L1 IMPACTS ON L2 COMPONENT READING SKILLS, WORD SKILLS, AND OVERALL READING ACHIEVEMENT

by

Katherine I. Martin

A.B., Brain, Behavior, and Cognitive Science, University of Michigan, 2009

M.A., Linguistics, University of Pittsburgh, 2011

Submitted to the Graduate Faculty of

The Kenneth P. Dietrich School of Arts & Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

Kenneth P. Dietrich School of Arts & Sciences

This dissertation was presented

by

Katherine I. Martin

It was defended on

April 6, 2015

and approved by

Keiko Koda, Professor, Department of Modern Languages, Carnegie Mellon University

Charles Perfetti, Professor, Department of Psychology, University of Pittsburgh

Yasuhiro Shirai, Professor, Department of Modern Languages and Literatures, Case Western Reserve University

Natasha Tokowicz, Associate Professor, Department of Psychology, University of Pittsburgh

Dissertation Advisor: Alan Juffs, Professor, Department of Linguistics, University of Pittsburgh

L1 IMPACTS ON L2 COMPONENT READING SKILLS, WORD SKILLS, AND OVERALL READING ACHIEVEMENT

Katherine I. Martin, Ph.D.

University of Pittsburgh, 2015

Learning to read in a second language as an adult is different in many ways from learning to read in a first language. Unlike children, adult second language (L2) learners have limited knowledge of the target language but may already have fluent reading skills in their first language (L1). These initial reading skills develop to be specifically tuned to the characteristics of the L1 writing system, and may not be optimized for literacy in the L2 (e.g., Frost, 2012; Koda, 2004). This dissertation consists of a program of research designed to examine the impacts that these L1 writing system characteristics have on the development of literacy skills in English as a second language (ESL). Study 1 examined performance on two fundamental literacy skills, phonological awareness and orthographic knowledge, as a function of L1 background and task demands. These data were collected abroad from native French, Hebrew, and Mandarin Chinese speakers, as well as native English speakers, and show clear influences of both L1 orthography and phonology on literacy skill performance. The large differences in performance associated with varying task demands have implications for accurately measuring and understanding students' underlying abilities. Study 2 examined the contributions of phonological awareness and orthographic knowledge to three measures of word identification: lexical decision, word naming, and pseudoword decoding, as well as global reading comprehension. These data reveal differential performance on the word identification tasks across L1s, as well as differential contributions of phonological awareness and orthographic knowledge to word identification.

Study 2 again revealed the effects of task demands on the relationships between sub-lexical literacy skills and word identification. Finally, Study 3 examined the development of language and literacy in adult ESL classroom learners who received either traditional reading instruction or a set of supplemental lessons providing a phonics-based instructional intervention. The results show influences of L1 background as well as different developmental patterns for phonological and orthographic skills based on the type of curriculum students received. The discussion highlights the contributions of this work to understanding cross-linguistic literacy skills and the importance of considering task demands when choosing language assessment measures.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This dissertation would not have been possible without the support of a large number of individuals and institutions. First, I acknowledge the support of the National Science Foundation: I was supported by a National Science Foundation Graduate Research Fellowship during the completion of this dissertation, and the international portion of data collection (Study 1 and Study 2) was supported by a National Science Foundation Doctoral Dissertation Research Improvement Grant, BCS-1421354. This research was also supported in part by a research grant from the Pittsburgh Science of Learning Center to Dr. Alan Juffs, which is funded by National Science Foundation grant number SBE-0836012 (PSLC, http://www.learnlab.org). It was previously funded by NSF award number SBE-0354420. I also acknowledge financial support from a *Language Learning* Dissertation Grant, the Israel Heritage Room Committee Scholarship (part of the Nationality Rooms Scholarship Program at the University of Pittsburgh), and a departmental research grant from the University of Pittsburgh Department of Linguistics. I also thank the English Language Institute at the University of Pittsburgh and the Pittsburgh Science of Learning Center for providing funds that allowed me to present parts of this work at the 2014 Second Language Research Forum in Charleston, South Carolina, the 2014 annual meeting of the American Association for Applied Linguistics in Portland, Oregon, the Sixth Annual inter-Science of Learning Center and Post-Doc Conference at the Spatial Intelligence and Learning

Center in Philadelphia, Pennsylvania, and the Seventh Annual inter-Science of Learning Center and Post-Doc Conference at the Pittsburgh Science of Learning Center in Pittsburgh, Pennsylvania.

I especially thank Philippe Prévost from the Université François-Rabelais in Tours, France; Anat Prior from the University of Haifa in Haifa, Israel; and Wei-Chang Shann and David Wible from National Central University in Zhongli City, Taiwan, for their support, assistance, and hospitality in helping to arrange my data collection visit to their institutions.

I also thank the teachers and students at the University of Pittsburgh English Language Institute for their enthusiasm and willingness to partake in my research for Study 3. In particular I acknowledge the contributions of Megan Reiley, the curriculum coordinator for the high-intermediate reading classes in the ELI, to the design, refinement, and implementation of the phonics-based instructional intervention.

I gratefully acknowledge the assistance of a large number of undergraduate research assistants from the Department of Linguistics and the Department of Psychology at the University of Pittsburgh and the Faculty of Education at the University of Haifa: Ranem Atia, Deja Bulluck, Joelle Chasse, Chun Chen, Rundi Guo, Luis Hernandez, Avani Kolla, Erika Latham, Yingqi Liu, Christine Matt, Jessica Packer, Ryan Pudlowski, Muna Rohana, Rachel Rosenberg, Daniel Sonnenberg, and Victoria Yuhas. Without there efforts I would still be collecting and coding data. I also heartily thank a large number of friends and colleagues who provided assistance in numerous forms as well as moral support when it was most needed: Adeetee Bhide, Wendy Li-Yun Chang, Chelsea Eddington, Xiaoping Fang, Elizabeth Hirshorn, Michelle Holcomb, Evelyn Milburn, Noriyasu Li, Kimberly Muth, Nora Presson, Caitlin Rice, Alba Tuninetti, Zhaohong Wu, and the many members of the PLUM lab and the Pittsburgh Science of Learning Center.

I thank my dissertation committee members for their thoughtful and helpful comments, advice, and feedback throughout this project: Alan Juffs, Keiko Koda, Chuck Perfetti, Natasha Tokowicz, and Yasuhiro Shirai. I also thank my statistics instructors at the University of Pittsburgh, particularly Scott Fraundorf, Kevin Kim, and Feifei Ye for their patience assistance and advice at many stages. I dedicate this dissertation to the memory of Kevin H. Kim, my stats adviser for life.

1.0 INTRODUCTION

Literacy comprises a crucial set of skills that often serve as a gateway for access to educational and employment opportunities (Kirsch, 1993). These skills comprise a foundational component for education for speakers of any language, and are also critical for learners of a second language (L2). Reports suggest that a majority of the world's population is bilingual (Baker & Jones, 1998), and English in particular is a crucial language for many seeking personal and professional advancement. A growing number of international companies use English as their language of business, and it is widely used as a *lingua franca* across the world in educational, professional, and social contexts (Graddol, 1997). Therefore, understanding the skills underlying literacy in a second language, and in English as a second language (ESL) in particular, is crucial for language educators and language learners, as well as those interested in the cognitive bases of literacy more generally.

1.1 COMPARING FIRST AND SECOND LANGUAGE READING

Adult learners face a unique set of challenges when learning to read in their L2. When children learn to read in their first language (L1), they generally already have high levels of spoken language proficiency and an extensive vocabulary (e.g., Birch, 2015; Grabe, 2009; Koda, 2004). The focus for children learning to read is therefore on learning the mappings between the phonological and the orthographic units, refining their awareness of their L1's phonological units, and building fluency and automaticity through practice. Adult L2 learners, on the other hand, generally have limited target language proficiency, but in many cases have already developed fluent L1 reading processes. An extensive body of research has documented that L1 reading processes develop so that they are tuned to the particular characteristics of the L1 orthography and morphophonological structure (e.g., Frost, 2012; Katz & Frost, 1992; Ziegler & Goswami, 2005). For example, readers of languages with consistent and transparent mappings between written and spoken language units (such as Spanish or Serbo-Croatian) tend to rely relatively more on phonological information and letter-by-letter decoding for word reading, although readers of languages with inconsistent and opaque mappings (such as Chinese) tend to rely relatively more on visual and orthographic information (e.g., Katz & Frost, 1992).

Generally speaking, obtaining literacy in the L1 benefits the development of literacy in the L2 (e.g., Cummins, 1979, 1984). This is particularly true for languages with broadly similar spoken and written language structures, because these speakers of these languages are able to take advantage of the fact that literacy skills and reading processes that have been developed for reading in the L1 can be employed for reading in the L2. This transfer can occur for component skills and reading processes at multiple levels, ranging from the visual perception of the written language units (graphemes) to phonological skills to expectations of syntactic and discourse structure (Durgunoğlu, Nagy, & Hancin-Bhatt, 1993; Grabe, 2009; Koda, 2004). Thus, when the L1 and L2 have shared structures, the skills developed for L1 can be directly transferred and successfully applied to the L2. However, differences in the spoken and written structures of the L1 and L2 may result in the transfer and use of reading skills and cognitive processes that are not optimized for the L2 (Koda, 2004). For example, individuals who have L1s with consistent and transparent grapheme-phoneme correspondences (GPCs), such as Spanish or Korean, tend to use more phonological information and decoding strategies while reading. However, when speakers of these languages learn to read in an L2 with inconsistent GPCs, such as English, transferring their phonology-focused L1 reading processes to the L2 results in regularization errors when they (incorrectly) apply the regular GPCs to words with exceptional spelling-pronunciation mappings.

1.2 COMPONENT SKILLS APPROACHES TO READING RESEARCH

Much of the research on the transfer of reading skills from L1 to L2 over the past 25 years has been conducted within the framework of the component skills approach (Carr & Levy, 1990; Grabe, 2009; Koda, 2004). This framework emphasizes the importance of separable cognitive and linguistic skills and their interactions for supporting reading, and its usefulness derives particularly from the consideration of reading as comprising a group of separable skills that can be examined individually as well as in conjunction with others. This view of reading thus allows for a detailed consideration of where evidence of transfer is manifested and where it is not. Although these skills may include both reader-driven, top-down processes as well as text-driven, bottom-up processes, the component skills approach has been particularly useful for researchers interested in the bottom-up, sub-lexical skill sets that support literacy development and which have been shown to develop differentially in response to a language's morphophonological and orthographic characteristics (Frost, 2012; Katz & Frost, 1992; Ziegler & Goswami, 2005, 2006). Among the many component skills that have been identified as crucial for underpinning literacy, phonological awareness and orthographic knowledge are two of the most widely understood.

1.2.1 Phonological awareness

Phonological awareness is an individual's ability to segment and manipulate phonological units of varying sizes (Goswami, 1999; Morais, Alegria, & Content, 1987). It is a crucial skill for literacy development not only in English but also cross-linguistically. Regardless of the writing system employed by a particular language, phonological awareness supports the ability to segment speech and build robust phonological representations of lexical items (McBride-Chang, 1995a, 1996; J. Zhang & McBride-Chang, 2010). However, it is particularly crucial for the development of automatic word recognition in alphabetic languages because it supports the extensive practice required for automatizing the conversion of letters to sounds and blending them together (Adams, 1990; Ehri, 1997). Although phonological awareness impacts the development of literacy, literacy experiences themselves also have a reciprocal impact on the development of phonological awareness. Awareness of syllables as phonological units develops even in individuals without literacy experience (Goswami & Bryant, 1990), but the further refinement of phonological awareness to smaller unit sizes apparently depends heavily on a speaker gaining literacy experience. In particular, speakers of a language adjust their awareness of phonological units down to the size of the phonological unit that is represented in the writing system. For example, speakers from alphabetic L1s tend to develop high levels of phonemic awareness (awareness of individual phonemes), supported by the fact that phonemes are generally represented individually in the writing system, but Chinese speakers without *pinyin*[1] experience generally have poor phonemic awareness. Instead, syllables are the phonological unit

[1] *Pinyin* in the system of romanization for Chinese characters that is currently in use in mainland China, to some extent and with variations in Taiwan, and Singapore.

for which they show greatest awareness (McBride-Chang, Bialystok, Chong, & Li, 2004; Tolchinsky, Levin, Aram, & McBride-Chang, 2012).

The general pattern of results from this research demonstrates that speakers of a language develop greater phonological awareness for phonological units that correspond to the representational unit of their L1 orthography (see Ziegler & Goswami, 2005 for a review). Beyond this, however, relatively little is known about how factors such as item structure and task demands may impact performance. Additionally, most research has focused on testing phonological awareness by examining awareness of onset and rime units or individual phonemes at the beginning or end of a word. However, these units are only some of the wider range of phonological units that exist and could be tested, and it is unclear what patterns of performance might be found for different units (such as body and rime) tested in speakers from different L1s.

1.2.2 Orthographic Knowledge

Similar to phonological awareness, orthographic knowledge is crucial for reading and spelling across languages. However, there is evidence that it is relatively more important for reading and other literacy skills in non-alphabetic languages (Arab-Moghaddam & Sénéchal, 2001; McBride-Chang & Ho, 2005; Tong, McBride-Chang, Shu, & Wong, 2009). Thus, differences may exist among readers from different L1s not only in terms of their level of performance on these two sets of skills, but also in terms of their relative reliance on these skills for supporting higher-level reading processes. In general, readers of alphabetic languages tend to rely relatively more on phonological information and their highly developed phonological skills for word identification and reading. In contrast, readers of non-alphabet languages tend to rely relatively more on visual and orthographic information for the same tasks (McBride-Chang et al., 2005; McBride-Chang

& Ho, 2005; Nassaji & Geva, 1999; Wade-Woolley, 1999). Considering that the skills and processes developed for reading in the L1 are often initially transferred and employed in the L2, these differences in the relative importance of phonological and orthographic skills in the L1 mean that in addition to quantitative differences, qualitative differences in L2 literacy skills may be found between learners from different L1s. This proposal has been supported by much of the research to date on L2 literacy sub-skills. In general, previous research has shown that learners coming from an alphabetic L1 background tend to rely more on phonological information, and learners coming from a non-alphabetic L1 background tend to rely more on orthographic information, in order to support their L2 reading performance (e.g., Chikamatsu, 1996; Wade-Woolley, 1999; Wang & Koda, 2005; Wang, Koda, & Perfetti, 2003).

Beyond these general patterns, however, much of the literature does not contribute to a fine-grained view of the various dimensions of orthographic knowledge, particularly in L1. This skill set can be measured in a number of different ways (see Berninger, 1994), for example, considering precise knowledge of individual word spellings versus general patterns of orthotactic constraints on letter combinations or positional bigram statistics for particular letter sequences. The finding of superior performance on L2 orthographic knowledge tasks for speakers with non-alphabetic L1s holds for tasks requiring visual knowledge of whole-word spellings; for example, Wang and Geva (2003) found that L1 Chinese speakers had superior performance on a forced-choice test of ESL spelling knowledge than L1 Korean speakers. However, other studies measuring orthographic knowledge in terms of sensitivity to letter combination legality and positional frequency have found more accurate performance by L1 Korean speakers than L1 Chinese speakers (Koda, 1999). These differing results highlight the importance of considering task structure and task demands when examining orthographic knowledge, particularly in an L2.

1.3 RELEVANCE OF L2 COMPONENT SKILLS READING RESEARCH

From the standpoint of advancing basic science, understanding the quantitative differences in literacy skills performance by L1 speakers reading in their L1, and L2 learners with varying L1 backgrounds reading in their L2, can help us understand the broad cognitive consequences of learning to read. It is also relevant for understanding how the brain optimizes its processing in response to different types of (language) input and whether or how these processes may subsequently be adapted in response to new and varying types of (language) input. This work additionally has important implications for applied language research. A careful examination of the quantitative differences in literacy skills performance by L2 learners with different L1 backgrounds is important for understanding the particular challenges that they are most likely to face, based on the characteristics of the particular L1 and L2. In addition, understanding qualitative differences in how learners rely on different sets of skills to support higher-level literacy behaviors has implications for understanding how these learners may be able to overcome particular processing challenges in order to obtain the same level of literacy success by relying on different skills.

Although much work has already been done, continuing to build a strong body of research documenting these differences with careful, controlled empirical research is a crucial step for growing our understanding of these issues and contributing to both basic and applied research goals. However, the measurement of literacy skills in careful, controlled laboratory conditions using individual test administration is much different from the reality and complexity of real classrooms, where students actually employ these skills. Additionally, individualized practice or training activities as used in some research studies are not always feasible for being implemented by real teachers with a classroom full of students with widely varying levels of ability. Thus, in order for the results of this research to be most directly relatable to real-life language learners in realistic learning situations, it is important for the same types of tasks and language measures to be brought into and used in these types of realistic learning environments. It is this kind of work that can help build a more robust understanding of how literacy skills can be efficiently measured in classroom environments and then instruction and practice activities developed that can support learners in the complexities of real learning situations.

1.4 INTERVENTION RESEARCH

Although there is a growing body of literature documenting the types of processing challenges that ESL learners may face as a result of the structural characteristics of their L1, there is relatively little work exploring ways to actually address these challenges, particularly for adult learners. The vast majority of research on literacy interventions is focused on normally developing and at-risk children learning to read in their native language (Gaskins et al., 1988; Gaskins, Ehri, Cress, O'Hara, & Donnelly, 1996; Kyle, Kujala, Richardson, Lyytinen, & Goswami, 2013; Lovett, Lacerenza, & Borden, 2000; Lovett, Lacerenza, Borden, et al., 2000; Stahl, Duffy-Hester, & Stahl, 1998; Walton, Bowden, Kurtz, & Angus, 2001; Walton & Walton, 2002; Walton, Walton, & Felton, 2001). In general, this body of work suggests that sub-lexical literacy skills, in particular phonological awareness, are trainable and have a causal relationship with both reading and spelling performance (e.g., Ball & Blachman, 1991; Gaskins et al., 1988; Kyle et al., 2013; Lovett, Lacerenza, & Borden, 2000; Lovett, Lacerenza, Borden, et al., 2000; Tunmer & Hoover, 1993; Wagner & Torgesen, 1987). A much smaller body of work shows similar results for bilingual children, as well as evidence for the transfer of literacy skills from the L1 to the L2 (Bialystok, McBride-Chang, & Luk, 2005; Durgunoğlu et al., 1993). In contrast, there is relatively little work looking at literacy interventions in adult L2 learners who are developing literacy in their L2 for the first time. The results of these few studies are also markedly different from the results of the research with children: rather than showing a clear trainability of sub-lexical literacy skills that supports growth in higher-level literacy processes, the intervention research with adults shows little evidence of improvement in literacy skills or their transfer to other literacy processes (e.g., M. Taylor, 2008). Although these initial results have not proven as promising as the work with children, so little work has been done with adults that a wide range of possible learning and training activities remain unexplored.

To reiterate, previous research demonstrates that the morphophonological and orthographic structures of a language interact to have an impact on the development of sub-lexical literacy skills and the relative importance of these skills for supporting higher-level reading processes. At least some of these sub-lexical literacy skills, such as phonological awareness, also develop concurrently with literacy experience and are responsive to direct instruction and focused practice, particularly in children learning to read in their L1. After developing to support L1 reading, readers' specific literacy strategies and processes are often transferred from the L1 into an L2, even if they are not optimal for reading in the L2. This can contribute to specific types of reading and spelling errors that are at least partially predictable for learners with particular L1-target language pairings.

1.5 LIMITATIONS OF PREVIOUS L2 COMPONENT SKILLS READING RESEARCH

There are a number of gaps in the previous research on L2 literacy skills that still need to be addressed. As mentioned above, relatively little consideration has been given to the implications of linguistic structure, task structure, and task demands for the measurement of a speaker's literacy skills (however see McBride-Chang, 1995b for a consideration of the impact of linguistic features on phonological awareness task performance). These factors are relevant for both phonological awareness and orthographic knowledge, each of which has multiple dimensions and possible units of analysis that need to be considered for a comprehensive understanding of task performance. These skill sets can also be measured using either receptive or productive tasks, and a clear examination of the impacts of these different task demands has not yet been made.

An additional shortfall in the literature is the relatively small number of studies comparing L2 literacy skills across L2 learners from multiple L1s with the same materials and the same controlled experimental conditions. Most studies that include more than one L1 only compare between two groups, generally one alphabetic (Spanish, Korean) and one non-alphabetic (often Chinese). Although studies with more than two groups are logistically challenging, they are important for building a comprehensive picture of L2 literacy skills and when and how a learner's L1 background does or does not impact their L2 performance. Additionally, more research is needed to compare language pairings other than Chinese and Korean, the most well-represented comparison in the literature to date.

Finally, more classroom-based research on literacy skills is needed, particularly for adult language learners. It is important to relate the measurement of literacy skills and task performance in a laboratory setting to the realities of language learners in a real language learning situation, such as a classroom with learners from a wide range of L1s. It is not clear whether the patterns of performance and relationships among skills that are found in a controlled setting will be maintained in a less-controlled classroom learning environment. Laboratory studies also provide only a snapshot of student performance, rather than a detailed picture of skill development across time. A better understanding of the natural development of L2 literacy skills, particularly in adult learners, may in turn lead to the development of more effective and age-appropriate instructional techniques and practice activities for these learners. Lastly, although the careful measurement of literacy skills in a controlled setting is important for clearly establishing overall patterns of performance, this way of measuring literacy sub-skills is not feasible for language instructors who might be interested in using similar tasks to understand their students' literacy skill strengths and weaknesses. If measures of sub-lexical literacy skills such as phonological awareness and orthographic knowledge are indeed predictive of higher-level reading processes and can reveal learners' relative skill strengths and weaknesses, they will be an important set of tools for language instructors. However, this can only be the case if they are also made accessible based on the resources available to the average instructor.

1.6 GOALS OF THE CURRENT RESEARCH

This dissertation aims to address a number of these gaps in our understanding of L2 reading. In the first two studies, a large battery of literacy and cognitive tasks was used to measure sub-lexical component literacy skills, word identification, and reading comprehension in adult ESL readers from three representative L1 writing systems: French (an alphabet), Hebrew (a consonant-based abjad), and Mandarin Chinese (a morphosyllabary). Together these two studies expand on previous work by including three different L1 groups, strategically chosen to be representative of the range of possible writing systems, and focusing on adult learners rather than children. In terms of their individual contributions, Study 1A focused on ESL speakers' performance on a number of phonological awareness and orthographic knowledge tasks. The stimuli were design to test a more comprehensive range of linguistic structures and structural characteristics than considered in much previous research, allowing for a more fine-grained understanding of the impact of L1 background on L2 phonological and orthographic skills. Study 1B focused on differences across L1 groups in regards to their word identification, both receptive and productive, as well as pseudoword decoding and overall reading comprehension performance. This study also considered how phonological awareness and orthographic knowledge support these reading skills differentially across L1s. Together, the data from Study 1A and Study 1B allowed for a careful analysis of the impact of L1 on L2 literacy skills, as well as how the linguistic structure of a test item and the demands of a particular task may interact to impact performance in a more detailed way, and across a wider and more strategically chosen set of L1 groups, than much previous research.

Study 2 expands this investigation of ESL literacy skills to adult ESL classroom learners. Data on students' phonological awareness and orthographic knowledge were collecting using similar tasks, adapted for group administration in a classroom setting, once near the beginning and once near the end of an academic semester. The collection of data at two different time points allowed for an examination of the development of students' ESL literacy and language skills and reading behaviors across a semester. The development of these skills was also considered in relation to the type of reading instruction that students received, specifically, whether students received just their traditional reading instruction or whether they additionally received four supplemental lessons on English spelling and pronunciation patterns using a phonics-based approach. Study 2 also served as an exploration of the feasibility and validity of measuring literacy skills in a group-administration, classroom setting, and a comparison of the pattern of results obtained under these administration conditions to those obtained from a more controlled laboratory setting.

2.0 WRITING SYSTEMS

One of the advantages of adopting a component skills approach to reading is that it allows researchers and educators to isolate particular processes and to narrowly pinpoint specific possible underlying causes of specific individuals' reading difficulties. This approach has been particularly useful for cross-linguistic research on reading because a component view allows for a detailed consideration of when and how the written structures of different languages impact the processes involved in reading (and when they do not).

In order to study the effect of linguistic structure on reading performance, it is necessary to specify the defining characteristics of those structures so that languages can be accurately categorized and grouped. At this point it is useful to distinguish between the meanings of two closely related terms: writing system and orthography. A writing system is defined by the mapping relationship between the written unit and the particular (size of) spoken language unit that it represents. An orthography, on the other hand, is the way that a particular spoken language is implemented in a type of writing system (Perfetti & Dunlap, 2008). For example, English and Spanish are both written using an alphabet (the type of writing system), but they have slightly different orthographies (the language is rendered in an alphabet in slightly different ways, with largely overlapping but somewhat different sets of symbols).

2.1 TYPES OF WRITING SYSTEM

There are five major types of writing system: alphabet, abugida, abjad, syllabary, and morphosyllabary. In addition, writing systems may also be described in terms of their orthographic depth, based on the consistency of the mappings between written and spoken language units. These ways to classify writing systems are described below.

2.1.1 Alphabets

In an alphabet, both consonant and vowel phonemes are represented by their own full graphemic symbol or symbols (letters). Although the alphabets that are usually the most familiar to speakers of English use the Roman (Latin) alphabet, as does English, other alphabets do exist. These include the Cyrillic alphabet (Russian, Serbian, Macedonian, Bulgarian), the Georgian alphabet, and the Greek alphabet. Although it is non-Roman and also non-linear, Korean Hangul is also an alphabetic system because each block unit, representing a syllable, can be decomposed into individual representing single consonants and vowels (which themselves represent phonetic features, Sampson, 2015).

2.1.2 Abjads and Abugidas

Similar to alphabets, abugidas and abjads are also segmental writing systems, but they differ from alphabets, and from one another, with regard to how they represent vowel sounds. In an abugida, also known as an alphasyllabary (e.g., Bhide, Gadgil, Zelinsky, & Perfetti, 2013), consonant sounds are represented by full graphemic symbols, similar to an alphabet. However, vowel sounds are not represented by full symbols, but rather with diacritics or regular modifications to the consonant symbols. An example can be seen in Figure 1, where the consonant symbols appear in black and the vowel symbols appear in red. Languages written with an abugida include languages written with the Devanagari script, such as Hindi, Nepali, and Marathi (Bhide et al., 2013).

ती एका लहान मुलाला शिकवते

Figure 1. The sentence ‘She teaches a young boy’ in Marathi (an abugida).

In an abjad, consonant sounds are also represented by full symbols, again with vowels represented as diacritics or small modifications to the consonant symbols. The difference is that although the vowel markings are always included in the written form for an abugida, they are normally excluded from a written text in an abjad. The most well-known modern languages using abjads are Hebrew, in which the vowel markings are referred to as ‘pointing’, and Arabic, in which the vowel markings are referred to as ‘voweling’. An example can be seen in Figure 2, where the top line displays a sentence without voweling, as it would normally be written or read by a native speaker, and the bottom line displays the same sentence with voweling, similar to how it would be written in literary texts or those targeted to foreigners or children learning to read (e.g., Abu-Rabia, 1998; Abu-Rabia, 2001, 2002; Abu-Rabia & Awwad, 2004; Heywood & Nahmad, 1965).

هي تدرس ولد صغير.

هِيَ تُدَرِّس وَلَد صَغير.

Figure 2. The sentence 'She teaches a young boy' in Arabic (an abjad).

2.1.3 Syllabaries

In a syllabary, the written units represent whole syllables in the spoken language, and cannot be decomposed into the component phonemes of the syllable. The hiragana and katakana scripts of Japanese are the most well-known examples of syllabary writing systems (although note that Japanese also uses the character-based *kanji* script). Figure 3 shows an example sentence in Japanese. Other languages that use a syllabary writing system include Cherokee and the African language Vai (Daniels & Bright, 1996).

彼女は男の子を教える

Figure 3. The sentence 'She teaches a young boy' in Japanese (a syllabary).

2.1.4 Morphosyllabaries

In a morphosyllabary, also known as a logography (e.g., Koda, 1990), the written units represent whole morphemes, the minimal unit of meaning in a language (Matthews, 2007). These units generally correspond to syllables, as in a syllabary. The difference is that in a syllabary the number of graphemes is limited by the number of syllables in the spoken language, and there is a conventional relationship between the written symbol(s) and the pronunciation. In contrast, in a morphosyllabary, there are many more graphemes than syllables, and there is generally only limited (or often no) phonological information encoded in the graphemes. The most well-known morphosyllabaries include the Chinese languages Mandarin and Cantonese. Figure 4 shows an example sentence in Chinese.

她教一个小男孩

Figure 4. The sentence 'She teaches a young boy' in Chinese (a morphosyllabary).

2.1.5 Orthographic Depth

In addition to being categorized with regard to their mapping principle, writing systems can also be categorized with regard to the consistency of the mapping that exists between the written and the spoken units, referred to as the ‘depth’. This distinction is particularly relevant for the alphabet and abjad writing systems, for which contrasting instances of shallow and deep orthographies can readily be distinguished. For example, English is well-known as a deep alphabet, with highly inconsistent mappings both from graphemes to phonemes and from phonemes to graphemes (Share, 2008; Ziegler, Stone, & Jacobs, 1997). On the other hand, Italian, Finnish, and Serbo-Croatian are relatively shallow, with highly consistent mappings between graphemes and phonemes and between phonemes and graphemes (e.g., Katz & Frost, 1992). Abjads such as Arabic and Hebrew are generally considered to be shallow when the voweling or pointing is included, but extremely deep when it is left out of the written text (e.g., Abu-Rabia, 1997a, 1998).

2.2 DISAGREEMENTS ABOUT CLASSIFYING WRITING SYSTEMS

Unfortunately, not all researchers agree on how to categorize languages on the basis of their orthography, or even how many types of writing system exist. As mentioned above, some authors group the abjad and abugida writing systems with true alphabets and thus recognize only three main categories of writing system: alphabet, syllabary, and morphosyllabary (Birch, 2015; Perfetti & Dunlap, 2008; I. Taylor, 1981). Consequently, some research has categorized languages such as Arabic and Farsi as alphabets, and occasionally even combined speakers of these languages with speakers of true alphabetic languages such as Spanish for the purpose of analyzing results (Akamatsu, 1999, 2002, 2003; Birch, 2015; T. L. Brown & Haynes, 1985; Koda, 1990). Other researchers have categorized Korean, which uses a non-Roman, non-linear alphabet, as non-alphabetic (Birch, 2015; Perfetti & Dunlap, 2008; Ziegler & Goswami, 2005), presumably on the basis of the syllable-based blocks in the writing system (I. Taylor & Taylor, 1995). Because of these discrepancies in how languages are categorized, general conclusions about the effects of different writing system types on reading processes are often difficult to define and careful analysis must be made of research that describes results solely on the basis of language type.

2.3 DEFINITIONS USED IN THE CURRENT RESEARCH

For the purposes of this dissertation, the five categories of alphabet, abugida, abjad, syllabary, and morphosyllabary will be observed. In addition, the distinction between Roman and non-

Roman alphabets will be made when relevant, and the distinction between shallow and deep alphabets will also be referred to as appropriate.

3.0 LANGUAGE INFLUENCES ON COMPONENT READING SKILLS

Two of the fundamental component skills underlying literacy acquisition are phonological awareness and orthographic knowledge. Phonological awareness is a general awareness of and ability to manipulate (sub-lexical) phonological units. It is a general term that can refer to awareness of varying phonological unit sizes, although it often used to refer to the awareness of individual phonemes (more specifically referred to as ‘phonemic awareness’) (Goswami, 1999; Goswami & Bryant, 1990; Morais et al., 1987). Note that phonological awareness is distinct from the awareness of speech as consisting of separate sounds (Morais et al., 1987; Swingley & Aslin, 2000; Ziegler & Goswami, 2005)[2] and refers rather to the awareness of different linguistic groupings of sounds and the ability to recognize and manipulate these groupings. The second of these skills, orthographic knowledge, involves conscious and unconscious knowledge of orthotactic constraints (legal letter combinations), general spelling patterns, and precise word spellings (Berninger, 1994).

Each of these skills is multidimensional and supports word recognition, spelling, and other aspects of L1 and L2 literacy success in a number of ways. However, many of the details regarding how phonological awareness and orthographic knowledge support L1 and L2 literacy

[2] Although there is some evidence that phonological awareness may be influenced by the phonetics and phonotactics of a language (Caravolas & Bruck, 1993; McBride-Chang, 1995b; McBride-Chang et al., 2008), most research in this domain does not focus explicitly on this aspect of phonological awareness.

outcomes differ somewhat based on the characteristics of the individual reader and both the morphophonological structure and the orthographic characteristics of the written language. These differences have important implications for understanding the cognitive processes underlying literacy acquisition, how the brain adapts to processing linguistic input of a specific type, and understanding the possible sources of reading difficulties that some readers face.

3.1 FIRST LANGUAGE RESEARCH ON SUB-LEXICAL LITERACY SKILLS

3.1.1 Phonological awareness in L1

A substantial body of research over the past 50 years has demonstrated both that phonological awareness is crucial for learning to read and that it is also reciprocally influenced by the process of literacy acquisition. Phonological awareness initially develops for large phonological units, particularly whole syllables, prior to any literacy experiences. Finer-grained awareness of phonological units smaller than the syllable then begins to develop if individuals start to gain literacy in a language whose graphemes represent units smaller than a syllable. This reciprocal relationship has been demonstrated both in cross-sectional research with adult speakers of a language who have not learned to read (e.g., Morais, Bertelson, Cary, & Alegria, 1986; Morais, Cary, Alegria, & Bertelson, 1979), cross-linguistic comparisons of readers with alphabetic vs. morphosyllabic L1s (e.g., Holm & Dodd, 1996; Huang & Hanley, 1995; McBride-Chang et al., 2004; Read, Zhang, Nie, & Ding, 1986), and research with children before and during the process of learning to read (Burgess & Lonigan, 1998; Perfetti, Beck, Bell, & Hughes, 1987; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001).

In research with adult speakers of (Iberian) Portuguese, Morais and colleagues found that individuals without literacy experience performed significantly worse on multiple phonological tasks (sound detection, rhyme identification, and sound deletion) than individuals who had gained literacy as adults (Morais et al., 1987; Morais et al., 1986). Although the illiterate adults had a disadvantage on tasks testing awareness of syllables as well as phonemes, their deficit in performance was noticeably larger for phonemes than for whole syllables (Morais et al., 1986). Durgunoğlu and Oney (2002) similarly found that adult speakers of Turkish without formal schooling showed very low levels of phonological awareness, letter recognition, word recognition, and spelling abilities. However, following only 90 hours of literacy instruction, these adult learners made significant gains in all lexical and sub-lexical literacy skills. Finally, in a study examining illiterate and semi-literate speakers of Brazilian Portuguese, de Santos Loureiro, Braga, do Nascimiento Souza, Nunes Filho, Queiroz, and Dellatolas (2004) found very low levels of phonemic awareness which was highly dependent on speakers' level of letter knowledge and word reading abilities. In contrast, these speakers were generally able to detect phonological rimes, and this ability was not dependent on the speakers' letter and word knowledge.

A very similar pattern of results has been found in comparisons individuals who have learned to read in an alphabetic language, which encourages an awareness of words as consisting of individual phonemes, and individuals who have learned to read without using an alphabet. In a study looking at adult readers of Chinese, Read et al. (1986) found that individuals who had learned to read without the use of alphabetic *pinyin* had significantly lower awareness of phonemes than those who had experience reading *pinyin*. A similar pattern of results was found by Holm and Dodd (1996) and McBride-Chang et al. (2004), who compared adult and child readers from mainland China and from Hong Kong who differed in their use of *pinyin* when learning to read in Chinese. The readers from mainland China, who had learned *pinyin* during the schooling, performed just as well or better than other individuals with an alphabetic L1 (including native English speakers) on multiple different phonological awareness tasks. In contrast, the readers from Hong Kong, who had learned to read both Chinese and English through a "look and say" method rather than being mediated by the use of an alphabetic system such as *pinyin* (Taft & Chen, 1992), performed significantly worse than multiple L1 groups who had come from an alphabetic background or had alphabetic literacy experience. Finally, Mann (1986) compared awareness of phonemes and syllables (English) or morae[3] (Japanese) in children learning to read in the United States and Japan. She found that six-year-old children learning to read in Japanese, whose written units represent whole syllables rather than single phonemes, had much worse performance than six-year-old children learning to read in English on tasks requiring the identification or deletion of phonemes (as compared to morae or syllables).

Further evidence for the reciprocal development of phonological awareness (of increasingly smaller phonological units) and literacy experience comes from both cross-sectional and longitudinal studies of children learning to read in English. In general, this research has found that awareness of syllables and onset and rime units is separable from awareness of phonemes (Høien, Lundberg, Stanovich, & Bjaalid, 1995; Yopp, 1988). Awareness of these larger phonological units develops prior to literacy instruction and also prior to phoneme awareness (Anthony, Lonigan, Driscoll, Phillips, & Burgess, 2003; Bradley & Bryant, 1983; Burgess & Lonigan, 1998; Goswami & Bryant, 1990; Lenel & Cantor, 1981; Stanovich, 1992;

[3] Morae (singular: mora) are phonological units, very similar to syllables, in languages such as Japanese that distinguish between heavy and light syllables (Kubozono, 2002; Matthews, 2007; Vance, 1987).

Treiman & Zukowski, 1990, 1991). In addition, children, and particularly younger children, generally demonstrate more accurate performance on phonological tasks that require them to identify or manipulate syllables or onsets and rimes as compared to phonemes (Carroll & Snowling, 2001; Goswami & East, 2000; Greaney, Tunmer, & Chapman, 1997; Liberman, Shankweiler, Fischer, & Carter, 1974; McClure, Ferreira, & Bisanz, 1996; K. Nation & Hulme, 1997; Stahl & Murray, 1994; Treiman & Baron, 1981). In fact, prior to any literacy experiences, children are generally unable to isolate, count, or delete single phonemes unless they also correspond to the onset of the word (Bruce, 1964; Calfee, 1977; Liberman et al., 1974; Treiman & Baron, 1981). This awareness of small phonological units increases with age and particularly with literacy instruction (e.g., Burgess & Lonigan, 1998; Stuart & Coltheart, 1988; Wagner, Torgesen, & Rashotte, 1994), but continues to present some challenges even to normally developing children as old as age 8 or 9 (Bruce, 1964; Fox & Routh, 1975). Further evidence of the impact of literacy experience on the development of phoneme awareness comes from studies showing that as phoneme awareness beings to appear, it is also influenced by knowledge of the spelling of stimulus items, in particular the number of letters (e.g., Ehri & Wilce, 1980; Tunmer & Nesdale, 1985).

It should be noted that phonological awareness and its development are impacted by the morphophonological structure of a speaker's language, in addition to the characteristics of its orthographic system (see for example the Psycholinguistic Grain Size Theory, Ziegler & Goswami, 2005). This is a natural consequent of the idea that a language gets the orthography that it deserves based on the structural characteristics of the spoken system. Awareness of syllables is particularly prominent in Chinese because syllables form the minimal unit of meaning as well as the minimal unit of writing and most syllables receive relatively equal levels of stress (McBride-Chang et al., 2004; Tolchinsky et al., 2012). In spoken form, the syllable is the major perceptual unit for word identification in Spanish for both children and adults (Alvarez, Carreiras, & Taft, 2001; Carreiras & Perea, 2002; Goikoetxea, 2005). The relatively simple syllabic structures and the number of open syllables in Italian, Turkish, and Greek similarly support the early development of syllable awareness in speakers of these languages (e.g., Cossu, Shankweiler, Liberman, Katz, & Tola, 1988; Ziegler & Goswami, 2005). Although syllables are salient phonological units cross-linguistically, other phonological units may also hold a special status or receive a processing advantage due to the characteristics of the linguistic structures in a particular language. For example, awareness of onsets and rimes is crucial in English and French in addition to awareness of phonemes because of the prevalence of phonological neighbors based around rhymes (Bruck, Genesee, & Caravolas, 1997; Cutler, Mehler, Norris, & Segui, 1986; De Cara & Goswami, 2003). Phonemic awareness also develops in English as a response not just to the alphabetic writing system but also in response to the large number of consonant clusters in the spoken language (e.g., Cheung, Chen, Lai, Wong, & Hills, 2001).

Similar effects occur in Turkish and in Czech. In Turkish, the existence of features such as vowel harmony in plural formation require speakers to become aware of phonemic-level sound changes early in their linguistic development, which later supports the development of their phonemic awareness and knowledge of grapheme-phoneme correspondences (Durgunoğlu & Öney, 2002). In Czech, the prevalence of consonant clusters in word onsets encourages the development of phoneme awareness in children learning to speak Czech (Caravolas & Bruck, 1993). Although not as widely recognition, other units such as morae and phonological bodies (CV strings) may also be privileged. In Japanese, morae are more salient than syllables for spoken word recognition, and this preference is enhanced by gaining literacy in kana (Inagaki, Hatano, & Otake, 2000; Tamaoka & Terao, 2004). Finally, Korean speakers often show better recognition of shared units between words when those shared units comprise phonological bodies as opposed to rhymes, and also tend to rely on the consistencies of bodies rather than rhymes for decoding unfamiliar words (Yoon, Bolger, Kwon, & Perfetti, 2002). These additional levels of phonological awareness supplement and complement the awareness that develops as a direct result of learning to read, for example the development of phonemic awareness in readers of alphabet languages.

The evidence to date suggests that phonological awareness supports literacy acquisition in a number of ways. Awareness of rhymes and the categorization of words by similar sounds, such as rhyming endings, are a salient part of children's language in speakers of English (Bryant, Bradley, Maclean, & Crossland, 1989; Chukovsky, 1963; Dowker, 1989; Maclean, Bryant, & Bradley, 1987). Children at the early stages of learning to read are able to use this knowledge of similar-sound words to help them recognize that common sequences of letters correspond to these sounds. In fact, children's knowledge of nursery rhymes is a reliable predictor of their later phonological awareness (Bryant et al., 1989; Maclean et al., 1987). Awareness of onset and rime units continues to be crucial for literacy skills, particularly in English (Goswami & Bryant, 1990; Treiman, 1983, 1985; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). This awareness allows for analogies to be made to other known words that have similar letter strings (e.g., G. D. A. Brown & Deavers, 1999; Goswami, Gombert, & de Barrera, 1998), and an awareness of onset and rime units combined with the overlap between onsets and initial phonemes in many words also provides a foothold for children to begin segmenting words into smaller units and developing phonemic awareness (Goswami & Bryant, 1990).

