010). For instance, larger P1-OT amplitudes reflected additional orthography-phonology checking when a switch of VWR strategy is required, such as between a direct lexical route for known real words (e.g., RW) and grapheme-phoneme conversion to read pseudohomophones (e.g., PH1) or between orthographic and phonological contexts (e.g., the oLDT (PH1) and the pLDT (PH2) of Study 1). The generally accepted difference in processing between these stimulus types concerns direct orthographic/lexical recognition for known real words and phonological grapheme-phoneme conversion for pseudohomophones. This distinction between conditions, therefore, reflects the more automatic, likely whole-word identification of the real word stimuli as being orthographically familiar (whereas the pseudohomophones were not), if not in fact as recognized words, as posited in the DRC (Coltheart et al., 2001) and BIAM (Grainger & Holcomb, 2009).

Although increased cognitive effort is expected in the phonological task of Study 1 (and, therefore, to its pseudohomophones, PH2) relative to the orthographic task (and its pseudohomophone, PH1), both pseudowords and pseudohomophones (pLDT conditions PW,PH2) require some level of grapheme-phoneme conversion. Meanwhile, routes to read real words and pseudohomophones (oLDT conditions RW,PH1) differ: whole-word direct lexical route for real words and sub-lexical grapheme-phoneme conversion for the pseudohomophones. Based on the oLDT strategy being to use the direct lexical route, which should provide the answer of whether the stimulus is a real word (successful direct lexical route) or not (unsuccessful direct lexical route), increased P1-OT amplitudes to the oLDT pseudohomophones (PH1) reflect additional orthographic checking of these stimuli according to the orthographic and direct lexical context of the task (i.e., orthographic processing is theoretically sufficient to complete it). Furthermore, if querying the direct lexical route is the strategy for the oLDT, it is also a factor for not observing orthographic lexicality effects in this timeframe, as the initial response and processing would be the same (or similar) for both stimuli (RW,PH1).

The increased effort reflected by increased P1-OT amplitudes during L2 processing of orthographically and phonologically illegitimate pseudowords relative to phonologically legitimate pseudohomophones demonstrates the involvement of activity at the very early ~100ms timeframe over occipitotemporal sites related to orthography-phonology mapping. This finding is in stark contrast to the lack of expected orthographic effect found for P1-OT in the oLDT and to the theorized orthographic associations of the P1-OT (Dien, 2009). Instead, this observation suggests a phonological (not orthographic) facet to the P1-OT based on the phonological lexicality effect it represents through a difference between orthographically matched stimuli that differ only in phonological legitimacy i.e., whether their phonology matched that of a real word (pseudohomophones) or not (pseudowords). Importantly, however, this phonological lexicality effect (larger bilateral P1-OT amplitude to PW compared with PH2) was only found in Spanish-English, showing a clear distinction between groups at this very early timeframe of VWR processing. One of the major distinguishing factors of the Spanish-English group compared with the others in this study is the shallow orthography and relative phonology-dependence of their L1 and, by extension, their theoretical basis for learning and using other languages (Tan et al., 2005). It is, therefore, not unreasonable that previously considered orthographic associations of the P1-OT in populations with orthographically deep native languages (e.g., English, Dutch) are phonological or, at least, more phonological than purely orthographic in such a population as Spanish-English late bilinguals. Considering the parallel processing mechanisms for sublexical processing and phonological onset timing being somewhat lacking in theories of bilingualism (Dijkstra et al., 2019), findings should be considered in developing the sublexical orthography and phonology elements of the Word Identification System of the BIA+ and Multilink models (Dijkstra et al., 2019; van Heuven & Dijkstra, 2010).

The investigation of frontal-central negativity (i.e., N100-FC) was centred on it being posited as an index of early (~100ms) phonological processing during VWR based on the somewhat controversial prior reports of phonological activation in this timeframe over these sites (e.g., Klein et al., 2015; Pammer et al., 2004; Wheat et al., 2010). While the current research did not explicitly replicate the findings of these studies, most likely due to such methodological differences as the experimental paradigms used in these other studies (e.g., masked priming), the findings do perpetuate the potential of early frontal-central negativity (N100-FC) reflecting parallel phonological processing. For instance, larger bilateral N100-FC amplitudes were observed in the non-native alphabetic L1 (Spanish-English) group through phonological lexicality effects in Study 1 showing increased effort for processing pseudowords (orthographically and phonologically illegitimate) compared with pseudohomophones (only orthographically illegitimate). Considering the phonological difference between pseudowords and pseudohomophones, it is not unreasonable to suggest that this directly links the N100-FC with phonological processing and somewhat supports the reports of ~100ms frontal-central activity during VWR. Such evidence should, therefore, be considered in developing implementations of parallel sublexical orthographic and phonological processing in the BIA+ and Multilink models (Dijkstra et al., 2019; van Heuven & Dijkstra, 2010).

It is acknowledged that effects on N100-FC amplitudes could be an electrophysiological reflection or even paradoxical lateralization of the effect on P1-OT amplitudes (or vice versa), as has been found for the later N400 (Lau, Phillips, & Poeppel, 2008). However, associations of frontal-central negativity and occipitotemporal positivity at ~100ms post-stimulus via ERPs with early processing of phonology and orthography, respectively, are compelling. These effects taken in tandem tentatively suggest parallel processing of orthography and phonology at this early ~100ms timeframe and necessitate deeper investigation. Furthermore, this queries the assumption of BIA (Dijkstra & van Heuven, 2002), BIA+ (van Heuven & Dijkstra, 2010), and Multilink (Dijkstra et al., 2019) models of bilingualism as well as the DRC (Coltheart et al., 2001) and BIAM (Grainger & Holcomb, 2009) for monolingual VWR that orthography is evaluated first in order for it to be linked to phonological representations (Dijkstra et al., 2019). While this fundamental premise of VWR theory more broadly that orthographic processing occurs before phonological processing is not in question per se, such findings do further emphasize the likelihood of cascaded/overlapping processing of orthography and phonology (Coltheart et al., 2009; Grainger & Holcomb, 2009).

Considering the psycholinguistic associations with the intrinsic processing of visually-presented language in terms of orthography-phonology mapping as well as of experience of language type and script expertise with the N170 (as discussed in detail in §3.1.3), there was a distinct and somewhat unexpected lack of effects on N170-OT amplitude, latency, or lateralization overall. Indeed, finding no psycholinguistic effects, while not unprecedented, is not the norm, but means the current research does not support previously reported evidence VWR-related N170 lateralization reflecting language expertise (Amora et al., 2022; Ma et al., 2022; Maurer, Brandeis, et al., 2005; Maurer, Brem, et al., 2005; Maurer et al., 2008; Yum & Law, 2021). It is possible that the strict stimulus control of conditions, including the relative simplicity and high frequency of the stimuli, explains why such effects found in previous studies were not observed here. For instance, N170 activity has been linked with bigram analysis (Dien, 2009), but all conditions were matched for bigrams, trigrams, and their relative frequencies. N170 activity has also been attributed to being sensitive to real orthography over “fake” (Simon, Bernard, Largy, Lalonde, & Rebai, 2004) or “nonorthographic” (Bentin et al., 1999) linguistic stimuli, distinguishing low-level visual stimuli from more linguistic inputs and potentially being a critical point between pre-linguistic and linguistic processing. However, all conditions in the current study were matched on legal orthography, differing only in legitimate orthography. Therefore, the lack of differences in N170-OT amplitude reported here follows previous studies in that the main findings have concerned manipulations of orthographic legality, not legitimacy. It is also possible that such N170 effects (on lateralization and/or amplitude) are more dynamic in task-sensitivity or even task-specificity than previously posited (Amora et al., 2022). Indeed, lexical decision tasks do not tap into the same processing dynamic as tasks used in previous studies e.g., repetition detection using real words, symbol strings, and pseudowords with picture distractors (Maurer, Brandeis, et al., 2005). Essentially, N170-OT amplitude can respond to linguistic manipulations in some contexts , but the distinctions between real words and pseudohomophones and between pseudohomophones and pseudowords, respectively, were not sufficient in the current study to elicit any amplitude differences. The extent that these effects were not observed in the current research suggests that at least such VWR-related N170 effects are not consistent and further replication either supporting or refuting is required (Amora et al., 2022; Yum & Law, 2021).

The current research further supports orthography-phonology mapping and cascaded/overlapping orthographic/phonological processing within ~200ms, the potential for early phonological activation at ~100ms, and distinctions between groups within these findings. While there were some parallels observed between the matched ESL groups, the current research identified various characteristics in behaviour and electrophysiology based on language profiles and whether L1 is native, non-native and alphabetic, or non-native and non-alphabetic.

8.2 Distinctions in VWR processing between language profiles

Regarding group differences, it is important to first highlight the native group as a viable control and the contrast of behavioural responses with the ESL groups. This showed the expected performance distinction of the native group consistently responding more quickly and accurately, which is broadly explainable by the late bilingual status and lesser experience and proficiency of the ESL groups (compared with natively proficient monolinguals). For instance, orthographic incongruence in real words was sufficiently and similarly inhibitive for both ESL groups, who were slower and less accurate to the phonologically congruent but orthographically incongruent P+O- condition compared with congruent controls (in contrast with the equivalent accuracy to P+O- and controls of the native group in the rhyme recognition task). Likewise in the pLDT, phonological effects of facilitative legitimate phonology (in pseudohomophones) and an inhibitive lack of legitimate phonology (in pseudowords) were observed in the native group, while both ESL groups responded similarly slowly to pseudowords and pseudohomophones. The simplest explanation for why the ESL groups responded similarly to English-based pseudohomophones (e.g., PH2) and pseudowords (e.g., PW) and struggled similarly with orthographic incongruence in real words is that their approach to processing unfamiliar L2 words was similar in that the primary focus for both groups appeared to be the legal but not legitimate orthography (present in both stimulus types) despite the overtly phonological objective of the task (the pLDT). In contrast, the differences observed in the native monolingual group (but not in the ESL groups), such as between pseudohomophones (e.g., PH2) and pseudowords (e.g., PW), are representative of a native/non-native distinction in processing non-legitimate orthography, which can be attributed to the greater proficiency with English phonotactic and syntagmatic rules in the native group. These findings illustrate the potential impact of orthographic variation in VWR, signifying a more orthography-weighted sublexical approach to VWR for ESL readers and that theories of bilingualism need to better take the variety of orthographies into account. Furthermore, it suggests orthography/phonology mapping during VWR of English can take similar processing times in alphabetic L1 and non-alphabetic L1 late bilingual ESL readers, especially when VWR requires overt phonology and/or grapheme-phoneme conversion as in rhyme judgement and phonological lexical decisions. The consistent (but expected) disparity in behavioural performance between native and ESL groups alongside comparisons between ESL groups not always being significantly different (e.g., behavioural performance to RW in the oLDT or to any RRT condition P+O+, P+O-, P-O+, and P-O-) show that there are parallels in VWR processing between alphabetic L1 and non-alphabetic L1 ESL readers, somewhat irrespective of their language profile. Furthermore, the overall similarity of behavioural performance between ESL groups supports the groups being well-matched on English proficiency, bolstering interpretations that any group differences can be attributed to language profiles (the distinguishing factor of the ESL groups).