The ongoing development of phonological awareness is also crucial in helping students discover and exploit grapheme-phoneme correspondences (GPCs), which in turn helps them to decode unfamiliar words (e.g., Olson, Davidson, Kliegl, & Foltz, 1985; Snowling, 1980; Tunmer, Herriman, & Nesdale, 1988) and to make connections between they have heard but have not read (e.g., Ehri, 1992; Share, 1995). An extensive body of research on (phonological) developmental dyslexia also demonstrates the strong relationship between phonological skills and the ability to decode unfamiliar or nonsense words: individuals with dyslexia generally struggle with phonological awareness and also have lower abilities in other skills requiring the use of phonological information, such as decoding pseudowords (Baddeley, Ellis, Miles, & Lewis, 1982; Catts, 1993; Frith & Snowling, 1983; Kochnower, Richardson, & DiBenedetto, 1983; Rack, Snowling, & Olson, 1992; Siegel & Ryan, 1988; Snowling, 2000; Stackhouse & Wells, 1997; Stanovich & Siegel, 1994; Szeszulski & Manis, 1987). As students gain reading experience, phonological awareness and related decoding abilities continue to support the extensive practice needed for the automatization of lower-level processes, such as word recognition, which frees up cognitive resources to be dedicated to the higher-level components of reading (Adams, 1990; Ehri, 1992, 1999; Perfetti, 1992). Finally, phonological awareness is also crucial for supporting other literacy skills, such as spelling: in general, poor phonological skills are associated with lower spelling performance in children with a wide range of ages (Bruck & Treiman, 1990; Bryant & Bradley, 1983; Frith, 1980; Perin, 1983; Rohl & Tunmer, 1988; Treiman, 1983; Waters, Bruck, & Seidenberg, 1985).

Overall, lower performance on a wide range of phonological tasks is associated with lower reading performance (see Brady & Shankweiler, 1991; Goswami & Bryant, 1990; H. S. Scarborough, 1998; Stanovich, 1986; Wagner & Torgesen, 1987 for reviews). This is true across oddity tasks (e.g., Bradley & Bryant, 1978), rhyme judgments (e.g., Ellis & Large, 1987; Holligan & Johnston, 1988), phoneme counting or tapping tasks (e.g., Perfetti et al., 1987; Share, Jorm, Maclean, & Matthews, 1984), and deletion tasks (e.g., Fox & Routh, 1975; Perfetti et al., 1987). The relationship continues to hold when students' IQ and vocabulary levels are controlled (e.g., Rosner & Simon, 1971; Tunmer & Nesdale, 1985). Importantly, the relationship also holds longitudinally: students' performance on phonological tasks is associated with their future performance on both real and nonsense word reading (e.g., Calfee, 1977; Mann, 1986; Perfetti et al., 1987) and their rate of progress in learning to read and spell (Bradley & Bryant, 1983, 1985; Bryant, Maclean, & Bradley, 1990; Calfee, Lindamood, & Lindamood, 1973; Liberman et al., 1974).

Although a substantial portion of the research on phonological awareness and its role in literacy acquisition has been done in speakers of English (see Share, 2008, for a discussion of the Anglocentricity in reading research), research with readers of other languages has also demonstrated the importance of phonological awareness for learning to read. There may be variation across L1s in terms of speakers' absolute levels of performance on different phonological awareness tasks, the level of awareness of different phonological units, and the use of different phonological unit sizes for decoding (Goswami, Ziegler, Dalton, & Schneider, 2001, 2003; Ziegler & Goswami, 2005). However, despite this variation, phonological skills remain crucial for literacy development and the pattern of their development remains largely the same. Thus, syllable and onset/rime awareness develop prior to phonemic awareness, and phonological awareness in general is broadly predictive of performance on other literacy skills such as word decoding and of gains in reading and spelling in languages as diverse as German, Dutch, Norwegian, Swedish, Finnish, Danish, Italian, Spanish, French, Portuguese, Turkish, Hungarian, Greek, and Hebrew (e.g., Bertelson, de Gelder, & van Zon, 1997; Bosman, 1996; Bruck et al., 1997; Carrillo, 1994; Cossu et al., 1988; Cossu, Shankweiler, Liberman, Tola, & Katz, 1987; de Jong & van der Leij, 2003; Demont & Gombert, 1996; Domínguez, 1996; Durgunoğlu & Öney, 1999; M. Harris & Giannouli, 1999; Høien et al., 1995; Jackson, Lu, & Ju, 1994; Jiménez González & Ortiz González, 2000; Lundberg, Frost, & Petersen, 1988; Lundberg, Olofsson, & Wall, 1980; McBride–Chang & Kail, 2002; Wolfgang Schneider, Küspert, Roth, Visé, & Marx, 1997; Share & Levin, 1999; Shatil, Share, & Levin, 2000; Siok & Fletcher, 2001; Wimmer, Landerl, & Schneider, 1994; Ziegler et al., 2010). This is true even for Chinese readers, for whom higher phonological awareness is associated with better character recognition and word reading at various ages (e.g., Ho & Bryant, 1997; Hu & Catts, 1998; Huang & Hanley, 1995, 1997; McBride-Chang & Ho, 2000; McBride–Chang & Kail, 2002; Siok & Fletcher, 2001; So & Siegel, 1997; Y. Zhang et al., 2013).

Finally, evidence for the continuing importance of phonological coding in fluent reading is seen in the automatic activation of phonological representations, even when not required by the task (see Frost, 1998 for an overview). For example, evidence of the automatic activation and processing of phonological information has been found using picture-word interference paradigms, (e.g., Alario, De Cara, & Ziegler, 2007), priming (e.g., Perfetti & Bell, 1991), silent reading (e.g., Folk, 1999; McCutchen & Perfetti, 1982; Perfetti & McCutchen, 1982; Rayner, Sereno, Lesch, & Pollatsek, 1995), letter search (e.g., Ziegler, van Orden, & Jacobs, 1997) and even semantic judgment tasks (e.g., V. Coltheart, Patterson, & Leahy, 1994; Luo, Johnson, & Gallo, 1998; van Orden, 1987). Although again most of this research has been conducted with L1 English speakers, the same pattern of findings extends both to other alphabetic languages (e.g., Ferrand & Grainger, 1992) but also to non-alphabetic languages such as Chinese (Hung &

Tzeng, 1981; Lam, Perfetti, & Bell, 1991; Perfetti & Zhang, 1995; Perfetti, Zhang, & Berent, 1992; Tan & Perfetti, 1998; Tzeng, Hung, & Wang, 1977; Ziegler, Tan, Perry, & Montant, 2000).

3.1.2 Orthographic knowledge in L1

Orthographic knowledge is also crucial for L1 literacy acquisition and plays a major role in supporting literacy. Generally speaking, orthographic knowledge encompasses the ability to discover, learn, form, store, and use information about the orthographic form of words and the typical characteristics of orthographic representations in a given language (Stanovich & West, 1989). It comprises a set of related skills that may or may not be under conscious control (see Berninger, 1994; Wagner & Barker, 1994 for an overview), including knowledge of orthographic structure (knowledge of positional frequencies of letters, for example, see Venezky, 1979; Venezky & Massaro, 1979), orthographic-linguistic mappings (including knowledge of multiple codes, such as whole-word codes or letter cluster to syllable or rime mappings, see Seidenberg & McClelland, 1989; van Orden, Pennington, & Stone, 1990), stored whole word images (Ehri, 1980a, 1980b), as well as more general metacognitive information about a writing system, such as conventions for capitalization and punctuation (see Clay, 1979).

Compared to phonological awareness, there has been somewhat less research on the component skills involved in orthographic awareness, its development, and the exact ways by which it contributes to literacy abilities. Nevertheless, a number of studies have revealed the importance of orthographic knowledge for literacy success in reading and spelling cross-linguistically. Developmentally, children know about words that rhyme from their spoken language use, and many words that rhyme phonologically also share a common spelling in their orthographic rime. This connection provides a way for children to start learning about whole spelling patterns, rather than memorizing the spelling of all words letter-by-letter (Goswami & Bryant, 1990). This is particularly useful for the many spelling patterns where there is NOT a one-to-one correspondence between letters and phonemes, and there is extensive evidence that children do use analogy to whole spelling patterns in words they already know to help them read and spell unfamiliar words (e.g., Baron, 1979; Campbell, 1985; Goswami, 1986, 1988a, 1988b, 1990; Marsh, Friedman, Desberg, & Saterdahl, 1981).

In addition to supporting the use of analogies for reading and spelling unfamiliar words, orthographic knowledge is also critical for the development of word-level fluency and automaticity in word recognition. Although phonological skills are crucial for being able to decode and assemble unfamiliar words in a text, rapid word recognition that frees up cognitive resources for higher-level processing requires that words be recognized as whole units rather than as a series of decomposed sub-units (e.g., Bear, 1991; Torgesen, 2002). The relationship is reciprocal: knowledge of and familiarity with orthographic sequences is acquired through extensive exposure to print, and this facilitates readers' ability to rapidly extract lexical information from printed forms (Adams, 1990; Ehri, 1994, 1998; Seidenberg & McClelland, 1989). At the same time, as readers become more fluent and automatic in their processing of text, they begin to rely relatively more on orthographic information as opposed to grapheme-by-grapheme decoding (Stage & Wagner, 1992; Wagner & Barker, 1994). Overall, greater orthographic awareness is associated with better decoding skills, more accurate word reading, and greater reading comprehension (e.g., Cunningham & Stanovich, 1993; Ehri, 1986; Goswami, 1986, 1988a, 1988b; Katzir et al., 2006; Manis, Doi, & Bhadha, 2000; Martinet, Valdois, & Fayol, 2004; Nagy, Berninger, Abbott, Vaughan, & Vermeulen, 2003; Olson, Forsberg, & Wise, 1994). Similar to phonological awareness, orthographic knowledge has been shown as important for literacy development cross-linguistically (Georgiou, Parrila, & Papadopoulos, 2008; Share, 2004; Templeton & Bear, 1992; Trabasso, Sabatini, Massaro, & Calfee, 2005). However, there is also evidence that orthographic knowledge is more closely related to reading skill than is phonological awareness for readers of non-alphabetic languages, likely because of the relative lack of phonological information available in a non-alphabetic written form (Arab-Moghaddam & Sénéchal, 2001; McBride-Chang & Ho, 2005; Nassaji & Geva, 1999; Tong et al., 2009).

It is important to note that some researchers have raised the question of whether orthographic knowledge is truly separable from phonological awareness, spelling knowledge, and print exposure (e.g., Vellutino, Scanlon, & Tanzman, 1994). Some of this concern is derived from the difficulty of designing tasks that are truly pure measures of orthographic knowledge. This can be appreciated by considering, for example, that many common measures of orthographic knowledge such as knowledge of letter names (e.g., Barron, 1991; Berninger & Alsdorf, 1989; Ehri, 1989) have a strong overlap with phonological knowledge and skills. However, this question has been the target of multiple studies that have repeatedly shown the unique predictive value and separability of orthographic knowledge for a variety of literacy tasks, including word reading, extended text reading, and spelling (e.g., Cunningham & Stanovich, 1990; Cunningham & Stanovich, 1993; Ehri, 1992; McCarthy & Warrington, 1990; Olson et al., 1994; Siegel, Share, & Geva, 1995; Stanovich, West, & Cunningham, 1991; Wagner & Barker, 1994).

In addition to its role in supporting literacy acquisition and the automatization of word recognition, the impact of orthographic knowledge on language processing more generally can be seen by means of the widespread effects that knowledge of orthographic forms has on tasks that are not inherently orthographic in nature. A wide body of evidence has demonstrated that orthographic form impacts spoken language perception and production, including auditory word recognition (e.g., Dehaene et al., 2010; Jakimik, Cole, & Rudnicky, 1985; Perre, Pattamadilok, Montant, & Ziegler, 2009; Slowiaczek, Soltano, Wieting, & Bishop, 2003; Tanenhaus, Flanigan, & Seidenberg, 1980; Ventura, Kolinsky, Pattamadilok, & Morais, 2008; Ventura, Morais, Pattamadilok, & Kolinsky, 2004; Ziegler & Muneaux, 2007; Ziegler, Muneaux, & Grainger, 2003), auditory lexical decision (e.g., Chéreau, Gaskell, & Dumay, 2007; Ferrand & Grainger, 1992; Ziegler & Ferrand, 1998), and even speech production (e.g., Damian & Bowers, 2003; Roelofs, 2006). As reviewed above, knowledge of orthography also directly impacts phonological awareness; this can be seen when, for example, children have more difficulty saying that *kite* [kaɪt] has three phonemes than saying that *kit* [kɪt] does (Ehri & Wilce, 1980; Hallé, Chéreau, & Segui, 2000; Scholes, 1998). Similar effects of orthography can be found in syllable detection tasks, rhyme judgments, and in what answers are provided when individuals are asked to generate phonological neighbors for a word (e.g., Muneaux & Ziegler, 2004; Seidenberg & Tanenhaus, 1979; Taft & Hambly, 1985).

Finally, similar to phonological awareness, the morphophonological structure of a language impacts the characteristics of that language's orthography, and the characteristics of the specific orthography in turn impact the way that orthographic information is processed by native readers of that language (see Frost, 2012 for details). One demonstration of these cross-linguistic differences in orthographic processing comes from studies examining the transposed-letter effect, which is the phenomenon in which transposing letters within a word shows minimal disruption for lexical processing. For example, one of the original demonstrations of this effect showed that a form prime with a transposed pair of letters, such as anwser-ANSWER, results in similar levels of priming as true identity primes, such as answer-ANSWER (Forster, Davis, Schoknecht, & Carter, 1987). This pattern of results has been found in multiple languages including English (e.g., Perea & Lupker, 2003, 2004), French (e.g., Schoonbaert & Grainger, 2004), Spanish (Perea & Carreiras, 2006b, 2006c; Perea & Lupker, 2004), Basque (Duñabeitia, Perea, & Carreiras, 2007; Perea & Carreiras, 2006a), and even Japanese (Perea & Pérez, 2008). In contrast, transposed letters are noticeably detrimental for processing in Hebrew and Arabic, in which the relative position of (root) consonants in relation to one another has a meaningful impact on the core semantics of what is being conveyed by the form (e.g., Frost, 2012; Perea, Mallouh, & Carreiras, 2010; Velan & Frost, 2007, 2009).

3.1.3 The contributions of phonological and orthographic skills to broader literacy outcomes

In addition to impacting the level and type of skill that readers develop for phonological awareness and orthographic knowledge, the particular characteristics of an L1 writing system also influence the degree to which readers rely on one of these skill sets versus another in order to accomplish reading comprehension. One of the most well-known frameworks for understanding these differences is the Orthographic Depth Hypothesis (Frost, Katz, & Bentin, 1987; Katz & Feldman, 1983; Katz & Frost, 1992). According to this hypothesis, readers of shallow orthographies, particularly shallow alphabets, rely relatively more on decoding during word recognition and reading and access phonological information through assembly directly from the textual information on a page. In contrast, readers of deep alphabets rely relatively more on visual and orthographic information for word recognition and access phonological information after word identification, through the process of lexical access. A similar idea has been put forth by Ziegler and Goswami (2005) in their Psycholinguistic Grain Size Hypothesis. This proposes that all languages rely on phonological information during reading, but that the phonological unit(s) that are most relevant will differ among languages based on the size of the phonological unit represented in the orthography and the consistency of the mappings between orthography and phonology. Broadly speaking, readers of consistent orthographies can effectively rely on small units and unit-by-unit decoding, whereas readers of inconsistent orthographies (such as English) must develop sensitivity to multiple units of varying sizes and be able to use them flexibly.

Evidence for this viewpoint comes from a wide range of studies. For example, Frost, Katz, and Bentin (1987) found that L1 speakers of Hebrew, English, and Serbo-Croatian were differentially affected by word frequency, with readers of the deepest orthography (unpointed Hebrew) most impacted by word frequency for performance on the naming task. This suggests that the Hebrew readers were most reliant on lexical information for gaining access to the correct phonological form in order to perform the naming task. Similar results have been found across multiple other languages, showing greater use of phonological information and decoding during word identification by speakers of relatively transparent languages such as German, Greek, and Korean (e.g., Chitiri & Willows, 1994; Kang & Simpson, 1996; Näslund & Schneider, 1996). Even within English, there are individual differences regarding the use of strategies that readers use for reading (e.g., Baron, 1979; Bryant & Impey, 1986; Treiman, 1984; Treiman & Baron, 1981).

Further evidence comes from an examination of the interrelationships among literacy-related skills in readers of different languages. For example, oral reading accuracy generally does not correlate with reading comprehension for native readers of Arabic and Farsi, in contrast to the commonly found relationship for native readers of English (e.g., Abu-Rabia, 2001; Arab-Moghaddam & Sénéchal, 2001; Nassaji & Geva, 1999; Saiegh-Haddad, 2003). Similarly, Abu-Rabia (1997b) found that orthographic processing skill did not correlate with reading in children learning Hebrew, again in direct contrast to the pattern for English (e.g., Cunningham, Perry, & Stanovich, 2001). Finally, a within-language comparison using Japanese kanji (morphosyllabic) and kana (syllabic) reading found that phonological interference (implemented via concurrent vocalization) did not affect kanji reading in 7-year-old Japanese children, although phonological interference did affect kana reading and visual interference affected kanji reading (Kimura & Bryant, 1983).

As a whole, researchers generally agree that the relevant question is not whether readers rely on phonological or orthographic information for literacy tasks, but rather the extent to which they rely on these skills differentially for varying tasks across languages. Thus, the question is more one of a division of labor (Seidenberg, 1992) and the answer may depend on the language(s) involved (Frost, 2012) and the particular demands of the task being performed (Goswami, 2012).

A final point needs to be acknowledged regarding the contributions of sub-lexical literacy skills, such as phonological awareness and orthographic knowledge, to reading more generally. This is that the task of reading does not consist solely of word reading, but also requires a wide range of other text and discourse processing and integration skills (e.g., Anderson & Pearson, 1984; Perfetti & Lesgold, 1977; van Dijk & Kintsch, 1983). Sub-lexical and lexical skills are therefore a necessary, but not sufficient, condition for reading success (Johnston & Kirby, 2006). In addition to their contributions to the initial development of literacy reviewed above, understanding the roles that sub-lexical skills and word reading play in wider reading abilities requires consideration of both word reading accuracy and efficiency. Regarding accuracy, there is wide individual variation in word knowledge and in the quality of this knowledge, with parallel variability in reading comprehension (e.g., Perfetti, 2007; Perfetti & Stafura, 2013; Share, 1995). The two are connected via the importance of stable, high-quality lexical representations for accurate word identification (the Lexical Quality Hypothesis, Perfetti & Hart, 2001; Perfetti & Hart, 2002). If a reader cannot accurately identify the correct word, comprehension will suffer. In addition, word recognition must also be efficient. Especially in experienced readers, therefore, a key characteristic of lower-level skills is that they are automatized to allow for the reader's limited cognitive resources to be dedicated to higher-level processes and comprehension (the Verbal Efficiency Hypothesis, Perfetti, 1988; see also Verhoeven, 1999). The development of high-quality lexical representations and fluent lexical access therefore constitute a key connection between phonological awareness, orthographic knowledge, word and pseudoword decoding, isolated word recognition, and reading comprehension (Adams, 1990; Perfetti & Hart, 2002; Pikulski & Chard, 2005).

3.2 SECOND LANGUAGE RESEARCH ON LITERACY AND SUB-LEXICAL LITERACY SKILLS

3.2.1 Phonological awareness in L2 and L2 literacy

As reviewed above, there is substantial evidence that the type of phonological awareness that an individual develops is heavily influenced by the morphophonological and orthographic characteristics of their L1. An additional body of research has also established that the level of phonological awareness that an individual develops in their L1 generally transfers into their L2 and can constrain the level of their L2 phonological awareness. Particularly for language pairs that have similar morphophonological and orthographic structures, this transfer of L1 phonological awareness to the L2 can in fact support and facilitate their L2 literacy skills and performance (Bialystok et al., 2005; Comeau, Cormier, Grandmaison, & Lacroix, 1999; Gottardo, Yan, Siegel, & Wade-Woolley, 2001; McBride-Chang et al., 2008). For example, Durgunoğlu and colleagues (Durgunoğlu & Hancin, 1992; Durgunoğlu et al., 1993) found that the best predictors of literacy acquisition (reading outcomes as well as phonological awareness skills) in Spanish-speaking learners of English were their literacy skills in Spanish, and that these learners were able to transfer the phonological skills they had developed in their L1 and use them for their new task of reading in L2 English. Similarly, McBride-Chang et al. (2004) found that on a measure of English syllable awareness, native Chinese speakers with *pinyin* experience performed even better than the L1 English speakers, despite performing the task in their L2. This is likely because of the very strong syllable-level skills that native Chinese speakers develop due to the preeminence of the syllable in both spoken and written Chinese. This evidence for positive transfer between languages with similar structures for the task at hand is fully consistent with broader understandings of the impact that language distance has on the transfer of language processing strategies (e.g., the Competition Model, see Li & MacWhinney, 2012; MacWhinney, 2005, 2008, 2012).

On the other hand, if the level of phonological awareness that an individual develops in their L1 is at a different phonological unit size than is typically developed by native readers of their L2, this may hinder the development of L2 literacy skills and the use of appropriate processes for literacy tasks. Holm and Dodd (1996) found that native Chinese speakers from

Hong Kong, who had no experience with alphabetic writing systems, did significantly worse on measures of English phoneme and rhyme awareness than L1 Vietnamese speakers who came from an alphabetic background, as well as L1Chinese speakers from mainland China, who had experience with an alphabetic system by means of using *pinyin* (see also McBride-Chang et al., 2004). Similar results for Chinese speakers have been found in a number of studies comparing them to native speakers of alphabetic languages as diverse as English, Korean, and Bahasa Indonesia (Liow & Poon, 1998; Wang et al., 2003). Finally, Wade-Woolley (1999) found a similar advantage in L2 English phonemic awareness for L1 Russian speakers (an alphabetic L1), as compared to L1 Japanese speakers (a non-alphabetic L1).

In addition to impacting individuals' level of performance on phonological tasks themselves, L1 background also influences the degree to which an individual relies on phonological skills during literacy tasks. Consistent with L1 work, the general finding is that L2 learners who come from an alphabetic L1 tend to rely relatively more on phonological skills than orthographic skills during literacy tasks (e.g., McBride-Chang et al., 2005). For example, Koda (1990) found that compared to L1 Japanese speakers, L1 English, Spanish, and Arabic speakers were significantly hindered in their comprehension and recall when phonological information was not available in a text, demonstrating their reliance on this phonological information for reading. In another study, Koda (1998) found that L1 Korean ESL speakers showed strong relationships among phonological awareness, decoding, and reading comprehension. Although a comparable group of L1 Chinese ESL speakers did not differ in their level of performance on the phonological awareness and decoding tasks, they also showed no significant relationships among phonological awareness, decoding, and reading comprehension. A similar pattern of results was found by Wade-Woolley (1999). In her study, a group of L1 Japanese ESL learners and a group of L1 Russian ESL learners were carefully matched on English vocabulary, reading comprehension, and pseudoword decoding. However, despite their equivalent performance on these literacy outcome measures, the two groups showed differential relationships among reading comprehension, phonological awareness, and orthographic sensitivity, with stronger relationships between phonological awareness and reading comprehension for the L1 Russian speakers (and stronger relationships between orthographic sensitivity and reading comprehension for the L1 Japanese speakers). The complementary pattern has also been found, in which native readers of non-alphabetic languages (e.g., Chinese) continue to rely on visually-based strategies for reading in English rather than adopting a new, decoding-based strategy (e.g., Perfetti et al., 2007).

Although much of this research has been conducted with English as the target language, a small number of studies have also confirmed this pattern of results in learners of other target languages, such as Chinese and Japanese (e.g., Chitiri, Sun, Willows, & Taylor, 1992; Koda, 1989b). For example, Chikamatsu (1996) found that L2 learners of Japanese coming from L1 English backgrounds relied more on phonological coding for Japanese word recognition, while the L2 learners from L1 Chinese backgrounds relied more on visual coding. At least some studies have shown that although this may be true for early stages of literacy development in a new language, at least some adjustments to processing strategies are possible and some learners are able to become more native-like over time (e.g., Liu, Dunlap, Fiez, & Perfetti, 2007; Liu, Perfetti, & Wang, 2006)

3.2.2 Orthographic knowledge in L2 and L2 literacy

Similar to the case for phonological awareness, experience learning to read in a given language impacts the development of orthographic knowledge and relative reliance on this knowledge for broader literacy performance. These L1 orthographic skills can also transfer from L1 to L2 (e.g., Deacon, Wade-Woolley, & Kirby, 2009), and may thus affect literacy performance in the L2. At a basic level, a number of studies have demonstrated that L2 readers of English process letters and texts and perform tasks such as visual search in a fundamentally different way from native readers (e.g., Green, Liow, Tng, & Zielinski, 1996; Green & Meara, 1987; Ktori & Pitchford, 2008; Randall & Meara, 1988). Similar research has also shown that L2 readers of English are differentially sensitive to changes in the orthographic form. For example, L1 Arabic readers tend to have a relative lack of sensitivity to vowel letters compared to consonant letters, which is manifested in more difficulty detecting missing vowels than consonants (Hayes-Harb, 2006; Ryan & Meara, 1991, 1996).

Other work has shown differential sensitivity to sub-lexical orthographic information, such as letter frequency and letter sequence legality, in readers from different L1s. For example, Haynes and Carr (1990) found that L1 Chinese learners of English performed significantly worse than L1 English speakers on a pseudoword (but not a real word) visual matching task. They interpreted this difference in performance to the L1 Chinese speakers' their greater reliance on whole visual forms for processing real words and their consequent decreased attention to component letters and letter combination constraints which adversely affected their performance for the pseudowords. Koda (1999) also investigated L2 English speakers' sensitivity to intraword lexical information, specifically positional bigram frequency and letter sequence legality. She found that L1 Chinese speakers were less successful than L1 Korean speakers at judging whether a non-word stimulus could be a possible English word, demonstrated reduced sensitivity to the intralexical characteristics of the items. In a very similar task, Fender (2003) found that L1 Arabic speakers performed worse than L1 Japanese speakers on judging unpronounceable non-words. Although the Arabic speakers had the benefit of a segmental system, Fender attributed their difficulties to a relative lack of sensitivity to vowels due to the orthographic characteristics of their L1 writing system (see also K. I. Martin & Juffs, In revision).

A related body of work has shown that L2 readers from non-alphabetic or non-segmental languages tend to rely relatively more on whole-word visual information and overall orthographic shapes for a wide range of linguistic tasks including visual matching, word recognition, semantic judgments, spelling, and reading comprehension. Akamatsu (1999, 2003) examined L2 English readers' sensitivity to aLtErNaTiNg cAsE for both isolated word recognition and extended text comprehension. He found that ESL readers with a non-segmental L1 (Chinese, Japanese) were more adversely affected by alternating case than those from a segmental L1 (Persian), and interpreted this as evidence that the non-segmental L1 readers depended more on the specific orthographic form for word recognition and were less able to process the component letters within a word.

In contrast to the research demonstrating reduced sensitivity to sub-lexical orthographic information in ESL readers from non-segmental or non-alphabetic languages, a different body of work suggests that these individuals rely more on whole-word orthographic and visual information for reading and show less sensitivity to characteristics of the component graphemes within words. Evidence for this pattern comes from studies of word naming, semantic categorization, phoneme deletion, and spelling. Wang and colleagues (Wang & Koda, 2005;

Wang et al., 2003) compared L1 Chinese and L1 Korean learners of English and found that the L1 Korean speakers had more difficulty with homophone distractor items than the L1 Chinese speakers and had more difficulty correctly pronouncing English words with inconsistent and exceptional grapheme-phoneme correspondences. On the other hand, the L1 Chinese speakers made more orthographically plausible errors in phoneme deletion than the L1 Korean speakers while performing relatively better on naming words with inconsistent and exceptional grapheme-phoneme correspondences. A similar detriment for processing irregular grapheme-phoneme correspondences was found by Hamada and Koda (2008) for both pseudoword decoding and retention. Looking at spelling, Wang and Geva (2003) found that L1 Chinese speakers were significantly worse than L1 English speakers at spelling pseudowords from dictation (although they did relatively better in a forced-choice confrontation spelling task).

3.3 INTERVENTION RESEARCH IN LITERACY

In addition to trying to understand the developmental underpinnings of various literacy skills, a large portion of the research on literacy development has focused on the trainability of sub-lexical literacy skills and the subsequent impact on broader literacy abilities. A number of studies of L1 reading have demonstrated the trainability of phonological awareness in children and adults, and with a variety of different tasks (e.g., Ball & Blachman, 1991; Content, Morais, Alegria, & Bertelson, 1982; Gaskins et al., 1988; Kyle et al., 2013; Lovett, Lacerenza, & Borden, 2000; Lovett, Lacerenza, Borden, et al., 2000; Morais, Content, Bertelson, Cary, & Kolinsky, 1988; Tunmer & Hoover, 1993). This training leads to improvements in word, non-word, and prose reading as well as spelling (Bradley & Bryant, 1983, 1985, 1991; Lundberg et al., 1988;

Wallach & Wallach, 1976; Williams, 1980). The impact of instruction on literacy can be seen particularly well in comparisons of learners who have been taught exclusively with a whole-word method versus those who have received some training in phonics (e.g., Alegria, Pignot, & Morais, 1982; Ehri et al., 2001). Although not uncontroversial, the most widely accepted scientific position is that phonics-based literacy instruction methods provide a reliable advantage in reading instruction in children learning to read in English as their native language, as summarized by the report from the National Reading Panel (NRP, 2000).

An interest in intervention and training research is also applicable to L2 contexts. As has been seen in previous research, L2 readers do not necessarily or automatically develop the most appropriate text processing strategies in their L2, especially without the benefit of direct instruction and guided practice (e.g., Birch, 2015; Jones, 1996; Koda, 2004). As observed by Holm and Dodd's (1996) research with highly proficient L1 Chinese readers of English without *pinyin* experience, strong phonological awareness skills are not strictly necessary for the acquisition of English literacy. However, the task of decoding new words is made substantially easier with the development of phonological skills. In addition, the cognitive maturity of adult learners makes them particularly well suited to take advantage of phonics-based interventions that include a meta-analytic component. Knowledge of English spellings is helpful not only for being able to sound out unfamiliar words, but also for understanding morphological relationships among words and being able to make inferences about their meanings (Chomsky, 1970, 1971, 1976; Jones, 1996).

Unfortunately, however, very few studies have tested the effectiveness of any kind of instruction or intervention for improving ESL students' literacy skills and outcomes, particularly at the adult level. Although this research need has been noted by many authors, little work has been done in this area (Adesope, Lavin, Thompson, & Ungerleider, 2011; Jones, 1996). A recent meta-analysis of pedagogical strategies for primary-school ESL students (Adesope et al., 2011) was able to analyze data from only 26 total studies, from which only 14 effect sizes were available for estimating the efficacy of phonics instruction in ESL learners. These authors additionally reported that they had to constrain their analysis to consider only students in grades K-6 because they were only able to locate three studies involving students in grades 7-12. Despite the relatively small number of effect sizes available for analysis, Adesope et al. found that phonics instruction had a significant positive effect on reading and writing outcome measures, and that it was in fact one of the three most effective instructional strategies examined in the analysis. This finding is consistent with additional, more recent studies, which continue to show strong, positive effects of phonics instruction on the development of ESL literacy skills in children enrolled in primary school (Nishanimut, Johnston, Joshi, Thomas, & Padakannaya, 2013).

In one of the few studies testing an intervention with adult ESL learners, Taylor (M. Taylor, 2008) tested the effectiveness of 16 weeks of web-based direct explicit phonics instruction for increasing English phonological and orthographic awareness in adult female L1 Arabic speakers. She found no overall effect of the training on either skill, although there was a small increase in the number of students who scored at or above a predetermined minimum level on the phonological awareness task. Despite the lack of research in this area, the success of phonics-based instructional training and intervention for L1 readers suggests that it is a worthwhile area for future research. As stated by Jones (1996, p. 3), "The question should no longer be whether to teach phonics as part of adult ESL instruction, but how this might be done most effectively."

4.0 OVERVIEW OF THE CURRENT RESEARCH

A large body of research has already established the importance of phonological awareness and orthographic knowledge for literacy development, both in L1 and L2, as well as the impacts of L1 experiences on sub-lexical literacy skills and broader literacy outcomes. However, there are still many gaps in our understanding of the cross-linguistic details of these skills, and the current research was designed to begin addressing these areas. Data for Study 1 and Study 2 were collected in a controlled, individual testing environment, allowing for maximum control over experimental conditions and minimal outside influence. These data were also collected abroad, in the home countries of the non-native English participants, in order to help reduce variability in English language experience and contact with native speakers (Kurvers, Stockmann, & van de Craats, 2010; Meisel, Clahsen, & Pienemann, 1981). Data for Study 3 were collected from adult ESL students in a classroom environment, thus reflecting the realities of immersion language learning. The inclusion of both types of data provide the current research with broad applicability, methodological strength, and external validity by approaching from multiple perspectives the issues of how linguistic background, language structure, and task demands jointly influence performance on a variety of L2 literacy skills.

4.1 STUDY 1

The goal of Study 1 was to measure L2 phonological awareness and L2 orthographic knowledge in ESL learners from three L1 writing systems: an alphabet (French), an abjad (Hebrew), and a morphosyllabary (Chinese). These languages were chosen to be representative of the continuum of possible writing systems, yet also to provide new insight on ESL literacy skills in speakers of these types of languages. French is a relatively deep alphabetic orthography, and although some literacy research has been conducted using French (Comeau et al., 1999; Goswami et al., 1998), much of the work looking at phonological awareness in non-English alphabets has examined speakers of shallow alphabetic systems such as Greek, German, Spanish, and Korean (Anthony et al., 2011; Chitiri et al., 1992; Chitiri & Willows, 1994; Durgunoğlu et al., 1993; Goswami et al., 2003; McBride-Chang et al., 2004; McBride-Chang et al., 2005; Näslund & Schneider, 1996). In contrast to the extensive work examining phonological skills in alphabetic language readers, as well as in morphosyllabary readers, much less work has focused on abjad languages. These orthographies have characteristics similar to both alphabetic and non-alphabetic writing systems, and our understanding of the literacy skills and reading behaviors of these readers is still relatively limited. Finally, a large and growing body of work has investigated the roles of phonology, orthography, and morphology in literacy development for L1 Chinese speakers, much of it conducted by Catherine McBride-Chang and her colleagues. However, much of this work has focused on comparisons of Chinese speakers from Hong Kong and mainland China, as well as heritage speakers in Canada, rather than L1 Chinese speakers from Taiwan. The Chinese speakers from Taiwan are a unique and interesting case because they receive some phonological training via the use of a syllabary-type phonological system, called *zhuyin fuhao*, during literacy instruction (Wang & Yang, 2008). However, unlike *pinyin* used in mainland China, this is not a true alphabetic system and thus may provide some support for the development of phonological awareness without providing direct support for phonemic awareness. Targeted work with this population is needed to understand the role that experience with *zhuyin fuhao* has in literacy skills development.

In addition to targeting learners from these three representative L1 backgrounds, and comparing their performance to that of L1 English speakers, another goal of Study 1 was to provide a more fine-grained analysis of phonological awareness skill. Words can be divided into component parts in multiple ways. Syllables, phonemes, and onset/rime divisions are relatively familiar to native English speakers, but research with other languages has suggested that another possible way to divide words is into body/coda units. In this type of division, the nucleus of a syllable is associated with the preceding, rather than the following, consonant (CV | C rather than C | VC) (Ben-Dror, Frost, & Bentin, 1995; Kim, 2007, 2009; Tolchinsky et al., 2012; Yoon et al., 2002). Thus in Study 1 we tested L1 English, L1 French, L1 Hebrew, and L1 Chinese speakers' phonological awareness for a comprehensive range of phonological units: first and second syllables within disyllabic words; onsets and rimes; bodies and codas; and initial, final, and word-medial single phoneme segments.

We predicted that English phonological awareness would be greatest for the phonological unit that corresponded to the representational unit of the L1 orthography for each group. In other words, we predicted that the preferred phonological grain size for L1 reading (Ziegler & Goswami, 2005) would transfer to the L2. Specifically, we predicted that the alphabetic L1 (French) speakers would perform better than the non-alphabetic L1 (Hebrew, Chinese) speakers at the phoneme level because alphabetic L1 literacy supports the development of phonemic awareness (Goswami & Bryant, 1990; Homer, 2009). We also expected that the morphosyllabic

L1 (Chinese) speakers would perform better than the abjad L1 (Hebrew) speakers, but only if their experience with *zhuyin fuhao* is able to support fine-grained phonological processing (Holm & Dodd, 1996). We predicted that the abjad L1 (Hebrew) speakers would perform better than the non-abjad L1 (French, English, Chinese) speakers at the body/coda divisions because the consonant-based nature of an abjad, both orthographically and morphologically, emphasizes CV as an important sub-lexical unit (Ben-Dror et al., 1995; Frost, 2012; Tolchinsky et al., 2012). We predicted that the alphabetic L1 (French, English) speakers would perform better than the non-alphabetic speakers at the onset/rime divisions due to the important of both the phonological and orthographic rime in (deep) alphabetic languages and the prevalence of phonological neighbors at the rime level in these languages (De Cara & Goswami, 2003; Ziegler & Goswami, 2005). Finally, we predicted that all L1 groups would have high performance at the syllable level because it is a functional unit in all three languages and because syllable awareness develops prior to literacy acquisition (Goswami & Bryant, 1990). However, we predicted the highest performance on syllables for the morphosyllabic L1 (Chinese) speakers because the syllable also maps to the unit of meaning in this orthography (McBride-Chang et al., 2004; Tolchinsky et al., 2012).

The final goal of Study 1 was to evaluate the effect of task on participants' level of performance. Specifically, different language skills may be needed to perform productive versus receptive phonological awareness tasks, and different types of orthographic knowledge may be needed to have good whole-word versus intralexical knowledge. Therefore we included two different tasks of phonological awareness and orthographic knowledge so that we could examine patterns of performance across them. For the orthographic knowledge tasks, we predicted that the morphosyllabic L1 (Chinese) speakers would perform best on the task targeting whole-word orthographic knowledge, followed by the alphabetic L1 (French, English) and the abjad L1 (Hebrew) speakers. We expected the morphosyllabic L1 speakers to do best because they are accustomed to using whole-word orthographic images in L1 reading, which may transfer to L2 and support better L2 whole-word spelling knowledge (Wang & Geva, 2003). We expected that the alphabetic L1 (French, English) speakers would be better than the abjad L1 (Hebrew) speakers because previous works suggests that abjad speakers are relatively less sensitive to written vowel information (K. I. Martin & Juffs, In revision; Ryan & Meara, 1991) and vowels are a key component of English pseudohomophones due to the large inventory of vowel phonemes but small number of vowel letters (Share, 2008).

Finally, for the task targeting intralexical orthographic knowledge, we predicted that the alphabetic L1 (French, English) speakers would perform best followed by the abjad L1 (Hebrew) and then the morphosyllabic L1 (Chinese) speakers. We expected the alphabetic L1 speakers to perform best because their L1 alphabetic experience should make them the most adept at processing component letters inside a word (Koda, 1999). We expected that the abjad L1 (Hebrew) speakers would perform better than the morphosyllabic L1 (Chinese) speakers because their L1 orthography is still relatively segmental, again encouraging greater intraword segmental attention (Akamatsu, 1999, 2003).

4.2 STUDY 2: CROSS-LINGUISTIC CONTRIBUTIONS OF PHONOLOGICAL AWARENESS AND ORTHOGRAPHIC KNOWLEDGE TO WORD IDENTIFICATION

The goal of Study 2 was to extend the examination of L1 influence and task effects to the next hierarchical level of literacy skill: word identification. Specifically, we were interested in comparisons among the four L1 groups regarding the relative contributions of phonological awareness and orthographic knowledge to word identification and broader reading comprehension in English. To examine these comparisons we used composite scores from the phonological awareness and orthographic knowledge data collected in Study 1 as predictors of three measures of word identification: lexical decision, word naming, and pseudoword decoding, as well as a measure of global reading comprehension.

The choice of these literacy outcomes was also strategic and was driven by our interest in examining the influence of task demands on performance outcomes at the lexical level, as well. Although both phonological and visual-orthographic strategies can be used in lexical decision and word naming, previous research comparing these tasks supports the idea that differences between oral and non-oral reading skill can be at least partially tapped through comparisons of these two tasks (Share, 2008). These tasks do show differences in terms of the relative impact of phonology and frequency. For example, regularity effects are consistently found in naming (Andrews, 1982; Baron & Strawson, 1976; Weekes, 1997), but not lexical decision (Seidenberg & McClelland, 1989; Seidenberg, Waters, Barnes, & Tanenhaus, 1984). On the other hand, frequency effects are more robust in lexical decision than in naming (Balota & Chumbley, 1984; Forster & Chambers, 1973; Frederiksen & Kroll, 1976; McCann & Besner, 1987; D. L. Scarborough, Cortese, & Scarborough, 1977).

Given the importance of both phonological and orthographic skills for literacy, we predicted that both phonological awareness and orthographic knowledge composite scores would make significant, independent contributions to each word identification task and to global reading comprehension. However, we also expected that phonological awareness and orthographic would have different strengths of relationship with these literacy outcomes. Specifically, we predicted that phonological awareness would be more strongly related to decoding for the alphabetic L1 (French, English) speakers, but that orthographic knowledge would be more strongly related to decoding for the morphosyllabic L1 (Chinese) and abjad L1 (Hebrew) speakers, based on speakers' word identification strategies in L1 (McBride-Chang et al., 2005; Nassaji & Geva, 1999; Wade-Woolley, 1999). Additionally, based on the different task demands of lexical decision and word naming, we expected to see stronger relationships between phonology and word naming (which requires the use of phonological information) than between phonology and lexical decision (which does not require this phonological involvement).

4.3 STUDY 3: PHONOLOGICAL AWARENESS AND ORTHOGRAPHIC KNOWLEDGE AND THEIR DEVELOPMENT IN ESL CLASSROOM STUDENTS

The goal of Study 3 was to extend this body of work to real learners, and to bring the use of these types of assessments from a laboratory setting to the classroom environment. One of the motivations for this study was an interest in evaluating whether it was effective to use the same types of literacy skill measures in a group setting, similar to what would be available for language instructors, rather than the traditional individual assessment. We were again interested in examining potential differences in literacy skills among different groups of L1 learners, in this case based on the enrollment of ESL students in an intensive English program at a large university in the United States. The broader aim of this study, though, was to examine any potential development of literacy skills that occurs in classroom learners over the course of a semester of intensive English instruction. If there is evidence that these skills do change, this would provide an avenue for English instructors to help develop their students' literacy abilities. We also designed and tested the effectiveness of a phonics-based intervention, in order to evaluate whether this type of instruction can boost these crucial literacy skills.