Similarities notwithstanding, there were various distinctions between groups. Differences in behaviour between ESL groups support the notion of differential VWR processing based on language profiles being better integrated into theories of bilingualism, such as that phonology is relied on more in ESL readers with an alphabetic L1 profile e.g., Spanish-English and orthography is relied on more in ESL readers with a non-alphabetic L1 profile e.g., Chinese-English. For instance, orthography/phonology-related brain activity in orthographic or phonological task contexts showed a distinction between groups with occipitotemporal P1 responses being larger to pseudohomophones in an orthographic task (PH1) than to pseudohomophones in a phonological task (PH2) in English and Spanish-English groups, but right-lateralized in English and left-lateralized in Spanish-English, and no difference in Chinese-English. This distinction in P1-OT lateralization is especially important because it differentiates between all three groups and links to prior accounts of the occipitotemporal P1 for orthographic processing, which these effects, especially with the right-lateralization in the native English group, firmly support, (Dien, 2009). More specifically, this difference in lateralization, considering the orthographic task effect it is linked to, likely reflects a difference in processing strategy that can be attributed to the differences in language profiles between groups and should be extended upon in theories such as Multilink (Dijkstra et al., 2019).

Such occipitotemporal P1 amplitudes being larger in an overtly orthographic context compared with a phonological one not only shows the sensitivity of this ~100ms post-stimulus timeframe at occipitotemporal sites to VWR-related and, specifically, orthographic processing, but that top-down influences of task can impact VWR processing as early as ~100ms and, vitally, in non-native readers too. Furthermore, this pattern of occipitotemporal P1 amplitude between orthographic and phonological contexts is shared by the two groups with an alphabetic native language and, especially as it was not found in the non-native non-alphabetic L1 group, reflects a more naturally alphabetic approach to VWR and could be an effect specific to readers with an alphabetic L1. Such occipitotemporal P1 lateralization, meanwhile, distinguished between native alphabetic L1 (English) and non-native alphabetic L1 (Spanish-English) groups, indicating another native/non-native distinction, this time specific to readers with an alphabetic L1 and reflecting right-lateralized native processing and left-lateralized non-native processing in overtly orthographic contexts.

Processing English-based pseudowords was also found to distinguish between groups through a behavioural speed-accuracy trade-off with the native monolingual group being more accurate and the non-alphabetic L1 ESL group being faster, while not involving the alphabetic L1 ESL group. Overall, these distinctions between groups in processing pseudohomophones and pseudowords relate to the Word Identification System (where orthography, phonology, and semantics are processed) and Task Decision System (which convolves top-down information with the input of the Word Identification System) of bilingual VWR models e.g., BIA/BIA+/Multilink and BIA+/Multilink, respectively. Specifically, these distinctions query the BIA+ assertion that the task/decision subsystem is precluded from influencing the word identification subsystem (Dijkstra et al., 2019; Dijkstra & van Heuven, 2002b; van Heuven & Dijkstra, 2010). The apparent effect of the overarching phonological task (the pLDT) on ESL readers' responses to pseudohomophones and pseudowords instead suggests that there is room for such top-down processing in (bilingual) VWR. These same findings also indirectly query the importance of language membership for bilingual VWR, as also posited by the BIA+ model (van Heuven & Dijkstra, 2010). By no means is this suggesting language selective access to be the case, but that the involvement of language membership in the L2 VWR process is relative to the readers and their language profiles, being somewhat dependent on the relationship between L1 and L2, such as the similarity of scripts or linguistic proximity between languages.

Several further distinctions of VWR processing were strongly linked to either the phonological L1 profile of Spanish-English ESL readers in terms of Spanish having a particularly shallow orthography (especially in contrast with English and Chinese/Mandarin) or the more orthographic L1 profile of Chinese-English ESL readers with logographic languages (e.g., Chinese/Mandarin) being primarily orthographic in the sense of being especially visual and not having the orthography/phonology association found in alphabetic languages. For instance, a key phonological effect of phonological lexicality was found only in the Spanish-English group, where bilateral P1-OT amplitudes were larger to pseudowords (PW) than to pseudohomophones (PH2). Furthermore, behavioural accuracy and response times showed orthographic-phonology incongruence (e.g., P+O- and P-O+ conditions) to be more problematic for the alphabetic L1 group (ESL readers used to a shallow orthography in their L1) than the non-alphabetic L1 group with their logographic L1 that has little to no direct orthography/phonology mapping as found in alphabetic languages. Due to this heavier focus on visual/orthographic factors in logographic scripts e.g., Chinese/Mandarin, orthographic processing in the non-alphabetic L1 group was distinct from alphabetic L1 groups. For instance, group comparisons showed a differential response to dissimilar and incongruent orthography, as observed for bilateral P1-O latency to the overall relative control condition in the rhyme judgement task (P-O-). This was especially notable in the Chinese-English group where the contrast of congruent and incongruent orthography was reversed and facilitation (as in English and Spanish-English) was replaced by inhibition. Such a finding again highlights the variance in cognitive control mechanisms between bilinguals with different language profiles for resolving psycholinguistic conflicts, showing further research is required, especially with non-alphabetic populations (van Heuven & Wen, 2018) and especially because inclusion of language background, relevant L1 VWR skills, and similarities between L1 and L2 are lacking in theories of bilingualism (Dijkstra et al., 2019). Furthermore, it should be noted that this P1-O latency finding was echoed through slower behavioural response times in Chinese-English than English in the P-O- condition, again proposing a potential link between early electrophysiology and related behavioural responses during VWR. Overall, this suggests some dependence on the reader via their language profile (as opposed to just the target language or act of reading in general) for the extent that phonological activation is required to complete the task at hand i.e., whether a strong or weak phonological approach is applicable.

Observing such phonological effects as the larger bilateral N100-FC amplitudes to pseudowords than pseudohomophones only in the alphabetic L1 (Spanish-English) bilingual group links to their orthographically shallow L1 background and suggests the increased sensitivity to phonology from the Spanish language is why N100-FC effects were found in the alphabetic L1 group but not the native or non-alphabetic L1 groups. The relative ease of the stimuli/tasks for the native monolingual group might also explain why N100-FC effects were not found in the native group. Not observing them in the non-alphabetic L1 group could indicate an alternative processing approach in the non-alphabetic bilingual group or at least an indication that frontal-central activity is not so directly recruited by the more orthographic approach proposed for non-alphabetic L1 ESL readers. Such findings of differential processing dependent on language profiles again highlight the potential variance in cognitive control between bilinguals with different language profiles and, following the notion of an integrated lexicon in bilinguals being chiefly for vocabulary/semantics and not the whole story for syntax/grammar (Brysbaert & Dijkstra, 2006; Brysbaert & Duyck, 2010), suggest the cognitive resources required for L2 processing differ based on task and/or L1-L2 relationship.

Alongside the phonological effects on the N100-FC being observed only in the alphabetic L1 bilingual group (Spanish-English) and not in the native monolingual (English) or non-alphabetic L1 bilingual (Chinese-English) groups, further distinctions based on language profiles were found. For instance, bilateral N100-FC latencies were shorter in Spanish-English than in Chinese-English to both real words (RW) and pseudohomophones (PH1), which is likely a reflection of expertise in L2 language type considering the target language being alphabetic and the frontal-central negativity being earlier in the alphabetic L1 group. Furthermore, frontal-central N100 activity differed in the left hemisphere between English and Spanish-English groups but in the right hemisphere between English and Chinese-English groups (e.g., to P-O-). Relative to the native monolingual control group, therefore, alphabetic L1 and non-alphabetic L1 ESL bilinguals exhibit frontal-central activity for the same new phonology in different hemispheres, indicating different L1-oriented strategy used for L2 VWR. This difference in N100-FC lateralization between ESL groups could also suggest that left frontal-central negativity reflects phonological activity for the native language type (i.e., alphabetic for the alphabetic L1 Spanish-English group) and the right hemisphere for non-native language type (i.e., alphabetic for the non-alphabetic L1 Chinese-English group).

Despite the lack of psycholinguistic effects on the N170-OT, as mentioned in the previous section, there were aspects of the N170-OT that did distinguish between groups. Brain and behavioural responses linked to orthographic processing were observed through effects of orthographic lexicality in the native group only, where N170-OT latency was later to pseudohomophones (PH1) than to real words (RW). Based on the processing requirements for real words and pseudohomophones and considering the orthography-phonology mapping hypothesis associated with N170 activity during VWR, this key finding does support the N170-OT as an index for dynamic orthography-phonology mapping via grapheme-phoneme conversion, at least within the native group as described in the DRC (Coltheart et al., 2001) and BIAM (Grainger & Holcomb, 2009). Moreover, behavioural response times, which were faster to real words (RW) than to pseudohomophones (PH1), echoed this effect on N170-OT latency, providing a direct link between the precursory electrophysiological activity and resulting behaviour during VWR. This link needs even more specific investigation, but already further evidences the neural underpinnings of the occipitotemporal N170 as heavily involved in orthography-phonology mapping (Amora et al., 2022). Indeed, the different response time and N170-OT activity being based on a direct lexical pathway (i.e., real words, RW) or grapheme-phoneme conversion (i.e., pseudohomophones, PH1) supports the notion of multiple routes to VWR as outlined in the DRC (Coltheart et al., 2001) and BIAM (Grainger & Holcomb, 2009) with both behavioural and electrophysiological data, which can be extended to the BIA+ and Multilink.

As mentioned above, N170 activity and lateralization in particular has been reported to reflect familiarity and expertise with a particular script (Maurer, Brandeis, et al., 2005; Maurer et al., 2008). However, this effect on N170-OT latency did not favour the left (as expected of N170 effects in native alphabetic VWR) or right hemisphere, something that requires further investigation. That said, the link with expertise also fits here, as this effect was not present in either Spanish-English or Chinese-English, despite the relatively high frequency of the stimuli and the relative high English proficiency of the ESL groups. Although not how the N170 association with orthographic script-based linguistic expertise has previously been described in the literature (e.g., Maurer, Brandeis, et al., 2005; Maurer et al., 2008), the divergence in N170-OT latency between groups feeds a similar notion and suggests the properties of the N170-OT (whether lateralization or latency) can be associated with linguistic expertise. It follows that the observed sensitivity of N170-OT latency to orthographic familiarity through real words compared with pseudohomophones was due, at least in part, to the lifelong expertise and exposure to English of the native English readers.

Ultimately, findings of the current research confirm that the language profile and, specifically, the type and orthographic depth of ESL readers' L1 does have an intrinsic impact on ESL L2 processing. Furthermore, based on early ERP effects within and between groups, the influence is more fundamental than previously acknowledged. Again, such evidence of processing dependent on language profiles firmly suggests the cognitive resources required for L2 processing differ based on L1-L2 relationship and should be considered in development of bilingualism theories e.g., BIA+ and Multilink models (Dijkstra et al., 2019; van Heuven & Dijkstra, 2010).

8.3 Limitations and future research

Findings from the current research, especially those showing processing divergences related to language profiles, not only apply to learning to read a second language but reinforce how readers with different L1 profiles might require alternative TESOL approaches. L2 proficiency has been highlighted as a vital factor in understanding the impact of early, late, and native-like L2 proficiencies on behavioural and ERP measures during VWR. The use of just two late-bilingual groups, however, to represent the much more complex range of language profiles is acknowledged as a technical limitation for generalizing the findings and making specific predictions or recommendations regarding TESOL. Nonetheless, the current research provides a foundation for further examining how the spectrum of language profiles might relate to and influence L2 ESL processing, encouraging comparisons using other languages with different properties. For instance, native Korean ESL readers with the Hangul alphabet and Japanese ESL readers with the hiragana syllabary could be used as contrasts to the Chinese/Mandarin logographic system, while a non-native non-shallow alphabetic L1 bilingual groups (e.g., German-English, Dutch-English) involving different levels of orthographic depth in L2 (e.g., French and German ESL readers for deeper alphabetic orthographies) would help bridge gaps between Spanish-English and Chinese-English bilinguals.

The relative lack of differences observed between conditions in such overtly phonological activities as visual rhyme judgement and the phonological lexical decision task was somewhat unexpected when taken at face value. This was especially notable for the N170-OT and not for the preceding P1-OT, querying the links between them, including to what extent they overlap serially in terms of time and cognitively in terms of VWR as well as whether their underlying neural sources are as similar as their occipitotemporal scalp locations. Results support the notion that N170-OT amplitude modulation from linguistic input is dynamic and requires a distinction in legal orthography (Amora et al., 2022), such as between stimuli using real letters of the target language and other non-letter linguistic stimuli e.g., symbols. However, this was not a manipulation included in the current research, which likely contributed to the limited findings for the N170-OT component. In a similar vein, comparability across stimuli and tasks using less variation of task methodologies and more stimulus variants (e.g., Go/No-go with a variety of stimulus types), but also increased robustness through higher trial counts per condition, might have helped clarify some findings while potentially identifying others. As with many areas of psychology, future research should ultimately focus on replicating such behavioural and electrophysiological findings, confirming to what extent the reported effects were phonological (as opposed to epiphenomenal effects of, for example, some element of a methodology).

8.4 Conclusion

Processing of orthography and phonology can be considered separately but are intrinsically linked and ultimately two parts of the same prelexical whole. This connection and its serial, parallel, and cascaded nature can depend on a variety of psycholinguistic factors, including lexicality status, orthographic congruence, and the language background of the reader. Phonology can be especially dynamic and context-dependent in the extent it is involved in VWR (Fischer-Baum et al., 2014; Huettig & Ferreira, 2022), often dependent on the extent orthography is congruent in context. As such, VWR processing appears to be susceptible to interference when dealing with incongruent orthography/phonology. As expected, this was observed to be the case when different orthographic inputs lead to similar phonology (as in P+O-) and, especially, when similar orthographic inputs (e.g., same rime, as in P-O+) result in different phonological outputs (Weber-Fox et al., 2013). Essentially, some VWR circumstances require phonology more than others and phonological activation becomes prominent when necessary or gets backgrounded when another process takes over. It would, therefore, follow that phonological processing does not only occur when necessary but always occurs, just to varying extents in a dynamic, context-dependent way in order for evaluation of its relative importance to orthography. Ultimately, emphasis on phonological processing increases with the necessity for it and phonological processing does always occur when not obstructed, but other factors, such as grapheme-phoneme conversion of unfamiliar orthography or simply incongruent orthography, can suppress it. It is, therefore, not unreasonable to suggest that seemingly contrary elements of different theories are each sometimes accurate in different contexts, such as support in the literature for both strong (where phonology is deemed necessary) and weak (where phonology has only a mediating role) phonological theories (Frost, 1998, 2003; Lukatela & Turvey, 1994; Rastle & Brysbaert, 2006).

Psycholinguistic features of the linguistic stimuli (e.g., orthography/phonology incongruence), however, are not the only influence on orthographic and phonological processing during VWR and the language profile of the reader is vital. Akin to bilingualism theories/models needing to extend to languages other than English (van Heuven & Wen, 2018), bilingualism theories/models (e.g., BIA (Dijkstra & van Heuven, 2002), BIA+ (van Heuven & Dijkstra, 2010), and Multilink (Dijkstra et al., 2019) models) should also take language profiles more deeply into account, as L2 processing approaches have some dependence on differences in language profiles, such as a natural tendency to process phonologically as in Spanish-English ESL readers. To what extent and how this might vary based on both L1 and L2 properties, however, is still unclear and should be a consideration of future work (Poeppel & Idsardi, 2022). Furthermore, language background, relevant L1 VWR skills, and similarities between L1 and L2 should be integrated into theories/models of bilingualism, such as Multilink (Dijkstra et al., 2019).

This thesis has documented some of the first research into VWR-related early pre-200ms ERP-based brain activity and consequent behaviour that is explicitly from the perspective of bilingual language backgrounds using alphabetic L1 and non-alphabetic L1 bilinguals as target populations. Findings of the current research pertaining to occipitotemporal ERP activity before ~200ms (i.e., P1-OT and N170-OT) reflected aspects of orthography-phonology mapping and support the notion of cascaded/overlapping processes as per the DRC (Coltheart et al., 2001) and BIAM (Grainger & Holcomb, 2009). Meanwhile, frontal-central negativity showed the potential of early processing of phonology in parallel with occipital/occipitotemporal visual/orthographic processing at ~100ms post-stimulus, which also supports a multi-route approach to VWR (e.g., DRC (Coltheart et al., 2001) and BIAM (Grainger & Holcomb, 2009)). Through patterns of such findings, the current research points to the prospect of different cognitive strategies for VWR being employed between ESL groups dependent on language profiles. Such evidence strongly supports the current research as a viable framework and worthwhile avenue of future investigation. As important as the ~500ms post-stimulus timeframe is for reading, the current research, following other work (e.g., Cornelissen et al., 2009; Hauk et al., 2006; Fu et al., 2020), shows that the first ~200ms after seeing a word is the critical foundation for VWR in both monolingual and bilingual readers. This is true for orthographic and phonological processing but has implications of timing and nature for preceding visual/orthographic processing as well as for the semantic, syntactic, and pragmatic processing that typically follows in reading. Significantly more research should, therefore, be conducted into the range of VWR processes at such early timeframes in populations across the language spectrum.

References

Abramson, M., & Goldinger, S. D. (1997). What the reader's eye tells the mind's ear: Silent reading activates inner speech. *Perception & Psychophysics*, *59*(7), 1059–1068. https://doi.org/10.3758/BF03205520

Acha, J., & Perea, M. (2008). The effects of length and transposed-letter similarity in lexical decision: evidence with beginning, intermediate, and adult readers. *British Journal of Psychology (London, England : 1953)*, *99*(Pt 2), 245–264. https://doi.org/10.1348/000712607X224478

Acunzo, D. J., MacKenzie, G., & van Rossum, M. C. W. (2012). Systematic biases in early ERP and ERF components as a result of high-pass filtering. *Journal of Neuroscience Methods*, *209*(1), 212–218. https://doi.org/10.1016/j.jneumeth.2012.06.011

Alario, F.-X., De Cara, B., & Ziegler, J. C. (2007). Automatic activation of phonology in silent reading is parallel: evidence from beginning and skilled readers. *Journal of Experimental Child Psychology*, *97*(3), 205–219. https://doi.org/10.1016/j.jecp.2007.02.001

Almeida, D., & Poeppel, D. (2013). Word-specific repetition effects revealed by MEG and the implications for lexical access. *Brain and Language*, *127*(3), 497–509. https://doi.org/10.1016/j.bandl.2013.09.013

Alvarez, R. P., Holcomb, P. J., & Grainger, J. (2003). Accessing word meaning in two languages: An event-related brain potential study of beginning bilinguals. *Brain and Language*, *87*(2), 290–304. https://doi.org/10.1016/S0093-934X(03)00108-1

Amora, K. K., Tretow, A., Verwimp, C., Tijms, J., Leppänen, P. H. T., & Csépe, V. (2022). Typical and Atypical Development of Visual Expertise for Print as Indexed by the

Visual Word N1 (N170w): A Systematic Review. *Frontiers in Neuroscience*, *16*. https://doi.org/10.3389/fnins.2022.898800

Amsel, B. D. (2011). Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia*, *49*(5), 970–983. https://doi.org/10.1016/j.neuropsychologia.2011.01.003

ANT Neuro. (2015). *asa*.

Appelbaum, L., Liotti, M., Perez III, R., Fox, S. P., & Woldorff, M. G. (2009). The temporal dynamics of implicit processing of non-letter, letter, and word-forms in the human visual cortex. *Frontiers in Human Neuroscience*, *3*(November), 1–11. https://doi.org/10.3389/neuro.09.056.2009

Ashby, J. (2010). Phonology is fundamental in skilled reading: Evidence from ERPs. *Psychonomic Bulletin & Review*, *17*(1), 95–100.

Ashby, J., Sanders, L. D., & Kingston, J. (2009). Skilled readers begin processing sub-phonemic features by 80ms during visualword recognition: evidence from ERPs. *Biological Psychology*, *80*, 84–94.

Assadollahi, R., & Pulvermuller, F. (2003). Early influences of word length and frequency: a group study using MEG. *NeuroReport*, *14*(8), 1183–1187.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX lexical database*. Linguistic Data Consortium, University of Pennsylvania.

Baddeley, A., Chincotta, D., Turk, D., & Stafford, L. (2002). Is the word length effect in STM entirely attributable to output delay ? Evidence from serial recognition. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *55A*(2), 353–369. https://doi.org/10.1080/02724980143000523

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual Word Recognition of Single-Syllable Words. *Journal of Experimental Psychology: General*, *133*(2), 283–316.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, *29*(4), 639–647. https://doi.org/10.3758/BF03200465

Balota, D. A., & Spieler, D. (1999). Word Frequency, Repetition, and Lexicality Effects in Word Recognition Tasks: Beyond Measures of Central Tendency. *Journal of Experimental Psychology: …*, *128*(1), 32–55.

Barber, H. a, & Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. *Brain Research Reviews*, *53*(1), 98–123. https://doi.org/10.1016/j.brainresrev.2006.07.002

Baron, J. (1973). Phonemic stage not necessary for reading. *The Quarterly Journal of Experimental Psychology*, *25*(2), 241–246. https://doi.org/10.1080/14640747308400343

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *Journal of Cognitive Neuroscience*, *11*(3), 235–260. http://www.ncbi.nlm.nih.gov/pubmed/10402254

Bijeljac-Babic, R., Biardeau, A., & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory & Cognition*, *25*(4), 447–457.

Bitan, T., Burman, D. D., Chou, T.-L., Lu, D., Cone, N. E., Cao, F., Bigio, J. D., & Booth, J. R. (2007). The interaction between orthographic and phonological information in children: An fMRI study. *Human Brain Mapping*, *28*(9), 880–891.

Bitan, T., Cheon, J., Lu, D., Burman, D. D., & Booth, J. R. (2009). Developmental Increase in Top–Down and Bottom–Up Processing in a Phonological Task: An Effective Connectivity, fMRI Study. *Journal of Cognitive Neuroscience*, *21*(6), 1135–1150. https://doi.org/10.1162/jocn.2009.21065.Developmental

Botezatu, M. R., Miller, C. A., & Misra, M. (2015). An event-related potential study of visual rhyming effects in native and non-native English speakers. *NeuroReport*, *26*(3), 118–123. https://doi.org/10.1097/WNR.0000000000000311

Boukrina, O., Hanson, S. J., & Hanson, C. (2014). Modeling activation and effective connectivity of VWFA in same script bilinguals. *Human Brain Mapping*, *35*(6), 2543–2560. https://doi.org/10.1002/hbm.22348

Braun, M., Hutzler, F., Münte, T. F., Rotte, M., Dambacher, M., Richlan, F., & Jacobs, A. M. (2015). The neural bases of the pseudohomophone effect: Phonological constraints on lexico-semantic access in reading. *Neuroscience*, *295*, 151–163. https://doi.org/10.1016/j.neuroscience.2015.03.035

Braun, M., Hutzler, F., Ziegler, J. C., Dambacher, M., & Jacobs, A. M. (2009). Pseudohomophone effects provide evidence of early lexico-phonological processing in visual word recognition. *Human Brain Mapping*, *30*(7), 1977–1989. https://doi.org/10.1002/hbm.20643

Brem, S., Halder, P., Bucher, K., Summers, P., Martin, E., & Brandeis, D. (2009). Tuning of the visual word processing system: Distinct developmental ERP and fMRI effects. *Human Brain Mapping*, *30*(6), 1833–1844. https://doi.org/10.1002/hbm.20751

Briesemeister, B. B., Hofmann, M. J., Tamm, S., Kuchinke, L., Braun, M., & Jacobs, A. M. (2009). The pseudohomophone effect: evidence for an orthography-phonology-conflict. *Neuroscience Letters*, *455*(2), 124–128.