5.0 STUDY 1: LABORATORY-BASED INVESTIGATION OF THE IMPACTS OF L1 ON L2 PHONOLOGICAL AWARENESS AND ORTHOGRAPHIC KNOWLEDGE

As detailed above, the goal of Study 1 was to provide a detailed cross-linguistic comparison of performance on L2 phonological awareness and orthographic knowledge tasks in ESL learners from representative L1 writing system backgrounds.

5.1 METHOD

5.1.1 Participants

A total of 46 native French speakers, 46 native Hebrew speakers, 46 native Chinese speakers, and 46 native English speakers were included in the final analyses for this study. An additional 14 French speakers, 21 Hebrew speakers, 27 Chinese speakers, and 14 English speakers were excluded from analyses for a variety of reasons. These reasons included: participants did not complete all tasks (9; 3 French, 4 Hebrew, 2 English), scored 65 or below on the reading comprehension test (see below) (4; 1 French, 1 Hebrew, 1 Chinese, 1 English), self-reported a neurological condition (1 French), listed an L1 other than French, Hebrew, Chinese, or English as appropriate to their language group (8; 3 French, 5 Hebrew), indicated that they had spoken another language at home with their parents from birth (8; 3 French, 3 Hebrew, 1 Chinese, 1

English), grew up in an environment with a majority language different from their L1 (2; 1 Chinese, 1 English), had extensive reading experience in a non-English language with a different writing system from their L1 (4; 2 Hebrew, 2 English), were over the age of 60 (1 French), were directly observed by the experimenter to be extensively distracted throughout the study procedures (4; 1 French, 1 Hebrew, 2 English), or due to experimenter error or equipment failure (3 English). A further 33 participants (1 French, 5 Hebrew, 24 Chinese, 2 English) were excluded in order to match sample sizes across the four groups and English proficiency levels among the non-native speakers.

All participants who were L2 speakers of English had studied English throughout their primary and secondary schooling. Native French participants were recruited at the Université François-Rabelais in Tours, France. Native Hebrew participants were recruited at the University of Haifa in Haifa, Israel. Native Chinese participants were recruited at National Central University in Zhongli City, Taiwan. The comparison group of native English speakers comprised undergraduate students at the University of Pittsburgh. Demographic information about these participants, including their age, years studying English, and the average age at which they began studying English/their L2 can be found in Table 1. This table also includes participants' self-rated proficiency in both L1 and L2 skills, as well as scores on the objective measures of English proficiency used to match the non-native samples (details below).

5.1.2 Materials and Procedure

Real word stimuli for all tasks were chosen from the 5,000 most frequent words of English (Davies, 2008-) in order to increase the likelihood that non-native participants would be familiar with them. All word and pseudoword stimuli were additionally checked to ensure that they were not cognates and did not share a substantial amount of form overlap with any other words in

Table 1. Demographic and language proficiency information for participants in Study 1 and Study 2.

	L1 English	L1 French	L1 Hebrew	L1 Chinese
Age	18.43 (0.89)	24.96 (9.13)	27.70 (5.60)	22.07 (2.79)
Number of years studying English	-- [a]	11.77 (3.33)	12.75 (4.83)	12.87 (3.10)
% of time speaking English	-- [a]	10.54 (12.86)	10.19 (11.19)	7.70 (9.15)
Self-rated English reading proficiency *(1 = not literate, 10 = very literate)*	-- [a]	6.60 (1.61)	7.60 (1.33)	6.67 (1.45)
Self-rated English writing proficiency *(1 = not literate, 10 = very literate)*	-- [a]	5.94 (1.49)	6.67 (1.71)	5.50 (1.60)
Self-rated English conversational proficiency *(1 = not fluent, 10 = very fluent)*	-- [a]	6.15 (1.62)	7.07 (2.02)	5.70 (1.66)
Self-rated English spoken language comprehension *(1 = no comprehension, 10 = perfect comprehension)*	-- [a]	6.90 (1.61)	7.33 (2.09)	5.76 (1.74)
Reading comprehension[b]	98.85 (11.66)	101.28 (13.48)	105.98 (12.80)	105.20 (10.46)
Listening comprehension[c]	--	21.07 (5.17)	22.46 (3.82)	22.07 (3.16)
Vocabulary levels test (productive)[d]	58.96 (6.08)	21.57 (15.02)	21.50 (15.44)	21.33 (9.50)
Vocabulary size test (receptive)[e]	59.38 (14.20)	63.33 (10.27)	53.54 (11.67)	45.76 (7.46)
Vocabulary size test – cognates removed[d]	22.28 (2.18)	13.52 (4.24)	13.04 (3.69)	12.74 (3.08)

Note. [a] The L1 English speakers were not asked this question. [b] The maximum possible score for this test was 120. [c] The maximum possible score for this test was 30. [d] The maximum possible score for this test was 72. [e] The maximum possible score for this test was 100. [d] The maximum possible score for this test with cognates removed was 26.

French, Hebrew, and Mandarin. All stimuli were presented on a PC laptop using E-Prime 2.0 software (Walt Schneider, Schman, & Zuccolutto, 2002) and responses were collected using a serial-response button box and a digital voice recorder. Spoken stimuli were recorded by a female native English speaker from northern California who used a precise enunciation (as opposed to casual speech). Oddity items and word-pseudohomophone stimuli were recorded as single words in the sentence frame "Say [word] now" and later spliced together to eliminate list intonation in the recordings. Deletion prompts were recorded as whole items, each following the form "Say [word] without saying [phonological unit]". Recordings for the auditory stimuli were made with a Marantz solid-state recorder at 16 bits per sample with a sample rate of 48,000 Hz saved into an uncompressed .wav file.

5.1.3 Phonological Awareness

Participants completed two tasks measuring their phonological awareness: oddity (receptive phonological awareness) and deletion (productive phonological awareness).

5.1.4 Oddity Task

The oddity task was used as a receptive measure of phonological awareness. In this task, participants heard three spoken words, two of which shared a phonological unit at the same position in the word. To expand on previous research, which has often tested only phoneme or onset/rime awareness, a wide range of phonological units was tested: first syllable, second syllable, onset, rime (VC), body (CV), coda, and single phonemes word-initially, word-medially, and word-finally. For word-medial phonemes both consonants and vowels were tested, although only consonants were tested word-initially and word-finally so that complex onsets and codas could be used to distinguish between onset trials and initial-phoneme trials and between coda trials and final-phoneme trials. Specifically, in much previous research, onsets and codas have been confounded with word-initial and word-final single phonemes; for example, in *cat* the /k/ sound is both an onset and a word-initial single phoneme. In order to avoid this problem in the current study, all stimuli testing onsets, codas, or word-initial or word-final phonemes contained complex onsets and codas (CC). This allowed for the differentiation of stimuli that shared a whole onset or coda (CC) from stimuli that were distinguished by sharing only a single word-initial (Cx) or word-final (xC) phoneme. For example, in the triplet *break-bring-speak*, *break* and *bring* share the first two phonemes (onset), both of which differ from the first two phonemes (onset) of the oddity item *speak*. In contrast, in the triplet *brain-breath-dress*, all three items share the second phoneme /ɹ/ although *brain* and *breath* are distinguished from the oddity item *dress* by just the first phoneme /b/. This strategy has been used successfully in some previous research in order to differentiate these types of phonological units (see for example Goswami & East, 2000).

Trials were constructed such that two of the three items shared a phonological unit, while the third (oddity item) also shared one sound with one of the two non-oddity items at a different location in the word. For example, in the trial with *shape-tape-shoot* the first two items share a rime and thus the third is the oddity item. However, this word still shares one sound (/ʃ/) with the first word, *shape*. Roughly equal numbers of these sounds shared with distractors were consonants and vowels. This manipulation is similar to distractor manipulations used with other phonological judgment tasks (e.g., Lenel & Cantor, 1981) and was included in order to increase the difficulty of the test to make sure that it was appropriate for use with adult speakers. However, this manipulation did require the participants to be advised what part of the word they should pay attention to. Therefore the screen always told participants whether they should indicate the word that was different at the beginning, different at the middle, or different at the end of the word, but participants were never told which specific phonological unit was being tested. Items were presented in these groupings; thus, participants first completed all trials in which they were to pay attention to the beginning of the word (initial phoneme, onset, body, or first syllable), followed by a short break, then all trials in which they were to pay attention to the middle of the word (medial consonant or medial vowel), followed by another short break, and finally all trials in which they were to pay attention to the end of the word (final phoneme, coda, rime, or second syllable). Each group of items began with a set of 4-5 practice trials with feedback. The order of trials within each group was randomized for each participant and no feedback was given during the test trials.

Stimuli were presented to participants via noise-cancelling headphones. After hearing the three words, participants indicated which item they thought was different from the other two by pressing a button labeled ‘1’, ‘2’, or ‘3’ on the response box. Participants were given up to 7500 ms to respond after the end of the presentation of the third stimulus item. There were eight trials for each phonological unit, with an interstimulus interval of 1100 ms between a participant’s button-press response and the beginning of the next trial. A full list of stimuli can be found in Appendix A. The outcome variables were accuracy and reaction time, measured from the end of the presentation of the third stimulus item. Reaction times were analyzed for correct trials only, and RTs more than three standard deviations from an individual participant’s mean RT were excluded from consideration. This resulted in the exclusion of 5.5% of the data.

Note that although some previous studies have used the oddity task with four items rather than three we used only three items to reduce the memory load requirement (see Schatschneider, Francis, Foorman, Fletcher, & Mehta, 1999 for a critique of the memory load requirements in four-item oddity tasks). With just one exception, all three words in a trial were from either the same or consecutive 1000-word frequency bands (e.g., all three words were among the 1,000 most frequent, or two of them were among the 1,000 most frequent and the third was in the 2000 most frequent) so that all stimuli in a trial were of similar familiarity to participants.

5.1.5 Deletion Task

The deletion task was used as a productive measure of phonological awareness. In this task participants heard an item and then were asked to say that same item with a specific part of it removed, for example, 'Say *wildlife* without saying *wild*'. Participants were given up to 15000 ms to respond after the end of the prompt and the interstimulus interval was 1100 ms between a participant's button-press to indicate they were done providing their answer and the beginning of the next trial. As in the deletion task a wide range of phonological units with varying sizes and locations were tested: first syllable, second syllable, onset, rime (VC), body (CV), coda, and single phonemes word-initially, word-medially, and word-finally. Unlike the oddity task, only medial consonants (and not medial vowels) were targeted for deletion, due to restrictions on pronounceable syllables in English. Items that required the deletion of an onset, coda, or word-initial or word-final phoneme again contained complex onsets and codas to differentiate these phonological units.

Half of the stimuli for each phonological unit were real words that required the deletion of a subunit that also comprised a word, such as *mike* without *my* (correct answer: /k/), or that formed another real word when the correct unit was deleted, for example, *bold* without /b/ (correct answer: *old*). The other half were pseudowords that required the deletion of a subunit that also comprised pseudoword, for example /zaɪg/ without /zaɪ/ (correct answer: /g/), or that formed another pseudoword when the correct unit was deleted, for example, /vaɪt/ without /v/

(correct answer: /aɪt/). Due to the productive nature of participant responses for this task, items were chosen to minimize the inclusion of lax vowels and other sounds or sound combinations that would be particularly problematic for native speakers of French, Hebrew, or Chinese (Nilsen & Nilsen, 2010). No words from the oddity task were used in the deletion task.

Stimuli were presented to participants via noise-cancelling headphones. After hearing each item, participants spoke their answer aloud while a digital voice recorder recorded their response. There were eight trials for each phonological unit, half of which were words and half of which were pseudowords. Participants began with five practice trials, during which they heard the trial, were given an opportunity to answer, and then heard the correct answer. Following the practice, participants completed all trials testing first or second syllables, followed by a short break, then all trials testing onsets, rimes, bodies, or codas, followed by another short break, and finally all trials testing single phonemes. Stimuli were grouped and ordered in this way based on previous research demonstrating that if participants begin with the most difficult (generally the single phoneme) items in a deletion task, they may struggle to understand the task and perform at floor (Anthony et al., 2011). Within the three groups items were randomized for each participant and no feedback was given during the test trials. A full list of stimuli can be found in Appendix B. The outcome variables were participant accuracy and time on item. A strict accuracy score was calculated for each participant. In order to receive credit for an item, participants had to perform the deletion correctly and provide the correct answer with no allowance for mispronunciation. Response time was measured as the total amount of time that participants remained on each item's presentation screen after hearing the prompt and before pressing the space bar to indicate that they had finished providing their spoken response. Response time was analyzed for correct trials only, and times more than three standard deviations from an individual participant's mean time on item were excluded from consideration. This resulted in the exclusion of 4.62% of the data.

5.1.6 Orthographic Knowledge

Participants also completed two tasks measuring their orthographic knowledge: a word/pseudohomophone discrimination task and a wordlikeness judgment task.

5.1.7 Word/pseudohomophone discrimination

The word-pseudohomophone task was used as a measure of participants' whole-word orthographic (spelling) knowledge (Berninger & Abbott, 1994). In this task participants heard a spoken English word and saw two possible spellings, one presented on the left side of the computer screen and the other presented on the right. Participants were asked to press a button on the serial response box as quickly as they could to indicate which was the correct spelling of the word that they heard. For example, participants heard /kloʊd/ and saw 'cloud' and 'kloud' on the screen; in this case, the correct answer was 'cloud' so participants should have pressed the far left button on the response box. Participants were given up to 7500 ms to respond after the end of the auditory presentation of the item and the interstimulus interval was 1200 ms between a participant's button-press response and the beginning of the next trial. There were a total of 80 monosyllabic items taken from a variety of previous studies using this task: Borowsky, Owen, and Masson (2002); Davis (1998); Laxon, Masterson, Pool, and Keating (1992); Lukatela and Turvey (1991); Lukatela, Eaton, Lee, Carello, and Turvey (2002); Lupker and Pexman (2010); Manis, Seidenberg, Doi, McBride-Chang, and Keating (1996); R. C. Martin (1982); Rastle and

Coltheart (1999); Reynolds and Besner (2005); Seidenberg, Petersen, MacDonald, and Plaut (1996); Taft and Russell (1992); and Yates, Locker, and Simpson (2003).

Half of the items (40) had misspellings that primarily involved a consonant (added, deleted, or misspelled) and the other half (40) involved misspellings exclusively affecting the vowel. Using data available from the E-Lexicon (Balota et al., 2007), all real word (correct answer) items were matched across the consonant and vowel trials on word length, frequency, age of acquisition, concreteness, imageability, orthographic and phonological neighborhood sizes and the word frequencies of orthographic and phonological neighbors, bigram sum and mean, number of phonemes, and number of morphemes, *p*s > .10. All pseudohomophones (distractors) were also matched across consonant and vowel trials on length, orthographic neighborhood size, and bigram sum and mean, *p*s > .15. Thus, any differences in performance on items involving misspelling of a consonant versus misspelling of a vowel phoneme could not be due to differences in these lexical characteristics.

Participants completed six practice trials with feedback, followed by all test trials with a short break at the halfway point. The side on which the correct answer appeared was counter-balanced across stimuli and the order of trials was randomized for each participant. A full list of stimuli can be found in Appendix C. The outcome variables were participant accuracy and reaction times. Reaction times were analyzed for correct trials only, and RTs more than three standard deviations from an individual participant's mean RT were excluded from consideration. This procedure resulted in the exclusion of 3.90% of the data.

5.1.8 Wordlikeness judgments

The wordlikeness judgment task was used as a measure of participants intralexical orthographic knowledge, specifically, their sensitivity to positional bigram frequencies of letters and legality of letter strings. In this task participants saw two pseudoword and/or nonword letter strings, one on the left of the computer screen and the other on the right, and had to choose which one they thought looked more like an English word. Participants were advised that none of the items would be real words, and that they should simply choose the word that they thought looked more like it *could be* an English word. Participants were given up to 7500 ms to provide their answer and were encouraged to compare the two words before responding. The interstimulus interval was 1000 ms between a participant's button-press response and the beginning of the next trial.

All stimuli were six letters long and were taken from Massaro, Venezky, and Taylor (1979) and also used by Koda (1999). The items were created by Massaro et al. by crossing two lexical characteristics: letter combination legality and positional frequency. Four types of stimuli were thus used: legal pseudowords with high positional frequencies, legal pseudowords with low positional frequencies, illegal nonwords with high positional frequencies, and illegal nonwords with low positional frequencies. Twenty trials each were created for all six possible pairwise comparisons between these four item types, for a total of 120 trials, with no single letter string occurring in more than one trial. Participants completed six practice trials without feedback, followed by all test trials with a short break at the halfway point. A full list of stimuli can be found in Appendix D. The outcome variables were again participant accuracy and reaction times. Reaction times were analyzed for correct trials only, and RTs more than three standard deviations from an individual participant's mean RT were excluded from consideration. This resulted in the exclusion of 1.85% of the data.

5.1.9 Cognitive abilities

Participants completed three tasks measuring their general cognitive abilities: operation span, measuring working memory; flankers, measuring executive control, and rapid digit naming, measuring their overall processing speed. These tasks were completed in the participants' native language so that performance was not influenced by English proficiency and were included so that the samples of participants from different L1 backgrounds could either be matched on cognitive abilities or their cognitive abilities could be included as covariates for all analyses.

5.1.10 Working memory

The operation span task (Turner & Engle, 1989), participants saw an arithmetic operation, such as (12/3) – 2 = 2, for 2500 ms. After the operation disappeared, participants had 5000 ms to press a 'Yes' or 'No' button on a serial response box to indicate whether the answer provided for the operation was correct or incorrect. Following this, they saw a single concrete noun in their native language, such as 'uncle', for 1250 ms. Participants were told to remember the word. They then saw another arithmetic operation for 2500 ms, followed by another correctness judgment and the presentation of a different concrete noun. This procedure continued for a set number of iterations, after which participants were asked to recall all the words and either typed (French and English) or wrote down on a piece of paper (Hebrew and Chinese) as many of the words from the list as they could remember. Participants were given as much time as needed to recall the words. After two practice trials, participants completed three sets each with two to six interleaved operations and words. Participants' scores were the maximum set size at which participants correctly recalled all words for at least two of the three trials, before failing to get at least two full trials correct. All participants saw the same arithmetic operations; the words presented to each language group appeared in the participants' L1 and the particular translations chosen were based on the advice of native speakers of each language. A full list of the English words and their translations in French, Hebrew, and Chinese are provided in Appendix E.

5.1.11 Executive control

The flankers task (Eriksen, 1995) was used to measure executive control. Participants saw a linear sequence of five arrows, of which the first two and the last two always pointed in the same direction (either left or right). On congruent trials the middle arrow pointed the same direction as the others; on incongruent trials the middle arrow pointed the opposite direction as the others. Participants were asked to press either the far left or far right key on a serial response box as quickly as possible to indicate which direction the middle arrow was pointing, ignoring the arrows on either side. Participant were given up to 3000 ms to respond, and the interstimulus interval was 500 ms plus a randomly generated waiting time of between 1 and 1000 ms. There were eight practice trials followed by 120 test trials. Scores were calculated based on participants' accuracy and a standardized difference in their reaction times to incongruent and congruent trials (incongruent RT – congruent RT/ average RT).

5.1.12 Rapid naming

A rapid digit naming task was used as a measure of rapid naming ability. The specific task used was taken from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999). Participants were shown a series of single digits and asked to read the digits aloud as quickly as possible in their L1. Participants began with one practice line of six digits, followed by two pages with four rows of nine digits each. Scores were calculated based on the total number of errors that participants made and the total time participants took to read the first test page plus the total time participants took to read the second test page (excluding the time required to turn pages). Note that because of different word lengths and articulation times for digit names across languages, reading times were standardized within each L1 group before analysis so that performance was directly comparable across groups.

5.1.13 English proficiency tests

Participants completed four tests of English proficiency: reading comprehension, listening comprehension, productive vocabulary knowledge, and receptive vocabulary knowledge. Details of the scores on these tests for each group are given in Table 1 (above).

5.1.14 Reading comprehension

For all participants, the reading comprehension test was produced by AccuPlacer (The College Board). For the L1 English speakers, the specific test was the "Reading Comprehension" test, and for the non-native English speakers, the specific test was the "ESL Reading Skills" test. Each test was an adaptive, computer-administered multiple-choice test with 20 questions, and students were given up to 30 minutes to complete the test. The Reading Comprehension test focused on paragraph-level comprehension and the function of specific linguistic elements within a broader discourse. The ESL Reading Skills test focused on sentence and short paragraph comprehension. The maximum score for each test was 120.

5.1.15 Listening Comprehension

The listening comprehension test was an in-house test developed by the English Language Institute at the University of Pittsburgh and is used to place intensive English students in appropriate class levels. There are three types of listening: general listening (descriptive narratives), academic listening (lecture excerpts), and conversations. There are a total of 30 multiple-choice questions on the test. Although the answer choices were written, the questions themselves were not provided in writing to the students; rather, then were presented aurally after the end of the listening passage. The questions were only given once, but participants were allowed to take notes while listening if they wanted to. The maximum score for this test was 30.

5.1.16 Vocabulary Knowledge

Two different tests of vocabulary knowledge were used. The first was an adaptation of the Vocabulary Levels Test (Laufer & Nation, 1999) that was done by Tom Cobb in order to remove cognates with French from the test (Cobb, 2000). This was a productive, cloze-style test. For each question, participants read a sentence with a blank space in it and were asked to write the word that best fit in the blank space. For all but one question, one or more letters of the beginning of the correct answer were provided as a clue for students. There were a total of 72 items of increasing difficulty (decreasing word frequency). Students were given up to 30 minutes to complete this test. Acceptable answers were discussed and agreed upon by two native English speakers (one American and one British). Answers had to have exactly correct spelling in order to be accepted. Participants were strongly encouraged to guess even if they were unsure of their answer. The maximum possible score for this test was 72.

The second vocabulary test was the Vocabulary Size Test (I. S. P. Nation & Beglar, 2007). This was a receptive, multiple-choice test. It is important to note that unlike the productive Vocabulary Levels Test, the Vocabulary Size Test contained a large number of cognates with French. For each question, participants saw an English word, which was also used in a neutral example sentence. They were given four possible definitions or meanings for the word, and were asked to choose the best one. There were a total of 100 items on the test, and participants were strongly encouraged to guess even if they were unsure of their answer. Participants were given up to 30 minutes to complete the test. The maximum possible score for this test was 72.

5.1.17 Overall Procedure

Participants completed the oddity task first, followed by word/pseudohomophone discrimination, deletion, and the wordlikeness judgments. The cognitive tasks were completed next; RAN was done first, followed by operation span and then flankers. Participants ended with a language history questionnaire that elicited basic demographic information and details regarding participants' language learning background and experiences as well as self-ratings of L1 and L2 proficiency (Tokowicz, Michael, & Kroll, 2004). The oddity task took approximately 10-15 minutes; word/pseudohomophone discrimination took approximately 5 minutes; deletion took approximately 12-17 minutes; the wordlikeness judgments took approximately 7-10 minutes; RAN took approximately 2 minutes; operation span took approximately 10-12 minutes; and flankers took approximately 5 minutes. All participants were tested individually.

5.2 RESULTS

5.2.1 Cognitive Abilities

The means and standard deviations for each L1 group's performance on the operation span (working memory), flankers (executive control), and rapid digit naming (rapid naming) tasks are displayed in Table 2.

Table 2. Descriptive statistics for performance on the tests of general cognitive abilities.

	L1 English	L1 French	L1 Hebrew	L1 Chinese
Operation span size	4.59 (1.26)	4.54 (1.24)	5.24 (1.10)	5.22 (0.84)
Flankers accuracy	99.00% (1.15)	98.84% (1.61)	96.85% (8.60)	99.09% (1.20)
Flankers standardized congruency RT difference	0.13	0.12	0.10	0.10
Rapid digit naming - # errors	0.13	0.26	0.30	0.24
Rapid digit naming – total time[a]	24.09 ms	29.24 ms	28.06 ms	21.38 ms

Note. [a]The raw times given here were standardized within each L1 group for the purposes of analysis.

There was a significant main effect of L1 on operation span performance, $F(3, 180) = 5.37$, $p = .001$. Follow-up comparisons were performed using *t*-tests with a Šídák correction for multiple comparisons. These comparisons revealed that native Hebrew and native Chinese speakers had significantly higher maximum item spans than native French and native English speakers (Hebrew versus French, $p < .05$; Hebrew versus English, $p < .05$; Chinese versus French, $p < .05$; Chinese versus English, $p < .05$). There was no significant difference between the native Hebrew and native Chinese speakers or between the native French and native English speakers. Because a significant main effect of L1 was found for the working memory task, these scores were included as a covariate in all analyses of the target variables (phonological awareness and orthographic knowledge).

There was also a significant main effect of L1 on the standardized incongruent-congruent reaction time difference for the flankers task, $F(3, 180) = 3.62$, $p < .05$. The main effect of L1 on flankers accuracy was marginally significant, $F(3, 180) = 2.63$, $p = .05$. For each main effect follow-up comparisons were again performed using *t*-tests with the Šídák correction. For the standardized reaction time difference, the native Hebrew and native Chinese speakers showed a smaller congruency effect than the native English speakers (Hebrew versus English, $p < .05$; Chinese versus English, $p = .05$). There were no other significant group differences. For accuracy, the native Chinese speakers had marginally higher accuracy than the native Hebrew speakers, $p = .095$; no other group differences approached significance. Because of the limited amount of variability in accuracy performance (all groups scored above 96% correct on average), only the standardized incongruent-congruent reaction time difference was retained as a covariate in all analyses of the target variables.

There was no significant main effect of L1 on the rapid digit naming task, either for the total number of errors or for the (standardized) total reading time. Thus, neither measure was retained as a covariate in analyses of the target variables.

It is interesting to note that the pattern of performance on these measures of general cognitive abilities, in participants matched on English proficiency (among non-native speakers) and broadly similar in age and educational background, conform broadly to the general finding of a bilingual advantage in executive control (e.g., Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok & Martin, 2004; Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009; Costa, Hernández, & Sebastián-Gallés, 2008; Gold, Kim, Johnson, Kryscio, & Smith, 2013;

Prior & MacWhinney, 2010). Specifically, the two groups of bilinguals with the highest English proficiency before matching samples, the native Hebrew and native Chinese speakers, showed the smallest difference in reaction times between the inconsistent and consistent trials on the flankers task. In contrast, there was no difference in performance between the native French speakers, who used English to a much more limited extent, and the native English speakers, who were specifically recruited to have minimal proficiency in an additional language. Working memory also requires the use of executive control for successful task performance (Kane et al., 2004). Here again we see an advantage for the native Hebrew and native Chinese speakers, compared to the native French and native English speakers, with no difference between the native French and English speakers.

5.2.2 Phonological Awareness

5.2.3 Oddity task

Recall that the oddity task was used as a receptive measure of phonological awareness. Mixed ANOVAs were used to analyze mean accuracies and reaction times. Item type was a within-subjects variable, L1 was a between-subjects variable, and the standardized incongruent-congruent RT difference on the flankers task and the operation span maximum span length score were covariates. Only data from the non-native speakers were included in the ANOVA; data from the native English speakers are included in the tables and figures for comparison purposes only. Two different ways of defining item type were considered. In the first, performance was analyzed in relation to whether the participants were making judgments about phonological units at the beginning, in the middle, or at the end of a word (location). In the second, performance was analyzed in relation to the specific phonological unit being targeted: first syllable, second syllable, onset, rime, body, coda, initial phoneme (consonants only), medial consonant, medial vowel, or final phoneme (consonants only).

5.2.4 Oddity word location accuracy

The results for word location (beginning, middle, or end) were examined first; descriptive statistics are presented in Table 3. There was a significant main effect of location, $F(2, 266) = 12.28$, $p < .001$, $\eta_p^2 = .09$ and a significant main effect of L1, $F(2, 133) = 5.71$ $p < .01$, $\eta_p^2 = .08$. The interaction between location and L1 was not significant. There was no main effect of flankers, but there was a significant effect of operation span size, $F(1, 178) = 17.25$, $p < .001$, $\eta_p^2 = .09$. Correlations revealed that the relationship between overall accuracy on the oddity task and operation span size was positive for all L1 groups. It was marginally significant for the L1 English speakers, $r = .27$, $p = .07$, and the L1 Chinese speakers, $r = .28$, $p = .06$, and significant for the L1 French speakers, $r = .35$, $p < .05$, and the L1 Hebrew speakers, $r = .35$, $p < .05$. Neither flankers nor operation span size interacted with location. Pairwise comparisons were used to compare the effect of location on performance, using the Šídák correction for multiple comparisons. These comparisons confirmed that, as seen in Figure 5, performance on items at the beginning of a word and performance on items at the end of a word was each higher than performance on items in the middle of a word, ps < .001. There was no difference between performance on items at the beginning and at the end of a word.

Table 3. Proportion correct on the oddity task, by location of the phonological unit inside the word and L1.

	L1			
Location	English	French	Hebrew	Chinese
Beginning	97.62 (5.36)	90.83 (14.63)	88.79 (18.39)	96.94 (11.05)
Middle	55.57 (12.09)	50.68 (13.39)	48.64 (13.17)	50.54 (15.69)
End	97.01 (3.75)	93.55 (9.67)	85.67 (15.90)	94.50 (9.80)

Note. Standard deviations are presented in parentheses. Results shown are prior to adjustment for differences in scores on the flankers and operation span tasks.

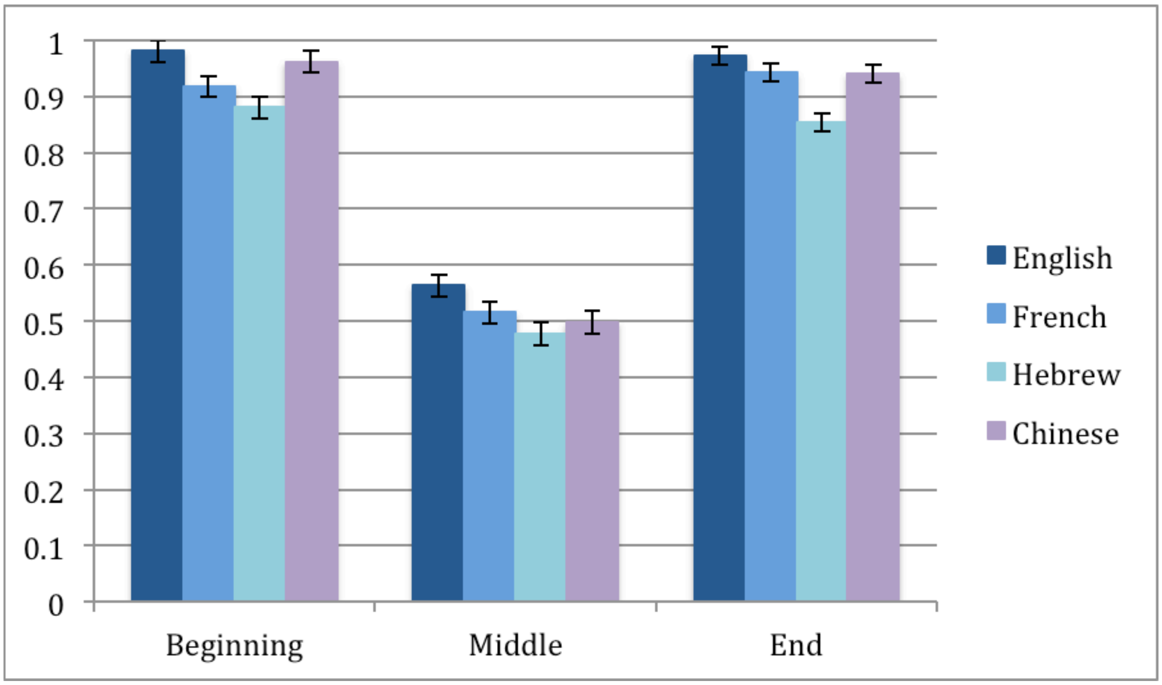

Figure 5. Mean proportion correct on the oddity task, by location of the phonological unit inside the word and L1, adjusted for performance on flankers and operation span. Error bars represent one standard error.

5.2.5 Oddity phonological unit accuracy

The results for phonological unit were examined next; descriptive statistics are given in Table 4. There was a significant main effect of unit, $F(9, 1197) = 6.13$, $p < .001$, $\eta_p^2 = .04$ and a significant main effect of L1, $F(2, 133) = 6.56$, $p < .01$, $\eta_p^2 = .09$. The interaction between unit and L1 was significant, $F(18, 1197) = 1.89$, $p = .01$, $\eta_p^2 = .08$. There was no main effect of flankers, but there was a significant effect of operation span size, $F(1, 133) = 15.31$, $p < .001$, $\eta_p^2 = .10$. As detailed above, this relationship was positive for all L1 groups; it was marginally significant for the L1 English and the L1 Chinese speakers and significant for the L1 French and the L1 Hebrew speakers. Neither flankers nor operation span size interacted with phonological unit. To follow up on the interaction between phonological unit and L1, simple main effects of L1 and simple pairwise comparisons among L1 groups were examined separately for each phonological unit. These analyses revealed that there was a significant simple main effect of L1 for accuracy on items testing the first syllable, $F(2, 133) = 3.32$, $p < .05$, $\eta_p^2 = .05$; the second syllable, $F(2, 133) = 8.84$, $p < .001$, $\eta_p^2 = .12$; onset, $F(2, 133) = 3.90$, $p < .05$, $\eta_p^2 = .06$; coda, $F(2, 133) = 7.84$, $p = .001$, $\eta_p^2 = .11$; and last phoneme, $F(2, 133) = 8.33$, $p < .001$, $\eta_p^2 = .11$. There was a marginal main effect of L1 for accuracy on items testing the first phoneme, $F(2, 133) = 2.85$, $p = .06$, $\eta_p^2 = .04$. There was no simple main effect of L1 for accuracy on items testing rime, body, medial consonant, or medial vowel units.

Table 4. Proportion correct on the oddity task, by phonological unit and L1.

	L1			
Location	English	French	Hebrew	Chinese
First Syllable	97.28 (10.51)	88.59 (18.60)	88.04 (21.07)	97.01 (11.84)
Second Syllable	98.64 (3.93)	98.10 (5.87)	89.13 (15.73)	95.92 (10.56)
Onset	98.64 (3.93)	92.39 (16.56)	87.50 (22.20)	97.01 (12.41)
Rime	98.91 (3.56)	96.74 (11.91)	94.02 (14.61)	97.28 (11.46)
Body	96.74 (6.68)	92.93 (15.06)	92.12 (18.14)	97.55 (11.67)
Coda	95.65 (6.57)	87.77 (16.14)	78.26 (23.49)	92.12 (13.01)
Initial Phoneme	97.83 (5.47)	89.40 (19.18)	87.50 (22.36)	96.20 (12.32)
Medial Consonant	33.43 (16.68)	39.40 (16.87)	41.30 (16.64)	43.21 (19.84)
Medial Vowel	77.72 (16.65)	61.96 (21.40)	55.98 (18.96)	57.88 (21.30)
Final Phoneme	95.11 (8.53)	91.58 (12.65)	81.25 (21.85)	92.66 (11.66)

Note. Standard deviations are presented in parentheses. Results shown are prior to adjustment for differences in scores on the flankers and operation span tasks.

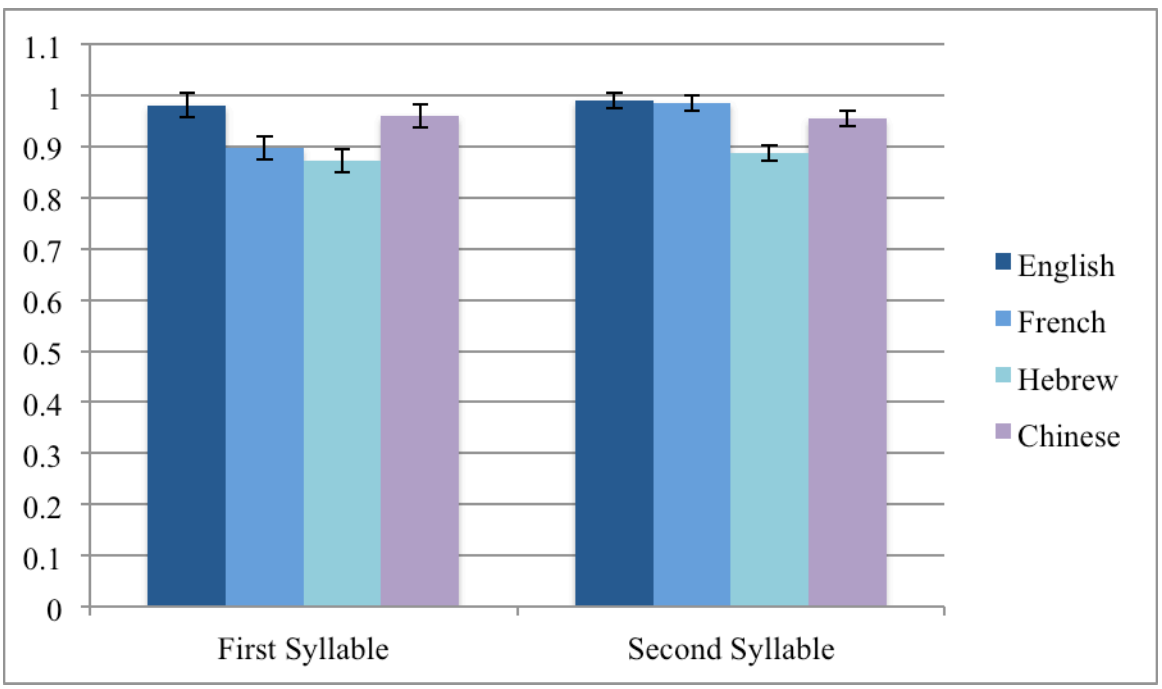

Figure 6. Mean proportion correct on the oddity task, by syllable and L1, adjusted for performance on flankers and operation span. Error bars represent one standard error.

Simple pairwise comparisons were next examined for those phonological units that had a significant or marginal simple main effect of L1. The Šídák correction for multiple comparisons was used. For items testing awareness of the first syllable, the L1 Hebrew speakers had the lowest accuracy (88.0%), followed by the L1 French speakers (88.6%), the L1 Chinese speakers (97.0%), and the L1 English speakers (97.3%). The L1 Hebrew and L1 French speakers did not differ from one another. The L1 Hebrew speakers were significantly less accurate than the L1 Hebrew speakers, $p < .05$; no other differences were significant. For items testing awareness of the second syllable, the L1 Hebrew speakers again had the lowest accuracy (89.1%), followed by the L1 Chinese speakers (95.9%), the L1 French speakers (98.1%), and the L1 English speakers (98.6%). The L1 Hebrew speakers were significantly less accurate than the L1 Chinese speakers, $p < .05$, and the L1 French speakers, $p < .001$, but the difference between the L1 Chinese and the L1 French speakers was not significant. The oddity accuracy results for the first and second syllable, adjusted for differences in flankers and operation span, are displayed in Figure 6.

For items testing awareness of the onset, the L1 Hebrew speakers had the lowest accuracy (87.5%), followed by the L1 French speakers (92.4%), the L1 Chinese speakers (97.0%), and the L1 English speakers (98.6%). The L1 Hebrew speakers were significantly less accurate than the L1 Chinese speakers; no other differences were significant. For items testing awareness of the coda, the L1 Hebrew speakers again had the lowest accuracy (78.3%), followed by the L1 French speakers (87.8%), the L1 Chinese speakers (92.1%), and the L1 English speakers (95.6%). The L1 Hebrew speakers were significantly less accurate than the L1 French speakers, $p < .05$, and the L1 Chinese speakers, $p = .001$, but the difference between the L1 French and the L1 Chinese speakers was not significant. The accuracy results for the onset, rime, body, and coda, adjusted for differences in flankers and operation span, are displayed in Figure 7.

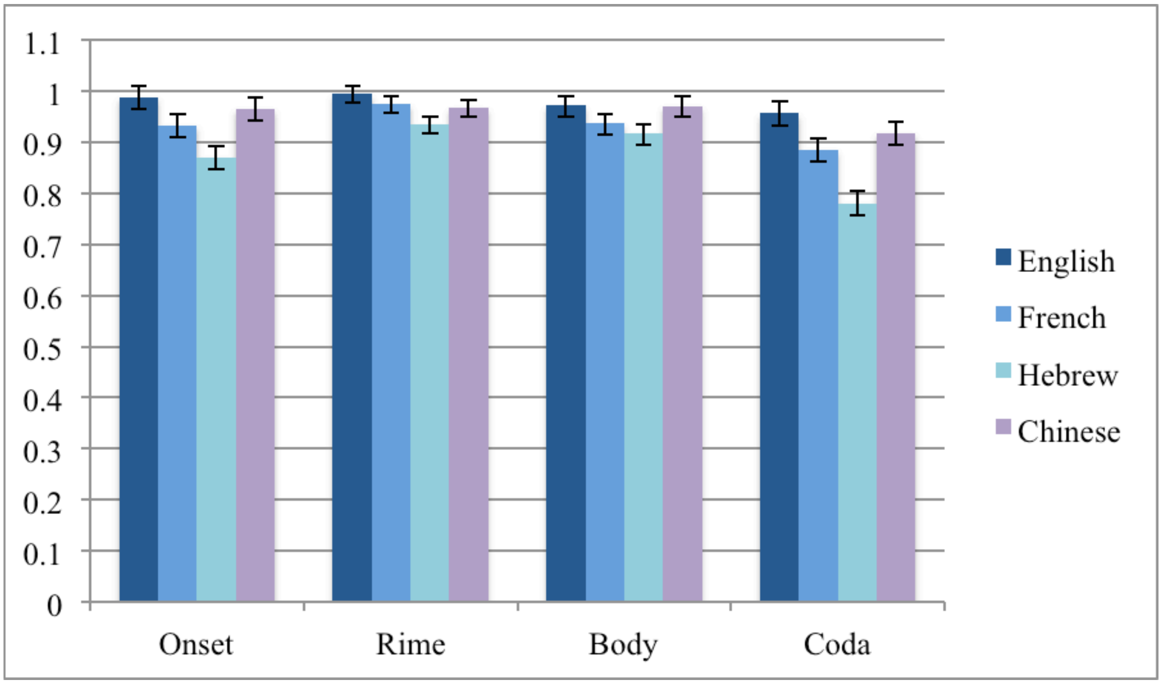

Figure 7. Mean proportion correct on the oddity task, by onset, rime, body, and coda unit and L1, adjusted for performance on flankers and operation span. Error bars represent one standard error.