Brown, C., & Hagoort, P. (1993). The Processing Nature of the N400: Evidence from Masked Priming. *Journal of Cognitive Neuroscience*, *5*(1), 34–44.

Brunswick, N. (2010). The Functional Neuroanatomy of Reading. In P. L. Cornelissen, P. C. Hansen, M. L. Kringelbach, & K. R. Pugh (Eds.), *The Neural Basis of Reading*. Oxford University Press.

Brysbaert, M. (2001). Prelexical phonological coding of visual words in Dutch: Automatic after all. *Memory & Cognition*, *29*(5), 765–773.

Brysbaert, M., & Dijkstra, T. (2006). Changing views on word recognition in bilinguals. In J. Morais & G. D'Ydewalle (Eds.), *Bilingualism and second language acquisition* (pp. 25–37). Royal Academes for Science and the Arts of Belgium.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism*, *13*(3), 359–371. https://doi.org/10.1017/S1366728909990344

Brysbaert, M., van Dyck, G., & van de Poel, M. (1999). Visual word recognition in bilinguals: evidence from masked phonological priming. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 137–148.

Brysbaert, M., van Wijnendaele, I., & Duyck, W. (2002). On the temporal delay assumption and the impact of non-linguistic context effects. *Bilingualism: Language and Cognition*, *5*(3), 199–201. https://doi.org/10.1017/s1366728902213012

Brysbaert, M., & Wijnendaele, I. Van. (2003). The importance of phonological coding in visual word recognition: Further evidence from second-language processing. *Psychologica Belgica*, *43*(4), 1–13.

Carrasco-Ortiz, H., Midgley, K. J., & Frenck-Mestre, C. (2012a). Are phonological representations in bilinguals language specific? An ERP study on interlingual homophones. *Psychophysiology*, *49*(4), 531–543. https://doi.org/10.1111/j.1469-8986.2011.01333.x

Carrasco-Ortiz, H., Midgley, K. J., & Frenck-Mestre, C. (2012b). Are phonological representations in bilinguals language specific? An ERP study on interlingual homophones. *Psychophysiology*, *49*(4), 531–543. https://doi.org/10.1111/j.1469-8986.2011.01333.x

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*, *18*(2), 90–98. https://doi.org/10.1016/j.tics.2013.11.005

Carreiras, M., Duñabeitia, J. A., & Molinaro, N. (2009). Consonants and Vowels Contribute Differently to Visual Word Recognition: ERPs of Relative Position Priming. *Cerebral Cortex*, *19*, 2659–2670.

Chen, Y., Lee, J. R., Kuo, W.-J., Hung, D. L., & Cheng, S. (2010). An ERP study of Chinese speakers' rhyme judgments to Chinese and English words. *NeuroReport*, *21*(9), 636–640. https://doi.org/10.1097/WNR.0b013e32833a5d2c.An

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564–570. https://doi.org/10.1016/j.tics.2006.10.002

Clahsen, H., Felser, C., Neubauer, K., & Silva, R. (2010). Morphological Structure in Native and Nonnative Language Processing. *Language Learning*, *March*, 21–43. https://doi.org/10.1111/j.1467-9922.2009.00550.x

Classon, E., Rudner, M., Johansson, M., & Rönnberg, J. (2013). Early ERP Signature of Hearing Impairment in Visual Rhyme Judgment. *Frontiers in Psychology*, *4*(May), 241. https://doi.org/10.3389/fpsyg.2013.00241

Coch, D., Hart, T., & Mitra, P. (2008). Three kinds of rhymes: An ERP study. *Brain and Language*, *104*(3), 230–243. https://doi.org/10.1016/j.bandl.2007.06.003

Coch, D., & Meade, G. (2016). N1 and P2 to words and wordlike stimuli in late elementary school children and adults. *Psychophysiology*, *53*(2), 115–128. https://doi.org/10.1111/psyp.12567

Coch, D., & Mitra, P. (2010). Word and pseudoword superiority effects reflected in the ERP waveform. *Brain Research*, *1329*, 159–174. https://doi.org/10.1016/j.brainres.2010.02.084

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505.

Coltheart, M. (2005). Modelling Reading: The Dual-Route Approach. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 6–23). Blackwell Publishing.

Coltheart, M., & Curtis, B. (1993). Models of reading aloud: Dual-route and parallel distributed processing approaches. *Psychological Review*, *100*(4), 589–608.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256.

Coltheart, M., Woollams, A., Kinoshita, S., & Perry, C. (1999). A position-sensitive stroop effect: Further evidence for a left-to-right component in print-to-speech conversion. *Psychonomic Bulletin & Review*, *6*(3), 456–463. https://doi.org/10.3758/BF03210835

Compton, P. E., Grossenbacher, P., Posner, M. I., & Tucker, D. M. (1991). A cognitive-anatomical approach to attention in lexical access. *Journal of Cognitive Neuroscience*, *3*(4), 304–312. https://doi.org/10.1162/jocn.1991.3.4.304

Cornelissen, P. L., Kringelbach, M. L., & Hansen, P. C. (2010). Visual Word Recognition: The First 500 Milliseconds. In P. L. Cornelissen, P. C. Hansen, M. L. Kringelbach, & K. Pugh (Eds.), *The neural basis of reading acquisition* (pp. 192–219). Oxford University Press.

Cornelissen, P. L., Kringelbach, M. L. M. L., Ellis, A. W., Whitney, C., Holliday, I. E., & Hansen, P. C. (2009). Activation of the left inferior frontal gyrus in the first 200 ms of reading: evidence from magnetoencephalography. *PLoS One*, *4*(4), 1–13.

Coulmas, F. (2003). *Writing Systems: An Introduction to Their Linguistic Analysis*. Cambridge University Press.

Cousineau, D., & Chartier, S. (2015). Outliers detection and treatment: a review. *International Journal of Psychological Research*, *3*(1), 58–67. https://doi.org/10.21500/20112084.844

Crystal, D. (2010). *The Cambridge Encyclopedia of Language* (Third). Cambridge University Press.

Davis, C. J. (2005). N-watch: a program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, *37*(1), 65–70.

Dehaene, S. (2014). Reading in the Brain Revised and Extended: Response to Comments. *Mind and Language*, *29*(3), 320–335. https://doi.org/10.1111/mila.12053

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, *15*(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003

Dennis, L., & Key, A. F. (2006). Below-Average, Average, and Above-Average Readers Engage Different and Similar Brain Regions While Reading. *Journal of Learning Disabilities*, *39*(4), 352–363.

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, & Computers*, *30*(1), 34–43. https://doi.org/10.3758/BF03209414

Dien, J. (2009). The neurocognitive basis of reading single words as seen through early latency ERPs: a model of converging pathways. *Biological Psychology*, *80*(1), 10–22. https://doi.org/10.1016/j.biopsycho.2008.04.013

Dien, J. (2010). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods*, *187*(1), 138–145. https://doi.org/10.1016/j.jneumeth.2009.12.009

Dien, J. (2017). Best practices for repeated measures ANOVAs of ERP data: Reference, regional channels, and robust ANOVAs. *International Journal of Psychophysiology*, *111*, 42–56. https://doi.org/10.1016/j.ijpsycho.2016.09.006

Diependaele, K., Duñabeitia, J. A., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, *64*(4), 344–358. https://doi.org/10.1016/j.jml.2011.01.003

Diependaele, K., Ziegler, J., & Grainger, J. (2010). Fast phonology and the Bimodal Interactive Activation Model. *European Journal of Cognitive Psychology*, *22*(5), 764–778. https://doi.org/10.1080/09541440902834782

Dijkstra, T., & van Heuven, W. J. B. (2002a). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/s1366728902003012

Dijkstra, T., & van Heuven, W. J. B. (2002b). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–224. https://doi.org/10.1017/S1366728902003012

Dijkstra, T., Wahl, A., Buytenhuijs, F., van Halem, N., Al-Jibouri, Z., de Korte, M., & Rekké, S. (2019). Multilink: a computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, 1–23. https://doi.org/10.1017/S1366728918000287

Drieghe, D., & Brysbaert, M. (2002). Strategic effects in associative priming with words, homophones, and pseudohomophones. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 951–961. https://doi.org/10.1037//0278-7393.28.5.951

Duyck, W. (2005). Translation and associative priming with cross-lingual pseudohomophones: Evidence for nonselective phonological activation in bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(6), 1340–1359. https://doi.org/10.1037/0278-7393.31.6.1340

Eberhard, Simons, & Fennig (Eds.). (2021). *Ethnologue: Languages of the World* (24th ed.). SIL International. https://www.ethnologue.com/

Ehri, L. C., & Wilce, L. S. (1980). The influence of orthography on readers' conceptualization of the phonemic structure of words. *Applied Psycholinguistics*, *1*(04), 371. https://doi.org/10.1017/S0142716400009802

Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Daal, V. H. P., Polyzoe, N., Tsipa, M.-L., & Petalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly*, *39*(4), 438–468. https://doi.org/10.1598/RRQ.39.4.5

Emmorey, K., Midgley, K. J., Kohen, C. B., Sehyr, Z. S., & Holcomb, P. J. (2017). The N170 ERP component differs in laterality, distribution, and association with continuous reading measures for deaf and hearing readers. *Neuropsychologia*, *106*, 298–309. https://doi.org/10.1016/j.neuropsychologia.2017.10.001

Eulitz, C., Eulitz, H., Maess, B., Cohen, R., Pantev, C., & Elbert, T. (2000). Magnetic brain activity evoked and induced by visually presented words and nonverbal stimuli. *Psychophysiology*, *37*(4), 447–455. https://doi.org/10.1017/S0048577200981204

Ferrand, L., & Grainger, J. (1993). The time-course of phonological and orthographic code activation in the early phases of visual word recognition. In *Bulletin of the Psychonomic Society* (Vol. 31, pp. 119–122).

Ferrand, L., & Grainger, J. (1994). Effects of orthography are independent of phonology in masked form priming. *The Quarterly Journal of Experimental Psychology Section A*, *47*(2), 365–382. https://doi.org/10.1080/14640749408401116

Fiebach, C. J., Friederici, A. D., Muller, K., & von Cramon, D. Y. (2002). fMRI evidence for dual routes to the mental lexicon in visual word recognition. *J Cogn Neurosci*, *14*(1), 11–23. https://doi.org/10.1162/089892902317205285

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (Fourth). Sage.

Fiez, J. A., Balota, D. A., Raichle, M. E., Petersen, S. E., & Louis, S. (1999). Effects of Lexicality, Frequency, and Spelling-to-Sound Consistency on the Functional Anatomy of Reading. *Neuron*, *24*, 205–218.

Filik, R., & Barber, E. (2011). Inner speech during silent reading reflects the reader's regional accent. *PloS One*, *6*(10), e25782. https://doi.org/10.1371/journal.pone.0025782

Fischer-Baum, S., Dickson, D. S., & Federmeier, K. D. (2014). Frequency and regularity effects in reading are task dependent: evidence from ERPs. *Language, Cognition and Neuroscience*, *29*(July 2014), 1–14. https://doi.org/10.1080/23273798.2014.927067

Fitzpatrick, T., & Izura, C. (2011). Word Association in L1 and L2. *Studies in Second Language Acquisition*, *33*(03), 373–398. https://doi.org/10.1017/S0272263111000027

Frankish, C., & Turner, E. (2007). SIHGT and SUNOD: The role of orthography and phonology in the perception of transposed letter anagrams. *Journal of Memory and Language*, *56*(2), 189–211. https://doi.org/10.1016/j.jml.2006.11.002

Franklin, M. S., Dien, J., Neely, J. H., Huber, E., & Waterson, L. D. (2007). Semantic priming modulates the N400, N300, and N400RP. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *118*(5), 1053–1068. https://doi.org/10.1016/j.clinph.2007.01.012

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: true issues and false trails. *Psychological Bulletin*, *123*(1), 71–99.

Frost, R. (2003). The robustness of phonological effects in fast priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked Priming: The State of the Art* (1st ed., pp. 96–106). Psychology Press. https://doi.org/10.4324/9780203502846

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: a multilingual comparison. *Journal of Experimental*

Psychology. Human Perception and Performance, *13*(1), 104–115. https://doi.org/10.1037/0096-1523.13.1.104

Fu, S., Feng, C., Guo, S., Luo, Y., & Parasuraman, R. (2012). Neural adaptation provides evidence for categorical differences in processing of faces and Chinese characters: An ERP study of the N170. *PLoS ONE*, *7*(7), e41103. https://doi.org/10.1371/journal.pone.0041103

Galin, D., Ornstein, R., Herron, J., & Johnstone, J. (1982). Sex and handedness differences in EEG measures of hemispheric specialization. *Brain and Language*, *16*(1), 19–55. https://doi.org/10.1016/0093-934X(82)90070-0

Gibbons, H., Kirsten, H., & Seib-Pfeifer, L. E. (2022). Attention to affect 2.0: Multiple effects of emotion and attention on event-related potentials of visual word processing in a valence-detection task. *Psychophysiology*. https://doi.org/10.1111/psyp.14059

Glezer, L. S., Eden, G., Jiang, X., Luetje, M., Napoliello, E., Kim, J., & Riesenhuber, M. (2016). Uncovering phonological and orthographic selectivity across the reading network using fMRI-RA. *NeuroImage*, *138*, 248–256. https://doi.org/10.1016/j.neuroimage.2016.05.072

Glezer, L. S., & Riesenhuber, M. (2013). Individual Variability in Location Impacts Orthographic Selectivity in the “Visual Word Form Area.” *Journal of Neuroscience*, *33*(27), 11221–11226. https://doi.org/10.1523/JNEUROSCI.5002-12.2013

Grainger, J., & Ferrand, L. (1994). Phonology and Orthography in Visual Word Recognition: Effects of Masked Homophone Primes. *Journal of Memory and Language*, *33*, 218–233.

Grainger, J., & Holcomb, P. J. (2009). Watching the Word Go by: On the Time-course of Component Processes in Visual Word Recognition. *Language and Linguistics Compass*, *3*(1), 128–156. https://doi.org/10.1111/j.1749-818X.2008.00121.x.Watching

Grainger, J., & Jacobs, A. M. (1996). Orthographic Processing in Visual Word Recognition: A Multiple Read-Out Model. *Psychological Review*, *103*(3), 518–565. https://doi.org/10.1037/0033-295X.103.3.518

Grainger, J., Kiyonaga, K., & Holcomb, P. J. (2006). The Time Course of Orthographic and Phonological Code Activation. *Psychological Science*, *17*(12), 1021–1026. https://doi.org/10.1111/j.1467-9280.2006.01821.x.The

Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In M. Kail & M. Hickmann (Eds.), *Language Acquisition across Linguistic and Cognitive Systems* (Vol. 52, pp. 267–283). John Benjamins Publishing Company.

Grainger, J., & van Heuven, W. J. B. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.), *The Mental Lexicon* (pp. 1–24). Nova Science Publishers.

Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2*(April), 54. https://doi.org/10.3389/fpsyg.2011.00054

Green, D. W. (1986). Control, Activation, and Resource: A Framework and a Model for the Control of Speech in Bilinguals. *Brain and Language*, *27*, 210–223.

Gronau, N., & Frost, R. (1997). Prelexical phonologic computation in a deep orthography: Evidence from backward masking in Hebrew. *Psychonomic Bulletin & Review*, *4*(1), 107–112. https://doi.org/10.3758/BF03210781

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*, 1711–1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Grossi, G., Coch, D., Coffey-Corina, S., Holcomb, P. J., & Neville, H. J. (2001). Phonological processing in visual rhyming: a developmental ERP study. *Journal of Cognitive Neuroscience*, *13*(5), 610–625.

Guo, J., Guo, T., Yan, Y., Jiang, N., & Peng, D. (2009). ERP evidence for different strategies employed by native speakers and L2 learners in sentence processing. *Journal of Neurolinguistics*, *22*(2), 123–134. https://doi.org/10.1016/j.jneuroling.2008.09.001

Haigh, C. A., & Jared, D. (2007). The Activation of Phonological Representations by Bilinguals While Reading Silently: Evidence From Interlingual Homophones. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(4), 623–644. https://doi.org/10.1037/0278-7393.33.4.623

Harley, T. A. (2010). Talking the talk. In *Talking the Talk*.

Hasko, S., Groth, K., Bruder, J., Bartling, J., & Schulte-Körne, G. (2013). The time course of reading processes in children with and without dyslexia: an ERP study. *Frontiers in Human Neuroscience*, *7*(October), 570. https://doi.org/10.3389/fnhum.2013.00570

Hauk, O., Coutout, C., Holden, A., & Chen, Y. (2012). The time-course of single-word reading: Evidence from fast behavioral and brain responses. *NeuroImage*, *60*, 1462–1477.

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

ERP data. *NeuroImage*, *30*(4), 1383–1400. https://doi.org/10.1016/j.neuroimage.2005.11.048

Hauk, O., Davis, M. H., & Pulvermüller, F. (2008). Modulation of brain activity by multiple lexical and word form variables in visual word recognition: A parametric fMRI study. *NeuroImage*, *42*(3), 1185–1195. https://doi.org/10.1016/j.neuroimage.2008.05.054

Hauk, O., Patterson, K., Woollams, A., Watling, L., Pulvermüller, F., & Rogers, T. T. (2006). [Q:] When would you prefer a SOSSAGE to a SAUSAGE? [A:] At about 100 msec. ERP correlates of orthographic typicality and lexicality in written word recognition. *Journal of Cognitive Neuroscience*, *18*(5), 818–832. https://doi.org/10.1162/jocn.2006.18.5.818

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, *115*, 1090–1103. https://doi.org/10.1016/j.clinph.2003.12.020

Hauk, O., Pulvermüller, F., Ford, M., Marslen-Wilson, W. D., & Davis, M. H. (2009). Can I have a quick word? Early electrophysiological manifestations of psycholinguistic processes revealed by event-related regression analysis of the EEG. *Biological Psychology*, *80*(1), 64–74. https://doi.org/10.1016/j.biopsycho.2008.04.015

Heim, S., Alter, K., Ischebeck, A. K., Amunts, K., Eickhoff, S. B., Mohlberg, H., Zilles, K., Von Cramon, D. Y., & Friederici, A. D. (2005). The role of the left Brodmann's areas 44 and 45 in reading words and pseudowords. *Cognitive Brain Research*, *25*(3), 982–993. https://doi.org/10.1016/j.cogbrainres.2005.09.022

Helenius, P., Tarkiainen, A., Cornelissen, P., Hansen, P. C., & Salmelin, R. (1999). Dissociation of normal feature analysis and deficient processing of letter-strings in dyslexic adults. *Cerebral Cortex (New York, N.Y. : 1991)*, *9*(5), 476–483. http://www.ncbi.nlm.nih.gov/pubmed/10450892

Hirshorn, E. A., Li, Y., Ward, M. J., Richardson, R. M., Fiez, J. A., & Ghuman, A. S. (2016). Decoding and disrupting left midfusiform gyrus activity during word reading. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 201604126. https://doi.org/10.1073/pnas.1604126113

Holcomb, P. J., & Grainger, J. (2006). On the time course of visual word recognition: an event-related potential investigation using masked repetition priming. *Journal of Cognitive Neuroscience*, *18*(10), 1631–1643. https://doi.org/10.1162/jocn.2006.18.10.1631

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*(6), 938–950. https://doi.org/10.1162/089892902760191153

Hoosain, R. (1991). *Psycholinguistic Implications for Linguistic Relativity: A Case Study of Chinese*. Taylor & Francis Group.

Huettig, F., & Ferreira, F. (2022). The Myth of Normal Reading. *Perspectives on Psychological Science*, 174569162211272. https://doi.org/10.1177/17456916221127226

Hulstijn, J. H. (2011). Language Proficiency in Native and Nonnative Speakers: An Agenda for Research and Suggestions for Second-Language Assessment. *Language Assessment Quarterly*, *8*(3), 229–249. https://doi.org/10.1080/15434303.2011.565844

IBM. (2020). *SPSS* (26).

Ille, N., Berg, P., & Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology*, *19*, 113–124.

Indefrey, P. (2006). It is time to work toward explicit processing models for native and second language speakers. *Applied Psycholinguistics*, *27*(1), 66–69. https://doi.org/10.1017.S0142716406060103

Itier, R. J., & Taylor, M. J. (2004). Source analysis of the N170 to faces and objects. *Neuroreport*, *15*(8), 1261–1265. https://doi.org/10.1097/01.wnr.0000127827.73576.d8

Izura, C., Pérez, M. A., Agallou, E., Wright, V. C., Marín, J., Stadthagen-González, H., & Ellis, A. W. (2011). Age/order of acquisition effects and the cumulative learning of foreign words: A word training study. *Journal of Memory and Language*, *64*(1), 32–58. https://doi.org/10.1016/j.jml.2010.09.002

Jalbert, A., Neath, I., Bireta, T. J., & Surprenant, A. M. (2011). When does length cause the word length effect? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(2), 338–353. https://doi.org/10.1037/a0021804

Jamal, N. I., Piche, A. W., Napoliello, E. M., Perfetti, C. a, & Eden, G. F. (2012). Neural basis of single-word reading in Spanish-English bilinguals. *Human Brain Mapping*, *33*(1), 235–245. https://doi.org/10.1002/hbm.21208

JASP Team. (2020). *JASP* (0.14.1).

Joshi, R. M., & Carreker, S. (2009). Spelling: Development, assessment, and instruction. In G. Reid, G. Elbeheri, J. Everett, D. Knight, & J. Wearmouth (Eds.), *The Routledge companion to dyslexia* (pp. 113–125). Routledge.