For items testing awareness of the first phoneme, the L1 Hebrew speakers had the lowest accuracy (87.5%), followed by the L1 French speakers (89.4%), the L1 Chinese speakers (96.2%), and the L1 English speakers (97.8%). The L1 Hebrew speakers were marginally less accurate than the L1 Chinese speakers, $p = .06$; no other differences were significant. Finally, for items testing awareness of the last phoneme, the L1 Hebrew speakers had the lowest accuracy (81.3%), followed by the L1 French speakers (91.6%), the L1 Chinese speakers (92.7%), and the L1 English speakers (95.1%). The L1 Hebrew speakers had significantly lower accuracy than the L1 French speakers, $p = .002$, and the L1 Chinese speakers, $p = .002$; no other L1 differences were significant. The accuracy results for the initial phoneme, medial consonant, medial vowel, and final phoneme, adjusted for differences in flankers and operation span, are displayed in Figure 8.

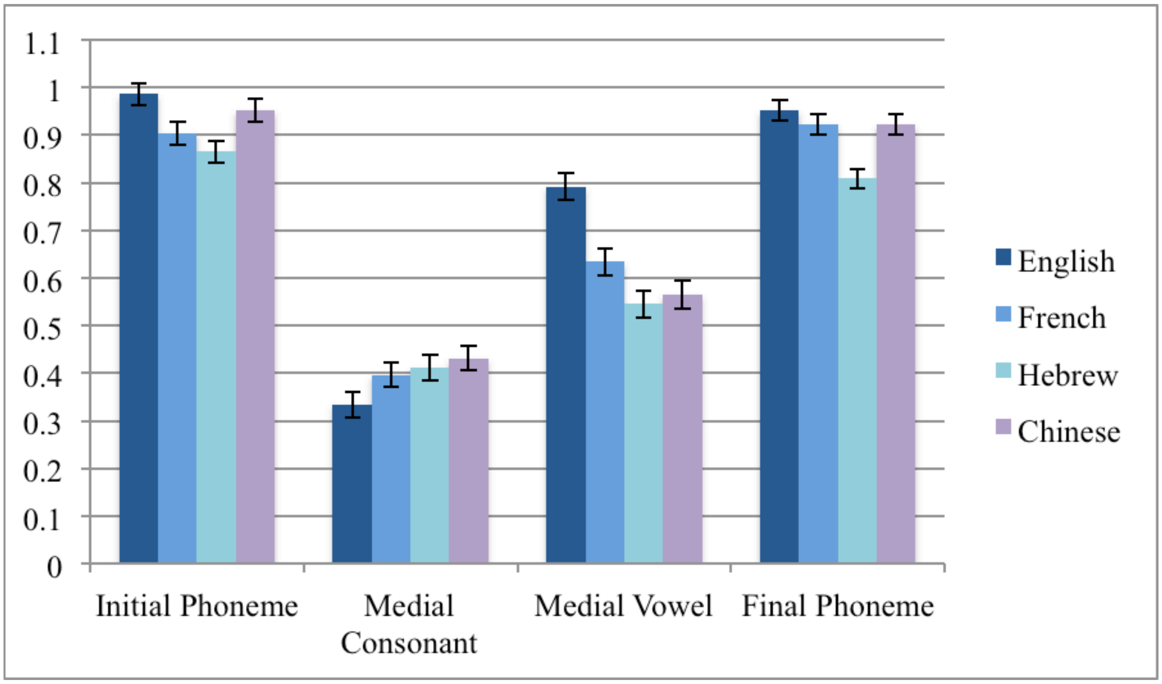

Figure 8. Mean proportion correct on the oddity task, by phoneme unit and L1, adjusted for performance on flankers and operation span. Error bars represent one standard error.

5.2.6 Oddity word location RTs

The results for word location (beginning, middle, or end) were again examined first; the raw means and standard deviations are presented in Table 5. There was a marginally significant main effect of location, $F(2, 266) = 2.80$, $p = .06$, $\eta_p^2 = .02$ and a significant main effect of L1, $F(2, 133) = 7.72$, $p = .001$ $\eta_p^2 = .10$. The interaction between location and L1 was not significant. There was no main effect of flankers or operation span size, and neither variable interacted with location.. Pairwise comparisons among the three locations revealed that participants were significantly slower in making judgments about shared phonological units in the middle of a word compared to either the beginning or the end of a word. In addition, participants were significantly slower in making judgments about phonological units at the end of a word than at the beginning of a word (see Figure 9). The finding that participants had faster RTs for units at the beginning of a word can be accounted for by the fact that participants did not need to wait until the end of the third oddity item had been pronounced before they responded about the beginning of the items, but they did need to hear the whole third item before they could respond about the end of the items. The faster RTs for judgments about the beginning of a word thus suggest that participants were prioritizing speeded responses during the task.

Table 5. Oddity RTs (in ms) by location of the phonological unit inside the word and L1.

	L1			
Location	English	French	Hebrew	Chinese
Beginning	-54.31 (376.74)	580.49 (587.65)	371.01 (729.60)	-31.94 (354.82)
Middle	1143.28 (750.78)	1394.12 (834.81)	1195.66 (845.36)	993.35 (530.09)
End	371.68 (572.21)	679.74 (431.17)	665.19 (655.82)	328.01 (277.63)

Note. Standard deviations are presented in parentheses. Results shown are prior to adjustment for differences in scores on the flankers and operation span tasks.

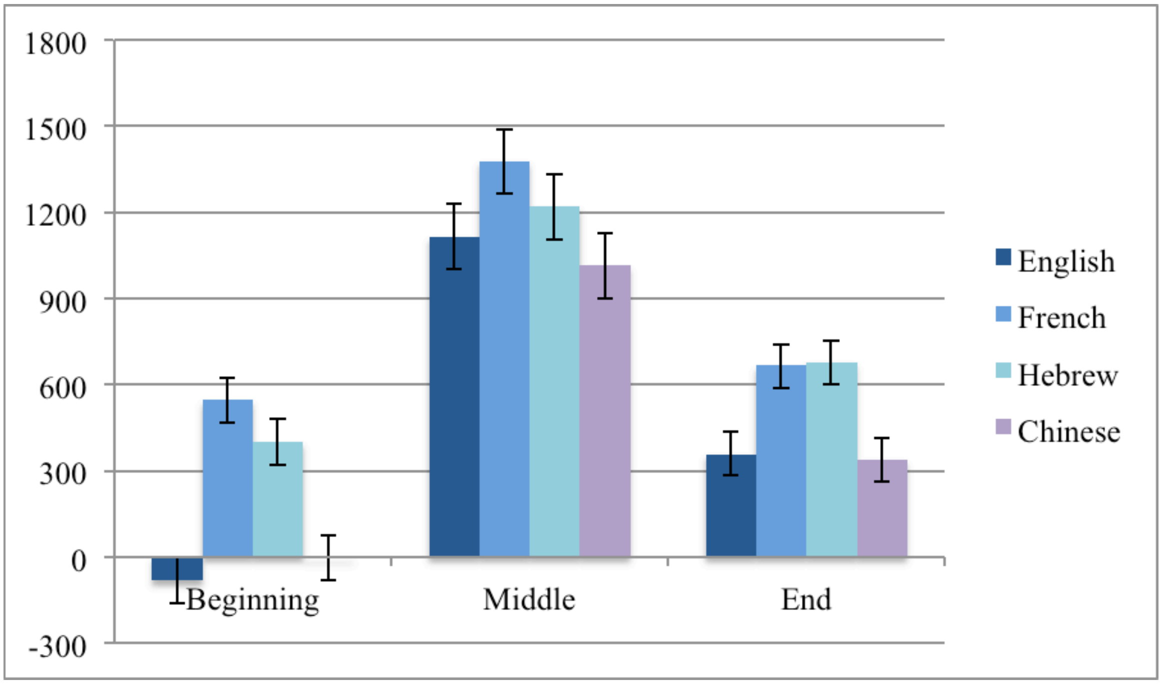

Figure 9. Mean RTs for performance on the oddity task, by location of the phonological unit inside the word and L1, adjusted for performance on flankers and operation span. Error bars represent one standard error.

5.2.7 Oddity phonological unit RTs

The results for phonological unit were examined next; descriptive statistics are presented in Table 6. There was no main effect of unit but there was a significant main effect of L1, $F(2, 127) = 9.10$, $p < .001$, $\eta_p^2 = .13$. The interaction between location and L1 was not significant. There was no main effect of flankers or operation span size, and neither variable interacted with location. Pairwise comparisons among the three non-native speaker groups revealed that the L1 French and the L1 Hebrew speakers did not differ significantly, but both groups had a significantly slower RT than the L1 Chinese speakers, *p*s < .01. The RTs for each group are given for the ten different phonological units in Figures 10, 11, and 12.

Table 6. Oddity RTs (in ms) by phonological unit and L1.

	L1			
Location	English	French	Hebrew	Chinese
First Syllable	-132.77 (429.18)	557.03 (634.13)	279.48 (759.63)	-82.33 (400.76)
Second Syllable	201.81 (624.18)	460.16 (458.66)	578.38 (919.58)	279.44 (324.80)
Onset	-47.34 (403.47)	555.68 (613.59)	339.44 (689.67)	-30.31 (388.33)
Rime	152.15 (602.15)	578.41 (518.95)	517.42 (640.48)	252.17 (329.29)
Body	-111.76 (471.25)	593.99 (643.21)	394.56 (932.36)	-54.76 (361.27)
Coda	517.65 (756.13)	900.81 (489.91)	814.36 (846.03)	399.70 (406.23)
Initial Phoneme	-38.83 (381.33)	724.85 (674.71)	486.21 (952.11)	42.79 (382.33)
Medial Consonant	1691.76 (1371.92)	1575.99 (1185.30)	1259.56 (1078.32)	1087.02 (718.78)
Medial Vowel	829.87 (636.22)	1310.36 (868.76)	1168.96 (928.24)	893.91 (598.70)
Final Phoneme	505.43 (548.97)	915.55 (553.29)	838.99 (639.55)	401.57 (279.43)

Note. Standard deviations are presented in parentheses. Results shown are prior to adjustment for differences in scores on the flankers and operation span tasks.

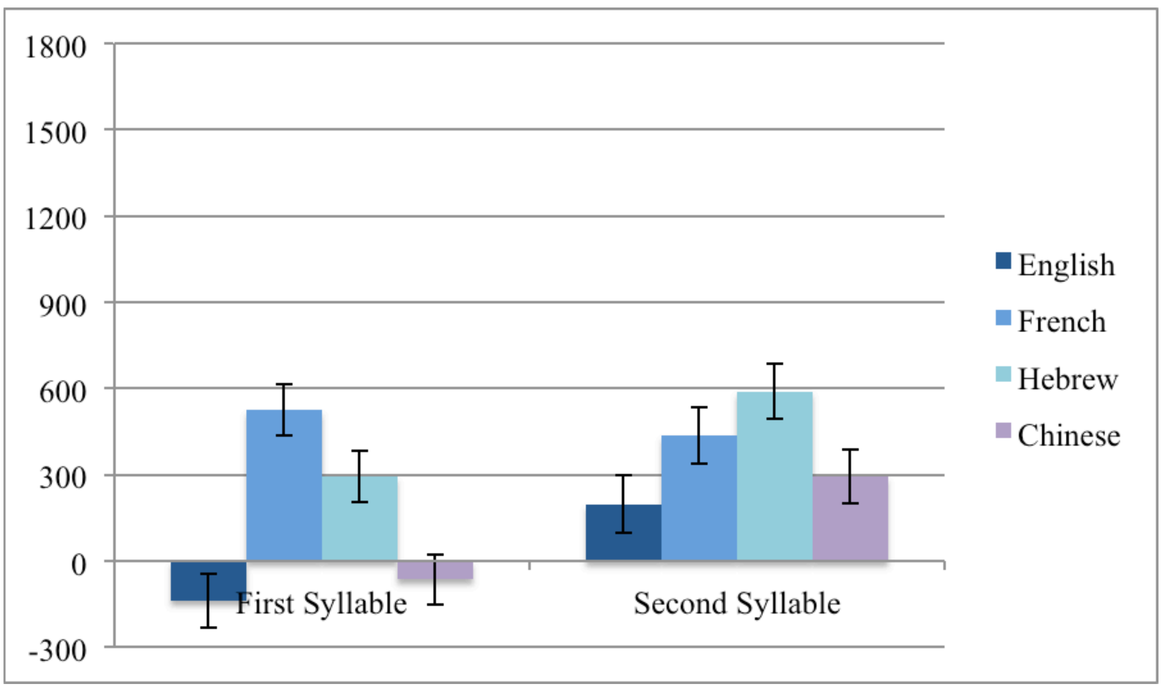

Figure 10. Mean RTs for performance on the oddity task, by syllable and L1, adjusted for performance on flankers and operation span. Error bars represent one standard error.

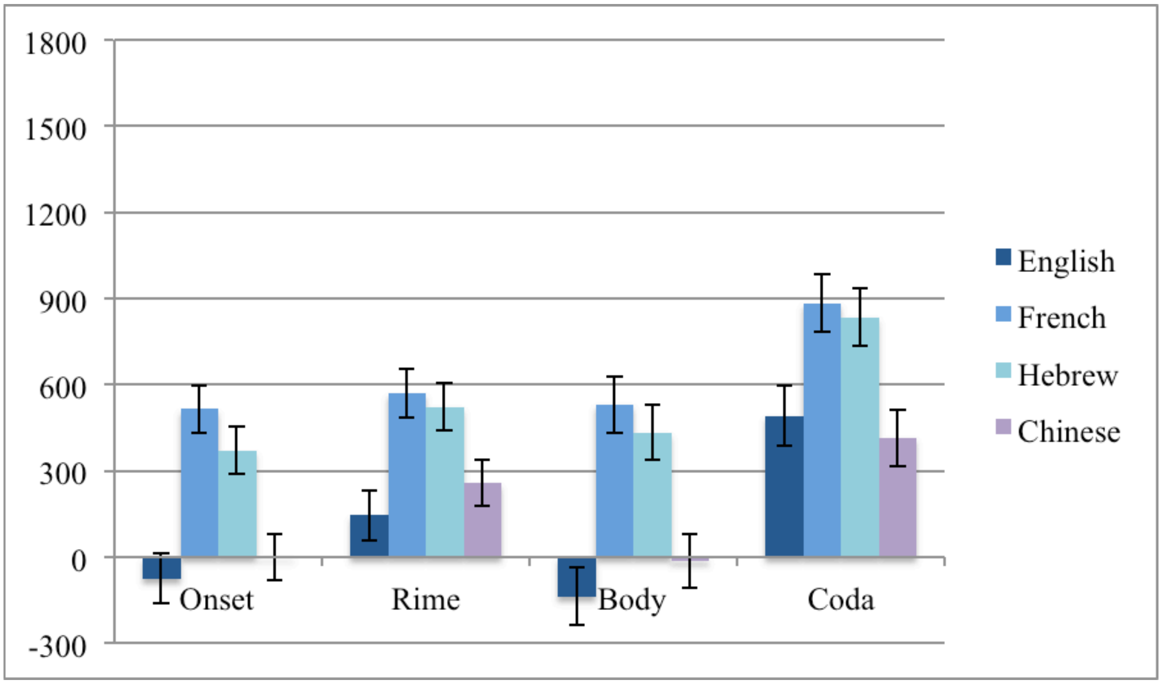

Figure 11. Mean RTs for performance on the oddity task, by onset, rime, body, and coda unit and L1, adjusted for performance on flankers and operation span. Error bars represent on standard error.

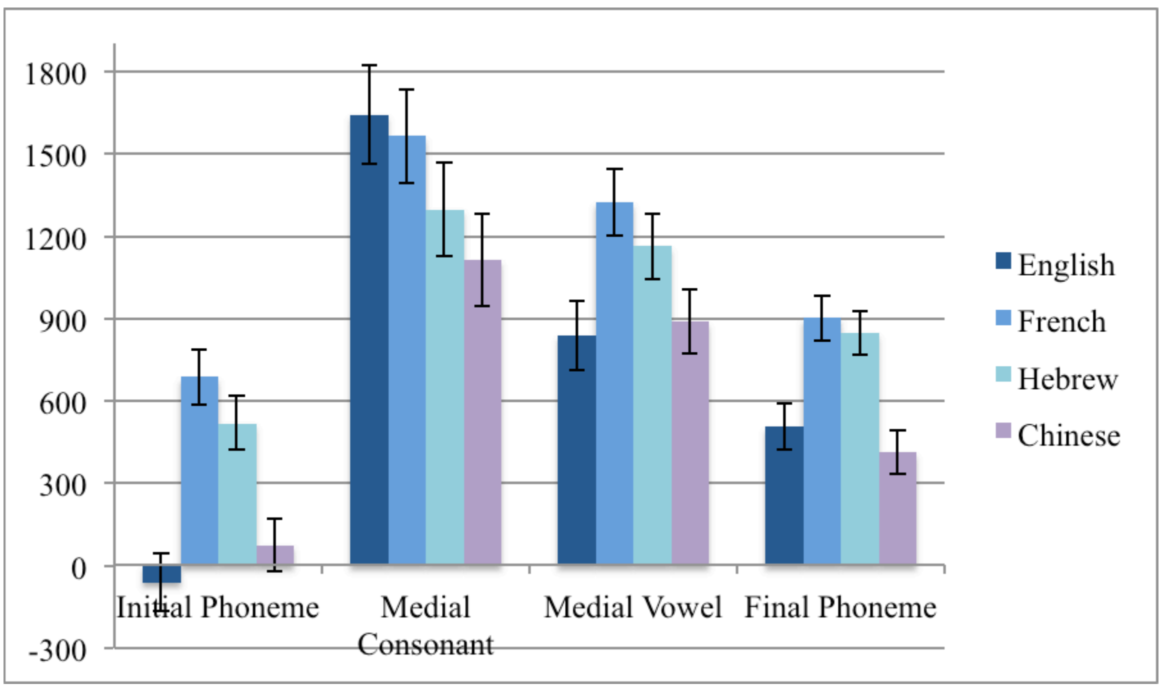

Figure 12. Mean RTs for performance on the oddity task, by phoneme unit and L1, adjusted for performance on flankers and operation span. Error bars represent one standard error.

5.2.8 Summary: Oddity task

Consistent with previous research testing phonological awareness for phonological units at different locations within in a word, performance was significantly lower on items targeting word-medial phonological units compared to word-initial or word-final (Stage & Wagner, 1992). Although we did find faster RTs to units at the beginning than at the end of the word, unlike Stage and Wagner (1992) we did not find a significant difference in accuracy on items testing phonological units at the beginning and at the end of a word.

The findings from the oddity task generally did not confirm our predictions. We expected that alphabetic L1 speakers (French) would out-perform those from a non-alphabetic L1 (Hebrew, Chinese) on items testing phonological awareness for single phonemes, regardless of position in the word, due to the support of their alphabetic L1 literacy for the development of phonemic awareness. However, this prediction was not borne out by the evidence. For the initial phoneme, the L1 French speakers were not significantly different from either non-alphabetic L1 group. In fact, the only significant difference for this unit was that the L1 Chinese speakers were *more* accurate than the L1 Hebrew speakers. Although the L1 French speakers outperformed the L1 Hebrew speakers on items testing the final phoneme, they did not differ from the L1 Chinese speakers, and there were no L1 differences at all for units testing awareness of a medial consonant or a medial vowel. The pattern of RTs was even less favorable for the L1 French speakers: although there was no interaction between L1 and phonological unit, the L1 French speakers had the slowest RTs overall and were significantly slower than the L1 Chinese speakers..

Of particular interest in the oddity results is the finding that the L1 Chinese speakers showed high levels of accuracy across items on the oddity task, often performing at a level similar to the L1 English speakers. This pattern is consistent with the hypothesis that these learners benefit from phonological awareness development as a consequent of their experience using the *zhuyin fuhao* phonological system during L1 literacy acquisition, and a similar pattern of strong performance on phonological awareness tasks by L1 Chinese speakers with some phonetic training has in fact been found in some previous work (Holm & Dodd, 1996; McBride-Chang et al., 2004).

Considering the body and coda units, we predicted that the L1 Hebrew speakers would show the best performance on these items due to the consonant-based nature of an abjad writing system, in which CV is an important sub-lexical unit (Ben-Dror et al., 1995; Frost, 2012; Tolchinsky et al., 2012). This prediction was also not borne out by the results. Considering accuracy, there was no significant effect of L1 for items testing body units, and the L1 Hebrew speakers in fact performed significantly worse than all other groups on the coda units. Although RTs were not analyzed for each phonological unit separately, visual inspection of performance on the body and coda units (see Figure 11) reveals that the L1 Hebrew speakers were much slower than the L1 Chinese (and the L1 English) speakers, although they did not differ substantially from the L1 French speakers.

Turning to the onset and rime units, we predicted that the L1 alphabetic speakers would perform better than the L1 non-alphabetic speakers on items testing awareness of these units, due to the importance of both the phonological and orthographic rime in alphabetic languages and the prevalence of phonological neighbors at the rime level in French and English in particular (De Cara & Goswami, 2003; Ziegler & Goswami, 2005). However, similar to the results for the body and coda units, the results did not support our prediction. The L1 French speakers were not significantly different from the other L1 groups for onset items, and there was no main effect of

L1 for accuracy on rime items. Looking at RTs, visual inspection again reveals that the L1 French speakers were the slowest overall, and were substantially slower than the L1 Chinese (and the L1 English) speakers for both onsets and rimes (see Figure 11).

Finally, considering performance for syllables, we expected the highest performance overall for these units because syllables are relatively large and salient phonological units and because syllable awareness develops prior to literacy cross-linguistically (Goswami & Bryant, 1990). Although performance was indeed high for items testing awareness of syllables, it was not noticeably higher than for items testing other phonological units. The results thus neither clearly confirm nor disconfirm this prediction. We also predicted that the L1moorphosyllabary speakers might perform at an even higher level on this task than the other L1 groups because the syllable is also the unit of meaning and the unit of writing for these languages (McBride-Chang et al., 2004; Tolchinsky et al., 2012). This prediction was partially supported. Numerically, the L1 Chinese speakers had the highest level of performance among the non-native English groups. However, although the L1 Chinese speakers were significantly more accurate than the L1 Hebrew speakers for units testing awareness of both the first and second syllable, the difference between the L1 Chinese and the L1 French speakers was not statistically significant for either syllable type.

5.2.9 Deletion task

Recall that the deletion task was used as a measure of productive phonological awareness skill (as opposed to the receptive oddity task). Mixed ANOVAs were again used to analyze mean accuracies and reaction times. Only data from the non-native speakers were included in the ANOVA; data from the native English speakers are included in the tables and figures for comparison purposes only. Item type and lexicality (pseudoword or real word) were within-subjects variables. L1 was a between-subjects variable, and the standardized incongruent-congruent RT difference on the flankers task and the operation span maximum span length score were covariates. Item type was again defined in relation to whether the participants were making judgments about phonological units at the beginning, in the middle, or at the end of a word (location); and also in relation to the specific phonological unit being targeted: first syllable, second syllable, onset, rime, body, coda, initial phoneme (consonants only), medial consonant, medial vowel, or final phoneme (consonants only).

5.2.10 Deletion word location accuracy

The results for word location (beginning, middle, or end) were examined first; descriptive statistics are presented in Table 7. The main effect of lexicality was not significant, however there was a significant main effect of location, $F(2, 266) = 13.94$, $p < .001$, $\eta_p^2 = .10$, and a significant main effect of L1, $F(2, 133) = 10.77$, $p < .001$, $\eta_p^2 = .14$. The interactions between lexicality and L1 and between location and lexicality were not significant, but the interaction between location and L1 was significant, $F(4, 266) = 8.06$, $p < .001$, $\eta_p^2 = .11$. There was no main effect of flankers, but there was a significant main effect of operation span size, $F(1, 133) = 23.00$, $p < .001$, $\eta_p^2 = .15$. Neither flankers nor operation span size interacted with lexicality, and flankers also did not interact with location. However, operation span size did interact with location, $F(2, 266) = 5.11$, $p = .01$, $\eta_p^2 = .04$.

To examine the pattern of this interaction, correlations were calculated for each L1 group between operation span size and accuracy for items testing the ability to delete a phonological unit at the beginning, in the middle, or at the end of a word. All correlations were positive, such that a higher operation span size was associated with higher accuracy on the deletion task. For the L1 English speakers, the correlation was significant for items testing the beginning of a word, $r = .41$, $p < .01$, and the end of a word, $r = .46$, $p = .001$; it was marginally significant for items testing the middle of a word, $r = .28$, $p = .06$. For the L1 French speakers, the correlation was significant for items testing all three locations: beginning, $r = .36$, $p = .01$; middle, $r = .35$, $p < .05$; end, $r = .52$, $p < .01$. For the L1 Hebrew speakers, the correlation was marginally significant for items testing the beginning of a word, $r = .27$, $p = .07$; and significant for items testing the middle of a word, $r = .39$, $p < .01$, and the end of a word, $r = .36$, $p < .05$. For the L1 Chinese speakers, the correlation was significant for items testing the beginning of a word, $r = .33$ $p < .05$, and the end of a word, $r = .36$, $p < .05$; it was not significant for items testing the middle of a word.

Neither the three-way interaction between lexicality, location, and flankers, nor the three-way interaction between lexicality, location, and operation span size, were significant. However, the three-way interaction between lexicality, location, and L1 was significant, $F(4, 266) = 3.87$, $p < .01$, $\eta_p^2 = .06$. To follow up on this three-way interaction, the L1 x location interaction was examined separately for pseudowords and for real words.

Table 7. Proportion correct on the deletion task, by location of the phonological unit inside the word and L1.

Lexicality	Location	English	French	Hebrew	Chinese
		L1			
Pseudo-words	Beginning	87.91 (10.66)	73.51 (14.74)	74.46 (15.30)	66.85 (13.68)
	Middle	78.26 (23.34)	62.50 (34.46)	69.02 (32.57)	50.54 (27.13)
	End	81.14 (13.9)	75.00 (16.56)	67.80 (18.11)	58.29 (15.26)
Real words	Beginning	93.61 (6.91)	85.05 (13.85)	81.79 (12.90)	75.68 (15.77)
	Middle	91.85 (14.96)	54.89 (27.19)	73.37 (28.09)	46.74 (29.63)
	End	94.97 (5.69)	85.33 (12.58)	78.40 (14.65)	79.21 (15.76)

Note. Standard deviations are presented in parentheses. Results shown are prior to adjustment for differences in scores on the flankers and operation span tasks.

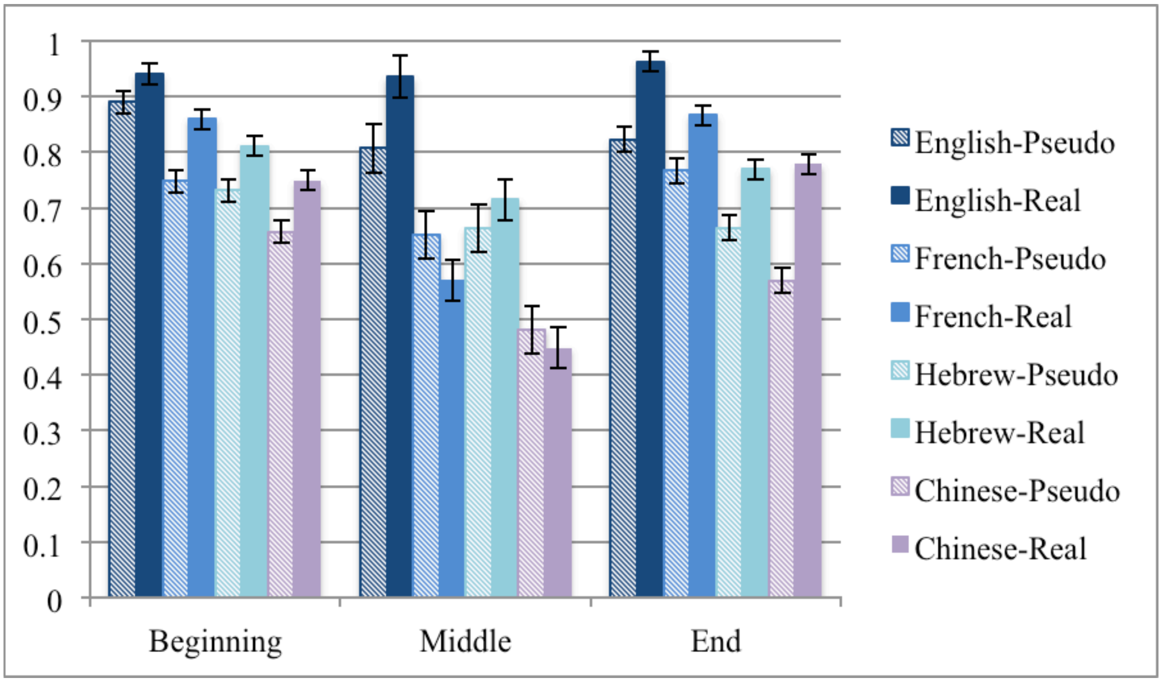

Figure 13. Mean proportion correct on the deletion task, by location of the phonological unit inside the word and L1, adjusted for performance on flankers and operation span. Pseudo = pseudoword; Real = real word. Error bars represent one standard error.

For pseudowords the L1 x location interaction was significant, Pillai's Trace = .11, $F(4, 266) = 4.03$, $p < .01$, $\eta_p^2 = .06$. There was a significant simple main effect of L1 for stimuli testing the ability to delete a phonological unit at the beginning of a pseudoword, $F(2, 133) = 5.26$, $p < .01$, $\eta_p^2 = .07$. The L1 Chinese speakers had the lowest accuracy (66.9%), followed by the L1 French speakers (73.5%), the L1 Hebrew speakers (74.5%), and the L1 English speakers (87.9%). Simple pairwise comparisons revealed that the L1 Chinese speakers were significantly less accurate than the L1 French speakers, $p = .003$, and the L1 Hebrew speakers, $p = .01$, but the difference between the L1 French and the L1 Hebrew speakers was not significant..

There was also a significant simple main effect of L1 for pseudoword stimuli testing the ability to delete a phonological unit in the middle of a pseudoword, $F(2, 133) = 5.12$, $p < .01$, $\eta_p^2 = .07$. The L1 Chinese speakers had the lowest accuracy (50.5%), followed by the L1 French speakers (62.5%), the L1 Hebrew speakers (69.0%), and the L1 English speakers (78.3%). Simple pairwise comparisons revealed that the L1 Chinese speakers were significantly less accurate the L1 French speakers, $p = .01$, and the L1 Hebrew speakers, $p = .005$, but the difference between the L1 French and the L1 Hebrew speakers was not significant.

Finally, there was also a significant simple main effect of L1 for pseudoword stimuli testing awareness of the ability to delete a phonological unit at the end of a pseudoword, $F(2, 133) = 17.31$, $p < .001$, $\eta_p^2 = .21$. The L1 Chinese speakers had the lowest accuracy (58.3%), followed by the L1 Hebrew speakers (67.8%), the L1 French speakers (75.0%), and the L1 English speakers (81.1%). Simple pairwise comparisons revealed that the L1 Chinese speakers were significantly less accurate than the L1 Hebrew speakers, $p = .005$, and the L1 French speakers, $p < .001$. The L1 Hebrew speakers were also significantly less accurate than the L1 French speakers, $p = .002$.

For real words the L1 x location interaction was also significant, Pillai's Trace = .20, $F(4, 266) = 7.49$, $p < .001$, $\eta_p^2 = .10$. There was a significant simple main effect of L1 for stimuli testing the ability to delete a phonological unit at the beginning of a word, $F(2, 133) = 7.68$, $p = .001$, $\eta_p^2 = .10$. The L1 Chinese speakers had the lowest accuracy (75.7%), followed by the L1 Hebrew speakers (81.8%), the L1 French speakers (85.1%), and the L1 English speakers (93.6%). Simple pairwise comparisons revealed that the L1 Chinese speakers were significantly less accurate than the L1 Hebrew speakers, $p = .03$, and the L1 French speakers, $p < .001$. The L1 Hebrew speakers were marginally less accurate than the L1 French speakers, $p = .07$.

There was also a significant simple main effect of L1 for real word stimuli testing the ability to delete a phonological unit in the middle of a word, $F(2, 133) = 11.07$, $p < .001$, $\eta_p^2 = .14$. The L1 Chinese speakers had the lowest accuracy (46.7%), followed by the L1 French speakers (54.9%), the L1 Hebrew speakers (73.4%), and the L1 English speakers (91.8%). Simple pairwise comparisons revealed that the L1 Chinese speakers were significantly less accurate than the L1 French speakers, $p = .02$, and the L1 Hebrew speakers, $p < .001$. The L1 French speakers were also significantly less accurate than the L1 Hebrew speakers, $p = .03$.

Finally, there was a significant simple main effect of L1 for real word stimuli testing the ability to delete a phonological unit at the end of a word, $F(2, 133) = 7.17$, $p = .001$, $\eta_p^2 = .10$. For these items it was the L1 Hebrew speakers who had the lowest accuracy (78.4%), followed by the L1 Chinese speakers (79.2%), the L1 French speakers (85.3%), and the L1 English speakers (95.0%). Simple pairwise comparisons showed that there was no significant difference between the L1 Hebrew and the L1 Chinese speakers, but both the L1 Hebrew and the L1 Chinese speakers were significantly less accurate than the L1 French speakers, $p = .001$ and $p < .01$, respectively.

5.2.11 Deletion phonological unit accuracy

The results for phonological unit were examined next; the raw means and standard deviations are presented in Table 8. There was a significant main effect of lexicality, $F(1, 133) = 10.58$, $p = .001$, $\eta_p^2 = .07$; as can be seen in Figure 14, Figure 15, and Figure 16, this effect reflects an accuracy advantage for real word stimuli over pseudoword stimuli. There was also a significant main effect of phonological unit, $F(8, 1064) = 6.15$, $p < .001$, $\eta_p^2 = .04$, and a significant main effect of L1, $F(2, 133) = 11.58$, $p < .001$, $\eta_p^2 = .15$. The interaction between lexicality and L1 was significant, $F(2, 133) = 4.57$, $p = .01$, $\eta_p^2 = .06$, as was the interaction between phonological unit and L1, $F(16, 1064) = 5.87$, $p < .001$, $\eta_p^2 = .08$. The interaction between lexicality and phonological unit was also significant, $F(8, 1064) = 4.64$, $p < .001$, $\eta_p^2 = .03$. There was no main effect of flankers but there was a significant main effect of operation span size, $F(1, 133) = 22.83$, $p < .001$, $\eta_p^2 = .15$. Neither flankers nor operation span size interacted with lexicality. However, the flankers x phonological unit interaction was marginally significant, $F(8, 1064) = 1.72$, $p = .09$, $\eta_p^2 = .01$, as was the operation span size x phonological unit interaction, $F(8, 1064) = 1.70$, $p = .09$, $\eta_p^2 = .01$.

Correlations revealed that the relationship between overall accuracy on the deletion task and operation span size was significant and positive for all L1 groups: L1 English, $r = .50$, $p < .001$; L1 French, $r = .41$, $p < .01$; L1 Hebrew, $r = .38$, $p = .01$; L1 Chinese, $r = .37$, $p = .01$. To examine the marginal interaction between flankers and phonological unit, correlations were calculated for each L1 group between flankers and accuracy for items testing the ability to delete each phonological unit. For the L1 English speakers, the L1 French speakers, and the L1 Chinese speakers, the correlation between flankers and deletion accuracy was not significant for any of the phonological units. However, for the L1 Hebrew speakers, there was a significant correlation between flankers and accuracy on the deletion task for rimes, $r = .31$, $p < .05$; bodies, $r = .33$, $p <$ 05; and initial phonemes, $r = .32$, $p < .05$; and a marginally significant correlation for onsets, $r = -.27$, $p = .07$. These correlations reveal that a larger difference in RTs to congruent and incongruent items (a larger congruency effect) on the flankers task was associated with higher accuracy on the deletion task for rimes, bodies, and initial phonemes, but that a smaller congruency effect was associated with lower accuracy for onsets.

The three-way interaction between lexicality, phonological unit, and flankers was not significant, but the three-way interaction between lexicality, phonological unit, and operation span size was significant, $F(8, 1064) = 2.09$, $p = .03$, $\eta_p^2 = .02$. To examine this interaction, correlations were calculated for each L1 group between operation span size and accuracy for items testing the ability to delete each phonological unit, for pseudoword and real word stimuli separately. For the L1 English speakers there was a significant correlation between operation span size and accuracy on the deletion task for pseudoword second syllables, $r = .35$, $p < .05$; real word second syllables, $r = .31$, $p < .05$; pseudoword rimes, $r = .31$, $p < .05$; real word rimes, $r = .37$, $p < .05$; pseudoword bodies, $r = .51$, $p < .001$; real word codas, $r = .37$, $p = .01$; real word initial phonemes, $r = .31$, $p < .05$; and pseudoword medial consonants, $r = .31$, $p < .05$; there was also a marginally significant correlation with pseudoword onsets, $r = .27$, $p = .07$; and real word bodies, $r = .28$, $p = .06$.

For the L1 French speakers there was a significant correlation between operation span size and accuracy on the deletion task for pseudoword second syllables, $r = .30$, $p < .05$; real word rimes, $r = .34$, $p < .05$; pseudoword bodies, $r = .34$, $p < .05$; real word medial consonants, $r = .35$, $p < .05$; and pseudoword final phonemes, $r = .40$, $p < .01$; there was also a marginally significant correlation with real word first syllables, $r = .26$, $p = .08$; real word codas, $r = .26$, $p =$

.08; real word initial phonemes, $r = .27$, $p = .07$; pseudoword medial consonants, $r = .27$, $p = .07$; and real word final phonemes, $r = .27$, $p = .07$.

For the L1 Hebrew speakers there was a significant correlation between operation span size and accuracy on the deletion task for pseudoword first syllables, $r = .40$, $p < .01$; pseudoword medial consonants, $r = .32$, $p < .05$; real word medial consonants, $r = .41$, $p < .01$; and pseudoword final phonemes, $r = .45$, $p < .01$; there was also a marginally significant correlation with real word second syllables, $r = .25$, $p = .09$; real word codas, $r = .26$, $p = .08$; pseudoword bodies, $r = .25$, $p = .09$; ad real word initial phonemes, $r = .27$, $p = .07$.

For the L1 Chinese speakers there was a significant correlation between operation span size and accuracy on the deletion task for real word onsets, $r = .34$, $p < .05$; pseudoword rimes, $r = .36$, $p < .05$; pseudoword bodies, $r = .33$, $p < .05$; pseudoword codas, $r = .38$, $p < .01$; and real word codas, $r = .35$, $p < .05$. Although no clear pattern is immediately discernable regarding which units do or do not show a significant correlation with operation span size in each L1, the overall finding is that in each case a larger operation span size is associated with higher accuracy on the deletion task, regardless of phonological unit or L1.

Table 8. Proportion correct on the deletion task, by phonological unit and L1.

Lexicality	Location	L1 English	L1 French	L1 Hebrew	L1 Chinese
Pseudo-words	First Syllable	86.96 (17.27)	65.76 (20.66)	75.54 (20.06)	73.91 (21.05)
	Second Syllable	78.26 (26.15)	68.48 (24.96)	58.15 (27.92)	40.76 (29.04)
	Onset	97.83 (7.12)	73.91 (23.54)	76.09 (25.80)	59.78 (28.62)
	Rime	64.67 (27.68)	66.30 (28.49)	43.48 (34.33)	42.93 (25.64)
	Body	77.17 (22.25)	82.07 (25.09)	67.93 (28.22)	67.93 (21.51)
	Coda	94.38 (14.17)	90.22 (17.06)	90.76 (17.76)	82.61 (17.38)
	Initial Phoneme	89.67 (14.51)	72.28 (23.70)	78.26 (22.12)	65.76 (24.36)
	Medial Consonant	78.26 (23.34)	62.50 (34.46)	69.02 (32.57)	50.54 (27.13)
	Final Phoneme	87.32 (16.64)	75.00 (22.36)	78.80 (26.34)	66.85 (19.04)
Real words	First Syllable	99.46 (3.69)	97.83 (7.12)	98.37 (6.24)	92.39 (12.77)
	Second Syllable	99.46 (3.69)	88.59 (17.25)	90.76 (12.30)	92.93 (13.60)
	Onset	97.28 (7.87)	81.52 (23.23)	82.61 (19.63)	71.74 (26.60)
	Rime	90.22 (16.23)	73.37 (25.50)	56.52 (35.13)	69.57 (31.58)
	Body	82.07 (19.48)	76.63 (28.58)	59.78 (31.39)	70.11 (26.15)
	Coda	94.57 (11.68)	87.50 (18.17)	85.33 (20.12)	82.07 (22.77)
	Initial Phoneme	95.65 (10.93)	84.24 (22.58)	86.41 (22.18)	68.48 (26.05)
	Medial Consonant	91.85 (14.96)	54.89 (27.19)	73.37 (29.09)	46.74 (29.63)
	Final Phoneme	95.65 (9.58)	91.85 (14.96)	80.98 (23.08)	72.29 (22.50)

Note. Standard deviations are presented in parentheses. Results shown are prior to adjustment for differences in scores on the flankers and operation span tasks.

Finally, the three-way interaction between lexicality, phonological unit, and L1 was also significant, $F(16, 1064) = 4.02$, $p < .001$, $\eta_p^2 = .06$. To follow up on the significant three-way interaction, the two-way interaction of L1 x phonological unit was tested separately for pseudoword and for word stimuli. This L1 x phonological unit interaction was significant for both pseudowords, Pillai's Trace = .33, $F(16, 254) = 3.11$, $p < .001$, $\eta_p^2 = .16$, and for real words, Pillai's Trace = .47, $F(16, 254) = 4.90$, $p < .001$, $\eta_p^2 = .24$. To follow up on these interactions further, the simple main effect of L1 and pairwise comparisons among L1s, if appropriate, were examined for each phonological unit separately for pseudowords and for real words.

There was no main effect of L1 for pseudoword stimuli testing the ability to delete the first syllable of an item. However, for pseudoword stimuli testing the ability to delete the second syllable of an item there was a significant simple main effect of L1, $F(2, 133) = 14.45$, $p < .001$, $\eta_p^2 = .18$. The L1 Chinese speakers had the lowest accuracy (40.8%), followed by the L1 Hebrew speakers (58.2%), the L1 French speakers (68.5%), and the L1 English speakers (78.3%). The L1 Chinese speakers were significantly less accurate than the L1 Hebrew speakers, $p < .01$, and the L1 French speakers, $p < .001$. The L1 Hebrew speakers were also significantly less accurate than the L1 French speakers, $p = .02$. The pseudoword deletion accuracy results for the first and second syllable, adjusted for differences in flankers and operation span, are displayed in Figure 14.