Katz, L., & Frost, R. (1992). The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis. *Advances in Psychology*, *94*, 67–84. https://doi.org/10.1016/S0166-4115(08)62789-2

Khateb, A., Pegna, A. J., Landis, T., Michel, C. M., Brunet, D., Seghier, M. L., & Annoni, J.-M. (2007). Rhyme processing in the brain - An ERP mapping study. *International Journal of Psychophysiology*, *63*, 240–250.

Kim, S. Y., Liu, L., & Cao, F. (2017). How does first language (L1) influence second language (L2) reading in the brain? Evidence from Korean-English and Chinese-English bilinguals. *Brain and Language*, *171*, 1–13. https://doi.org/10.1016/j.bandl.2017.04.003

Klein, M., Grainger, J., Wheat, K. L., Millman, R. E., Simpson, M. I. G., Hansen, P. C., & Cornelissen, P. L. (2015). Early Activity in Broca's Area during Reading Reflects Fast Access to Articulatory Codes from Print. *Cerebral Cortex*, *25*(7), 1715–1723. https://doi.org/10.1093/cercor/bht350

Knight, D., & Muncer, S. J. (2011). Type and token bigram frequencies for two-through nine-letter words and the prediction of anagram difficulty. *Behavior Research Methods*, *43*(2), 491–498. https://doi.org/10.3758/s13428-011-0068-x

Kotz, S. A., Holcomb, P. J., & Osterhout, L. (2008). ERPs reveal comparable syntactic sentence processing in native and non-native readers of English. *Acta Psychologica*, *128*(3), 514–527. https://doi.org/10.1016/j.actpsy.2007.10.003

Kroll, J. F., & Stewart, E. (1994). Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections between Bilingual Memory Representations. *Journal of Memory and Language*, *33*, 149–174.

Kujala, J., Pammer, K., Cornelissen, P., Roebroeck, A., Formisano, E., & Salmelin, R. (2007). Phase coupling in a cerebro-cerebellar network at 8-13 Hz during reading. *Cerebral Cortex*, *17*(6), 1476–1485. https://doi.org/10.1093/cercor/bhl059

Kuo, L.-J., Kim, T.-J., Yang, X., Li, H., Liu, Y., Wang, H., Hyun Park, J., & Li, Y. (2015). Acquisition of Chinese characters: the effects of character properties and individual differences among second language learners. *Frontiers in Psychology*, *6*(August), 1–10. https://doi.org/10.3389/fpsyg.2015.00986

Kuo, W.-J., Yeh, T.-C., Lee, J.-R., Chen, L.-F., Lee, P.-L., Chen, S.-S., Ho, L.-T., Hung, D., Tzeng, O., & Hsieh, J.-C. (2004). Orthographic and phonological processing of Chinese characters: An fMRI study. *NeuroImage*, *21*(4), 1721–1731. https://doi.org/10.1016/j.neuroimage.2003.12.007

Kutas, M., & Federmeier, K. D. (2009). *N400*.

Landauer, T., Foltz, P., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028

Landi, N., & Perfetti, C. A. (2007). An electrophysiological investigation of semantic and phonological processing in skilled and less-skilled comprehenders. *Brain and Language*, *102*(1), 30–45. https://doi.org/10.1016/j.bandl.2006.11.001

Laszlo, S., & Federmeier, K. D. (2014). Never seem to find the time: Evaluating the physiological time course of visual word recognition with regression analysis of single-item event-related potentials. *Language, Cognition and Neuroscience*, *29*(5), 642–661. https://doi.org/10.1080/01690965.2013.866259

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews. Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Lee, C. C., Imaizumi, K., Schreiner, C. E., & Winer, J. a. (2004). Concurrent tonotopic processing streams in auditory cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *14*(4), 441–451. https://doi.org/10.1093/cercor/bhh006

Lehtonen, M., Hultén, A., Rodríguez-Fornells, A., Cunillera, T., Tuomainen, J., & Laine, M. (2012). Differences in word recognition between early bilinguals and monolinguals: behavioral and ERP evidence. *Neuropsychologia*, *50*(7), 1362–1371. https://doi.org/10.1016/j.neuropsychologia.2012.02.021

Leinenger, M. (2014). Phonological coding during reading. *Psychological Bulletin*, *140*(6), 1534–1555. https://doi.org/10.1037/a0037830

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 12–31. https://doi.org/10.1037/0278-7393.34.1.12

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, *17*, 673–680. https://doi.org/10.1017/S1366728913000606

Lin, S. E., Chen, H. C., Zhao, J., Li, S., He, S., & Weng, X. C. (2011). Left-lateralized N170 response to unpronounceable pseudo but not false Chinese characters-the key role of orthography. *Neuroscience*, *190*, 200–206. https://doi.org/10.1016/j.neuroscience.2011.05.071

Liu, Y., & Perfetti, C. a. (2003). The time course of brain activity in reading English and Chinese: an ERP study of Chinese bilinguals. *Human Brain Mapping*, *18*(3), 167–175. https://doi.org/10.1002/hbm.10090

Lu, Q., Tang, Y.-Y., Zhou, L., & Yu, Q. (2011). The different time courses of reading Chinese characters: An ERP study. *Neuroscience Letters*, *498*, 194–198.

Luck, S. J. (2004). Ten Simple Rules for Designing and Interpreting ERP Experiments. In *Event-related Potentials: A Methods Handbook*.

Luck, S. J. (2005). The Design and Interpretation of ERP Experiments. In *An introduction to the event-related potential technique* (pp. 51–98).

Luck, S. J. (2014). Quantifying ERP Amplitudes and Latencies. In *An Introduction to the Event-Related Potential Technique* (2nd ed., pp. 283–308). MIT Press.

Luck, S. J., & Gaspelin, N. (2016). How to Get Statistically Significant Effects in Any ERP Experiment (and Why You Shouldn't). *Psychophysiology*, *4424*, 1–44.

Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., & Hawkins, H. L. (1994). Effects of spatial cuing on luminance detectability: Psychophysical and electrophysiological evidence for early selection. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 887–904. https://doi.org/10.1037/0096-1523.20.4.887

Lukatela, G., & Turvey, M. T. (1990). Automatic and pre-lexical computation of phonology in visual word identification. *European Journal of Cognitive Psychology*, *2*(4), 325–343.

Lukatela, G., & Turvey, M. T. (1994). Visual lexical access is intially phonological: I. Evidence from associativepriming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General*, *123*(2), 107–128. https://doi.org/10.1037/0096-3445.123.2.107

Luke, K. K., Liu, H. L., Wai, Y. Y., Wan, Y. L., & Tan, L. H. (2002). Functional anatomy of syntactic and semantic processing in language comprehension. *Human Brain Mapping*, *16*(3), 133–145. https://doi.org/10.1002/hbm.10029

Luo, C. R., Johnson, R. a, & Gallo, D. a. (1998). Automatic activation of phonological information in reading: evidence from the semantic relatedness decision task. *Memory & Cognition*, *26*(4), 833–843. https://doi.org/10.3758/BF03211402

Ma, X., Kang, J., Li, X., Maurer, U., Cao, X., & Sommer, W. (2022). Does learning different script systems affect configural visual processing? ERP evidence from early readers of Chinese and German. *Psychophysiology*, *59*(6). https://doi.org/10.1111/psyp.14006

MacSweeney, M., Goswami, U., & Neville, H. (2013). The neurobiology of rhyme judgment by deaf and hearing adults: An ERP study. *Journal of Cognitive Neuroscience*, *25*(7), 1037–1048. https://doi.org/10.1162/jocn

Mahé, G., Bonnefond, A., Gavens, N., Dufour, A., & Doignon-Camus, N. (2012). Impaired visual expertise for print in French adults with dyslexia as shown by N170 tuning. *Neuropsychologia*, *50*(14), 3200–3206. https://doi.org/10.1016/j.neuropsychologia.2012.10.013

Martin, C. D., Nazir, T., Thierry, G., Paulignan, Y., & Démonet, J.-F. (2006). Perceptual and lexical effects in letter identification: An event-related potential study of the word superiority effect. *Brain Research*, *1098*(1), 153–160. https://doi.org/10.1016/j.brainres.2006.04.097

Martin, K. I. (2017). The impact of L1 writing system on ESL knowledge of vowel and consonant spellings. *Reading and Writing*, *30*(2), 279–298. https://doi.org/10.1007/s11145-016-9673-5

Martín-Loeches, M., Hinojosa, J. a., Fernández-Frías, C., & Rubia, F. J. (2001). Functional differences in the semantic processing of concrete and abstract words. *Neuropsychologia*, *39*, 1086–1096. https://doi.org/10.1016/S0028-3932(01)00033-1

Martín-Loeches, M., Hinojosa, J. a, Gómez-Jarabo, G., & Rubia, F. J. (1999). The recognition potential: An ERP index of lexical access. *Brain and Language*, *70*, 364–384. https://doi.org/10.1006/brln.1999.2178

Maurer, U., Brandeis, D., & McCandliss, B. D. (2005). Fast, visual specialization for reading in English revealed by the topography of the N170 ERP response. *Behavioral and Brain Functions*, *1*, 13. https://doi.org/10.1186/1744-9081-1-13

Maurer, U., Brem, S., Bucher, K., & Brandeis, D. (2005). Emerging neurophysiological specialization for letter strings. *Journal of Cognitive Neuroscience*, *17*(10), 1532–1552. https://doi.org/10.1162/089892905774597218

Maurer, U., & McCandliss, B. D. (2007). The development of visual expertise for words: the contribution of electrophysiology. In *Single-word reading: Biological and behavioral perspectives* (pp. 43–63). https://doi.org/10.4324/9780203810064

Maurer, U., Zevin, J. D., & McCandliss, B. D. (2008). Left-lateralized N170 Effects of Visual Expertise in Reading: Evidence from Japanese Syllabic and Logographic Scripts. *Journal of Cognitive Neuroscience*, *20*(10), 1878–1891.

McCandliss, B. D., Posner, M. I., & Givón, T. (1997). Brain Plasticity in Learning Visual Words. *Cognitive Psychology*, *33*, 88–110. https://doi.org/10.1006/cogp.1997.0661

McClelland, J. L., & Rogers, T. T. (2003). The Parallel Distributed Processing Approach to Semantic Cognition. *Nature Reviews. Neuroscience*, *4*, 310–322.

McClelland, J. L., & Rumelhart, D. (1981). An Interactive Activation Model of Context Effects in Letter Perception: Part I. An Account of Basic Findings. *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*, *September*, 580–596. https://doi.org/10.1016/B978-1-4832-1446-7.50048-0

McCutchen, D., & Perfetti, C. A. (1982). The visual tongue-twister effect: Phonological activation in silent reading. *Journal of Verbal Learning and Verbal Behavior*, *21*(6), 672–687. https://doi.org/10.1016/S0022-5371(82)90870-2

McNorgan, C., Chabal, S., Young, D. O., Lukic, S., Booth, J. R., O'Young, D., Lukic, S., & Booth, J. R. (2015). Task dependent lexicality effects support interactive models of reading: A meta-analytic neuroimaging review. *Neuropsychologia*, *67*, 148–158. https://doi.org/10.1016/j.neuropsychologia.2014.12.014

Mechelli, A., Crinion, J. T., Long, S., Friston, K. J., Lambon Ralph, M. a, Patterson, K., McClelland, J. L., & Price, C. J. (2005). Dissociating reading processes on the basis of neuronal interactions. *Journal of Cognitive Neuroscience*, *17*(11), 1753–1765. https://doi.org/10.1162/089892905774589190

Menary, R. (2014). Neural Plasticity, Neuronal Recycling and Niche Construction. *Mind and Language*, *29*(3), 286–303. https://doi.org/10.1111/mila.12051

Meschyan, G., & Hernandez, A. E. (2006). Impact of language proficiency and orthographic transparency on bilingual word reading: An fMRI investigation.