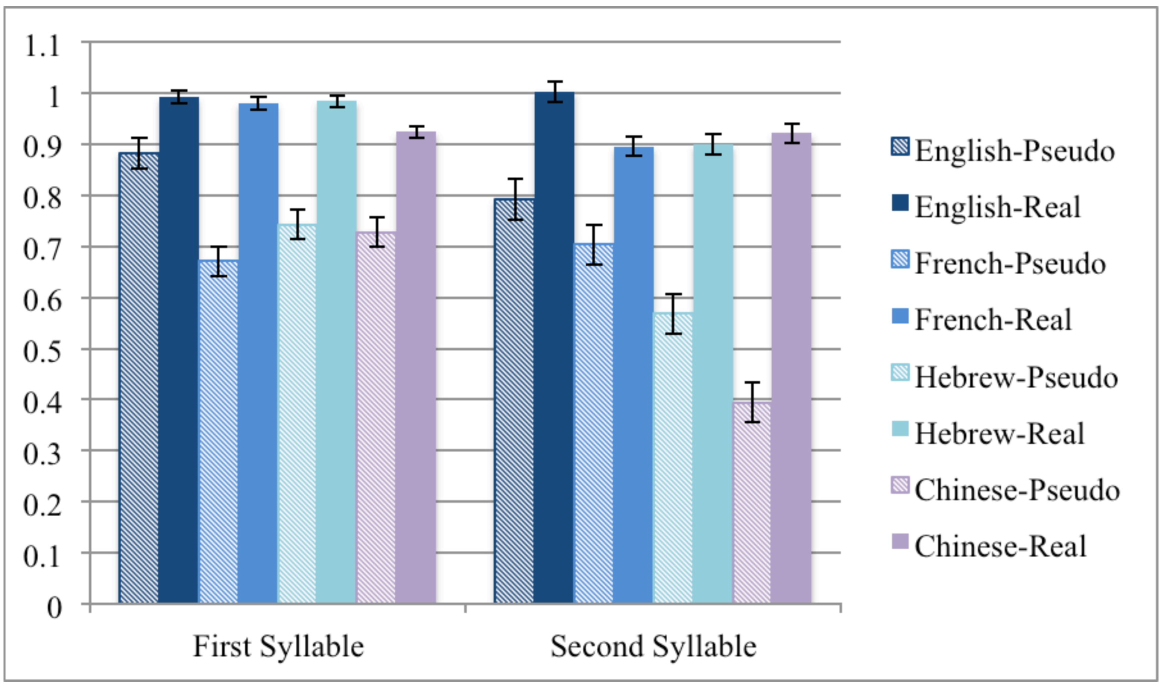

Figure 14. Mean proportion correct on the deletion task, by syllable and L1, adjusted for performance on flankers and operation span. Pseudo = pseudoword; Real = real word. Error bars represent one standard error.

For pseudoword stimuli testing the ability to delete the onset of an item there was a significant simple main effect of L1, $F(2, 133) = 6.08$, $p < .01$, $\eta_p^2 = .08$. The L1 Chinese speakers had the lowest accuracy (59.8%), followed by the L1 French speakers (73.9%), the L1 Hebrew speakers (76.1%), and the L1 English speakers (97.8%). The L1 Chinese speakers were significantly less accurate than the L1 French speakers, $p < .01$, and the L1 Hebrew speakers, $p < .01$, but the L1 French speakers did not differ from the L1 Hebrew speakers. For pseudoword stimuli testing the ability to delete the rime of an item there was a significant simple main effect of L1, $F(2, 133) = 10.80$, $p < .001$, $\eta_p^2 = .14$. The L1 Chinese speakers had the lowest accuracy (42.9%), followed by the L1 Hebrew speakers (43.5%), the L1 English speakers (64.5%), and the L1 French speakers (66.3%). The L1 Chinese speakers did not differ significantly from the L1 Hebrew speakers but were significantly less accurate than the L1 French speakers, $p < .001$. The L1 Hebrew speakers were also significantly less accurate than the L1 French speakers, $p < .001$.

For pseudoword stimuli testing the ability to delete the body of an item there was a significant simple main effect of L1, $F(2, 133) = 8.40$, $p < .001$, $\eta_p^2 = .11$. The L1 Chinese and the L1 Hebrew speakers had the lowest accuracy (67.9% for each group), followed by the L1 English speakers (77.2%) and the L1 French speakers (82.1%). The L1 Chinese and the L1 Hebrew speakers were not different from one another, but each was significantly less accurate than the L1 French speakers, *p*s < .001. For pseudoword stimuli testing the ability to delete the coda of an item there was a significant simple main effect of L1, $F(2, 133) = 4.79$, $p = .01$, $\eta_p^2 = .07$. The L1 Chinese speakers had the lowest accuracy (82.6%), followed by the L1 French speakers (90.2%), the L1 Hebrew speakers (90.8%), and the L1 English speakers (94.4%). The L1 Chinese speakers were significantly less accurate than the L1 French speakers, $p < .01$, and the L1 Hebrew speakers, $p < .05$, but the difference between the L1 French and the L1 Hebrew speakers was not significant.. The pseudoword deletion accuracy results for the onsets, rimes, bodies, and codas, adjusted for differences in flankers and operation span, are displayed in Figure 15.

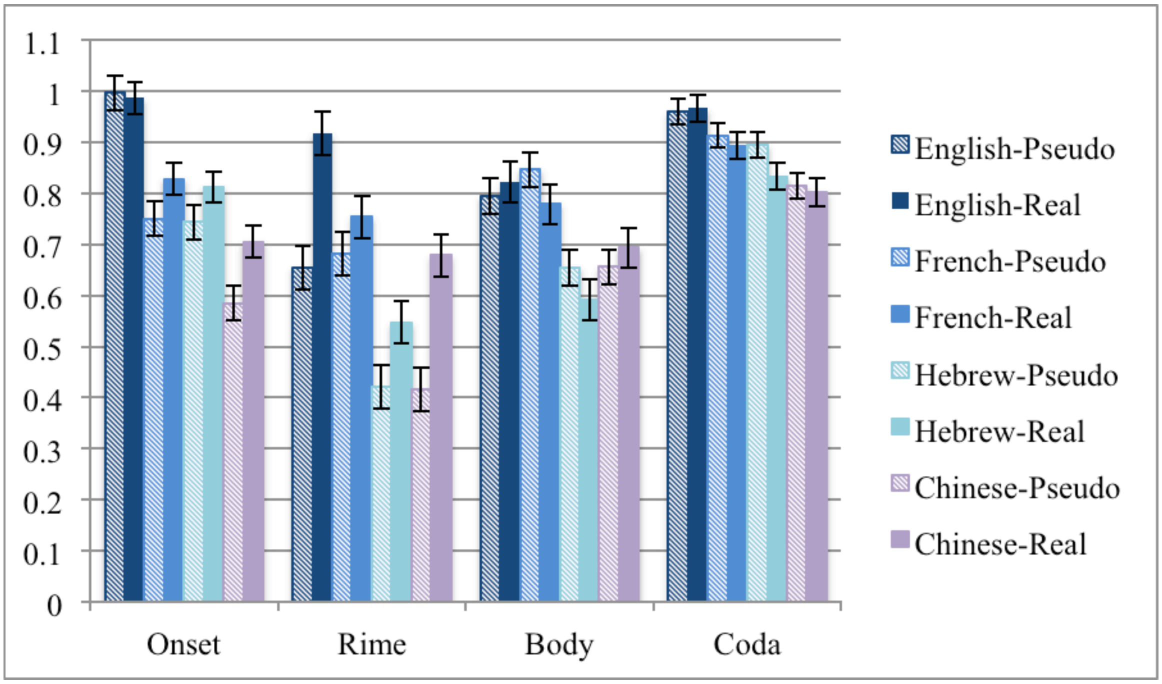

Figure 15. Mean proportion correct on the deletion task, by onset, rime, body, or coda unit, adjusted for performance on flankers and operation span. Pseudo = pseudoword; Real = real word. Error bars represent one standard error.

For pseudoword items testing the ability to delete the first phoneme of an item there was a significant simple main effect of L1, $F(2, 133) = 3.41$, $p < .05$, $\eta_p^2 = .05$. The L1 Chinese speakers had the lowest accuracy (65.8%), followed by the L1 French speakers (72.3%), the L1 Hebrew speakers (78.3%), and the L1 English speakers (89.7%). The L1 French speakers did not differ from the L1 Chinese or the L1 Hebrew speakers, but the L1 Chinese speakers were significantly less accurate than the L1 Hebrew speakers, $p = .01$. For pseudoword items testing the ability to delete a medial consonant in an item there was also a significant simple main effect of L1, $F(2\ 133) = 5.12$, $p < .01$, $\eta_p^2 = .07$. The L1 Chinese speakers had the lowest accuracy (50.5%), followed by the L1 French speakers (62.5%), the L1 Hebrew speakers (69.0%), and the L1 English speakers (78.3%). The L1 Chinese speakers were significantly less accurate than the L1 French speakers, $p = .01$, and the L1 Hebrew speakers, $p < .01$, but the L1 French speakers were not significantly different from the L1 Hebrew speakers. Finally, for pseudoword items testing the ability to delete the final consonant in an item there was a significant simple main effect of L1, $F(2, 133) = 4.61$, $p = .01$, $\eta_p^2 = .07$. The L1 Chinese speakers had the lowest accuracy (66.8%), followed by the L1 French speakers (75.0%), the L1 Hebrew speakers (78.8%), and the L1 English speakers (87.3%). The L1 Chinese speakers were significantly less accurate than the L1 French speakers, $p = .01$, and the L1 Hebrew speakers, $p = .01$, but the L1 French speakers were not significantly different from the L1 Hebrew speakers. The pseudoword deletion accuracy results for the phoneme units, adjusted for differences in flankers and operation span, are displayed in Figure 16.

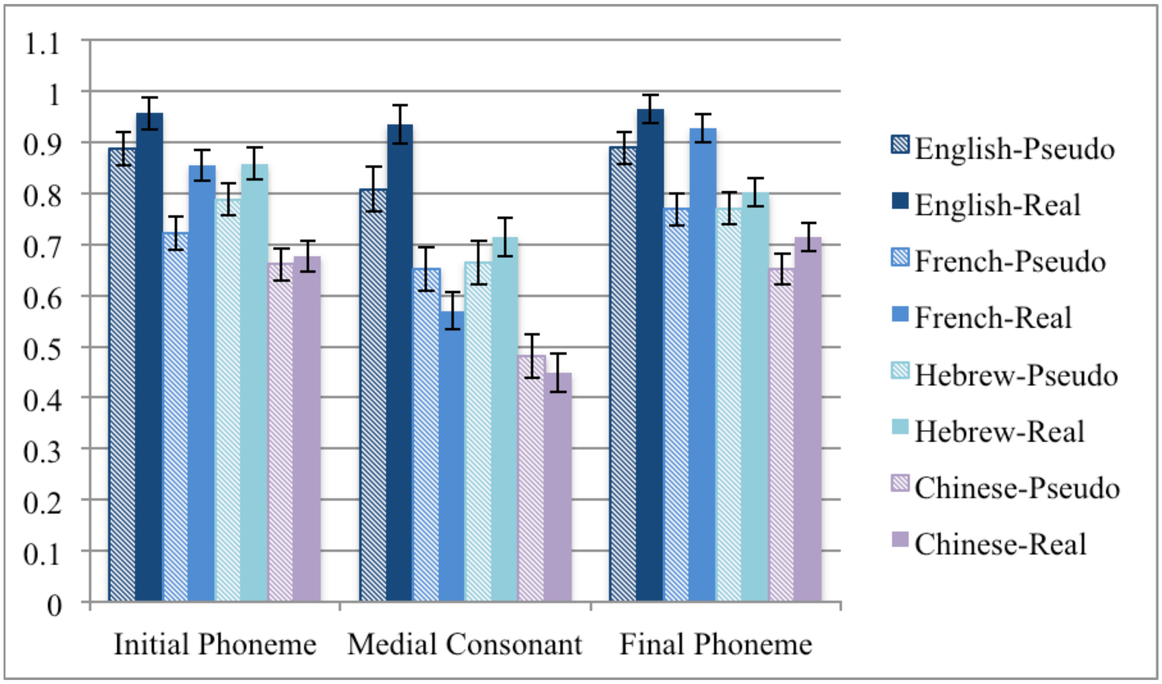

Figure 16. Mean proportion correct on the deletion task, by phoneme, adjusted for performance on flankers and operation span. Pseudo = pseudoword; Real = real word. Error bars represent one standard error.

For real words, there was a significant simple main effect of L1 on items testing the ability to delete the first syllable of a word, $F(2, 133) = 6.59$, $p < .01$, $\eta_p^2 = .09$. The L1 Chinese speakers had the lowest accuracy (92.4%), followed by the L1 French speakers (97.2%), the L1 Hebrew speakers (98.4%), and the L1 English speakers (99.5%). The L1 Chinese speakers were significantly less accurate than the L1 French speakers, $p < .01$, and the L1 Hebrew speakers, $p < .01$, but the L1 French and the L1 Hebrew speakers were not significantly different. There was no main effect of L1 on real word stimuli testing the ability to delete the second syllable of a word. The real word deletion accuracy results for the first and second syllable, adjusted for differences in flankers and operation span, are displayed in Figure 14.

There was a significant simple main effect of L1 on real word stimuli testing the ability to delete the onset of a word, $F(2, 133) = 4.01$, $p < .05$, $\eta_p^2 = .06$. The L1 Chinese speakers had the lowest accuracy (71.7%), followed by the L1 French speakers (81.5%), the L1 Hebrew speakers (82.6%), and the L1 English speakers (97.3%). The L1 Chinese speakers were significantly less accurate than the L1 French speakers, $p = .01$, and the L1 Hebrew speakers, $p < .05$, but the L1 French and the L1 Hebrew speakers were not significantly different from one another.. There was also a significant simple main effect of L1 on real word stimuli testing the ability to delete the rime of a word, $F(2, 133) = 5.29$, $p < .01$, $\eta_p^2 = .07$. The L1 Hebrew speakers had the lowest accuracy (56.5%), followed by the L1 Chinese speakers (69.6%), the L1 French speakers (73.4%), and the L1 English speakers (90.2%). The L1 Hebrew speakers were significantly less accurate than the L1 Chinese speakers, $p < .01$, and the L1 French speakers, $p < .01$, but the L1 Chinese speakers did not differ significantly from the L1 French speakers.

There was a significant simple main effect of L1 on real word stimuli testing the ability to delete the body of a word, $F(2, 133) = 4.68$, $p = .01$, $\eta_p^2 = .07$. The L1 Hebrew speakers had the lowest accuracy (59.8%), followed by the L1 Chinese speakers (70.1%), the L1 French speakers (76.6%), and the L1 English speakers (82.1%). The L1 Hebrew speakers were marginally less accurate than the L1 Chinese speakers, $p = .09$, and were significantly less accurate than the L1 French speakers, $p < .01$, but the L1 Chinese speakers were not significantly different from the L1 French speakers. The simple main effect of L1 was marginally significant for real word stimuli testing the ability to delete the coda of a word, $F(2, 133) = 2.53$, $p = .08$, $\eta_p^2 = .04$. The L1 Chinese speakers had the lowest accuracy (82.1%), followed by the L1 Hebrew speakers (85.3%), the L1 French speakers (87.5%), and the L1 English speakers (94.6%). The L1 Chinese speakers were significantly less accurate than the L1 French speakers; no other L1 differences were significant. The real word deletion accuracy results for the onsets, rimes, bodies, and codas, adjusted for differences in flankers and operation span, are displayed in Figure 15.

For real word stimuli testing the ability to delete the initial phoneme of a word there was a significant simple main effect of L1, $F(2, 133) = 9.59$, $p < .001$, $\eta_p^2 = .13$. The L1 Chinese speakers had the lowest accuracy (68.5%), followed by the L1 French speakers (84.2%), the L1 Hebrew speakers (86.4%), and the L1 English speakers (95.7%). The L1 Chinese speakers were significantly less accurate than the L1 French speakers, $p < .001$, and the L1 Hebrew speakers, $p < .001$, but the difference between the L1 French and the L1 Hebrew speakers was not significant. There was a significant simple main effect of L1 on real word stimuli testing the ability to delete a medial consonant in a word, $F(2, 133) = 11.07$, $p < .001$, $\eta_p^2 = .14$. The L1 Chinese speakers had the lowest accuracy (46.7%), followed by the L1 French speakers (54.9%), the L1 Hebrew speakers (73.4%), and the L1 English speakers (91.8%). The L1 Chinese speakers were significantly less accurate than the L1 French speakers, $p = .02$, and the L1 Hebrew speakers, $p < .001$. The L1 French speakers were also significantly less accurate than the L1 Hebrew speakers, $p < .05$. Finally, there was a significant simple main effect of L1 on real word stimuli testing the ability to delete the final phoneme in a word, $F(2, 133) = 12.53$, $p < .001$, $\eta_p^2 = .16$. The L1 Chinese speakers had the lowest accuracy (72.3%), followed by the L1 Hebrew speakers (81.0%), the L1 French speakers (91.8%), and the L1 English speakers (95.7%). The L1 Chinese speakers were significantly less accurate than the L1 Hebrew speakers, $p < .01$, and the L1 French speakers, $p < .001$. The L1 Hebrew speakers were marginally less significant than the L1 French speakers, $p = .074$. The real word deletion accuracy results for the phoneme units, adjusted for differences in flankers and operation span, are displayed in Figure 16.

5.2.12 Deletion word location RTs

The results for word location (beginning, middle, or end) were again examined first; the raw means and standard deviations are presented in Table 9. The main effects of lexicality, location, and L1 were all non-significant (see Figure 17). The interactions between lexicality and L1, location and L1, and lexicality and location were also non-significant. There was no main effect of flankers, although the main effect of operation span size was marginally significant, $F(1, 111) = 3.48$, $p = .07$, $\eta_p^2 = .03$. Correlations revealed that the relationship between overall RT and operation span size was negative for all L1s, although it was non-significant for the L1 English, L1 French, and L1 Chinese speakers, and only marginally significant for the L1 Hebrew speakers, $r = -.27$, $p = .07$. These correlations reveal that in general, participants with a higher working memory capacity had faster RTs on the deletion task.

Neither flankers nor operation span size interacted with lexicality or location. The three-way interaction between lexicality, location, and flankers was significant, $F(2, 222) = 4.08$, $p = .02$, $\eta_p^2 = .04$, although the three-way interaction between lexicality, location, and operation span size was not significant. To examine the significant lexicality x location x flankers interaction, correlations were calculated for each L1 group between flankers and accuracy for items testing the ability to delete a phonological unit at the beginning, in the middle, and at the end of a word, for pseudoword and real word stimuli separately. However, none of these correlations approached significance for any L1 group. Finally, the three-way interaction between lexicality, location, and L1 was significant, $F(4, 222) = 2.60$, $p = .04$, $\eta_p^2 = .05$. To follow up on this three-way interaction, the L1 x location interaction was examined separately for pseudowords and for real words.

Table 9. Deletion RTs (in ms) by location of the phonological unit inside the word and L1.

		L1			
Lexicality	Location	English	French	Hebrew	Chinese
Pseudo-words	Beginning	1606.01 (424.32)	2108.96 (651.43)	1805.10 (411.13)	1977.20 (750.56)
	Middle	2033.56 (774.62)	2507.63 (772.82)	2279.66 (905.26)	2782.76 (1156.73)
	End	1712.12 (490.08)	2141.99 (586.74)	2048.08 (622.97)	2221.72 (866.03)
Real words	Beginning	1559.62 (481.47)	1981.34 (602.47)	1733.95 (400.77)	2072.24 (713.66)
	Middle	1941.06 (600.84)	2540.54 (927.91)	2329.89 (889.90)	2481.77 (1025.48)
	End	1597.87 (449.26)	2043.00 (537.56)	1828.00 (535.62)	2084.54 (684.06)

Note. Standard deviations are presented in parentheses. Results shown are prior to adjustment for differences in scores on the flankers and operation span tasks.

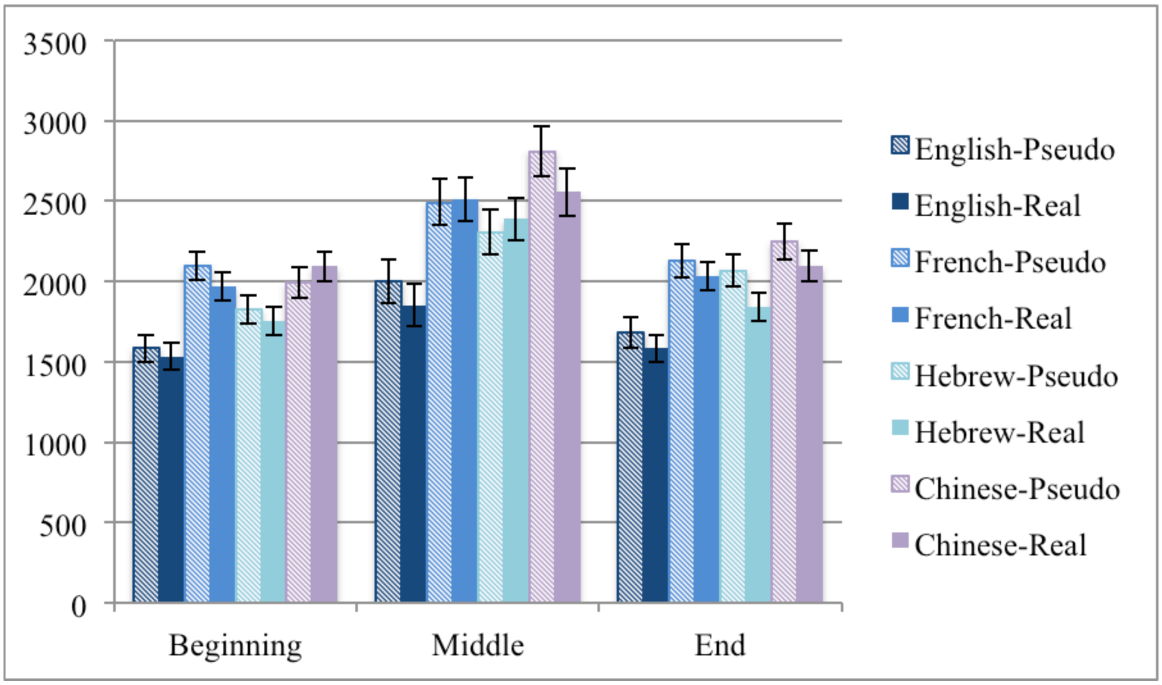

Figure 17. Mean RTs for performance on the deletion task, by location of the phonological unit inside the word and L1, adjusted for performance on flankers and operation span. Pseudo = pseudoword; Real = real word. Error bars represent one standard error.

For pseudowords the L1 x location interaction was significant, Pillai's Trace = .09, $F(4, 222) = 2.67$, $p = .03$, $\eta_p^2 = .05$. To follow up on the significant L1 x location interaction, the simple main effect of L1 was again examined separately for each location. The simple main effect of L1 was not significant for pseudoword stimuli testing awareness of phonological units at the beginning of a word or at the end of a word. However, the simple main effect of L1 was marginally significant for pseudoword stimuli testing awareness of phonological units in the middle of a word, $F(2, 111) = 2.62$, $p = .08$, $\eta_p^2 = .05$. The L1 Chinese speakers had the slowest response times (2782.76 ms), followed by the L1 French speakers (2507.63 ms), the L1 Hebrew speakers (2279.66 ms), and the L1 English speakers (2033.56 ms). Pairwise comparisons revealed that the L1 Chinese speakers were significantly slower than the L1 Hebrew speakers, $p = .02$; no other L1 comparisons were significant.

For real words the L1 x location interaction was not significant, thus the marginal main effect of location on performance was examined collapsing across L1. This analysis revealed no significant effect of location, thus no further comparisons were made.

5.2.13 Deletion phonological unit RTs

The response time results for phonological unit were examined next; the raw means and standard deviations are presented in Table 10. The main effects of lexicality, phonological unit, and L1 were not significant. The interaction between lexicality and L1 and the interaction between lexicality and phonological unit were not significant, but the interaction between phonological unit and L1 was significant, $F(16, 672) = 1.81$, $p < .05$, $\eta_p^2 = .04$. The main effects of flankers and operation span size were not significant, and neither variable interacted with lexicality or phonological unit. The three-way interaction between lexicality, phonological unit, and flankers was not significant, $F(8, 992) = 1.06$, $p = .39$. The three-way interactions between lexicality, phonological unit, and flankers and between lexicality, phonological unit, and operation span size were also not significant. However, the three-way interaction between lexicality, phonological unit, and L1 was significant, $F(16, 672) = 3.51$, $p < .001$, $\eta_p^2 = .08$.

To follow up on the significant three-way interaction, the two-way interaction of L1 x phonological unit was examined separately for pseudoword and for real word stimuli. This L1 x phonological unit interaction was significant for pseudowords, Pillai's Trace = .41, $F(16, 156) = 2.48$, $p < .01$, $\eta_p^2 =$,20. To follow up on this interaction, the simple main effect of L1 was examined for each phonological unit for pseudowords alone.

Table 10. Deletion RTs (in ms) by phonological unit and L1.

Lexicality	Location	L1 English	L1 French	L1 Hebrew	L1 Chinese
Pseudo-words	First Syllable	1749.58 (598.59)	2141.58 (706.35)	1994.99 (558.57)	2134.76 (905.15)
	Second Syllable	1909.49 (610.83)	2425.89 (855.88)	2683.17 (1101.44)	3286.55 (1237.11)
	Onset	1596.57 (469.64)	2153.97 (775.17)	1930.96 (519.95)	2228.13 (849.32)
	Rime	1676.91 (666.48)	2168.49 (673.96)	2173.45 (858.92)	2321.52 (1195.55)
	Body	1608.70 (570.40)	2170.13 (754.39)	1676.73 (486.54)	1948.07 (755.25)
	Coda	1701.96 (498.63)	2132.50 (561.05)	1789.48 (428.03)	2086.71 (934.06)
	Initial Phoneme	1519.98 (489.90)	1889.07 (654.49)	1728.96 (520.98)	1713.42 (790.63)
	Medial Consonant	1957.80 (656.83)	2481.29 (780.77)	2273.90 (969.62)	2715.69 (1087.56)
	Final Phoneme	1549.07 (513.82)	1998.61 (622.02)	1732.74 (604.80)	1925.17 (789.49)
Real words	First Syllable	1460.57 (544.18)	1864.31 (627.95)	1668.57 (455.56)	1884.10 (742.82)
	Second Syllable	1456.81 (526.27)	1925.99 (647.30)	1828.83 (544.63)	1911.27 (571.36)
	Onset	1452.37 (538.84)	2049.69 (806.00)	1815.57 (541.32)	2252.81 (976.03)
	Rime	1736.31 (591.98)	2217.64 (751.84)	2093.16 (1049.01)	2661.05 (984.66)
	Body	1866.82 (645.76)	2181.29 (734.63)	1866.95 (510.28)	2377.80 (1023.98)
	Coda	1719.17 (526.16)	2128.12 (675.96)	1832.41 (675.04)	2249.22 (845.06)
	Initial Phoneme	1504.90 (593.56)	1838.70 (539.63)	1677.39 (442.77)	1954.96 (838.00)
	Medial Consonant	1959.80 (616.52)	2560.32 (945.10)	2254.13 (702.04)	2411.83 (1078.35)
	Final Phoneme	1507.65 (452.10)	1948.67 (501.34)	1781.82 (524.15)	1758.19 (542.65)

Note. Standard deviations are presented in parentheses. Results shown are prior to adjustment for differences in scores on the flankers and operation span tasks.

For pseudoword stimuli the simple main effect of L1 was not significant for items testing the ability to delete the first syllable of an item. However, the simple main effect of L1 was significant for pseudoword stimuli testing the ability to delete the second syllable of an item, $F(2, 84) = 4.97$, $p < .01$, $\eta_p^2 = .11$. The L1 Chinese speakers had the slowest response times (3286.55 ms), followed by the L1 Hebrew speakers (2683.17 ms), the L1 French speakers (2425.89 ms), and the L1 English speakers (1909.49 ms). The L1 Chinese speakers were significantly slower than the L1 Hebrew speakers, $p < .05$, and the L1 French speakers, $p < .01$, but the difference between the L1 Hebrew and the L1 French speakers was not significant. The pseudoword deletion response time results for the first and second syllable, adjusted for differences in flankers and operation span, are displayed in Figure 18.

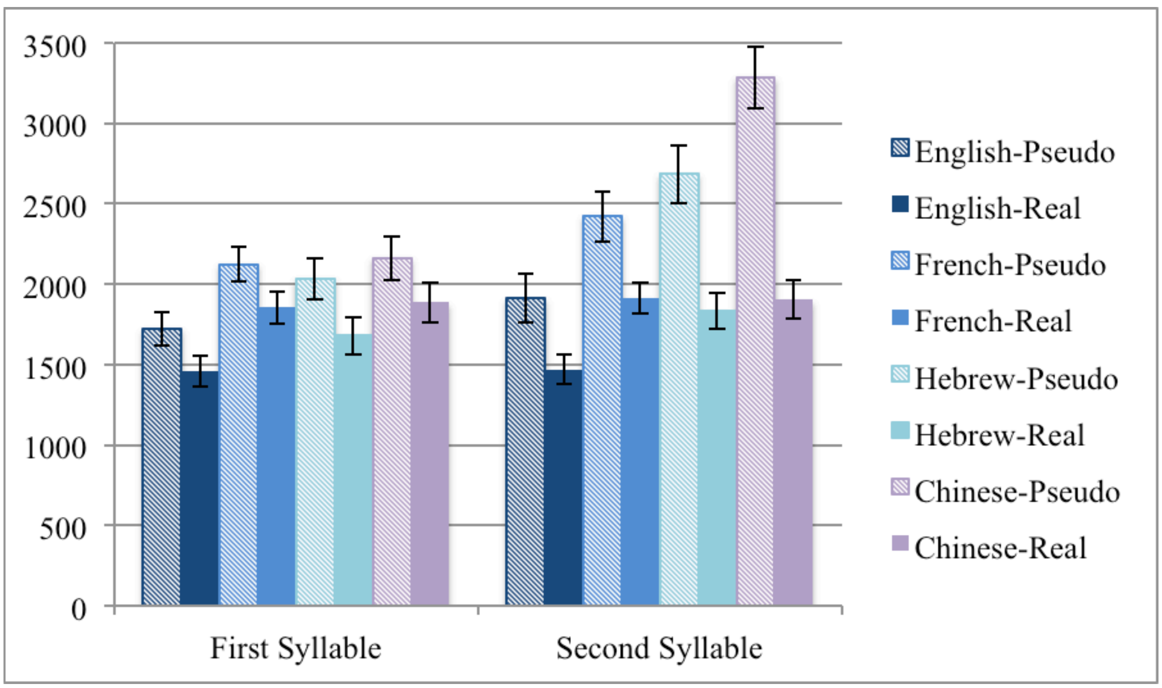

Figure 18. Mean RTs for performance on the deletion task, by syllable and L1, adjusted for performance on flankers and operation span. Pseudo = pseudoword; Real = real word. Error bars represent one standard error.

The simple main effect of L1 was not significant for pseudoword stimuli testing the ability to delete the onset or the rime of an item. However, the simple main effect of L1 was marginally significant for pseudoword stimuli testing the ability to delete the body of an item, $F(2, 84) = 3.10$, $p = .05$, $\eta_p^2 = .07$. The L1 French speakers had the slowest response times (2170.13 ms), followed by the L1 Chinese speakers (1948.07 ms), the L1 Hebrew speakers (1673.73 ms), and the L1 English speakers (1608.70 ms). The L1 French speakers were significantly slower than the L1 Hebrew speakers, $p < .05$; no other L1 differences were significant. The simple main effect of L1 was also not significant for pseudoword stimuli testing the ability to delete the coda of an item. The pseudoword deletion response time results for the onsets, rimes, bodies, and codas, adjusted for differences in flankers and operation span, are displayed in Figure 19.

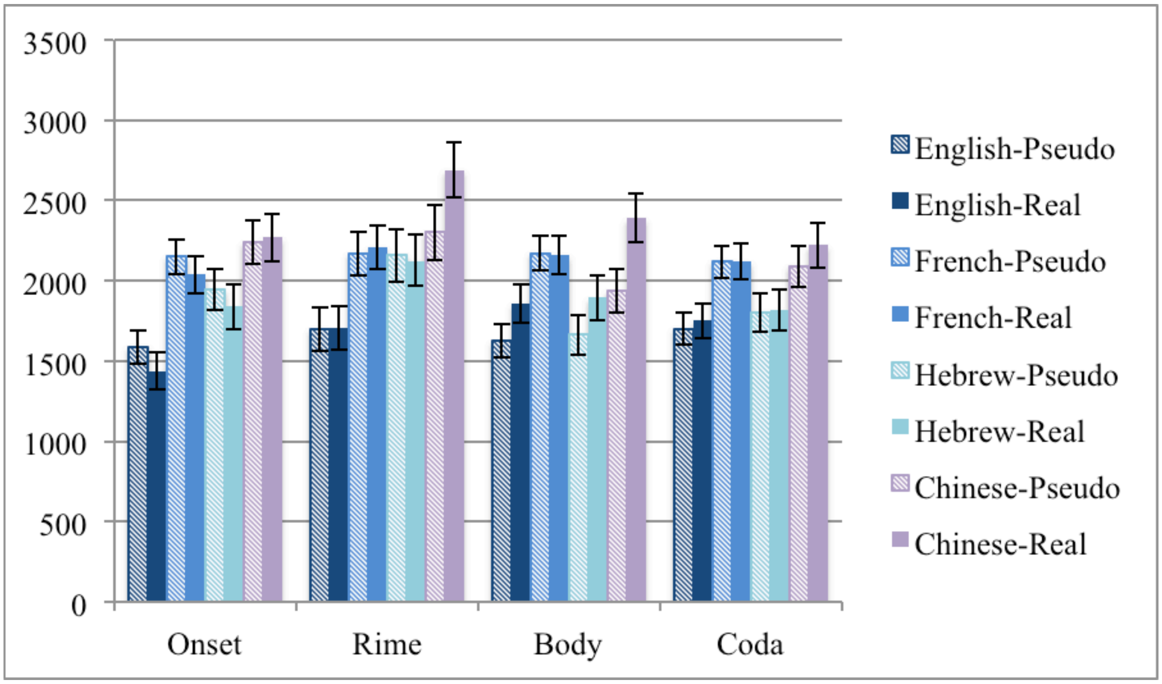

Figure 19. Mean RTs for performance on the deletion task, by onset, rime, body, and coda unit and L1, adjusted for performance on flankers and operation span. Pseudo = pseudoword; Real = real word. Error bars represent one standard error.

The simple main effect of L1 was not significant for pseudoword stimuli testing the ability to delete an initial phoneme, a medial consonant, or a final phoneme of an item. The pseudoword deletion response time results for the initial segment, medial consonant, and final segment, adjusted for differences in flankers and operation span, are displayed in Figure 20.

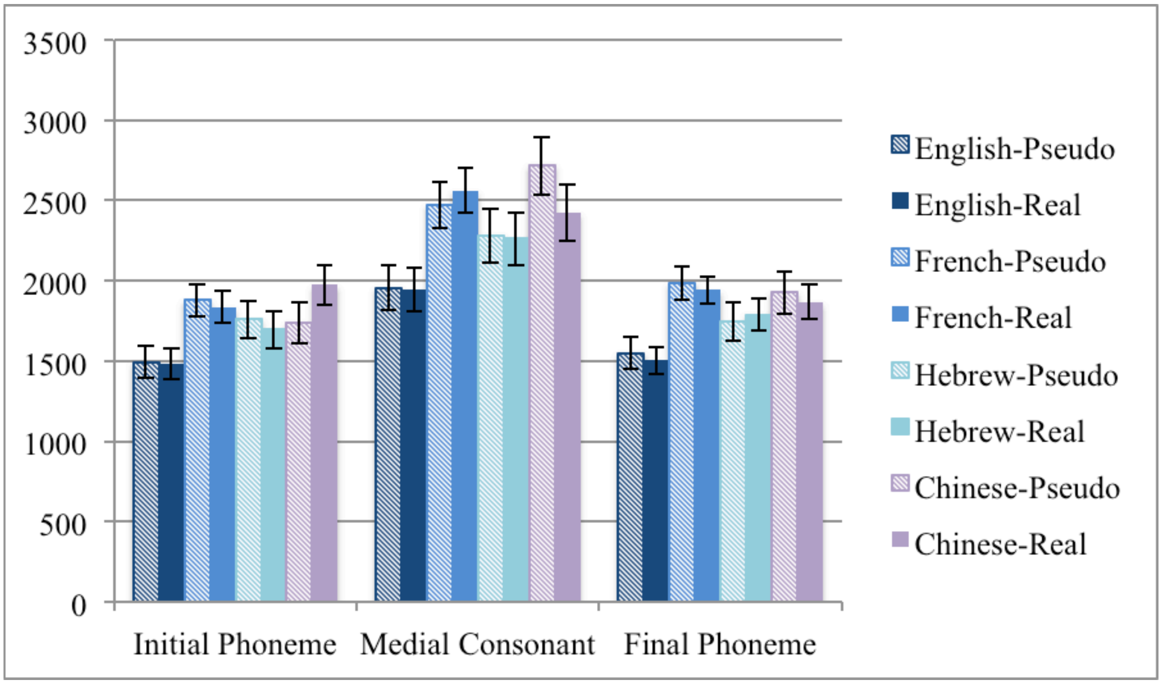

Figure 20. Mean RTs for performance on the deletion task, by phoneme and L1, adjusted for performance on flankers and operation span. Pseudo = pseudoword; Real = real word. Error bars represent one standard error.

For real words, the two-way interaction of L1 x phonological unit was not significant. Due to the lack of interaction, pairwise comparisons among the three non-native participant groups were examined collapsing across phonological units. These pairwise comparisons revealed no significant differences among L1 groups; no further comparisons were examined.

5.2.14 Summary: Deletion Task

Again, consistent with previous research testing phonological awareness for units at different locations within a word (Stage & Wagner, 1992), performance was significantly lower for items targeting word-medial consonants, compared to word-initial or word-final consonants. Although there was a slight trend for accuracy to be higher and RTs to be faster for units at the beginning of a word than those at the end of a word, similar to what was found by Stage and Wagner (1992), this pattern was stronger for pseudoword than real word items, and was generally quite small numerically..

Looking at the results for the different phonological units, the results were somewhat more consistent with our predictions than the findings from the oddity task. We again expected that the alphabetic L1 speakers (English, French) would out-perform the participants from a non-alphabetic L1 (Hebrew, Chinese) on items testing phonological awareness for single phonemes, regardless of position in the word (Goswami & Bryant, 1990; Homer, 2009). As on the oddity task, the L1 English speakers were the most accurate group for phonemes, consistent with our predictions. Again consistent with our predictions and in contrast to the oddity task, the L1 Chinese speakers had a significantly lower level of accuracy for single phonemes, compared to the other L1 groups. The L1 French speakers were thus more accurate than the L1 Chinese speakers, but tended to be numerically less accurate than the L1 Hebrew speakers, although this difference was generally not significant. The L1 French speakers did do better than the L1 Hebrew speakers on one comparison: final phonemes in real words. One interpretation of this pattern of results is that it is more informative about the L1 Hebrew speakers than the L1 French speakers: despite the consonant-focused characteristics of the Hebrew orthography and the general lack of written vowels, the segmental structure of the orthography supports the development of phonemic awareness at a level comparable to that of L1 readers of true alphabetic languages. Considering the RT results, there were generally few L1 differences for the deletion task. However, the numerical differences among L1 groups were consistent with the results with the oddity task: the L1 French speakers returned to their lower level of performance, tending to be the slowest to respond for both pseudoword and word items.

It is worth emphasizing the fact that in direct contrast to the oddity task, the L1 Chinese speakers had quite low levels of accuracy on the deletion task. Given that the two tasks are designed (and often used) to tap the same type of phonological skill using the same phonological units, it is likely that the difference in task demands is driving this discrepant pattern of results. The difference in performance across the two tasks varied both by participant L1 and by phonological unit, however. This reveals the large impact of task demands on the measurement of individual skill levels, as well as the complex interactions of task demands with other factors. The results also have important implications for researchers using these tasks in both theoretical and applied contexts, as discussed further below.

Moving to the body and coda units, we again predicted that the L1 Hebrew speakers would show the best performance on these items, due to the structure of their L1 orthography (Ben-Dror et al., 1995; Frost, 2012; Tolchinsky et al., 2012). This pattern was not evident in our data. Although the L1 Hebrew speakers did relatively well deleting codas from pseudowords (they had the second-highest level of accuracy, after the L1 English speakers), they were not significantly better than the L1 French speakers. The pattern was similar for real word stimuli: the L1 Hebrew speakers had a moderate level of accuracy that did not differ significantly from the L1 Chinese or the L1 French speakers. In addition, their ability to delete the body of an item was quite poor, tied for lowest accuracy for pseudoword items and significantly lower accuracy than all other groups for real word items. The only case in which they showed any evidence of an advantage for processing body/coda divisions was in RTs. Although numerically the L1 Hebrew speakers tended to respond to deletion items targeting body/coda divisions faster than the L1 French and the L1 Chinese speakers, these L1 differences were not significant.

Looking at the items testing awareness of onsets and rimes, as with the oddity task we predicted that the L1 alphabetic speakers would perform better than the L1 non-alphabetic speakers on items testing awareness of these units (De Cara & Goswami, 2003; Ziegler & Goswami, 2005). This prediction was only partially supported. The L1 English and the L1 French speakers had the highest accuracies for units that required deleting the rime, and in fact for pseudoword stimuli the L1 French speakers did numerically better than the L1 English speakers. This is consistent with previous evidence demonstrating the importance and salience of rime units in alphabetic languages, particularly those with a deep orthography like French and English (De Cara & Goswami, 2003; Ziegler & Goswami, 2005). However, the L1 French speakers did not show a distinct advantage on items that required them to delete an onset (thus isolating the rime), performing numerically (although not significantly) worse than the L1 Hebrew speakers for both pseudoword and real word items. This was also the pattern that was generally found for RTs: the L1 French speakers were numerically (but not significantly) slower than the L1 Hebrew speakers for onset/rime stimuli, similar to the pattern for all other phonological units.

Finally, looking at performance on syllables, we again expected high performance overall on these units because of their salience and early development (Goswami & Bryant, 1990). Although performance was generally high for the real word items, with all mean accuracies at 88.5% or above, accuracy was much lower for the pseudoword items, thus showing limited evidence of an advantage for syllables, without lexical support. In addition, RTs were not distinctly faster for syllables than for other phonological units, showing a further lack of performance advantage for syllable items. We also expected that the L1 Chinese speakers might do better than the other L1 groups on the syllable items in particular, due to their status as major phonological, morphological, orthographic, and semantic units in Chinese (McBride-Chang et al., 2004; Tolchinsky et al., 2012). This prediction was not supported. There was no significant L1 effect for first syllables with pseudoword items, or for second syllables with real word items; when there was a significant L1 effect, the L1 Chinese speakers were in fact less accurate than either the L1 Hebrew or the L1 French speakers. The L1 Chinese speakers thus demonstrated difficulty with the productive deletion task across phonological units, and not just for items testing awareness of single phonemes.

5.2.15 Summary and Discussion: Phonological Awareness

A number of interesting results were found in participants' performance on the two phonological awareness tasks. First, the results demonstrate that task demands have a large impact on task performance, particularly for non-native speakers of the language being used for testing. Accuracies on the oddity task, which required only listening and receptive judgments of shared phonological units, were generally quite high, at or above 80-85% correct, with the exception of items testing medial phonemes. Much lower levels of accuracy were found for the deletion task, ranging down to 40-50% accurate in some cases. The difference in performance can also be seen in RTs: RTs for the deletion task were often four times as long, or even longer, than RTs to comparable items on the oddity task.

This impact of task is thus an important consideration for researchers and educators who are interested in measuring students' phonological awareness. Receptive tasks may be somewhat faster and easier to administer and score, and are widely used in studies of phonological awareness because of the focus on initial literacy development in children. Due to differences in cognitive maturity, these types of receptive tasks are indeed appropriate for younger learners (e.g., Anthony et al., 2011). However, they may not be an effective assessment tool in adult learners. As was seen in the current study, using receptive tasks with adults (particularly those who have already gained literacy in their L1 and are used to functioning in an academic setting) risks leading to ceiling effects for accuracy, meaning that varying levels of phonological skill cannot be distinguished with these tasks. Although it may be theoretically possible to distinguish different levels of skill on the basis of RTs, examinations of phonological awareness in English have focused almost exclusively on accuracy rather than efficiency (Share, 2008), resulting in a general lack of information regarding typical or atypical RT patterns.

Based on the current results, therefore, it seems that productive tasks are the best fit for researchers working with adult participants. There are two brief caveats to this position that we would like to mention. First, the fact that the L1 Chinese speakers performed quite well on the receptive (oddity) task but not on the productive (deletion) task suggests that despite the risk of ceiling effects, receptive tasks may be capable of capturing weak and narrow but existent phonological skills. The current results suggest that the L1 Chinese speakers had a minimum level of phonological skill that allowed them to perform well on the oddity task, but that they struggled to use these abilities productively. Second, this is one of only a small number of studies to compare phonological awareness performance on the basis of different task types and task demands, thus this type of comparison needs to be made multiple times with different participant groups in order to determine the reliability and stability of the results. With these two caveats in mind, for most researchers using productive tasks of phonological awareness will likely be most appropriate for working involving literate adult L2 participants.

Interestingly, many of our specific predictions regarding which L1 groups would perform best on phonological awareness tasks targeting specific phonological units were not supported by the results. Although previous research has suggested that speakers of some languages, including Korean (Kim, 2007, 2009; Yoon et al., 2002) and Hebrew (Ben-Dror et al., 1995; Tolchinsky et al., 2012) show a preference for dividing words into body and coda units (as opposed to onset and rime units), this preference was not found in the current results. In fact, the L1 Hebrew speakers often performed quite poorly when the task required them to identify or manipulate words divided in this way. This result also challenges the existence of a preference for dividing words into body/coda units. Given the relatively small number of studies that have reported this type of preference (Ben-Dror et al., 1995; Frost, 2012; Kim, 2007, 2009; Tolchinsky et al., 2012; Yoon et al., 2002), much additional research is needed to establish whether, and in which languages, this type of word division preference exists.

Another possible explanation for pattern of results that we found is the fact that the tasks were performed in the participants' L2 (English) rather than their L1 (Hebrew). The only study we are aware of that specifically targeted body/coda units for manipulation in L1 Hebrew speakers (rather than finding a preference for these units through error analysis) required participants to perform the task in Hebrew, not in English. If this is indeed the reason for the lack of body/coda preference in the current study, it suggests that L1 phonological structure preferences do not necessarily transfer to an L2, particularly in readers with many years of study and exposure to that L2. This is evidence against our hypothesis that the preferred phonological grain size that develops in L1 literacy (Ziegler & Goswami, 2005) would also be the preferred phonological grain size for performing phonological and literacy tasks in the L2. This may be the case for two reasons: either the preferred unit for L1 does not automatically transfer to the L2, or the preferred unit for L1 transfers to L2 but this preference decreases in strength over time, so that more proficient learners no longer show evidence of this preference. Given the narrow range of proficiency levels targeted in the current study, the present data cannot be used to distinguish these two possibilities; future work involving a broader range of proficiency levels either in cross-sectional or longitudinal research is needed to clarify this point.

Our prediction that the L1 French speakers would perform well on the items targeting single phonemes was not strongly supported. In fact, the L1 French speakers did relatively poorly overall on the phonological awareness tasks, a finding that is inconsistent with previous research showing that readers of an alphabetic L1 do better on tasks of phonemic awareness than readers of a non-alphabetic L1 (e.g., Holm & Dodd, 1996; Liow & Poon, 1998; McBride-Chang et al., 2004; Wade-Woolley, 1999; Wang et al., 2003). Although it is not immediately clear why this pattern was found, at least two possibilities are brought to mind. First, the relatively deep orthography of French may play a role. Although phonological awareness and phonological skills in general are crucial for literacy across languages, and particularly alphabetic languages, these skills develop more quickly and more accurately in learners with more transparent writing systems (Seymour, Aro, & Erskine, 2003; Ziegler & Goswami, 2005). The relative depth of the French alphabet may thus be a contributing factor to their surprisingly low performance on these tasks. Second, anecdotal evidence from participants' level of English conversational fluency and comfort interacting with the experimenter in English suggests that although the learners in the four L1 groups were matched on four objective measures of English proficiency, the L1 French speakers may have relatively less fluency and confidence in their phonological skills, thus impacting their performance on these tasks.

Finally, it is important to recognize that although the L1 French speakers did not always perform as well as predicted, they were more accurate than the L1 Chinese speakers on the productive (deletion) task and generally not differ significantly from the L1 Hebrew speakers.

As mentioned above, it is possible that this pattern of results is more informative regarding the L1 Hebrew speakers than the L1 French speakers. Specifically, it suggests that the segmental nature of the Hebrew orthography does support the development of fine-grained, phonemic awareness. Given the relatively few studies that have included analyses of L1 abjad speakers, this finding is an important contribution for understanding the impact of these orthographies on L2 literacy skills. Although some studies have found that speakers of abjad L1s struggle with literacy activities that particularly target sensitivity to or knowledge of vowels (Fender, 2008; Hayes-Harb, 2006; K. I. Martin & Juffs, In revision; Ryan & Meara, 1991, 1996; Saigh & Schmitt, 2012), other studies have shown that readers of these L1s tend to behave more similar to speakers of alphabetic than non-alphabetic L1s (Akamatsu, 1999, 2003; Koda, 1989a). The current results from the phonological awareness tasks support this latter finding: with regard to phonological skills, L1 abjad speakers tend to pattern more like speakers of alphabetic L1s than morphosyllabic L1s.

The highly discrepant levels of performance by the L1 Chinese speakers on the oddity versus the deletion tasks are also an important finding. In addition to demonstrating the clear potential for task effects on performance, the high level of accuracy on the receptive oddity task demonstrates that L1 Chinese speakers with some phonological writing experience (in this case via *zhuyin fuhao*) can develop phonological awareness, even at the phoneme level. This is consistent with previous research showing that L1 Chinese speakers with *pinyin* experience also develop phonemic awareness (Holm & Dodd, 1996; McBride-Chang et al., 2004). However, the L1 Chinese speakers' poor performance on the productive deletion task suggests that not all learners are automatically able to apply these underlying phonological skills to productive tasks that require active manipulation of sub-lexical units and articulation of a response. Regardless of their relative lack of productive skills, though, it appears that foundational phonological skills do exist in these learners. This is an important point for curriculum and materials writers and language instructors to consider, because it provides a natural starting point to directly target the further development of phonological skills and the application of these skills to broader literacy tasks.

Finally, it is useful to note that we can also see some evidence of direct L1 phonological influence (as opposed to orthographic influence) in the results. Considering the oddity task and the pseudoword stimuli in the deletion task, we can see that L1 French speakers did noticeably better on items testing awareness of the second syllable in a disyllabic word than the first syllable. This advantage is consistent with the characteristics of French phonology: unless a word ends in a schwa-type vowel, French words generally receive stress on the final syllable of the word (Tranel, 1987). This stress placement may mean that L1 French speakers are relatively more attuned to the final syllable of words, thus explaining their relative advantage on second syllables compared to first syllables. In addition, although the difference was not always reliable, the L1 French speakers tended to do numerically better on items in both tasks that tested awareness of rimes. This finding is also consistent with the structure of the French lexicon and the salience of phonological rimes as an organizing structure for the language (e.g., De Cara & Goswami, 2003). There is also some evidence of L1 phonology for the L1 Chinese speakers' performance on the oddity task. Although this group of learners did quite well on the oddity task, they did relatively worse on detecting oddity codas than onsets, rimes, or body units. This relative lack of accuracy in detecting codas is consistent with the strict limitations on phonemes that can appear in coda position in Chinese syllables, which has the effect of reducing native Chinese speakers' need to attend to sounds in this position.

5.2.16 Orthographic Knowledge

5.2.17 Word/pseudohomophone discrimination

The word/pseudohomophone discrimination task required participants to determine which of two phonologically-plausible spellings of an English word was the correct spelling, and was designed to test whole-word orthographic (specifically, spelling) knowledge. Mixed ANOVAs were used to analyze mean accuracies and reaction times. Item type (misspelling of a consonant or a vowel grapheme) was a within-subjects variable, L1 was a between-subjects variable, and the standardized incongruent-congruent RT difference on the flankers task and the operation span maximum span length score were covariates. Only data from the non-native speakers were included in the ANOVA; data from the native English speakers are included in the tables and figures for comparison purposes only.

5.2.18 Accuracy

Means and standard deviations for participants' accuracy, by misspelling type (consonant or vowel) and L1, are in Table 11. There was no main effect of consonant versus vowel misspelling, but the main effect of L1 was marginally significant, $F(2, 133) = 2.72$, $p = .07$, $\eta_p^2 = .04$. There was no main effect of flankers, no main effect of operation span size, and neither variable interacted with misspelling type.. However, there was a significant interaction between misspelling type and L1, $F(2, 133) = 3.86$, $p < .05$, $\eta_p^2 = .06$.

Table 11. Proportion correct on the word-pseudohomophone discrimination task, by L1 and misspelling type.

L1 Group	Consonant	Vowel
English	99.02 (1.71)	98.59 (2.02)
French	92.61 (5.77)	91.52 (7.65)
Hebrew	94.95 (6.00)	91.03 (8.34)
Chinese	96.47 (4.27)	94.24 (5.47)

Note. Standard deviations are given in parentheses.

To follow up on the significant misspelling type x L1 interaction, paired samples *t*-tests were conducted to compare accuracy on items with consonant versus vowel misspellings for each L1 group. The *t*-tests revealed that the difference was not significant for the L1 English or the L1 French speakers, but that the L1 Hebrew and the L1 Chinese speakers each had a significantly higher accuracy on items with consonant misspellings than on items with vowel misspellings, $p < .001$ and $p = .001$, respectively.

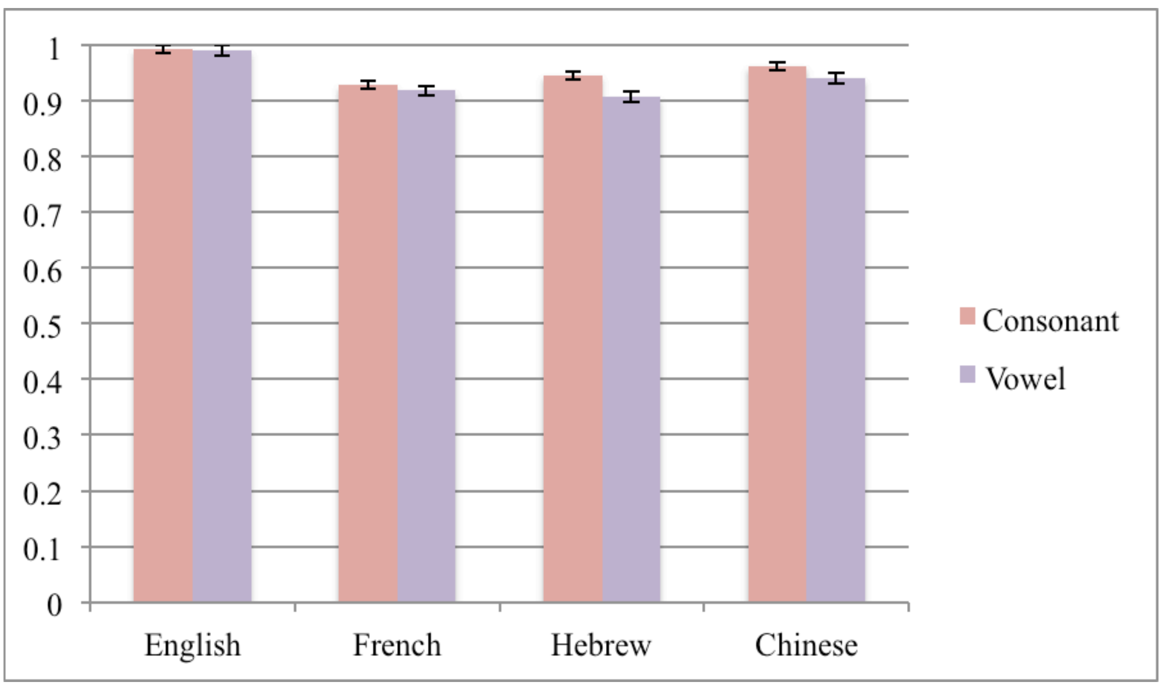

Figure 21. Mean proportion correct on the word-pseudohomophone discrimination task, by L1 and misspelling type. Error bars represent one standard error.

5.2.19 RTs

Means and standard deviations for participants' reaction times, by misspelling type (consonant or vowel) and L1, are in Table 12. There was no main effect of consonant versus vowel misspelling, L1, flankers, or operation span size. The interactions between misspelling type and flankers and between misspelling type and operation span size were not significant. In addition, the interaction between misspelling type and L1 was not significant. No further comparisons were made; the means for the RT results can be seen in Figure 22.

Table 12. RTs (in ms) on the word-pseudohomophone discrimination task.

L1 Group	Consonant	Vowel
English	-658.81 (218.02)	-695.58 (222.91)
French	33.51 (579.25)	16.81 (572.93)
Hebrew	-184.18 (363.22)	-189.93 (404.89)
Chinese	-221.12 (474.17)	-72.72 (337.26)

Note. Standard deviations are given in parentheses. RTs were measured from the end of the sound file that presented the word aurally; a negative RT therefore indicates that participants responded before the end of the sound file. This was possible because the word did not need to be heard in order for a judgment to be made.

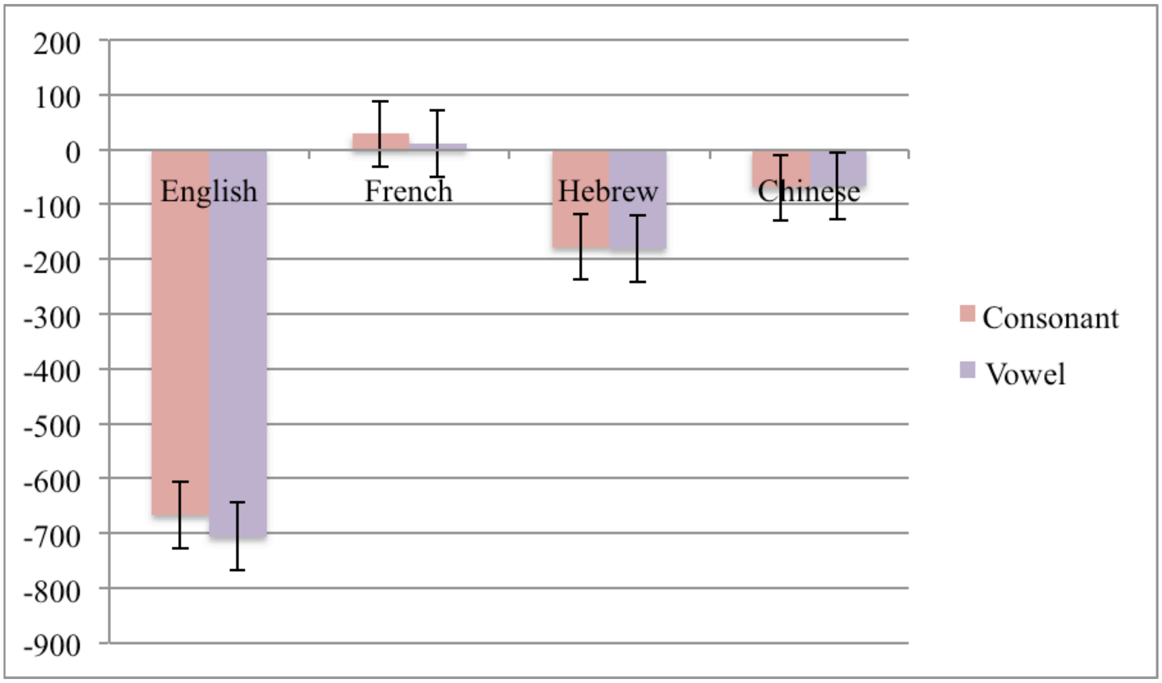

Figure 22. Mean RTs on the word-pseudohomophone discrimination task, by L1 and misspelling type. Error bars represent one standard error.

5.2.20 Wordlikeness Judgments

The wordlikeness judgment task presented participants with two non-word letter sequences which varied on positional letter bigram frequency and letter combination legality, and asked them to indicate which they thought looked more like a real English word. This task was designed to test intralexical orthographic knowledge, as opposed to whole-word orthographic knowledge. Mixed ANOVAs were used to analyze mean accuracies and reaction times. Only data from the non-native speakers were included in the ANOVA; data from the native English speakers are included in the tables and figures for comparison purposes only. Comparison type was a within-subjects variable with four levels: frequency (trials on which participants had to choose between an item with a high positional bigram frequency and an item with a low positional bigram frequency), lexicality (trials on which participants had to choose between an item with a legal letter combination and an item with an illegal letter combination), combined frequency and lexicality (trials on which participants had to choose between an item with a legal combination of letters that had a high positional bigram frequency and an item with an illegal combination of letters that had a low positional bigram frequency), and conflicting frequency and legality cues (trials on which participants had to choose between an item with a legal combination of letters that had a low positional bigram frequency and an item with an illegal combination of letters that had a high positional bigram frequency). L1 was a between-subjects variable, and the standardized incongruent-congruent RT difference on the flankers task and the operation span maximum span length score were covariates.

5.2.21 Accuracy

Descriptive statistics for participants' accuracy, by comparison type and L1, are in Table 13. There was a significant main effect of comparison type, $F(3, 399) = 4.24$, $p < .01$, $\eta_p^2 = .03$, and a significant main effect of L1, $F(2, 133) = 17.59$, $p < .001$, $\eta_p^2 = .21$. There was no main effect of flankers, but the main effect of operation span size was significant, $F(1, 133) = 6.19$, $p = .01$, $\eta_p^2 = .04$. The interaction between comparison type and flankers was not significant, but the interaction between comparison type and operation span size was significant, $F(3, 399) = 2.71$, $p < .05$, $\eta_p^2 = .02$. The interaction between comparison type and L1 was also significant, $F(6, 399) = 3.66$, $p = .001$, $\eta_p^2 = .05$.

Table 13. Proportion correct on the wordlikeness judgment task, by L1 and comparison type.

	High vs. low bigram frequency	Legal vs. illegal letter combination	Combined frequency and legality effects	Conflicting frequency and legality cues
English	68.70% (9.66)	84.13% (8.68)	93.59% (7.58)	74.78% (13.66)
French	69.40% (7.80)	84.46% (11.08)	89.02% (13.28)	73.15% (18.33)
Hebrew	64.29% (7.72)	72.50% (13.85)	82.28% (14.09)	60.22% (19.58)
Chinese	68.10% (7.94)	83.10% (9.52)	92.50% (8.48)	75.11% (14.47)

Note. Standard deviations are given in parentheses.

To follow up on this significant interaction, the simple main effect of L1 was examined for each comparison type, followed by pairwise comparisons among the L1 groups as appropriate. The significance of all pairwise comparisons was adjusted using the Šidák correction for multiple comparisons. There was a significant simple main effect of L1 for comparisons between an item with high versus low positional bigram frequency, $F(2, 133) = 4.77$, $p = .01$, $\eta_p^2 = .07$. The L1 Hebrew speakers had the lowest accuracy (64.3%), followed by the L1 Chinese speakers (68.1%), the L1 English speakers (68.7%), and the L1 French speakers (69.4%). The L1 Hebrew speakers were marginally less accurate than the L1 Chinese speakers, $p = .07$, and significantly less accurate than the L1 French speakers, $p = .01$; the difference between the L1 Chinese and the L1 French speakers was not significant. There was also a significant simple main effect of L1 for comparisons between an item with legal versus illegal letter combinations, $F(2, 133) = 16.36$, $p < .001$, $\eta_p^2 = .20$. The L1 Hebrew speakers had the lowest accuracy (72.5%), followed by the L1 Chinese speakers (83.1%), the L1 English speakers (84.1%), and the L1 French speakers (84.5%). The L1 Hebrew speakers were significantly less accurate than the L1 Chinese speakers, $p < .001$, and the L1 French speakers, $p < .001$; the difference between the L1 Chinese and the L1 French speakers was not significant.

Considering the comparisons that included a combination of different cues, there was a significant simple main effect of L1 for comparisons between an item with a legal combination of letters with a high positional bigram frequency and an item with an illegal combination of letters with a low positional bigram frequency (combined frequency and legality effects), $F(2, 133) = 8.86$, $p < .001$, $\eta_p^2 = .12$. The L1 Hebrew speakers had the lowest accuracy (82.3%), followed by the L1 French speakers (89.0%), the L1 Chinese speakers (92.5%), and the L1 English speakers (93.6%). The L1 Hebrew speakers were significantly less accurate than the L1 French speakers, $p = .01$, and the L1 Chinese speakers, $p < .001$; the difference between the L1 Chinese and the L1 French speakers was not significant. Finally, there was a significant simple main effect of L1 for comparisons between an item with a legal combination of letters with a low positional bigram frequency and an item with an illegal combination of letters with a high positional bigram frequency (conflicting frequency and legality cues), $F(2, 133) = 11.38$, $p < .001$, $\eta_p^2 = .15$. The L1 Hebrew speakers had the lowest accuracy (60.2%), followed by the L1 French speakers (73.2%), the L1 English speakers (74.8%), and the L1 Chinese speakers (75.1%). The L1 Hebrew speakers were again significantly less accurate than the L1 French speakers, $p < .001$, and the L1 Chinese speakers, $p < .001$; the difference between the L1 French and the L1 Chinese speakers was not significant. These L1 differences, broken down by L1 and comparison type, are displayed in Figure 23.

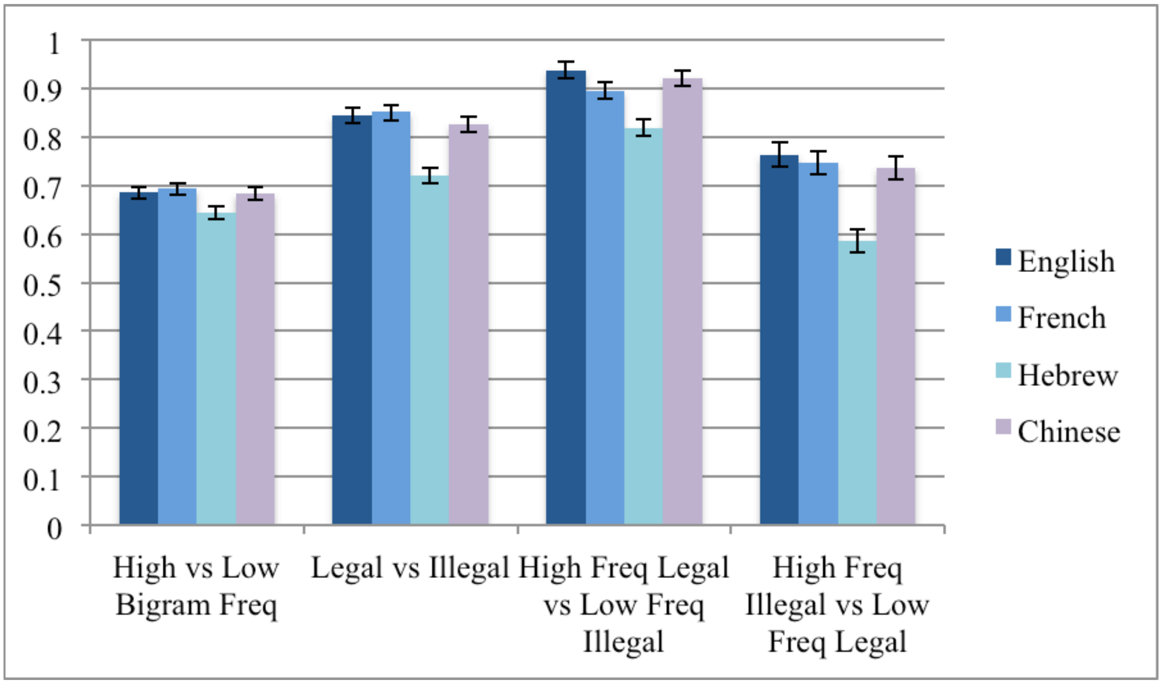

Figure 23. Proportion correct on the wordlikeness judgment task by L1 and comparison type. Error bars represent one standard error.

5.2.22 RTs

Descriptive statistics for participants' RTs, by comparison type and L1, are in Table 14. There was no main effect of comparison type, but the main effect of L1 was marginally significant, $F(2, 133) = 2.42$, $p = .09$, $\eta_p^2 = .04$. The main effects of flankers and operation span size were not significant, and neither variable interacted with comparison type. However, the interaction between comparison type and L1 was significant, $F(6, 399) = 3.01$, $p < .01$, $\eta_p^2 = .04$.

Table 14. RTs (in ms) on the wordlikeness judgment task, by L1 and comparison type.

	High vs. low bigram frequency	Legal vs. illegal letter combination	Combined frequency and legality effects	Conflicting frequency and legality cues
English	1718.58 (668.11)	1588.96 (576.32)	1529.33 (593.25)	1678.37 (495.61)
French	2651.71 (771.79)	2460.85 (709.36)	2284.12 (671.08)	2614.27 (834.50)
Hebrew	2188.69 (1090.95)	2095.79 (998.19)	2047.90 (1011.94)	2202.06 (1033.73)
Chinese	2521.66 (780.41)	2418.87 (737.07)	2225.23 (652.68)	2573.56 (812.97)

Note. Standard deviations are given in parentheses.

To follow up on this significant interaction, the simple main effect of L1 was examined for each comparison type, followed by pairwise comparisons among the L1 groups as appropriate. The significance of all pairwise comparisons was again adjusted using the Šidák correction for multiple comparisons. There was a significant simple main effect of L1 for comparisons between an item with high versus low positional bigram frequency, $F(2, 133) = 3.19$, $p < .05$, $\eta_p^2 = .05$. The L1 French speakers had the slowest RTs (2651.71 ms), followed by the L1 Chinese speakers (2521.66 ms), the L1 Hebrew speakers (2188.69 ms), and the L1 English speakers (1718.58 ms). The L1 French speakers were significantly slower than the L1 Hebrew speakers, $p < .05$; no other L1 comparisons were significant. The simple main effect of L1 was marginally significant for comparisons between an item with legal versus illegal letter combinations, $F(2, 133) = 2.57$, $p = .08$, $\eta_p^2 = .04$. The L1 French speakers had the slowest RTs (2460.85 ms), followed by the L1 Chinese speakers (2418.87 ms), the L1 Hebrew speakers (2095.79 ms), and the L1 English speakers (1588.96 ms). However, none of the L1 pairwise comparisons were significant.

Considering the comparisons that included a combination of different cues, the simple main effect of L1 was not significant for comparisons between an item with a legal combination of letters with a high positional bigram frequency and an item with an illegal combination of letters with a low positional bigram frequency (combined frequency and legality effects). The pattern of results was similar to other comparisons types, however: the L1 French speakers had the slowest RTs (2284.12 ms), followed by the L1 Chinese speakers (2225.23 ms), the L1 Hebrew speakers (2047.90 ms), and the L1 English speakers (1529.33 ms). Finally, the simple main effect of L1 was marginally significant for comparisons between an item with a legal combination of letters with a low positional bigram frequency and an item with an illegal combination of letters with a high positional bigram frequency (conflicting frequency and legality cues), $F(2, 133) = 2.81$, $p = .06$, $\eta_p^2 = .04$. The L1 French speakers again had the slowest RTs (2614.27 ms), followed by the L1 Chinese speakers (2573.56 ms), the L1 Hebrew speakers (2202.06 ms), and the L1 English speakers (1678.37 ms). However, none of the L1 pairwise comparisons were significant. These L1 differences, broken down by L1 and comparison type, are displayed in Figure 24.

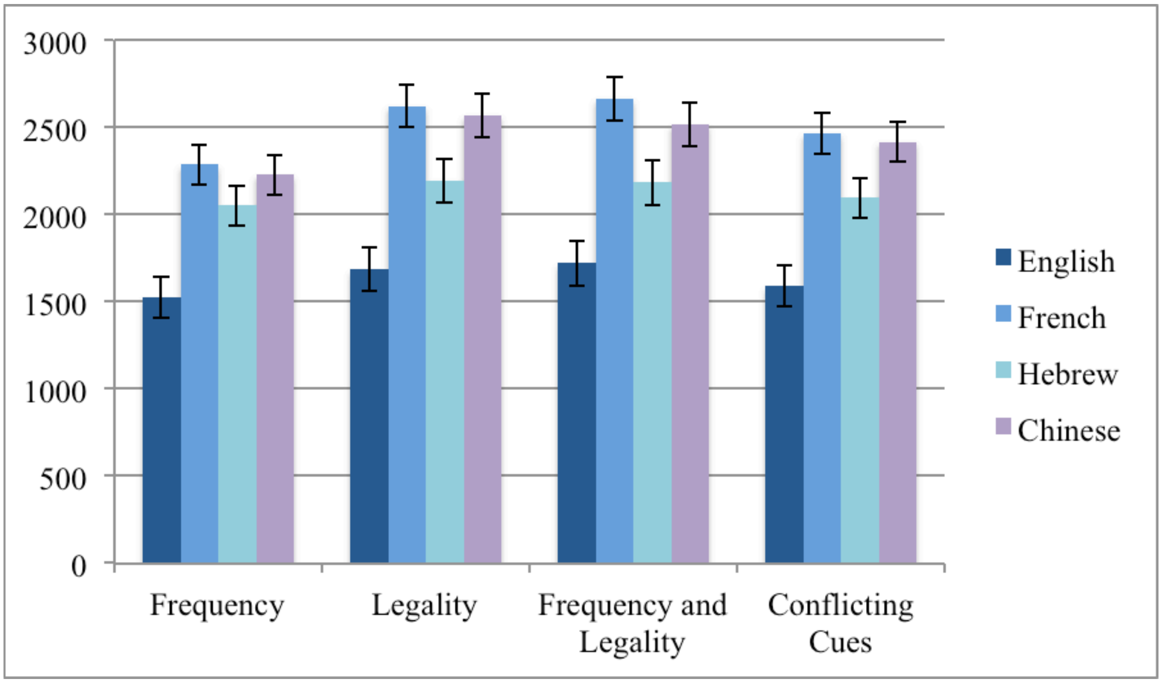

Figure 24. RTs on the wordlikeness judgment task by comparison type and L1. Error bars represent one standard error.

5.2.23 Summary and Discussion: Orthographic Knowledge

Similar to the results from the phonological awareness tasks, our predictions for the orthographic knowledge tasks received mixed support. The word/pseudohomophone discrimination task was designed to test participants' whole-word orthographic knowledge, particularly targeting their knowledge of precise spellings of English words. Based on this targeted skill, we predicted that we would see the highest performance for the L1 Chinese (morphosyllabary) speakers, followed by the L1 French (alphabet) and then the L1 Hebrew (abjad) speakers. We expected the morphosyllabic L1 speakers to do best because they are accustomed to using whole-word orthographic images for literacy in their L1, which may transfer to L2 and thus support better whole-word spelling knowledge (e.g., Wang & Geva, 2003). We expected the alphabetic L1 speakers to do better than the L1 abjad speakers based on previous work which has demonstrated that abjad speakers are relatively less sensitive to written vowel information (Hayes-Harb, 2006; K. I. Martin & Juffs, In revision; Ryan & Meara, 1991, 1996).

Our prediction that the L1 Chinese speakers would do best on this task was largely confirmed. Although the L1 English speakers had the highest accuracy overall, this was unsurprising given that they were performing the task in their native language. Among the non-native English speakers, the L1 Chinese speakers had the highest level of accuracy, consistent with our prediction. This provides further evidence that the strong visual-orthographic skills that speakers of morphosyllabic languages develop for L1 literacy can transfer and also help them develop precise spelling knowledge in their L2.

There was mixed support for our prediction that the L1 French speakers would perform better than the L1 Hebrew speakers. Overall, the level of accuracy was higher for the L1 Hebrew speakers than for the L1 French speakers, contrary to our prediction. However, an examination of performance on items with misspellings targeting consonants versus vowels reveals that the L1 Hebrew speakers had significantly lower accuracy for identifying misspellings that involved vowels compared to consonants. Thus, the non-alphabetic characteristics of the Hebrew speakers' L1 may help them develop relatively strong visual-orthographic processing skills, somewhat similar to the L1 Chinese speakers, leading to their higher overall accuracy than the L1 French speakers. At the same time, the lack of written vowel information in the Hebrew orthography encourages a learned inattention to vowel information that can transfer to L2 English and reduce their sensitivity to written vowels in English. It is interesting to note that the L1 Chinese speakers also showed a similar pattern. Although the size of the difference in accuracy between items with consonant versus vowel misspellings was smaller for the L1 Chinese speakers (2.3%) than the L1 Hebrew speakers (3.9%), both groups showed a significantly lower level of accuracy for identifying words that had misspelled vowels than misspelled consonants, although there was no significant difference for the L1 English or the L1 French speakers. This suggests that speakers of any language that does not represent vowels separately in the orthography, regardless of whether the language is segmental or morphosyllabic, will face a particular challenge with vowel spellings in ESL.

These results again have important implications for curriculum and materials writers and language instructors. The fact that both L1 Hebrew and L1 Chinese speakers showed a significantly lower level of accuracy identifying misspellings that involved vowels than consonants highlights the importance of emphasizing vowel spellings with ESL learners from non-alphabetic L1s in particular. It also suggests that other learners with similar L1s, such as Arabic, are likely to also face similar challenges. It is also possible that learners of other orthography types, such as syllabaries (i.e., Japanese) and abugidas (i.e., Marathi, Hindi, Thai) may also face particular challenges with spellings of vowels, although additional research specifically targeting these L1 groups is needed to confirm this prediction. Broadly speaking, this type of cross-linguistic research is crucial for developing an empirical understanding the factors contributing to certain patterns of difficulties that L2 learners of English may face and thus providing instructors a starting point for developing targeted instruction and literacy practice activities.

The wordlikeness task was designed to measure a different type of orthographic knowledge, this time focused on intralexical information relevant to the individual segments that make up a word. Based on the relative levels of segmental representation in written French, Hebrew, and Chinese, we predicted that the L1 French speakers would be most accurate on this task, followed by the L1 Hebrew speakers and then the L1 Chinese speakers (Akamatsu, 1999, 2002; Koda, 1999). Contrary to our prediction, for all comparison types the L1 Hebrew speakers were significantly less accurate than the L1 French and the L1 Chinese speakers, but there was no significant difference between the L1 French and the L1 Chinese speakers.

Another interesting pattern of results can be seen by examining the L1 Hebrew participants' relative level of accuracy, compared to the other L1 groups, on the two tasks. Specifically, the L1 Hebrew speakers did relatively well on the word/pseudohomophone discrimination task, but comparatively poorly on the wordlikeness judgment task. An examination of the particular non-word comparisons on which they struggled the most (relative to the other L1 groups) shows that they did relatively better at judging between items with high versus low bigram positional frequency – they were more sensitive to this factor and capable of using it to make an accurate wordlikeness judgment. However, they were less able to take advantage of letter combination legality to help them make judgments, particularly when the legality information conflicted with the positional bigram frequency information (such as when judging between a high bigram frequency, illegal letter sequence and a low bigram frequency, legal letter sequence).

Although unexpected, it may be possible to understand this particular behavioral pattern in relation to the morphological and orthographic characteristics of Hebrew. Hebrew has a templatic, or root-and-pattern, type of morphology (Ussishkin, 2006), in which consonant-based roots provide the core semantic meaning of words. However, there are a modest number of letters (22) and roots generally have only three (or occasionally four) consonants, meaning that most possible combinations of consonants are existing roots and therefore the relative order, or position, of these consonants is crucial for extracting meaning (Frost, 2012). The importance of consonant order or positioning in Hebrew may thus predispose these L1 speakers to pay relatively more attention to positional bigram frequency as a component of sub-lexical orthographic knowledge in their L2, as well. The fact that the majority of consonant combinations form valid roots in Hebrew may also mean that these learners are relatively less attuned to the possibility that certain letter combinations may by illegal. Regardless of the fundamental reason for the pattern, though, this difference in performance across different tasks again reveals the importance of considering task demands and the particular type of knowledge that is targeted by a specific task.

5.3 STUDY 1 SUMMARY

The goal of Study 1 was to complete a more comprehensive, linguistically-focused and also cross-linguistic analysis of L2 learners' performance on two crucial English literacy skills: phonological awareness and orthographic knowledge. The results from this study provide important evidence for the strong influences of linguistic structure, task demands and response type, and learner L1 background on performance on these skills, particularly when they are measured in the L2 (English).

Considering first the impact of test items' linguistic structure, there was strong evidence that participants' performance on the phonological awareness tasks differed according to the location of the targeted phonological unit within a word. Specifically, participants had higher accuracy and faster RTs for items testing awareness and manipulation of phonological units at the beginning and end of words, relative to the middle of words. This is consistent with the relatively greater linguistic salience of word beginnings and endings and confirms this factor as influential for performance on phonological awareness items, regardless of response modality or task demands. In addition, the specific phonological unit that was targeted for testing had a strong influence on participants' level of performance. Accuracy was generally higher for items that involved identifying or manipulating larger phonological units (syllables, rimes, bodies) compared to smaller phonological units (phonemes, onsets, codas) although this general tendency was complicated by the interaction of phonological unit and word location. Lastly, the inclusion of both real word and pseudoword items on the deletion task revealed the importance of lexical knowledge for supporting participants' performance on phonological tasks. Overall accuracy was much higher for items that involved real words compared to pseudowords, and this pattern was particularly emphasized for the L1 groups whose languages differed substantially from the language of testing (the L1 Hebrew and the L1 Chinese speakers). Although the three ESL samples in this study were equated on two measures of vocabulary knowledge, lexical knowledge of a language is broader than just the number of words that an individual knows. Thus, the similarity between French and English may have allowed the L1 French speakers to rely more on their general knowledge of linguistic structure and consequently perform more accurately on the pseudoword items than the speakers of languages that differed more substantially from English.

Our prediction that the preferred phonological grain size would transfer into the L2 and also be found as the preferred phonological grain size in English was generally not supported. Although the L1 French speakers did well on items testing individual phonemes, they were not different than the L1 Chinese speakers for the receptive task and not different than the L1 Hebrew speakers for the productive task. For onsets and rimes, again the L1 French speakers did relatively well, but were significantly better than the other L1 groups only for pseudoword items in the deletion task. Contrary to previous studies and our previous, we found no evidence for the L1 Hebrew speakers to prefer dividing words into body/coda units. Although performance was high for syllables overall, the L1 Chinese speakers showed a particular advantage only for the receptive oddity task and not the productive deletion task. Despite this, we did see evidence of the influence of L1 phonology and orthography on phonological awareness tasks in some cases.

The results from the orthographic knowledge tasks were also highly informative. Here our predictions were more strongly supported. There was evidence of strong orthographic skills in non-alphabetic ESL learners, consistent with previous research demonstrating high levels of visual-orthographic knowledge and skills and a heavy reliance on these skills for literacy tasks in learners from morphosyllabic L1s (e.g., McBride-Chang et al., 2005). Despite this, there was also a clear difference in performance on items testing knowledge of vowel versus consonant spellings, in which speakers of L1s without written vowels were much worse at identifying vowel misspellings. We also saw large differences in relative task performance based on the particular type of orthographic knowledge being targeted: whole-word orthographic knowledge versus sub-lexical orthographic characteristics. Although the L1 Hebrew speakers performed relatively well on the test of whole-word orthographic knowledge, they struggled compared to the other L1 groups on the test of intralexical orthographic knowledge.

The fact that may of our predictions were not fully supported underscores the complexity of the relationships among factors that impact performance on component literacy skills. There are a number of possible explanations for these results. First, as mentioned above, the results suggest that the preferred phonological grain size for L1 literacy is not necessarily maintained in the L2. This could be because it does not transfer in the first place, or, it could be because the reliance on that preferred L1 unit in the L2 decreases over time and with growing L2 proficiency.

Another possible explanation is the possibility that more advanced L2 learners begin to develop competence with and reliance on multiple phonological units (Ziegler & Goswami, 2005), thus showing less clear evidence of transfer of their preferred L1 phonological unit into the measurement of phonological awareness in the L2. A third possibility is that pure phonological awareness tasks, implemented without any reference to written forms, are not the most appropriate type of measure for capturing varying behaviors that are based on differences in the L1 orthography. If orthographic forms had been included in these tasks, or if additional tasks that targeted graphophonological awareness more broadly had been used, it is possible that stronger or more predictable effects of L1 may have been found. Overall, it is important to recognize that the complexity of the results is a natural outgrowth of the complexity of two languages interacting and competing within a single individual.