NeuroImage, *29*(4), 1135–1140. https://doi.org/10.1016/j.neuroimage.2005.08.055

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1974). Functions of graphemic and phonemic codes in visual word-recognition. *Memory & Cognition*, *2*(2), 309–321.

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: an electrophysiological exploration in both languages of Spanish-English bilinguals. *Brain Research. Cognitive Brain Research*, *22*(2), 205–220. https://doi.org/10.1016/j.cogbrainres.2004.08.010

Moscoso del Prado Martín, F., Hauk, O., & Pulvermüller, F. (2006). Category specificity in the processing of color-related and form-related words: an ERP study. *NeuroImage*, *29*(1), 29–37. https://doi.org/10.1016/j.neuroimage.2005.07.055

Nation, K., & Cocksey, J. (2009). Beginning readers activate semantics from sub-word orthography. *Cognition*, *110*, 273–278.

National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. *NIH Publication No. 00-4769*, *7*, 35. https://doi.org/10.1002/ppul.1950070418

Nelson, J. R., Liu, Y., Fiez, J., & Perfetti, C. A. (2009). Assimilation and accommodation patterns in ventral occipitotemporal cortex in learning a second writing system. *Human Brain Mapping*, *30*(3), 810–820. https://doi.org/10.1002/hbm.20551

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: new evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, *13*(1), 45–52. https://doi.org/10.3758/BF03193811

Norman, J. (1988). *Chinese* (First). Cambridge University Press.

Office of National Statistics. (2012). *Migration Statistics Quarterly Report, November 2012*.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Oppenheim, G., Wu, Y. J., & Thierry, G. (2018). Found in Translation: Late Bilinguals Do Automatically Activate Their Native Language When They Are Not Using It. *Cognitive Science*, *42*(5), 1700–1713. https://doi.org/10.1111/cogs.12618

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: near-homophony in nonnative visual word recognition. *Cognition*, *111*(2), 263–269. https://doi.org/10.1016/j.cognition.2008.12.007

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2010a). Is a FAN Always FUN? Phonological and Orthographic Effects in Bilingual Visual Word Recognition. *Language and Speech*, *53*(3), 383–403. https://doi.org/10.1177/0023830910371462

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2010b). Is a FAN Always FUN? Phonological and Orthographic Effects in Bilingual Visual Word Recognition. *Language and Speech*, *53*(3), 383–403. https://doi.org/10.1177/0023830910371462

Pammer, K., Hansen, P. C., Kringelbach, M. L., Holliday, I., Barnes, G., Hillebrand, A., Singh, K. D., & Cornelissen, P. L. (2004). Visual word recognition: the first half second. *NeuroImage*, *22*, 1819–1825.

Pas, M., Nakamura, K., Sawamoto, N., Aso, T., & Fukuyama, H. (2016). Stimulus-driven changes in the direction of neural priming during visual word recognition. *NeuroImage*, *125*, 428–436. https://doi.org/10.1016/j.neuroimage.2015.10.063

Pattamadilok, C., Bulnes, L. C., Devlin, J. T., Bourguignon, M., Morais, J., Goldman, S., & Kolinsky, R. (2015). How Early Does the Brain Distinguish between Regular Words, Irregular Words, and Pseudowords during the Reading Process? Evidence from Neurochronometric TMS. *Journal of Cognitive Neuroscience*, *27*(6), 1259–1274. https://doi.org/10.1162/jocn_a_00779

Pattamadilok, C., Chanoine, V., Pallier, C., Anton, J.-L., Nazarian, B., Belin, P., & Ziegler, J. C. (2017). Automaticity of phonological and semantic processing during visual word recognition. *NeuroImage*, *149*(September 2016), 244–255. https://doi.org/10.1016/j.neuroimage.2017.02.003

Paulesu, E., Bonandrini, R., Zapparoli, L., Rupani, C., Mapelli, C., Tassini, F., Schenone, P., Bottini, G., Perry, C., & Zorzi, M. (2021). Effects of orthographic consistency on bilingual reading: Human and computer simulation data. *Brain Sciences*, *11*(7). https://doi.org/10.3390/BRAINSCI11070878

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M., Pesenti, S., Gallagher, A., Perani, D., Price, C., Frith, C. D., & Frith, U. (2000). A cultural effect on brain function. *Nature Neuroscience*, *3*(1), 91–96. https://doi.org/10.1038/71163

Peeters, D., Dijkstra, T., & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language*, *68*(4), 315–332. https://doi.org/10.1016/j.jml.2012.12.003

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, *15*(2), 202–206. https://doi.org/10.1016/j.conb.2005.03.007

Perea, M., & Carreiras, M. (2008). Do orthotactics and phonology constrain the transposed-letter effect? *Language and Cognitive Processes*, *23*(1), 69–92. https://doi.org/10.1080/01690960701578146

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language*, *27*(1), 59–70. https://doi.org/10.1016/0749-596X(88)90048-4

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., Nelson, J., Liu, Y., Fiez, J., & Tan, L. H. (2012). *List of Graphemes of Commonly-used Chinese Characters*. Hong Kong Education Bureau.

Perfetti, C. A., Nelson, J., Liu, Y., Fiez, J., & Tan, L.-H. (2010). The Neural Bases of Reading: Universals and Writing System Variations. In P. L. Cornelissen (Ed.), *The Neural Basis of Reading* (pp. 147–172).

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson Jr., R., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*, 127–152.

Plaut, D. C., Mcclelland, J. L., Seidenberg, M. S., Patterson, K., & T--, G. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115. https://doi.org/10.1037/0033-295X.103.1.56

Poeppel, D., & Idsardi, W. (2022). We don't know how the brain stores anything, let alone words. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2022.08.010

Pogorzelski, S., & Wheldall, K. (2005). The importance of phonological processing skills for older low-progress readers. *Educational Psychology in Practice*, *21*(1), 1–22. https://doi.org/10.1080/02667360500035074

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception & Psychophysics*, *76*(2), 270–279. https://doi.org/10.3758/s13414-013-0605-z

Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage*, *19*(3), 473–481. https://doi.org/10.1016/S1053-8119(03)00084-3

Pritchard, S. C., Coltheart, M., Palethorpe, S., & Castles, A. (2012). Nonword reading: Comparing dual-route cascaded and connectionist dual-process models with human data. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1268–1288. https://doi.org/10.1037/a0026703

Proverbio, A. M., & Adorni, R. (2008). Orthographic familiarity, phonological legality and number of orthographic neighbours affect the onset of ERP lexical effects. *Behavioral and Brain Functions*, *4*(27). https://doi.org/10.1186/1744-9081-4-27

Proverbio, A. M., Vecchi, L., & Zani, A. (2004). From orthography to phonetics: ERP measures of grapheme-to-phoneme conversion mechanisms in reading. *Journal of Cognitive Neuroscience*, *16*(2), 301–317. https://doi.org/10.1162/089892904322984580

PST. (2017). *E-Prime* (2.0.10.353).

Pulvermüller, F., Shtyrov, Y., & Hauk, O. (2009). Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language*, *110*(2), 81–94. https://doi.org/10.1016/j.bandl.2008.12.001

Qualtrics. (2005). *Qualtrics* (2022). Qualtrics.

Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: Are they real? Do they matter? *Cognitive Psychology*, *53*, 97–145.

Rastle, K., & Coltheart, M. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin & Review*, *5*(2), 277–282. https://doi.org/10.3758/BF03212951

Riele, S. M. M. Te, Nooteboom, S. G., & Quene, H. (1996). Strategies used in rhyme-monitoring. *Proceeding of Fourth International Conference on Spoken Language Processing ICSLP 96*, *1*, 90–93.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*(3), 249–284. https://doi.org/10.1016/S0010-0277(97)00027-9

Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. *Cortex*, *59*(1874), 33–48. https://doi.org/10.1016/j.cortex.2014.07.001

Rueckl, J. G., Paz-Alonso, P. M., Molfese, P. J., Kuo, W.-J., Bick, A., Frost, S. J., Hancock, R., Wu, D. H., Mencl, W. E., Duñabeitia, J. A., Lee, J.-R., Oliver, M., Zevin, J. D., Hoeft, F., Carreiras, M., Tzeng, O. J. L., Pugh, K. R., & Frost, R. (2015). Universal brain signature of proficient reading: Evidence from four contrasting languages. *Proceedings of the National Academy of Sciences*, *112*(50), 15510–15515. https://doi.org/10.1073/pnas.1509321112

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory & Cognition*, *18*(4), 367–379.

Rugg, M. D., & Barrett, S. E. (1987). Event-related potentials and the interaction between orthographic and phonological information in a rhyme-judgment task. *Brain and Language*, *32*(2), 336–361.

Ryynanen, O., Hyttinen, J., & Malmivuo, J. (2004). Study on the spatial resolution of EEG - effect of electrode density and measurement noise. *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, *2*(9), 1547–1554. https://doi.org/10.1109/IEMBS.2004.1404226

Segalowitz, S. J., & Zheng, X. (2009). An ERP study of category priming: evidence of early lexical semantic access. *Biological Psychology*, *80*(1), 122–129. https://doi.org/10.1016/j.biopsycho.2008.04.009

Seghier, M. L., Maurer, U., & Xue, G. (2014). What makes written words so special to the brain? *Frontiers in Human Neuroscience*, *8*, 1–3. https://doi.org/10.3389/fnhum.2014.00634

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. In *Journal of Experimental Psychology Human Learning and Memory* (Vol. 5, Issue 6, pp. 546–554). American Psychological Association.

Sereno, S. C., Brewer, C. C., & O'Donnell, P. J. (2003). Context effects in word recognition: evidence for early interactive processing. *Psychol Sci*, *14*(4), 328–333. https://doi.org/psci_14471 [pii] ET - 2003/06/17

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*(11), 489–493. https://doi.org/10.1016/j.tics.2003.09.010

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*(10), 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

Simon, G., Bernard, C., Largy, P., Lalonde, R., & Rebai, M. (2004). Chronometry of visual word recognition during passive and lexical decision tasks: an ERP investigation. *International Journal of Neuroscience*, *114*(11), 1293–1324. https://doi.org/10.1080/00207450490

Simon, G., & Lalonde, R. (2004). Chronometry of visual word recognition during passive and lexical decision tasks: an ERP investigation. *International Journal of Neuroscience*, *114*, 1293–1324. https://doi.org/10.1080/00207450490

Slobin, D. (2003). Language and thought online: cognitive consequences of linguistic relativity. In *Language in mind: Advances in the study of language and thought* (pp. 157–192).

Snowling, M. J., & Hulme, C. (2005). *The Science of Reading: A Handbook* (M. J. Snowling & C. Hulme, Eds.). Wiley-Blackwell.

Sun-Alperin, M. K., & Wang, M. (2011). Cross-language transfer of phonological and orthographic processing skills from Spanish L1 to English L2. *Reading and Writing*, *24*(5), 591–614. https://doi.org/10.1007/s11145-009-9221-7

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics*. Pearson.

Taft, M. (1991). *Reading and the Mental Lexicon*. Lawrence Erlbaum.

Tagamets, M. a, Novick, J. M., Chalmers, M. L., & Friedman, R. B. (2000). A parametric approach to orthographic processing in the brain: an fMRI study. *Journal of Cognitive Neuroscience*, *12*(2), 281–297. https://doi.org/10.1162/089892900562101

Taha, H., & Khateb, A. (2013). Resolving the orthographic ambiguity during visual word recognition in Arabic: An event-related potential investigation. *Frontiers in Human Neuroscience*, *7*(DEC), 1–12. https://doi.org/10.3389/fnhum.2013.00821

Tan, L. H., Laird, A. R., Li, K., & Fox, P. T. (2005). Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: a meta-analysis. *Human Brain Mapping*, *25*(1), 83–91. https://doi.org/10.1002/hbm.20134

Tanaka-Ishii, K., & Terada, H. (2011). Word familiarity and frequency. *Studia Linguistica*, *65*(1), 96–116. https://doi.org/10.1111/j.1467-9582.2010.01176.x

Tanner, D., Morgan-short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, *00*, 1–13. https://doi.org/10.1111/psyp.12437

Tanner, D., Norton, J. J. S., Morgan-Short, K., & Luck, S. J. (2016). On high-pass filter artifacts (they're real) and baseline correction (it's a good idea) in ERP/ERMF analysis. *Journal of Neuroscience Methods*, *266*, 166–170. https://doi.org/10.1016/j.jneumeth.2016.01.002

Taroyan, N. A. (2015). Seeing is knowing? Visual word recognition in non-dyslexic and dyslexic readers: An ERP study. *Visual Cognition*, *23*(5), 577–596. https://doi.org/10.1080/13506285.2015.1055852

Taroyan, N. A., & Nicolson, R. I. (2009). Reading words and pseudowords in dyslexia: ERP and behavioural tests in English-speaking adolescents. *International Journal of*

Psychophysiology : Official Journal of the International Organization of Psychophysiology, *74*(3), 199–208. https://doi.org/10.1016/j.ijpsycho.2009.09.001

The jamovi project. (2021). *jamovi* (1.6.23).

Thiebaut De Schotten, M., Cohen, L., Amemiya, E., Braga, L. W., & Dehaene, S. (2014). Learning to read improves the structure of the arcuate fasciculus. *Cerebral Cortex*, *24*(4), 989–995. https://doi.org/10.1093/cercor/bhs383

Timmer, K., & Schiller, N. O. (2012). The role of orthography and phonology in English: an ERP study on first and second language reading aloud. *Brain Research*, *1483*, 39–53. https://doi.org/10.1016/j.brainres.2012.09.004

Tong, X., Maurer, U., Chung, K. K. H. H., & Mcbride, C. (2016). Neural specialization for print in Chinese-English language learners. *Journal of Neurolinguistics*, *38*, 42–55. https://doi.org/10.1016/j.jneuroling.2015.10.001

Trauzettel-Klosinski, S., & Dietz, K. (2012). Standardized assessment of reading performance: The new international reading speed texts IReST. *Investigative Ophthalmology and Visual Science*, *53*(9), 5452–5461. https://doi.org/10.1167/iovs.11-8284

Treiman, R., & Kessler, B. (2005). Writing Systems and Spelling Development. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 120–134). Blackwell Publishing.

Twomey, T., Kawabata Duncan, K. J., Hogan, J. S., Morita, K., Umeda, K., Sakai, K., & Devlin, J. T. (2013). Dissociating visual form from lexical frequency using Japanese. *Brain and Language*, *125*(2), 184–193. https://doi.org/10.1016/j.bandl.2012.02.003

Twomey, T., Kawabata Duncan, K. J., Price, C. J., & Devlin, J. T. (2011). Top-down modulation of ventral occipito-temporal responses during visual word recognition. *NeuroImage*, *55*(3), 1242–1251. https://doi.org/10.1016/j.neuroimage.2011.01.001

Vale, A. P. (2011). Orthographic context sensitivity in vowel decoding by Portuguese monolingual and Portuguese-English bilingual children. *Journal of Research in Reading*, *34*(1), 43–58. https://doi.org/10.1111/j.1467-9817.2010.01482.x

van der Mark, S., Bucher, K., Maurer, U., Schulz, E., Brem, S., Buckelmüller, J., Kronbichler, M., Loenneker, T., Klaver, P., Martin, E., & Brandeis, D. (2009). Children with dyslexia lack multiple specializations along the visual word-form (VWF) system. *NeuroImage*, *47*(4), 1940–1949. https://doi.org/10.1016/j.neuroimage.2009.05.021

van Heuven, W. J. B., & Dijkstra, T. (2010). Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain Research Reviews*, *64*(1), 104–122. https://doi.org/10.1016/j.brainresrev.2010.03.002

van Heuven, W. J. B., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language conflict in the bilingual brain. *Cerebral Cortex (New York, N.Y. : 1991)*, *18*(11), 2706–2716. https://doi.org/10.1093/cercor/bhn030

van Heuven, W. J. B., & Wen, Y. (2018). *The need for a universal computational model of bilingual word recognition and word translation*. https://doi.org/10.1017/S1366728918000688

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, *15*(3), 181–198.

van Orden, G. C., & Kloos, H. (2005). The Question of Phonology and Reading. In *The Science of Reading: A Handbook* (pp. 61–78).

Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(3), 616–627. https://doi.org/10.1037//0096-1523.28.3.616

Veale, J. F. (2014). Edinburgh Handedness Inventory - Short Form: a revised version based on confirmatory factor analysis. *Laterality*, *19*(2), 164–177. https://doi.org/10.1080/1357650X.2013.783045

Vingerhoets, G., van Borsel, J., Tesink, C., van den Noort, M., Deblaere, K., Seurinck, R., Vandemaele, P., & Achten, E. (2003). Multilingualism: An fMRI study. *NeuroImage*, *20*(4), 2181–2196. https://doi.org/10.1016/j.neuroimage.2003.07.029

Vogel, A. C., Petersen, S. E., & Schlaggar, B. L. (2014). The VWFA: it's not just for words anymore. *Frontiers in Human Neuroscience*, *8*(March), 88. https://doi.org/10.3389/fnhum.2014.00088

Vygotsky, L. (1986). *Thought and Language*. MIT Press.

Wang, F., & Maurer, U. (2017). Top-down modulation of early print-tuned neural activity in reading. *Neuropsychologia*, *102*, 29–38. https://doi.org/10.1016/j.neuropsychologia.2017.05.028

Wang, F., & Maurer, U. (2020). Interaction of top-down category-level expectation and bottom-up sensory input in early stages of visual-orthographic processing. *Neuropsychologia*, *137*. https://doi.org/10.1016/j.neuropsychologia.2019.107299

Wang, M., Koda, K., & Perfetti, C. A. (2003). Alphabetic and nonalphabetic L1 effects in English word identification: a comparison of Korean and Chinese English L2 learners. *Cognition*, *87*, 129–149. https://doi.org/10.1016/S0

Weber-Fox, C., Spencer, R., Cuadrado, E., & Smith, A. (2003). Development of neural processes mediating rhyme judgments: Phonological and orthographic interactions. *Developmental Psychobiology*, *43*(2), 128–145. https://doi.org/10.1002/dev.10128

Wheat, K. L., Cornelissen, P. L., Frost, S. J., & Hansen, P. C. (2010). During Visual Word Recognition, Phonology Is Accessed within 100 ms and May Be Mediated by a Speech Production Code: Evidence from Magnetoencephalography. *The Journal of Neuroscience*, *30*(15), 5229–5233. https://doi.org/10.1523/JNEUROSCI.4448-09.2010

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the Time Course of Phonological Encoding. *Journal of Memory and Language*, *34*(3), 311–334. https://doi.org/10.1006/jmla.1995.1014

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*, 475–482.

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data – a practical approach. *Journal of Neuroscience Methods*, *250*, 34–46. https://doi.org/10.1016/j.jneumeth.2014.08.002

Williams, C., & Bever, T. (2010). Chinese character decoding: A semantic bias? *Reading and Writing*, *23*(5), 589–605. https://doi.org/10.1007/s11145-010-9228-0

Wirth, M., Horn, H., Koenig, T., Stein, M., Federspiel, A., Meier, B., Michel, C. M., & Strik, W. (2007). Sex differences in semantic processing: event-related brain potentials distinguish between lower and higher order semantic analysis during word reading. *Cerebral Cortex (New York, N.Y. : 1991)*, *17*(9), 1987–1997. https://doi.org/10.1093/cercor/bhl121

Woodhead, Z. V. J., Barnes, G. R., Penny, W., Moran, R., Teki, S., Price, C. J., & Leff, A. P. (2012). Reading front to back: MEG evidence for early feedback effects during word recognition. *Cerebral Cortex*, *24*(3), 817–825. https://doi.org/10.1093/cercor/bhs365

Wu, Y. J., & Thierry, G. (2010). Chinese-English bilinguals reading English hear Chinese. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(22), 7646–7651. https://doi.org/10.1523/JNEUROSCI.1602-10.2010

Xu, Y., Pollatsek, a, & Potter, M. C. (1999). The activation of phonology during silent Chinese word reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(4), 838–857. https://doi.org/10.1037/0278-7393.25.4.838

Yeong, S. H. M., Fletcher, J., & Bayliss, D. M. (2014). Importance of phonological and orthographic skills for english reading and spelling: A comparison of english monolingual and mandarin-english bilingual children. *Journal of Educational Psychology*, *106*(4), 1107–1121. https://doi.org/10.1037/a0036927

Yu, D., Park, H., Gerold, D., & Legge, G. E. (2010). Comparing reading speed for horizontal and vertical English text. *Journal of Vision*, *10*(2), 1–17. https://doi.org/10.1167/10.2.21

Yule, G. (1996). *The Study of Language* (Second). University Press, Cambridge.

Yum, Y. N., & Law, S. P. (2021). N170 reflects visual familiarity and automatic sublexical phonological access in L2 written word processing. *Bilingualism*, *24*(4), 670–680. https://doi.org/10.1017/S1366728920000759

Zhan, J., Yu, H., & Zhou, X. (2013). fMRI evidence for the interaction between orthography and phonology in reading Chinese compound words. *Frontiers in Human Neuroscience*, *7*(November), 753. https://doi.org/10.3389/fnhum.2013.00753

Zhang, D., He, W., Wang, T., Luo, W., Zhu, X., Gu, R., Li, H., & Luo, Y. J. (2014). Three stages of emotional word processing: An ERP study with rapid serial visual presentation. *Social Cognitive and Affective Neuroscience*, *9*(12), 1897–1903. https://doi.org/10.1093/scan/nst188

Zhang, Q., Zhang, J. X., & Kong, L. (2009). An ERP study on the time course of phonological and semantic activation in Chinese word recognition. *International Journal of Psychophysiology*, *73*(3), 235–245. https://doi.org/10.1016/j.ijpsycho.2009.04.001

Ziegler, J. C., Jacobs, A. M., & Klüppel, D. (2001). Pseudohomophone effects in lexical decision: Still a challenge for current word recognition models. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 547–559. https://doi.org/10.1037//0096-1523.27.3.547

Zorzi, M. (2010). The connectionist dual process (CDP) approach to modelling reading aloud. *European Journal of Cognitive Psychology*, *22*(5), 836–860. https://doi.org/10.1080/09541440903435621