The final goal of Study 1 was to contribute to the body of cross-linguistic literacy research and to shed new light on the language skills of learners from representative L1s with particular characteristics that have been relatively underrepresented in the research literature. Much of the research on L2 phonological awareness in alphabetic learners has used participants from relatively shallow alphabets, such as Greek, German, Spanish, and Korean (Anthony et al., 2011; Chitiri et al., 1992; Chitiri & Willows, 1994; Durgunoğlu et al., 1993; Goswami et al., 2003; McBride-Chang et al., 2004; McBride-Chang et al., 2005; Näslund & Schneider, 1996). In contrast, in the current research we targeted L1 speakers of French, a substantially deeper (less transparent) alphabet. The results of the current research were also somewhat different than expected, suggesting that not only the mapping principle but also the orthographic depth of a writing system may impact the development of varying levels of phonological awareness and learners' ability to transfer this skill to L2. This possibility deserves further attention, particularly because the possibility of different developmental trajectories based on L1 orthographic depth has implications for our understanding of the universal and language-specific aspects of phonological skills development.

In addition, we specifically targeted L1 speakers of an abjad language (Hebrew). These languages have characteristics of both alphabetic and non-alphabetic writing systems and yet relatively little research has focused on speakers of these languages, meaning that relatively little is understood about the impact of the abjad orthographic characteristics on phonological and orthographic skills and L2 literacy development. The data from the current study provide a starting point for understanding these abilities in L1 speakers of abjad languages, and how they compare to speakers of other L1 orthographic types. Specifically, we found that the L1 Hebrew speakers generally performed quite well on the phonological awareness tasks, in many cases not significantly different from the L1 French speakers with a true alphabetic L1. This suggests that despite an abjad not being a true alphabet, the segmental nature of these writing systems encourages the development of fine-grained phonological awareness, even at the phoneme level, in a way that is comparable to alphabetic languages. At the same time, the L1 Hebrew speakers struggled relative to the other L1 groups on the wordlikeness discrimination task, which targeted intralexical orthographic knowledge. If the segmental nature of their L1 orthography supports the development of phonological awareness, it is not clear why it would not do the same to support the development of sensitivity to the intralexical characteristics of positional bigram frequency and particularly letter combination legality. Additional research with L1 speakers of both Hebrew and Arabic is needed to document this difficulty in more detail and understand its underlying cause(s).

Finally, we targeted L1 speakers of Chinese from Taiwan. Although a substantial amount of research has examined phonological and orthographic skills in L1 Chinese speakers, the majority of studies have used participants either from Hong Kong, mainland China, or heritage speakers of Chinese in Canada or the United States (e.g., Bialystok et al., 2005; Cheung et al., 2001; Gottardo, Chiappe, Yan, Siegel, & Gu, 2006; Gottardo et al., 2001; Ho & Bryant, 1997; Ho & Lai, 1999; Liow & Poon, 1998; McBride-Chang et al., 2004; McBride-Chang et al., 2005; Read et al., 1986). L1 Chinese speakers from mainland China learn *pinyin* and use it as an (alphabetic) phonetic system for learning character pronunciations, and L1 Chinese speakers from Hong Kong learn both characters and English words through a whole-form memorization, 'look-and-say' method (McBride-Chang et al., 2004; McBride-Chang et al., 2005). L1 Chinese speakers from Taiwan are different from either of these groups in that they learn a phonetic, syllable-based system called *zhuyin fuhao* to assist in the learning of character pronunciations. Although this does introduce these speakers to a phonologically-based system, it is primarily based on syllables rather than phonemes, and thus it is not clear whether or how it impacts the development of phonological skills in these speakers. Although this study did not have the advantage of directly comparing L1 Chinese speakers from Taiwan with those from Hong Kong or mainland China, the results from the phonological awareness tasks (specifically the oddity task) do suggest that familiarity with *zhuyin fuhao* does provide some support for the development of phonological awareness, even at the phoneme level. However, this ability appears to be relatively weak, because the L1 Chinese speakers were highly successful only for the receptive task that did not require active manipulation and articulation of a response.

The results of Study 1 have important implications both for language researchers and for language educators. The goal of many language researchers is to understand the cognitive underpinnings of literacy and its developmental differences across languages, and the current results highlight a number of important issues these researchers need to consider when choosing tasks for use in research. Although receptive phonological awareness tasks are widely used in this research area due to the prevalence of young children as research subjects, the results from the current study suggest that these tasks may not be appropriate for use in adult participants because they run the risk of ceiling effects. Although this finding needs to be confirmed with additional studies, and further work should be done comparing performance on different types of receptive and productive tasks, based on the current evidence it would be wise for researchers to focus primarily on productive tasks in order to maximize the ability of their tasks to differentiate levels of performance. However, one caveat to this recommendation reflects the fact that the receptive task, but not the productive task, demonstrated high phonological awareness abilities in the L1 Chinese speakers. Thus, it is possible that receptive tasks may still be useful for demonstrating the existence of relatively weak phonological skills in certain target populations.

The goal of language educators is to be able to understand and predict the varying challenges that their students will face so that they can target their instruction most effectively. The results of the current study demonstrate a number of areas where particular L1 groups may face difficulties in literacy tasks. Specifically, L1 Chinese speakers may develop relatively weak phonological awareness abilities, which will likely reduce their ability to decode unfamiliar words and to connect unfamiliar written forms with spoken forms that they already know. Despite this, these learners may be able to transfer their strong orthographic skills from L1 to L2 and thus develop relatively high spelling abilities. On the other hand, both the L1 Hebrew and the L1 Chinese speakers showed a significantly lower accuracy with identifying misspelled vowels than consonants. Thus, learners who come from L1s without written vowels may struggle with spelling, particularly for vowel sounds. These findings suggest specific areas of literacy skill that language instructors can directly target with their students if they come from an L1 with these characteristics. Additional work of this sort is needed to add to our empirical understanding of the basis of specific literacy challenges that ESL learners face as a result of the L1 background experiences.

6.0 STUDY 2: CROSS-LINGUISTIC CONTRIBUTIONS OF PHONOLOGICAL AWARENESS AND ORTHOGRAPHIC KNOWLEDGE TO WORD IDENTIFICATION

As detailed above, the goal of Study 2 was to compare performance on three different word identification tasks among ESL learners from representative L1 writing system backgrounds, as well as the compare the relative contributions of phonological awareness and orthographic knowledge (as measured in Study 1) to these word identification skills.

6.1 METHOD

6.1.1 Participants

The same participants that participated in Study 1 also participated in Study 2. The same subset of participants was also used for the analyses in Study 2.

6.1.2 Materials and Procedure

Participants completed three word-level reading tasks (lexical decision, naming, and pseudoword decoding) and one sentence- and paragraph-level reading comprehension test (detailed above, in Study 1 Method). All words used in the lexical decision and naming tasks were carefully selected from a small number of sources in order to increase the likelihood that participants would be familiar with them. These included the General Service List (West, 1953), a list of approximately 2000 headwords that, together with their word families provide approximately 80% coverage of English texts (Neufeld & Billuroğlu, 2005); the words covered in the ESL vocabulary series *Words for Students of English*, volumes 1-6 (Rogerson, Davis, Hershelman, & Jasnow, 1992; Rogerson, Esarey, Hershelman, Jasnow, Moltz, et al., 1992; Rogerson, Esarey, Hershelman, Jasnow, Schmandt, et al., 1992; Rogerson, Esarey, Schmandt, & Smith, 1992; Rogerson, Hershelman, & Jasnow, 1992; Rogerson, Hershelman, Jasnow, & Moltz, 1992); and the 2000 most frequent words of English (Davies, 2008-). All items were checked through online dictionary sources and by native speakers of the language to ensure that they were not cognates and did not share a substantial amount of form overlap with any words in French, Hebrew, and Mandarin.

6.1.3 Lexical Decision

Lexical decision was used as a measure of receptive word recognition. Participants saw one item at a time centered on a computer screen and were asked to respond as quickly as possible by pressing a labeled button on a serial response box to indicate whether the item was a real English word ('Yes') or not ('No'). Words were displayed for up to 7500 ms and disappeared upon participants' button press. The interstimulus interval was 1000 ms. Participants completed ten practice trials followed by 190 test trials, all without feedback, with a short break at the halfway point.

The test trials included 40 real words with consistent grapheme-phoneme correspondences, 40 real words with inconsistent grapheme-phoneme correspondences, 30 real words with exception spellings[4], 20 legal English pseudowords with high bigram frequencies, 20 legal English pseudowords with low bigram frequencies, 20 pseudohomophones of real English words, and 20 unpronounceable non-word letter strings. All stimuli were monosyllabic, and a complete listing of items by type can be found in Appendix F.

The consistency of the grapheme-phoneme correspondences was determined using the consistency statistics from Treiman, et al. (Treiman et al., 1995). Three statistics were used: the proportion of a word's body (CV) neighbors that have a consistent pronunciation, calculated over tokens (each neighbor weighted by its frequency of occurrence); the proportion of a word's rime (VC) neighbors that have a consistent pronunciation, calculated over tokens; and a sum of the *H* statistics of orthographic uncertainty[5] for the body (CV) and rime (VC) of a word, again calculated over tokens. As reported in Treiman et al. (Treiman et al., 1995), all consistency statistics were based on monosyllabic CVC words from the Merriam-Webster pocket dictionary of approximately 20,000 words.

Items selected as words with consistent grapheme-phoneme correspondences had an average of 99.38% of their body neighbors with a consistent CV pronunciation (range: 93% to 100%) and 99.95% of their rime neighbors with a consistent VC pronunciation (range: 99% to 100%). The sum of the CV and VC *H* statistics for these items was, on average, .04 (range: 0 to

[4] Although we attempted to include 40 exception words, as well, limitations on the number of well-defined exception words (see below) that were included within the set of words determined as likely to be known by non-native English speakers prevented us from including a full 40 items of this type.

[5] The *H* statistic is a measure of the uncertainty of a word's pronunciation based on the components of its written form and was originally proposed by Fitts and Posner (1967). The formula and a detailed explanation of the calculation of the statistic can be found on page five of Treiman et al. (1995). The value of *H* for a word with completely consistent grapheme-phoneme correspondences is 0 and *H* increases as the number of possible pronunciations for the orthographic unit(s) increases and as the relative frequencies of these possible pronunciations become more similar.

.50; recall that 0 indicates perfect grapheme-phoneme consistency). In contrast, items selected as words with inconsistent grapheme-phoneme correspondences had an average of only 41% of their body neighbors with a consistent CV pronunciation (range: 0% to 97%) and only 47.35% of their rime neighbors with a consistent VC pronunciation (range: 0% to 99%). The sum of the CV and VC *H* statistics for these items was, on average, 1.32 (range: 0 to 3.31). The difference between the consistency statistics for the consistent and inconsistent words was significant for all three measures: body consistency, $F(1, 78) = 87.75$, $p < .001$; rime consistency, $F(1, 78) = 70.42$, $p < .001$; sum of the *H* statistics, $F(1, 78) = 149.69$, $p < .001$. Items selected as words with exception spellings were taken from Ziegler, Stone, and Jacobs (1997) who identified them as words that were the only English words with their particular orthographic rime (spelling of the vowel plus following consonant(s))[6]. Using statistics available from the E-Lexicon (Balota et al., 2007), the consistent, inconsistent, and exception words were selected so that they were matched on their L1 (English) age of acquisition, concreteness, imageability, raw and log frequency from the hyperspace analogue to language (HAL) database (Lund & Burgess, 1996), raw and log frequency from the SUBTLEX database (Brysbaert & New, 2009), and bigram sum and mean statistics, all *p*s > .21.

The legal English pseudowords with high and low bigram statistics were generated from the E-Lexicon. They were CVC monosyllabic items of 4-5 letters in length. The high- and low-

---

[6] Note that this definition of exception words is different from some other uses of the term in the literature. For example, Coltheart, Besner, Jonasson, and Davelaar (1979), Gluskho (1979), Goswami and Bryant (1990), and Parkin (1982) define exception words as those having a less-frequent pronunciation than other words with the same orthographic spelling pattern (such as *have*, which is an exception compared to neighbors such as *gave* and *save*). Others, such as Seidenberg, Waters, Barnes, and Tanenhaus (1984) and Treiman et al. (1995), define exception words more generally as those that have ‘unusual’ or ‘atypical’ pronunciations for their particular spelling patterns.

bigram pseudowords were matched on length, $F(1, 39) < .001$, $p = 1.00$, although the high-bigram pseudowords had significantly higher bigram sum, mean, and frequency by position statistics than the low-bigram pseudowords, sums, $F(1, 39) = 632.82$, $p < .001$; means, $F(1, 39) = 604.75$, $p < .001$; frequency by position, $F(1, 39) = 196.05$, $p < .001$.

The pseudohomophones were taken from previous stimuli used by Besner and Davelaar (1983), Borowsky et al. (2002), Laxon et al. (1992), Lukatela and Turvey (1991, 1994a, 1994b, 2000), Lukatela et al. (2002), Lupker and Pexman (2010), Manis et al. (1996), Martin (1982), Reynolds and Besner (2005), Seidenberg et al. (1996), Taft and Russell (1992), and Yates et al. (2003). All items were 3-5 letters in length.

Finally, the unpronounceable non-words were taken from Fleming (1976), Siegel et al. (1995), and Solman (1988). All items were 4-5 letters in length and violated orthographic constraints for letter sequences in English.

The outcome variables were participant accuracy and reaction times. Reaction times were analyzed for correct trials only, and RTs more than three standard deviations from an individual participant's mean RT were excluded from consideration. This resulted in the exclusion of 3.02% of the data.

6.1.4 Naming

Naming was used as a measure of productive word recognition. Participants saw one item at a time centered on a computer screen and were asked to read the word aloud as quickly as possible. Responses were recorded with a digital voice recorded and response times were measured using a voice key attached to a serial response box. Words were displayed until the microphone registering participants' spoken response. After the voice key was triggered, a blank screen was displayed for 500 ms, followed by a screen that prompted participants to press a button for the next item. A microphone test was used prior to testing to ensure that participants were able to speak loudly and clearly enough to properly trigger the voice key. This was followed by ten practice trials and subsequently 110 test trials.

The test items consisted of 40 real words with consistent grapheme-phoneme correspondences, 40 real words with inconsistent grapheme-phoneme correspondences, and 30 real words with exception spellings, all defined in the same way as for the lexical decision task. Items selected as words with consistent grapheme-phoneme correspondences had an average of 99.63% of their body neighbors with a consistent CV pronunciation (range: 95% to 100%) and 100% of their rime neighbors with a consistent VC pronunciation. The sum of the CV and VC H statistics for these items was, on average, .02 (range: 0 to .21; recall that 0 indicates perfect grapheme-phoneme consistency). In contrast, items selected as words with inconsistent grapheme-phoneme correspondences had an average of only 27.15% of their body neighbors with a consistent CV pronunciation (range: 0% to 99%) and only 33.18% of their rime neighbors with a consistent VC pronunciation (range: 0% to 99%). The sum of the CV and VC H statistics for these items was, on average, 1.59 (range: .30 to 2.77). The difference between the consistency statistics for the consistent and inconsistent words was significant for all three measures: body consistency, $F(1, 78) = 175.76$, $p < .001$; rime consistency, $F(1, 78) = 123.83$, $p < .001$; sum of the H statistics, $F(1, 78) = 201.46$, $p < .001$. Items selected as words with exception spellings were again taken from Ziegler, et al. (1997). The consistent, inconsistent, and exception words were again selected so that they were matched on their L1 (English) age of acquisition, concreteness, imageability, raw and log frequency from the HAL database, raw and log frequency from the SUBTLEX database, and bigram sum and mean statistics, all *p*s > .25. A full listing of stimuli can be found in Appendix G.

The outcome variables were participant accuracy and reaction times. Acceptable pronunciations were determined by mutual agreement between a native American English and a native British English speaker, each with advanced training in linguistics. Both standard American and standard British English pronunciations were accepted, with minor variations due to foreign accent. Reaction times were analyzed for correct trials only, and RTs more than three standard deviations from an individual participant's mean RT were excluded from consideration. This resulted in the exclusion of 4.63% of the data.

6.1.5 Pseudoword Decoding

Decoding skill was measured using the pseudoword decoding subtest from the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999). This test consists of 63 pronounceable English pseudowords, beginning with simple and short monosyllabic items and continuing with items that are increasingly longer and more complex. Participants were shown all the words printed in double-spaced columns on a single page and were asked to read each item how they thought it should be pronounced, based on their knowledge of English. Participants were also asked to read through the list as quickly as possible. Their responses were recorded with a digital voice recorder and later transcribed using the International Phonetic Alphabet (IPA, 1999). A full listing of stimuli can be found in Appendix H.

Per the official instructions, the outcome variable was the number of pseudowords that participants read aloud correctly within 45 seconds. Acceptable pronunciations were again determined by mutual agreement between a native American English and a native British

English speaker, each with advanced training in linguistics, following the pronunciation guide in the test instructions. Both standard American and standard British English pronunciations were accepted, allowing for minor variations due to foreign accent.

6.1.6 Cognitive Abilities

Participants completed the same three cognitive tasks as described in Study 1. These variables were also used as covariates in the current study as appropriate.

6.1.7 Overall Procedure

Participants completed the lexical decision task first, followed by pseudoword decoding and then naming, deletion. The cognitive tasks and language history questionnaire were completed next. The oddity task took approximately 10-15 minutes; word/pseudohomophone discrimination took approximately 5 minutes; deletion took approximately 12-17 minutes; the wordlikeness judgments took approximately 7-10 minutes; rapid digit naming took approximately 2 minutes; operation span took approximately 10-12 minutes; and flankers took approximately 5 minutes. All participants were tested individually.

6.2 RESULTS

The primary goal of this study was to evaluate the relative contributions of phonological awareness and orthographic knowledge, described in Study 1, to varying measures of English word identification (lexical decision, word naming, pseudoword decoding) and global reading comprehension in speakers with representative L1 backgrounds. To determine the contributions of phonological awareness and orthographic knowledge to these tasks, composite scores for these two skills were calculated on the basis of performance on the tasks described in Study 1. First, a *z*-score for accuracy on each task was computed for each participant. Then, to create the composite phonological awareness score we averaged the *z*-scores for the oddity and the deletion tasks, and to create the composite orthographic knowledge score we averaged the *z*-scores for the word-pseudohomophone discrimination and the wordlikeness judgment tasks. Correlations among the composite phonological awareness and orthographic knowledge measures, lexical decision accuracy and RT, naming accuracy and RT, pseudoword decoding accuracy, global reading comprehension score, and measures of cognitive abilities are presented first. The results for the L1 English speakers are in Table 15, the L1 French speakers are in Table 16, the L1 Hebrew speakers are in Table 17, and the L1 Chinese speakers are in Table 18.

Table 15. Correlations among the variables in Study 2 for the L1 English speakers.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Phonological awareness composite	--									
2. Orthographic knowledge composite	.12	--								
3. Lexical decision d-prime	-.15	-.23	--							
4. Lexical decision RT	-.13	-.38**	.26†	--						
5. Naming accuracy	.35*	.16	.08	.10	--					
6. Naming RT	-.25†	.02	.03	.09	.07	--				
7. Pseudoword decoding score	.31*	-.17	-.03	-.20	.09	-.15	--			
8. Global reading comprehension	.23	.17	-.07	-.004	.11	-.25†	.28†	--		
9. Flankers	.07	.03	.05	.16	-.03	-.07	-.05	-.12	--	
10. Operation span size	.49**	.24	-.31*	-.06	-.09	-.08	.03	.06	.17	--

Note. †p < .10; *p < .05; **p < .01.

Table 16. Correlations among the variables in Study 2 for the L1 French speakers.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Phonological awareness composite	--									
2. Orthographic knowledge composite	.41**	--								
3. Lexical decision d-prime	.30*	.36*	--							
4. Lexical decision RT	.09	-.18	-.21	--						
5. Naming accuracy	.61***	.54***	.73***	-.07	--					
6. Naming RT	-.02	-.18	-.22	.18	-.28†	--				
7. Pseudoword decoding score	.31*	.32*	.49***	-.16	.59***	-.36*	--			
8. Global reading comprehension	.54***	.33*	.68***	.01	.72***	-.15	.24	--		
9. Flankers	.05	-.01	.03	.05	.08	-.21	.12	.04	--	
10. Operation span size	.44**	.11	.24	.05	.35*	-.02	.28†	.20	-.03	--

Note. †$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

Table 17. Correlations among the variables in Study 2 for the L1 Hebrew speakers.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Phonological awareness composite	--									
2. Orthographic knowledge composite	.60***	--								
3. Lexical decision d-prime	.44**	.58***	--							
4. Lexical decision RT	.18	-.06	-.02	--						
5. Naming accuracy	.57***	.53***	.63***	-.21	--					
6. Naming RT	.23	-.16	-.08	.55***	-.03	--				
7. Pseudoword decoding score	.22	.40**	.47**	-.31*	.62***	-.26†	--			
8. Global reading comprehension	.38**	.36*	.55***	-.14	.62***	.09	.41**	--		
9. Flankers	.36*	.16	.28†	-.08	.18	.02	.06	.12	--	
10. Operation span size	.44**	.32*	.28†	-.13	.41**	-.08	.20	.22	.07	--

Note. †$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

Table 18. Correlations among the variables in Study 2 for the L1 Chinese speakers.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Phonological awareness composite	--									
2. Orthographic knowledge composite	.55***	--								
3. Lexical decision d-prime	.25	.28†	--							
4. Lexical decision RT	-.06	-.29†	-.10	--						
5. Naming accuracy	.72***	.54***	.40**	-.23	--					
6. Naming RT	-.10	.01	-.01	.20	-.24	--				
7. Pseudoword decoding score	.54***	.31*	.23	-.30*	.65***	-.48**	--			
8. Global reading comprehension	.21	.30*	-.16	-.29*	.24	-.12	.16	--		
9. Flankers	-.24	-.28†	-.14	.08	-.36*	-.20	-.04	-.15	--	
10. Operation span size	.41**	.24	-.05	.02	.09	-.17	.20	.29†	.19	--

Note. †p < .10; *p < .05; **p < .01; ***p < .001.

6.2.1 Lexical Decision

Lexical decision was used as a receptive measure of word recognition. Two measures of lexical decision performance were evaluated: accuracy and reaction times. Accuracy was analyzed using d-prime, a measure of discrimination sensitivity appropriate for yes/no responses (Macmillan & Creelman, 2004), in order to account for the slightly different numbers of correct yes/no trials as well as any possible response bias. With this measure, a higher value indicates a greater ability to discriminate correctly between yes and no trials. For the first stage of analysis, performance on the lexical decision task was compared among the four L1 groups (English, French, Hebrew, and Chinese). For the second stage of analysis, the relative contributions of phonological awareness and orthographic knowledge to lexical decision performance were evaluated for each L1 group using correlations and regressions.

6.2.2 Lexical Decision Accuracy

Accuracy (d-prime) was analyzed using univariate ANOVA with L1 as a between-subjects factor and the standardized incongruent-congruent RT difference on the flankers task and the operation span maximum span length as covariates. Means and standard deviations for each L1 group are given in Table 19 and displayed in Figure 25. There was a significant main effect of L1, $F(2, 133) = 10.74$, $p < .001$, $\eta_p^2 = .14$. Post-hoc comparisons among the L1 groups were thus conducted, and the Šidák correction for multiple comparisons was used. The L1 French speakers had the lowest d-prime score (1.38), followed by the L1 Hebrew speakers (1.43), the L1 Chinese speakers (1.84), and the L1 English speakers (2.20). The L1 Chinese speakers were significantly more accurate than the L1 French speakers, $p < .001$, and the L1 Hebrew speakers, $p < .001$; the difference between the L1 French and the L1 Hebrew speakers was not significant.

Table 19. Accuracy on the lexical decision task.

L1 group	d-prime score
English	2.20 (.30)
French	1.37 (.47)
Hebrew	1.43 (.57)
Chinese	1.84 (.44)

Note. Standard deviations are given in parentheses.

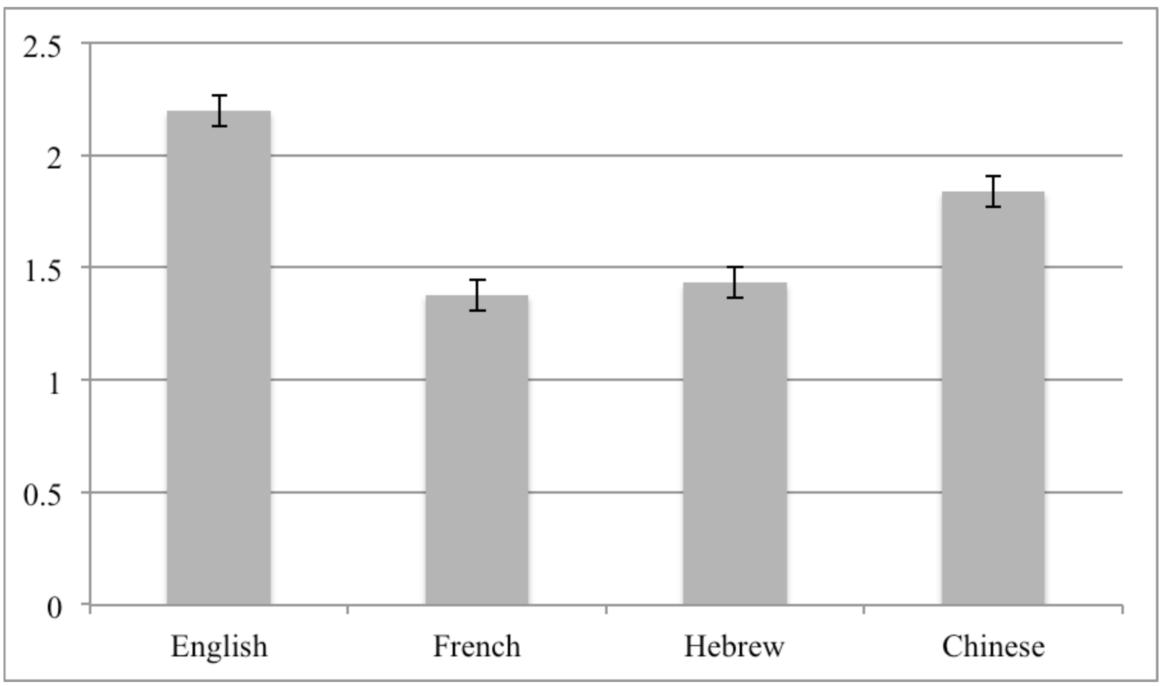

Figure 25. Accuracy on the lexical decision task, measured in d-prime units.

For the correlation and regression analyses, each L1 group was considered separately. The correlations between the composite phonological awareness and orthographic knowledge scores and lexical decision accuracy for each L1 group can be found in Tables 15 through 18. For the L1 English speakers, the correlation between lexical decision accuracy and lexical decision RTs was marginally significant, such that higher accuracy was somewhat associated with longer RTs. Lexical decision accuracy was significantly correlated with operation span size, such that higher accuracy was associated with lower working memory capacity. No other correlations were significant for the L1 English speakers.

For the L1 French speakers, lexical decision accuracy was significantly correlated with naming accuracy, pseudoword decoding, and global reading comprehension. These correlations indicate that higher lexical decision accuracy was associated with higher accuracy on the naming, pseudoword decoding, and global reading comprehension measures, as well. Lexical decision accuracy was also significantly correlated with the composite phonological awareness score, and with the composite orthographic knowledge score. In each case, greater phonological or orthographic skills were associated with higher lexical decision accuracy.

For the L1 Hebrew speakers, lexical decision accuracy was again significantly correlated with naming accuracy; pseudoword decoding, and global reading comprehension. These correlations again indicate that higher lexical decision accuracy was associated with higher accuracy on the naming, pseudoword decoding, and global reading comprehension measures, as well. Lexical decision accuracy was also significantly correlated with the composite phonological awareness score, and with the composite orthographic knowledge score. In each case, greater phonological or orthographic skills were associated with higher lexical decision accuracy, and these correlations were somewhat larger than those found with the L1 French speakers. For the L1 Hebrew speakers lexical decision accuracy was also marginally correlated with flankers, such that higher lexical decision accuracy was associated with a larger congruency effect; and operation span size, such that higher lexical decision accuracy was associated with higher working memory capacity.

Finally, for the L1 Chinese speakers, lexical decision accuracy was significantly correlated with naming accuracy. In contrast to the L1 French and the L1 Hebrew speakers, however, accuracy on the lexical decision task was not significantly correlated with accuracy on the other word- and text-level tasks. Lexical decision accuracy was marginally correlated with the composite orthographic knowledge score, such that greater orthographic skill was associated with higher lexical decision accuracy.

The regression analyses were conducted next. Flankers and operation span size were entered into the regression first. For the L1 English speakers, these variables together accounted for a marginally significant 10.6% of variance in lexical decision accuracy, $p = .09$. Flankers and operation span size accounted for a significant 14.5% of variance for the L1 Hebrew speakers, $p < .05$. However, they did not account for a significant amount of variance for the L1 French speakers or for the L1 Chinese speakers.

The composite phonological awareness and orthographic knowledge scores were entered in the next step. The results of these regressions are displayed in Table 20. For the L1 English speakers, the addition of these two variables accounted for a further non-significant 2.6% of variance in accuracy. For the L1 French speakers, these two variables together accounted for a marginally significant 11.8% of variance, $p = .06$. Considering the coefficients for the two variables, only orthographic knowledge was a marginally significant predictor of lexical decision accuracy for the L1 French speakers. For the L1 Hebrew speakers, phonological awareness and orthographic knowledge together accounted for a significant additional 23.9% of variance in lexical decision accuracy, $p = .001$. Considering the coefficients, orthographic knowledge was a stronger predictor of accuracy than phonological awareness for the L1 Hebrew speakers.

Finally, for the L1 Chinese speakers, phonological awareness and orthographic knowledge together accounted for a further non-significant 9.9% of variance in accuracy.

Taken as a whole, the model including flankers, operation span size, phonological awareness composite score, and orthographic knowledge composite score was not significant for the L1 English speakers, or for the L1 Chinese speakers. However, the model was marginally significant for the L1 French speakers, $F(4, 41) = 2.19$, $p = .09$, and significant for the L1 Hebrew speakers, $F(4, 41) = 6.41$, $p < .001$.

Table 20. Regression models of lexical decision d-prime scores.

	L1 English			L1 French			L1 Hebrew			L1 Chinese		
Predictors	ΔR^2	β	p	ΔR^2	β	p	ΔR^2	β	p	ΔR^2	β	p
Step 1	.106		.089	.058		.276	.145		.034	.020		.649
Flankers		.106	.473		.038	.800		.264	.069		-.134	.389
Operation span size		-.327	.031		.239	.114		.258	.075		-.026	.864
Step 2	.026		.552	.118		.064	.239		.001	.099		.113
Flankers		.105	.481		.034	.810		.184	.173		.004	.981
Operation span size		-.291	.100		.161	.317		.086	.534		-.187	.278
Phonological awareness composite		.007	.969		.106	.544		.034	.847		.205	.290
Orthographic knowledge composite		-.165	.279		.296	.065		.505	.002		.212	.244

6.2.3 Lexical Decision RTs

RTs were also analyzed using univariate ANOVA with L1 as a between-subjects factor and the standardized incongruent-congruent RT difference on the flankers task and the operation span maximum span length as covariates. Means and standard deviations for each L1 group are given in Table 21 and displayed in Figure 26. There was a significant main effect of L1, $F(2, 133) = 10.93$, $p < .001$, $\eta_p^2 = .14$. Post-hoc comparisons among the L1 groups were thus conducted, and the Šidák correction for multiple comparisons was again used. The L1 French speakers had the slowest RTs (922.13 ms), followed by the L1 Hebrew speakers (815.13 ms), the L1 Chinese speakers (730.89 ms), and the L1 English speakers (642.88 ms). The L1 French speakers were significantly slower than the L1 Hebrew speakers, $p < .05$, and the L1 Hebrew speakers, $p < .001$. The L1 Hebrew speakers were marginally slower than the L1 Chinese speakers, $p = .09$.

Table 21. RTs on the lexical decision task.

L1 group	RT (ms)
English	642.88 (120.94)
French	922.13 (194.05)
Hebrew	815.13 (221.28)
Chinese	730.89 (123.26)

Note. Standard deviations are given in parentheses.

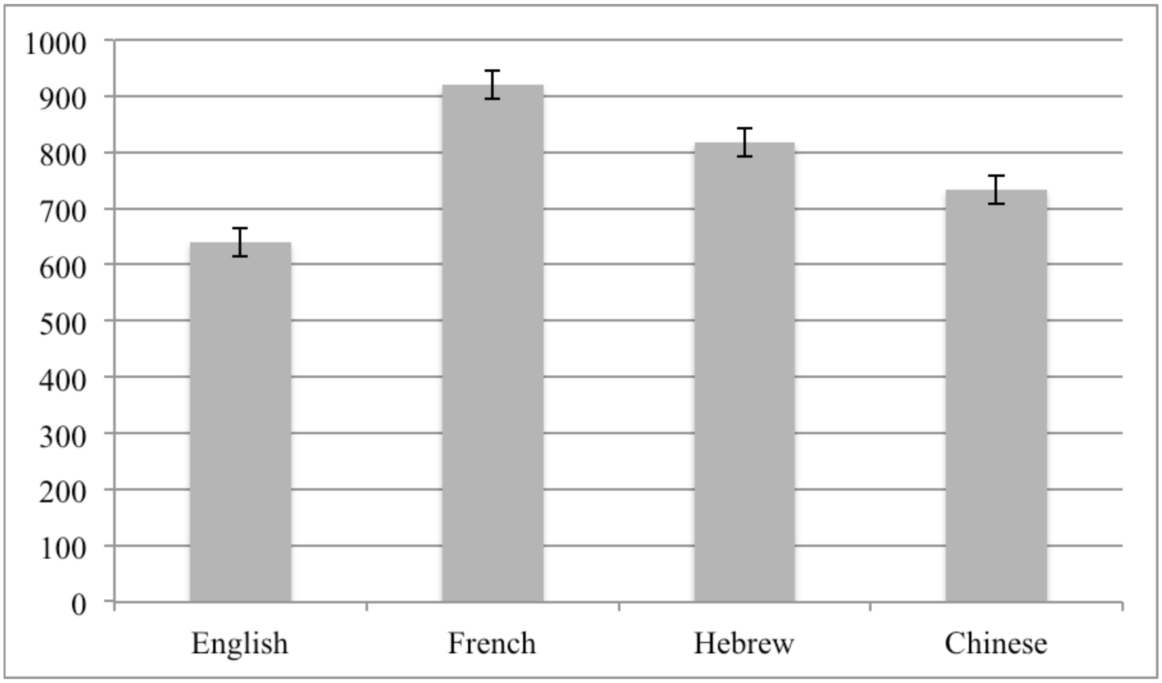

Figure 26. RTs on the lexical decision task.

For the correlation and regression analyses, each L1 group was again considered separately. The correlations between the composite phonological awareness and orthographic knowledge scores and lexical decision RTs for each L1 group can be found in Tables 15 through 18. For the L1 English speakers the only significant correlation was between lexical decision RTs and the composite orthographic knowledge score; the correlation between lexical decision RTs and lexical decision accuracy was marginally significant. These correlations indicate that greater orthographic knowledge and higher lexical decision accuracy are associated with faster lexical decision RTs. For the L1 French speakers there were no significant correlations between lexical decision RTs and any other variables, including the composite phonological awareness and orthographic knowledge scores. For the L1 Hebrew speakers, lexical decision RTs were significantly correlated with naming RTs; and with and pseudoword decoding. These correlations indicate that longer lexical decision RTs were also associated with longer naming

RTs, and that shorter lexical decision RTs were associated with higher pseudoword decoding scores. Finally, for the L1 Chinese speakers, lexical decision RTs were significant correlated with pseudoword decoding; and with global reading comprehension. These correlations indicate that shorter lexical decision RTs were associated with higher pseudoword decoding and global reading comprehension scores. For the L1 Chinese speakers there was also a marginally significant correlation between the orthographic knowledge composite score and lexical decision RTs, such that greater orthographic knowledge was associated with faster RTs.

The regression analyses were conducted next. Flankers and operation span size were entered into the regression first. Together these variables did not account for a significant amount of variance in lexical decision RTs for any of the L1 groups.

The composite phonological awareness and orthographic knowledge scores were entered in the next step. The results of these regressions are displayed in Table 22. For the L1 English speakers, the addition of these two variables accounted for a further significant 15.5% of variance in lexical decision RTs, $p < .05$. Considering the coefficients for the two composite scores reveals that orthographic knowledge was the stronger of the two predictors for the L1 English speakers. For the L1 French speakers, phonological awareness and orthographic knowledge together accounted for a non-significant additional 5.9% of variance in lexical decision RTs. For the L1 Hebrew speakers, phonological awareness and orthographic knowledge accounted for a significant additional 14.5% of variance in lexical decision RTs, $p < .05$. Considering the coefficients for the L1 Hebrew speakers reveals that in contrast to the L1 English speakers, phonological awareness was the stronger predictor for the L1 Hebrew speakers, and that the relationship was in an unexpected direction: higher phonological awareness was associated with longer, rather than shorter, RTs. Finally, for the L1 Chinese speakers, phonological awareness and orthographic knowledge accounted for a non-significant additional 9.5% of variance in lexical decision RTs.

Although an evaluation of the change in R^2 suggests that phonological awareness and orthographic knowledge explain significant variance in lexical decision RTs for both the L1 English and the L1 Hebrew speakers, only the overall model for the L1 English speakers was marginally significant, $F(4, 45) = 2.38$, $p = .07$. The models were not significant for the L1 French speakers, the L1 Hebrew speakers, or the L1 Chinese speakers.

Table 22. Regression models of lexical decision RTs.

Predictors	L1 English			L1 French			L1 Hebrew			L1 Chinese		
	ΔR^2	β	*p*	ΔR^2	β	*P*	ΔR^2	β	*p*	ΔR^2	β	*p*
Step 1	.034		.478	.005		.891	.023		.604	.007		.865
Flankers		.175	.257		.054	.725		-.072	.634		.080	.608
Operation span size		-.093	.543		.051	.739		-.129	.398		.008	.959
Step 2	.155		.028	.059		.287	.145		.037	.095		.128
Flankers		.170	.240		.039	.798		-.217	.167		-.005	.976
Operation span size		.063	.709		-.006	.971		-.273	.096		.064	.709
Phonological awareness composite		-.129	.429		.193	.302		.536	.011		.121	.535
Orthographic knowledge composite		-.387	.011		-.255	.132		-.258	.156		-.372	.046

6.2.4 Naming

Naming was used as a measure of productive word recognition. Two measures of naming performance were evaluated: accuracy and reaction times. For the first stage of analysis, naming performance was compared among the four L1 groups (English, French, Hebrew, and Chinese). For the second stage of analysis, the relative contributions of phonological awareness and orthographic knowledge to naming performance were evaluated for each L1 group using correlations and regressions.

6.2.5 Naming Accuracy

Accuracy was analyzed using univariate ANOVA with L1 as a between-subjects factor and the standardized incongruent-congruent RT difference on the flankers task and the operation span maximum span length as covariates. Means and standard deviations for each L1 group are given in Table 23 and displayed in Figure 27. There was a significant main effect of L1, $F(2, 133) = 10.54$, $p < .001$, $\eta_p^2 = .14$. Post-hoc comparisons among the L1 groups were thus conducted, and the Šidák correction for multiple comparisons was used. The L1 Chinese speakers had the lowest accuracy (66.4%), followed by the L1 French speakers (72.9%), the L1 Hebrew speakers (75.9%), and the L1 English speakers (97.6%). The L1 Chinese speakers were significantly less accurate than the L1 French speakers, $p = .001$, and the L1 Hebrew speakers, $p < .001$; the difference between the L1 French and the L1 Hebrew speakers was not significant.

Table 23. Proportion correct on the naming task.

L1 group	Accuracy
English	97.63% (2.00)
French	72.93% (10.18)
Hebrew	75.87% (11.17)
Chinese	66.41% (12.24)

Note. Standard deviations are given in parentheses.

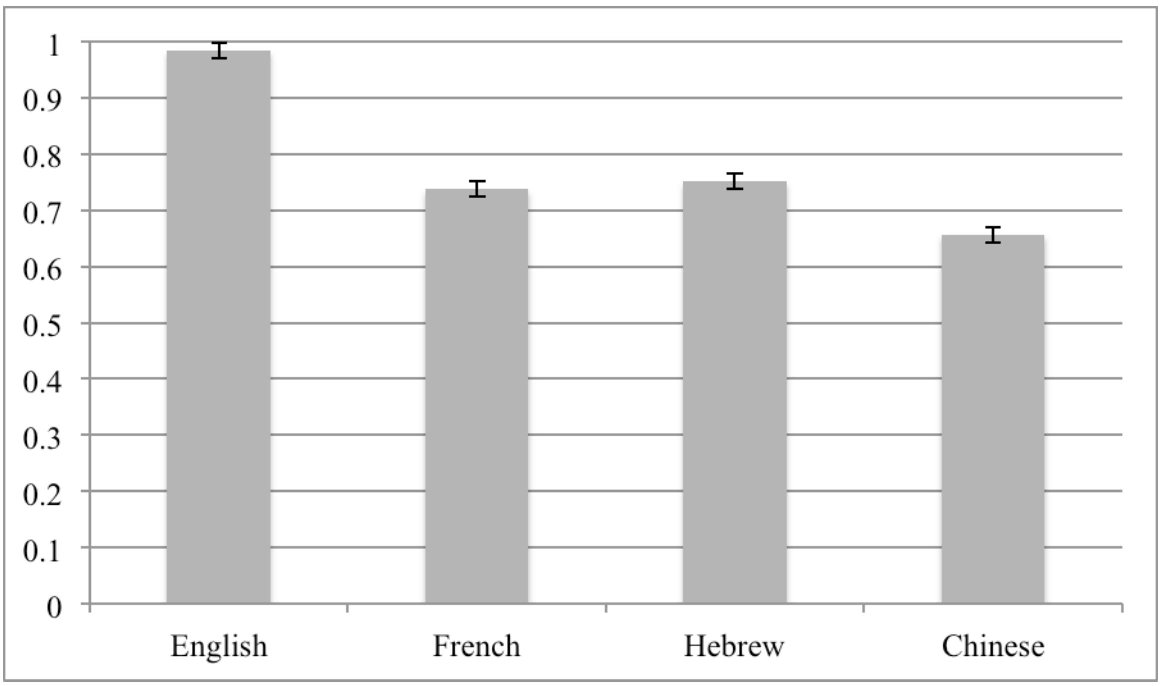

Figure 27. Proportion correct on the naming task.

For the correlation and regression analyses, each L1 group was considered separately. The correlations between the composite phonological awareness and orthographic knowledge scores and naming accuracy for each L1 group can be found in Tables 15 through 18. For the L1 English speakers the only significant correlation was between naming accuracy and the phonological awareness composite score, such that greater phonological awareness was associated with higher accuracy on the naming task. For the L1 French speakers naming accuracy was significantly correlated with lexical decision accuracy, pseudoword decoding, global reading comprehension, and operation span size. There was also a marginally significant correlation between naming accuracy and lexical decision RTs. These correlations indicate that higher accuracy on the naming task was associated with higher accuracy and faster RTs on lexical decision, higher accuracy on the pseudoword decoding and global reading comprehension measures, and also greater working memory capacity. There was also a significant correlation between naming accuracy and the composite phonological awareness and composite orthographic knowledge scores, indicating that greater phonological and orthographic skills were associated with higher naming accuracy for the L1 French speakers.

The results were broadly similar for the L1 Hebrew speakers. For this group naming accuracy was significantly correlated with lexical decision accuracy, pseudoword decoding, global reading comprehension, and operation span size. These correlations again indicate that higher accuracy on the naming task was associated with higher accuracy on lexical decision, higher accuracy on the pseudoword decoding and global reading comprehension measures, and also greater working memory capacity, and the strengths of these relationships were quite similar to those found for the L1 French speakers. There was also a significant correlation between naming accuracy and the composite phonological awareness and orthographic knowledge scores, indicating that greater phonological and orthographic skills were associated with higher naming accuracy for the L1 Hebrew speakers, as well.

Finally, for the L1 Chinese speakers, naming accuracy was significantly correlated with lexical decision accuracy, pseudoword decoding, and flankers. For this group higher naming accuracy was associated with higher lexical decision and pseudoword decoding accuracy, as before, and with a smaller congruency effect. There was again a significant correlation between naming accuracy and the composite phonological awareness and orthographic knowledge scores, indicating that greater phonological and orthographic skills were associated with higher naming accuracy.

The regression analyses were conducted next. Flankers and operation span size were entered into the regression first. For the L1 English speakers these two variables accounted for a non-significant 0.8% of variance in naming accuracy. However, flankers and operation span size accounted for a significant amount of variance in naming accuracy for all non-native English groups: L1 French, $R^2 = .132$, $p < .05$; L1 Hebrew, $R^2 = .19$, $p = .01$; L1 Chinese, $R^2 = .154$, $p < .05$.

The composite phonological awareness and orthographic knowledge scores were entered in the next step. The results of these regressions are displayed in Table 24. For the L1 English speakers, phonological awareness and orthographic knowledge together accounted for a significant additional 23.1% of variance in naming accuracy, $p < .01$. Examining the regression coefficients reveals that for the L1 English speakers, phonological awareness was a much stronger predictor of naming accuracy than was orthographic knowledge. For the L1 French speakers, phonological awareness and orthographic knowledge together accounted for a significant additional 35.7% of variance in naming accuracy, $p < .001$. In contrast to the L1 English speakers, the regression coefficients for the L1 French speakers indicate that phonological awareness and orthographic knowledge had similarly strong relationships with naming accuracy.

For the L1 Hebrew speakers, phonological awareness and orthographic knowledge again together accounted for a significant additional 21.7% of variance in naming accuracy, $p < .01$. Similar to the L1 French speakers, the regression coefficients for the L1 Hebrew speakers suggested a similar strength of relationship between phonological awareness and orthographic knowledge and naming accuracy. Finally, for the L1 Chinese speakers, phonological awareness and orthographic knowledge together accounted for a significant additional 44.7% of variance in naming accuracy, $p < .001$. Similar to the L1 English speakers, the regression coefficients for the L1 Chinese speakers indicate that phonological awareness had a much stronger relationship with naming accuracy than orthographic knowledge.

Overall, the models including flankers, operation span size, phonological awareness composite score, and orthographic composite score were significant for all four L1 groups: L1 English, $F(4, 41) = 3.22$, $p < .05$; L1 French, $F(4, 41) = 9.81$, $p < .001$; L1 Hebrew, $F(4, 41) = 7.01$, $p < .001$; L1 Chinese, $F(4, 41) = 15.44$, $p < .001$.

Table 24. Regression models of proportion correct for the naming task.

	L1 English			L1 French			L1 Hebrew			L1 Chinese		
Predictors	ΔR^2	β	*p*	ΔR^2	β	*P*	ΔR^2	β	*p*	ΔR^2	β	*p*
Step 1	.008		.838	.132		.047	.190		.011	.154		.028
Flankers		-.018	.906		.088	.537		.151	.279		-.388	.009
Operation span size		-.086	.581		.355	.016		.398	.006		.164	.258
Step 2	.231		.004	.357		<.001	.217		.002	.447		<.001
Flankers		-.010	.945		.068	.548		.002	.989		-.107	.332
Operation span size		-.380	.023		.144	.255		.173	.208		-.207	.078
Phonological awareness composite		.510	.002		.390	.007		.336	.054		.673	<.001
Orthographic knowledge composite		.185	.194		.364	.005		.270	.080		.186	.132

6.2.6 Naming RTs

RTs were also analyzed using univariate ANOVA with L1 as a between-subjects factor and the standardized incongruent-congruent RT difference on the flankers task and the operation span maximum span length as covariates. Means and standard deviations for each L1 group are given in Table 25 and displayed in Figure 28. There was a marginally significant main effect of L1, $F(2, 133) = 2.40$, $p = .095$, $\eta_p^2 = .04$. Post-hoc comparisons among the L1 groups were conducted using the Šidák correction for multiple comparisons. The L1 French speakers had the slowest RTs (841.33 ms), followed by the L1 Chinese speakers (808.39 ms), the L1 Hebrew speakers (766.61 ms), and the L1 English speakers (600.03 ms). The L1 French speakers were marginally slower than the L1 Hebrew speakers, $p = .09$; no other L1 differences were significant.

Table 25. RTs on the naming task.

L1 group	RTs (ms)
English	600.03 (100.33)
French	841.33 (179.97)
Hebrew	766.61 (143.83)
Chinese	808.39 (139.65)

Note. Standard deviations are given in parentheses.

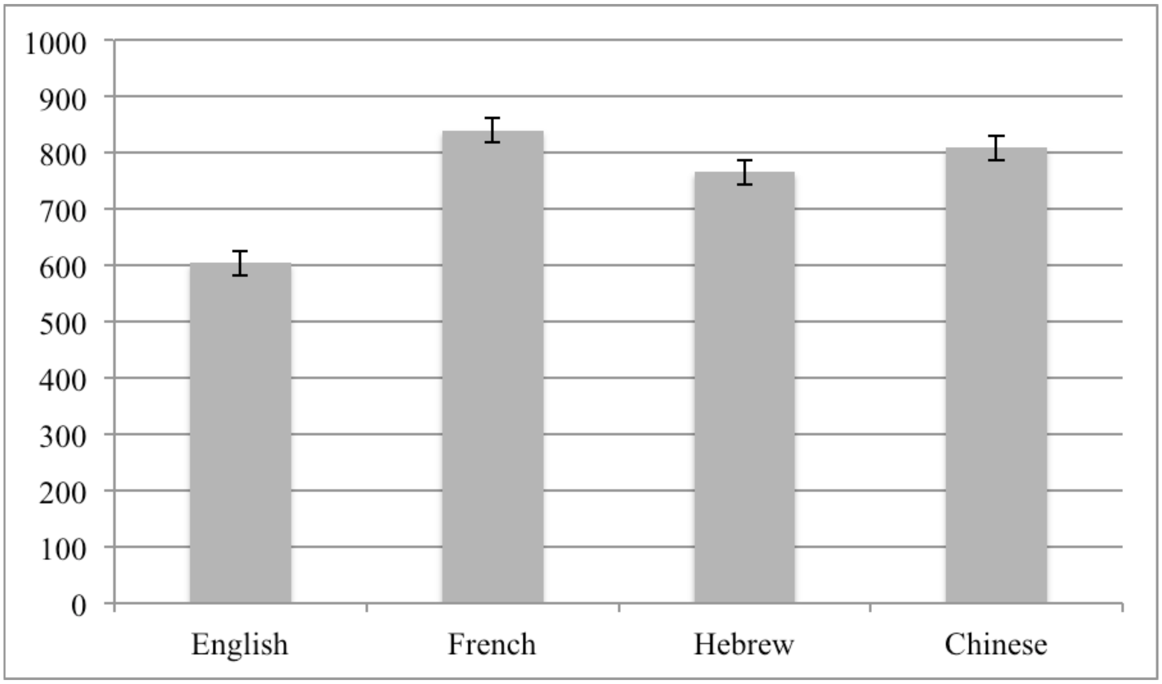

Figure 28. RTs on the naming task.

For the correlation and regression analyses, each L1 group was again considered separately. The correlations between the composite phonological awareness and orthographic knowledge scores and naming RTs for each L1 group can be found in Tables 15 through 18. For the L1 English speakers naming RTs were marginally correlated with global reading comprehension, such that shorter RTs were associated with higher reading comprehension scores. Naming RTs were also marginally correlated with the phonological awareness composite score, such that shorter RTs were associated with greater phonological awareness. For the L1 French speakers, naming RTs were marginally correlated with naming accuracy, such that shorter naming RTs were associated with higher naming accuracy. In addition, naming RTs were significant correlated with pseudoword decoding, such that shorter naming RTs were associated with higher pseudoword decoding scores. For the L1 Hebrew speakers, naming RTs were significantly correlated with lexical decision RTs, such that shorter naming RTs were also associated with shorter lexical decision RTs. Naming RTs were also marginally correlated with pseudoword decoding; shorter naming RTs were associated with higher pseudoword decoding scores. Finally, for the L1 Chinese speakers, naming RTs were only significantly correlated with pseudoword decoding; shorter naming RTs were again associated with higher pseudoword decoding scores.

The regression analyses were conducted next. Flankers and operation span size were entered into the regression first. These two variables did not account for any significant variance in naming RTs for any of the L1 groups. The composite phonological awareness and orthographic knowledge scores were entered in the next step. The results of these regressions are displayed in Table 26. For the L1 English speakers phonological awareness and orthographic knowledge together contributed a non-significant additional 6.1% of variance in naming RTs. The combined contribution of phonological awareness and orthographic knowledge to variance in naming RTs was also non-significant for the L1 French speakers and for the L1 Chinese speakers. However, phonological awareness and orthographic knowledge together explained a significant additional 22.3% of naming RTs for the L1 Hebrew speakers, $p < .01$. The regression coefficients indicate that phonological awareness had a strong positive relationship with RTs, similar to what was found for lexical decision, such that a higher phonological awareness score was associated with longer naming RTs. Orthographic knowledge had a strong negative relationship with RTs, such that a higher orthographic knowledge score was associated with shorter naming RTs.

Overall the models including flankers, operation span size, phonological awareness composite score, and orthographic knowledge composite score were not significant for the L1

English speakers, the L1 French speakers, or the L1 Chinese speakers. However, the model was significant for the L1 Hebrew speakers, $F(4, 41) = 3.06$, $p < .05$.

Table 26. Regression models of naming RTs.

Predictors	L1 English			L1 French			L1 Hebrew			L1 Chinese		
	ΔR^2	β	p	ΔR^2	β	P	ΔR^2	β	p	ΔR^2	β	p
Step 1	.010		.800	.046		.366	.006		.871	.058		.276
Flankers		-.059	.705		-.213	.160		.022	.884		-.170	.266
Operation span size		-.073	.636		-.022	.881		-.078	.609		-.143	.349
Step 2	.061		.273	.038		.432	.223		.005	.010		.796
Flankers		-.062	.685		-.222	.147		-.130	.387		-.197	.246
Operation span size		.053	.770		-.043	.800		-.205	.192		-.095	.588
Phonological awareness composite		-.278	.114		.099	.589		.635	.002		-.133	.503
Orthographic knowledge composite		.044	.776		-.215	.198		-.447	.013		.049	.793

6.2.7 Pseudoword Decoding

Accuracy on the pseudoword decoding task was calculated as the number of pseudowords each participant read aloud correctly within 45 seconds. No RT data were available for this task. As with the other variables, for the first stage of analysis, pseudoword decoding performance was compared among the four L1 groups (English, French, Hebrew, and Chinese). For the second stage of analysis, the relative contributions of phonological awareness and orthographic knowledge to pseudoword decoding were evaluated for each L1 group using correlations and regressions.

The means and standard deviations for accuracy on the pseudoword decoding task in each L1 group are given in Table 27 and displayed in Figure 29. There was a significant main effect of L1, $F(2, 133) = 4.08$, $p < .05$, $\eta_p^2 = .06$. Post-hoc comparisons among the L1 groups were thus conducted using the Šidák correction for multiple comparisons. The L1 Chinese speakers had the lowest score (17.63 words), followed by the L1 French speakers (19.91 words), the L1 Hebrew speakers (21.83 words), and the L1 English speakers (38.43 words). The L1 Chinese speakers had a significantly lower score than the L1 Hebrew speakers, $p < .05$; no other L1 differences were significant.

Table 27. Scores on the pseudoword decoding task. Maximum possible score was 63.

L1 group	Score
English	38.43 (7.13)
French	19.91 (7.91)
Hebrew	21.83 (7.58)
Chinese	17.63 (7.07)

Note. Standard deviations are given in parentheses. Maximum possible score was 63.

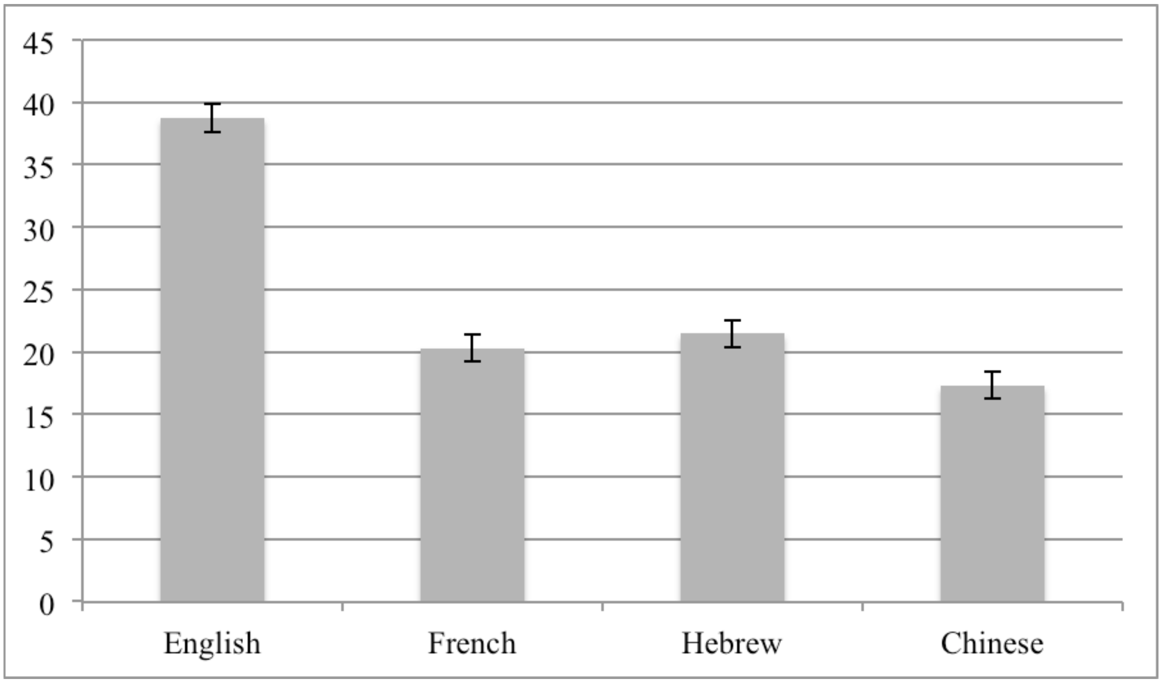

Figure 29. Scores on the pseudoword decoding task (number of pseudowords read aloud correctly in 45 seconds). Maximum possible score was 63.

For the correlation and regression analyses, each L1 group was considered separately. The correlations between the composite phonological awareness and orthographic knowledge scores and pseudoword decoding for each L1 group can be found in Tables 15 through 18. For the L1 English speakers pseudoword decoding was marginally correlated with global reading comprehension scores, so that higher pseudoword decoding scores were associated with higher global reading comprehension scores. In addition, pseudoword decoding was significantly correlated with the phonological awareness composite score, such that greater phonological awareness was associated with higher pseudoword decoding scores. For the L1 French speakers pseudoword decoding was significantly correlated with lexical decision accuracy, naming accuracy, and naming RTs, and was marginally correlated with operation span size. These correlations indicate that higher pseudoword decoding scores were associated with higher lexical decision and naming accuracy, shorter naming RTs, and higher working memory capacity. Pseudoword decoding was also significantly correlated with both the phonological awareness and the orthographic knowledge composite scores, such that greater phonological or orthographic skill was associated with higher pseudoword decoding scores.

For the L1 Hebrew speakers pseudoword decoding was significantly correlated with lexical decision accuracy and RTs, naming accuracy, and global reading comprehension, as well as marginally correlated with naming RTs. These correlations reveal that higher pseudoword decoding scores were associated with higher global reading comprehension and lexical decision and naming accuracy, as well as shorter lexical decision and naming RTs. Pseudoword decoding was also significantly correlated with the orthographic knowledge composite score, such that greater orthographic knowledge was associated with higher pseudoword decoding scores. Finally, for the L1 Chinese speakers, pseudoword decoding was significantly correlated with naming accuracy and RTs for both lexical decision and naming, such that higher pseudoword decoding scores were associated with higher naming accuracy and faster RTs on both the lexical decision and the naming tasks. Similar to the L1 French speakers, pseudoword decoding was also significantly correlated with both the phonological awareness and orthographic knowledge composite scores, such that greater phonological or orthographic skill was associated with higher pseudoword decoding scores.

The regression analyses were conducted next. Flankers and operation span size were entered into the regression first. These two variables accounted for a non-significant amount of variance in pseudoword decoding scores for all four L1 groups. The composite phonological awareness and orthographic knowledge scores were entered in the next step. The results of these regressions are displayed in Table 28. For the L1 English speakers phonological awareness and orthographic knowledge together accounted for a significant additional 14.7% of variance in pseudoword decoding scores, $p < .05$. Examining the regression coefficients reveals that phonological awareness had a significant positive relationship with pseudoword decoding, such that higher phonological awareness was associated with higher pseudoword decoding. However, the relationship between orthographic knowledge and pseudoword decoding scores was not significant. For the L1 French speakers, however, phonological awareness and orthographic knowledge accounted for only a non-significant 9.3% of additional variance in pseudoword decoding scores. For the L1 Hebrew speakers phonological awareness and orthographic knowledge accounted for a marginally significant additional 12.6% of variance in pseudoword decoding, $p = .06$. Examination of the regression coefficients reveals that in contrast to the L1 English speakers, only orthographic knowledge had a significant relationship with pseudoword decoding scores for the L1 Hebrew speakers. Finally, for the L1 Chinese speakers phonological awareness and orthographic knowledge together accounted for a significant additional 25.6% of variance in pseudoword decoding, $p < .01$. Examination of the regression coefficients for the L1 Chinese speakers reveals that similar to the L1 English speakers, phonological awareness had a strong positive relationship with pseudoword decoding scores, although there was no significant relationship for orthographic knowledge.

Overall the models including flankers, operation span size, phonological awareness composite score, and orthographic knowledge composite score were not significant for the L1 English speakers or for the L1 Hebrew speakers. However, it was marginally significant for the L1 French speakers, $F(4, 41) = 2.38$, $p = .07$, and significant for the L1 Chinese speakers, $F(4, 41) = 4.45$, $p < .01$.

Table 28. Regression models of pseudoword decoding scores.

Predictors	L1 English			L1 French			L1 Hebrew			L1 Chinese		
	ΔR^2	β	p	ΔR^2	β	p	ΔR^2	β	p	ΔR^2	β	p
Step 1	.004		.921	.095		.117	.043		.391	.046		.360
Flankers		-.055	.725		.132	.369		.042	.779		-.080	.601
Operation span size		.040	.798		.282	.058		.200	.189		.216	.162
Step 2	.147		.037	.093		.107	.126		.055	.256		.002
Flankers		-.050	.733		.127	.372		.013	.934		.121	.407
Operation span size		-.107	.533		.206	.198		.104	.519		-.064	.675
Phonological awareness composite		.390	.023		.112	.518		-.078	.700		.565	.002
Orthographic knowledge composite		-.185	.218		.252	.111		.411	.026		.051	.752

6.2.8 Reading Comprehension

In addition to examining performance on three word identification tasks, we also examined performance on reading comprehension and the relative strengths of the relations between phonological awareness and orthographic knowledge and reading comprehension. Because reading comprehension scores were also used to match participants on proficiency level, only the correction and regression analyses were conducted. The correlations between the composite phonological awareness and orthographic knowledge scores and global reading comprehension for each L1 group can be found in Tables 15 through 18. For the L1 English speakers, only two correlations were marginally significant: between reading comprehension and naming RTs, such that shorter RTs were associated with higher global reading comprehension scores, and between reading comprehension and pseudoword decoding, such that higher pseudoword decoding scores were also associated with higher global reading comprehension scores. For the L1 French speakers global reading comprehension was significantly correlated with lexical decision accuracy and naming accuracy, such that higher reading comprehension was associated with higher lexical decision and naming accuracy. Global reading comprehension was also significantly correlated with both the phonological awareness and orthographic knowledge composite scores, such that greater phonological or orthographic skill was associated with higher reading comprehension. This relationship was somewhat stronger for phonological awareness than it was for orthographic knowledge.

For the L1 Hebrew speakers global reading comprehension was significantly correlated with lexical decision and naming accuracy and pseudoword decoding; in each case, higher reading comprehension was associated with higher accuracy scores for lexical decision, naming, and pseudoword decoding. Global reading comprehension was also significantly correlated with both the phonological awareness and orthographic knowledge composite scores, such that greater phonological or orthographic skill was associated with higher reading comprehension. Finally, for the L1 Chinese speakers global reading comprehension was significantly correlated with lexical decision RTs and marginally correlated with operation span size. These correlations revealed that higher reading comprehension was associated with faster lexical decision RTs and higher working memory capacity. Global reading comprehension was also significantly correlated with the orthographic knowledge composite score, such that greater orthographic knowledge was associated with higher reading comprehension.

The regression analyses were conducted next. As previously, flankers and operation span size were entered first. These two variables accounted for a non-significant 2.2% of variance in reading comprehension scores for the L1 English speakers. For the L1 French speakers their contribution was a non-significant 4.4%, and for the L1 Hebrew speakers it was a non-significant 6.0%. However, for the L1 Chinese speakers, flankers and operation span size accounted for a marginally significant 12.6% of variance in reading comprehension scores, $p = .06$.

The composite phonological awareness and orthographic knowledge scores were entered in the next step. The results of these regressions are displayed in Table 29. For the L1 English speakers, phonological awareness and orthographic knowledge together explained a non-significant additional 7.4% of variance in reading comprehension scores. However, for the L1 French speakers, these two variables explained a significant additional 25.8% of reading comprehension scores, $p < .01$. Examination of the regression coefficients reveals that it was phonological awareness in particular that had a strong positive relationship with reading comprehension for these speakers. For the L1 Hebrew speakers, phonological awareness and orthographic knowledge together explained a marginally significant additional 11.2% of variance in reading comprehension scores, $p = .08$. However, examination of the regression coefficients for these speakers revealed no significant unique contribution for either variable. Similar to the L1 English speakers, for the L1 Chinese speakers phonological awareness and orthographic knowledge together explained only a non-significant 3.4% of variance in reading comprehension scores.

Overall the model of reading comprehension scores with flankers, operation span size, phonological awareness composite score, and orthographic knowledge composite score as predictors was only significant for the L1 French speakers, $F(4, 41) = 4.44$, $p < .01$. The model was marginally significant for the L1 Hebrew speakers, $F(4, 41) = 2.13$, $p = .095$, but it was not significant for the L1 Chinese speakers or for the L1 English speakers.

Table 29. Regression models of reading comprehension scores.

	L1 English			L1 French			L1 Hebrew			L1 Chinese		
Predictors	ΔR^2	β	p	ΔR^2	β	p	ΔR^2	β	p	ΔR^2	β	p
Step 1	.022		.626	.044		.380	.060		.263	.126		.055
Flankers		-.138	.373		.047	.753		.109	.466		-.208	.159
Operation span size		.080	.605		.206	.175		.212	.159		.329	.028
Step 2	.074		.197	.258		.002	.112		.075	.034		.448
Flankers		-.133	.383		.018	.893		.004	.978		-.157	.327
Operation span size		-.084	.635		-.026	.860		.052	.745		.296	.082
Phonological awareness composite		.256	.141		.495	.003		.231	.255		-.072	.701
Orthographic knowledge composite		.160	.300		.127	.383		.203	.259		.222	.213

6.3 STUDY 2 SUMMARY AND DISCUSSION

The purpose of Study 2 was to examine the performance of participants with representative L1 backgrounds on a variety of word identification tasks, as well as the relative contributions of phonological awareness and orthographic knowledge to predicting outcomes on these tasks. Three tasks that differed in their type of response and their demands on orthographic and phonological information were used in order to examine potential task effects in addition to this cross-linguistic comparison.

First, lexical decision was used as a receptive measure of word recognition. This task can be accomplished using either a lexical/orthographic or a phonological strategy, although phonological information does not necessarily need to be activated and lexical access was encouraged in this study via the inclusion of pseudohomophone distractors (e.g., *rane*) among the items. Second, word naming was used as a productive measure of word recognition. This task can also be accomplished using either a lexical/orthographic or a phonological strategy, although in this case phonological information must be accessed in some way for the task to be completed successfully. Third, pseudoword decoding was used as a measure of general phonological decoding skill. Although reading unfamiliar words can be accomplished via the use of an orthography-based analogy strategy (e.g., Goswami, 1986), pseudoword decoding is generally associated strongly with phonological awareness skills (e.g., Goswami & Bryant, 1990). Finally, although participants were matched across L1s on their reading comprehension scores, we also examined the relative contributions of phonological awareness and orthographic knowledge to global reading comprehension scores. Similar to the naming task, we expected stronger relationships between phonological awareness and reading comprehension for the L1 alphabetic participants (French, English) and stronger relationships between orthographic knowledge and reading comprehension for the L1 non-alphabetic participants (Hebrew, Chinese).

A comparison of the results across the three tasks reveals different patterns of performance among the L1 groups, supporting the hypothesis that learners with different L1 backgrounds will have different strengths in word identification and may rely on different underlying skills to perform the same tasks. Considering our specific predictions first, we expected that there would be stronger relationships between phonological awareness and word naming performance for the L1 French and the L1 English (alphabetic) speakers than the L1 Hebrew and the L1 Chinese (non-alphabetic) speakers, based on the common finding that readers of shallow and alphabetic languages rely relatively more on phonological information for word reading (e.g., Katz & Frost, 1992).

This prediction was only partially supported. The L1 English speakers did show a strong relationship between phonological awareness and naming accuracy (β = .51), which was noticeably stronger than the same relationship for the L1 Hebrew speakers (a marginally significant β = .34). However, the L1 French speakers showed a relationship between phonological awareness and naming accuracy that was only a bit stronger than the L1 Hebrew speakers (β = .39), and noticeably less strong than the L1 English speakers. Additionally, it was the L1 Chinese speakers who showed the strongest relationship between phonological awareness and naming accuracy (β = .67). It should also be noted that only the L1 French and the L1 Hebrew speakers showed a significant prediction of naming accuracy by orthographic knowledge in addition to phonological awareness. Although somewhat unexpected, these results also parallel the findings from Study 1, in which the L1 Hebrew speakers showed patterns of phonological awareness performance that were very similar to those of the L1 French speakers. These data therefore provide additional evidence that the segmental characteristics of the Hebrew orthography encourage the development of literacy skills and reading strategies that are similar to those developed by L1 readers of true alphabetic languages.

Although the data were not fully consistent with this original prediction, comparing the results from the lexical decision and naming tasks reveals a different pattern that was consistent with the logic of this prediction. Specifically, because of the stronger inherent demands on phonological information for the naming task compared to the lexical decision task, we expected a stronger relationship between phonological awareness and naming than between phonological awareness and lexical decision. This is what was found for all L1 groups, rather than just the L1 French and the L1 English speakers: stronger and more pervasive relationships between phonological awareness and naming accuracy than phonological awareness and lexical decision accuracy.

For the pseudoword decoding task, we predicted that phonological awareness would be more strongly related to decoding for the alphabetic L1 (French, English) speakers, but that orthographic knowledge would be more strongly related to decoding for the morphosyllabic L1 (Chinese) and abjad L1 (Hebrew) speakers. These predictions were based on typical word identification strategies in L1 for these types of writing systems (McBride-Chang et al., 2005; Nassaji & Geva, 1999; Wade-Woolley, 1999). This prediction was partially supported. The L1 English speakers showed a stronger relationship between decoding and phonological awareness than orthographic knowledge; however, the L1 French speakers did not show a significant relationship between decoding and either phonological awareness or orthographic knowledge.

The L1 Hebrew speakers showed a stronger relationship between decoding and orthographic knowledge than phonological awareness, as expected, but the L1 Chinese speakers showed a significant relationship between decoding and phonological awareness rather than orthographic knowledge.

The reason for the lack of significant relationships with decoding in the L1 French speakers is not clear, and future work is needed to determine whether this result is consistent across samples. For example, a useful comparison would be to include multiple samples of participants from alphabetic L1s that differ in their orthographic depth, to determine whether this may be a contributing factor. The finding of a stronger relationship between decoding and phonological awareness, rather than orthographic knowledge, for the L1 Chinese speakers was not expected. However, it may have resulted from the different levels of performance on the two tasks for these speakers. Performance on the orthographic knowledge tasks was relatively high, and the L1 Chinese speakers tended to be the best performing ESL group on these tasks. On the other hand, they struggled much more on the phonological awareness tasks, particularly the productive deletion task. The high levels of performance on the orthographic knowledge tasks may have been approaching ceiling, resulting in relatively little variability and thus little predictive value for these tasks when compared to the lower levels of performance on the phonological awareness tasks. In other words, the phonological awareness tasks may have been the better predictor of decoding precisely because they were more difficult, and therefore better and differentiating individuals on the basis of their phonological skills (which are also needed for decoding).

There are a number of additional findings that are worth mentioning. First, as mentioned above, the L1 Chinese speakers were significantly worse than the other L1 groups on the two tasks which required the productive use of phonological information: naming and pseudoword decoding. This finding of relatively low accuracy on the two tasks with high phonological demands is consistent with the results from Study 1, which demonstrated that the L1 Chinese speakers may have underlying phonological skills but that they are relatively weak and that these speakers are unable to apply them successfully to productive phonological tasks. These results are also consistent with our expectation that the L1 morphosyllabary speakers would have relatively lower levels of performance on tasks requiring the activation and use of phonological information. However, the L1 Chinese speakers did significantly better than the L1 French and the L1 Hebrew speakers on the lexical decision task, which did not necessarily require the use of phonological information (and in fact encouraged lexical access and the activation of orthographic details of words via the inclusion of pseudohomophones in the stimuli). This result is also consistent with the results from Study 1, demonstrating that the L1 Chinese speakers had strong orthographic skills and were able to use these skills to help them succeed at the lexical decision task.

Interestingly the L1 French speakers, who performed unexpectedly poorly on the phonological awareness tasks in Study 1, also performed relatively poorly on the naming and pseudoword decoding tasks. Although not a targeted prediction for this study, we had expected the best ESL performance from the L1 French speakers on these tasks, due to their experiences with alphabetic decoding in L1 and the strong relationships that are generally found between phonological skills and decoding (e.g., Durgunoğlu et al., 1993; Goswami & Bryant, 1990). As in Study 1, this was again in contrast to our prediction that the L1 alphabetic speakers would perform relatively well on word identification tasks that required the use of phonological information. It thus appears that despite coming from an alphabetic L1 background, the L1

French speakers had relatively weak phonological skills that also affected their word identification success. This suggests that strong phonological skills cannot necessarily be assumed in L2 learners, even when those learners come from a linguistic background that is generally more likely to encourage the development of those skills.

Regarding the reading comprehension test, only the L1 French speakers showed any significant relationship between performance on reading comprehension and either phonological awareness or orthographic knowledge. The lack of significant relationships is not consistent with our predictions, but it is likely that the wide range of skills needed to perform successfully on a global reading comprehension test simply washed out the potential effects of sub-lexical literacy skills. This interpretation is consistent with the findings of a recent meta-analysis of the correlates of L2 reading comprehension, which showed that L2 decoding has a weaker relationship with L2 reading comprehension than other factors such as L2 grammar and vocabulary knowledge (Jeon & Yamashita, 2014). It is worth noting, however, that the significant relationship we did find, between L1 French phonological awareness and reading comprehension, is of a type and direction of relationship that is consistent with our original prediction.

Finally, looking at the lexical decision data, we see that orthographic knowledge tended to have a stronger relationship with performance on this task than phonological awareness. The L1 French and the L1 Hebrew speakers both showed positive relationships between orthographic knowledge accuracy and lexical decision discrimination, while the L1 English speakers showed a negative relationship between orthographic knowledge accuracy and lexical decision RTs (such that higher orthographic knowledge scores were associated with faster lexical decision RTs). The most surprising result was that for the L1 Hebrew speakers phonological awareness was a significant *positive* predictor of lexical decision and naming RTs, so that higher phonological awareness accuracy was associated with longer RTs. It is not immediately clear why this may have been the case; one possibility is that participants who had stronger phonological awareness skills tended to activate more phonological information during lexical decision, thus resulting in more and slower lexical processing despite the fact that phonological information was not strictly necessary to perform the task.

Although the results from Study 2 did not fully support our original predictions, there is still strong evidence of the influence of both L1 and task demands. As in Study 1, these results have important implications for researchers and educators interested in L2 literacy skills. From a research perspective, it is again crucial to consider that different literacy tasks may naturally recruit different underlying skills, and that this may have a substantial influence on the results that are found. Thus, research tasks need to be chosen purposefully, with specific research questions in mind that consider response type and task demands. A similar consideration of task details needs to be used when interpreting results. Although certain L1 groups may naturally rely on some specific skills and strategies over others, differing task demands and levels of performance may interact with these tendencies to produce complex patterns of results. The results from Study 2 also have implications for educators who are interested in understanding the sources of the challenges that their students face, and how they may be able to most effectively target those challenges with instruction. The current findings provide additional support for the idea that speakers from non-alphabetic L1s may rely relatively more on orthography-based strategies for processing written forms, and speakers from alphabetic L1s relatively more on phonology-based strategies. The current findings also highlight the importance of targeting the appropriate skills for instruction based on what skills are needed to successfully complete a task:

despite their strong orthographic skills, or perhaps precisely because of them, phonological awareness was the best predictor of L1 Chinese decoding ability (a strongly phonological task). Thus, if instructors have students who are struggling with a specific task, a careful task analysis and consideration of task demands would be useful for delineating the underlying component skills that support performance on that task and allowing instructors to target those skills specifically.

7.0 STUDY 3: PHONOLOGICAL AND ORTHOGRAPHIC SKILLS AND THEIR DEVELOPMENT IN CLASSROOM ESL LEARNERS

The goal of Study 3 was to extend the work conducted in Study 1 and Study 2 with individual learners in a carefully controlled laboratory setting to active ESL students in a real classroom environment. This study allows us to examine performance on phonological awareness and orthographic knowledge tasks in a less controlled environment and compare the patterns of results that we see in each. This study also allows us to examine whether and how these literacy skills develop across a semester of intensive English instruction. Finally, we adapted and implemented a phonics-based instructional intervention for use with adult ESL learners to examine whether it was possible to impact the developmental trajectory of students' literacy skills.

7.1 METHOD

7.1.1 Participants

All participants were students enrolled in a higher-intermediate non-credit intensive English reading course at the English Language Institute at a large urban university in the United States. All students were enrolled in a high-intermediate skills-based class focused on reading during one of three consecutive academic semesters. Data were collected from a total of 177 participants across these three semesters; however, 18 participants repeated the course in multiple semesters during which data were collected. These repeating students were excluded from all analyses. In addition, there was only one student from a Cyrillic L1 background; this participant was also excluded, leaving a total of 158 students in the final sample.

For the purposes of analyses, participants were divided into one of five groups based on their L1 background. There were 17 students from a Roman alphabet L1 (1 French, 1 German, 1 Italian, 3 Portuguese, 9 Spanish, 1 Turkish, 1 Vietnamese), 18 students from a non-Roman alphabet L1 (Korean), 94 students from a consonant-based L1 (abjad or abugida: 89 Arabic and 5 Thai), 10 students from a syllabic L1 (Japanese), and 19 students from a morphosyllabary L1 (Chinese). In addition, data were collected from 29 native English speakers who were undergraduates at the University of Pittsburgh. Examination of their self-reported language learning histories revealed that 12 of these individuals had learned another language before the age of 12 or had studied a language with a different writing system (Japanese, Korean, or Chinese), and were therefore excluded. A total of 17 monolingual native English speakers were therefore used as a comparison group.

7.1.2 Test Materials

Participants completed a total of four tasks: the Elision and Blending sub-tests from the Comprehensive Test of Phonological Processing (CTOPP; Wagner et al., 1999), measuring phonological awareness; a spelling verification worksheet, measuring orthographic knowledge; and a sound discrimination task, measuring English sound discrimination abilities. Participants also completed a survey that asked them about their attitudes, habits, and experiences with regards to reading and spelling in both their L1 and in English (their L2). All data were collected in a group classroom setting in a language laboratory.

7.1.3 Phonological Awareness Tasks

Two productive tests of phonological awareness were used: Elision and Blending. In the Elision test, for each item participants heard instructions to say an English word, and then to say that word without a specific phoneme. For example, one item asked participants to say *tan* without saying /t/. The Elision test was therefore specifically a test of phoneme deletion. This test is normally administered individually; however, for the purposes of this study the administration was largely computerized and adapted to a group setting. The experimenter began by explaining to the class that they were going to play some word games, and that they were going to try some practice items to begin. The first three items were practice items; for these items the experimenter presented the prompt aurally to the whole class and elicited answers. After these three items, a further five items were given directly by the instructor, and participants were asked to record their answers individually on their computer. Following this, the correct answer was elicited from and confirmed for the whole class. After these five practice items with feedback, participants continued through the remaining 15 items on their own, listening individually to the prompts on their computer and recording their answers.

In the Blending test, for each item participants heard a question that provided them with the component sounds of an English word, broken apart and spaced out, and asked them to say what word those sounds made. For example, one item asked to participants to say what word was made by the sounds /t/-/oɪ/ (*toy*). The administration for the Blending test was very similar to the Elision test. The experimenter again began by leading the whole class through six practice items for which she elicited the answer from the class as a whole, followed by three further practice items for which participants were asked to record the answers individually on their computer. As with the Elision task, the correct answer was elicited from and confirmed for the whole class. Following these three items, participants continued through the remaining 17 items on their own, listening individually to the prompts on their computer and recording their answers. Throughout the administration of both tests, students were encouraged to work individually, avoid listening to others around them, and were monitored to make sure they were performing the tasks orally and not writing down the stimuli in order to work out the answers. For both tasks, the same items were presented in the same order for both pre-test and post-test.

7.1.4 Spelling Verification

Participants completed a spelling verification worksheet that as a measure of their orthographic (whole-word spelling) knowledge. There were a total of 120 test questions, preceded by 3 practice questions that were not scored. For each question, participants saw a single lexical item and had to mark whether they thought it was a correctly spelled English word or not. Participants were not required to make any corrections to items that they thought were misspelled, and they were encouraged to use what they knew about English words and English spelling to guess if they were unsure of their answer.

Items were specifically chosen to be words that the students would likely be familiar with. In order to accomplish this, words were chosen from a relatively small pool of possible items. These included all monomorphemic, one- or two-syllable words from the General Service List (West, 1953), the 2,000 most frequent words of English (Davies, 2008-), all words included in the ESL textbook series *Words for Students of English* volumes 1-6 (Rogerson, Davis, et al., 1992; Rogerson, Esarey, Hershelman, Jasnow, Moltz, et al., 1992; Rogerson, Esarey, Hershelman, Jasnow, Schmandt, et al., 1992; Rogerson, Esarey, Schmandt, et al., 1992; Rogerson, Hershelman, & Jasnow, 1992; Rogerson, Hershelman, Jasnow, et al., 1992), all verbs used in the ESL textbook series *Interchange*, books 1-3 (Richards, Hull, & Proctor, 2004a, 2004b, 2004c), the class vocabulary lists for students at or below the targeted class level, words that had been used productively by students in recorded language-lab activities at or below the targeted class level (Dunlap, personal communication), and words that occurred 100 or more times in a corpus of texts produced by students in the same language institute (Juffs, 2004-2015).

Words were chosen from this pool of potential items to cover a wide range of characteristics. There were 70 monosyllabic words and 50 disyllabic words. Half of each of the monosyllabic and disyllabic words were high frequency and half were low frequency; frequency level was determined using information available from the E-Lexicon (Balota et al., 2007). In addition, stimuli were chosen based on the consistency of their grapheme-phoneme correspondences. Similar to Study 2, the consistency of the words' grapheme-phoneme correspondences was determined using empirically-derived lists of consistent, inconsistent, and exception words (Treiman et al., 1995; Ziegler, Stone, et al., 1997). For the monosyllabic words, there were 24 words each with consistent, inconsistent, and exception words (taken from Ziegler, Stone, et al., 1997). For the disyllabic words, there were only 24 consistent and 24 inconsistent words because no empirically-based list of disyllabic words with exceptional spellings was available. The high- and low-frequency consistent, inconsistent, and exception monosyllabic and disyllabic words were matched as closely as possible on number of letters, number of phonemes, imageability, concreteness, bigram frequency, the number of orthographic neighbors, and frequency of those orthographic neighbors.

Finally, three different types of spellings were presented to students. For one third of each item type, the word that students saw was in fact correctly spelled, and thus they had to identify it as a correctly spelled English word. The other two thirds of each item type were misspelled. Two different types of misspelling were used: one that altered the pronunciation of the base word (e.g., *heelth* for 'health') and one that preserved the pronunciation of the base word (e.g., *portch* for 'porch'). These misspelled items were designed following the same procedure as Harris, Perfetti, and Rickles (2014). In addition, pilot-testing with 6 monolingual native English speakers and norming with 43 monolingual native English speakers was used to finalize the items. This was done to confirm that native English speakers with dialects representative of the varieties of English commonly heard in the local community of the ESL participants agreed with our categorization of the pronunciation similarity of the misspelled items and their base words.

The same items were used for both pre-test and post-test. However, the items were presented in a different pseudorandom order for each testing time, with the restriction that no more than three items of the same type were presented in a row. All items are presented in Appendix I.

7.1.5 English Sound Discrimination

Participants also completed an English sound discrimination task. A total of 27 English sound contrasts (23 vowels and 4 consonants) were tested, and each contrast was tested with two separate minimal pairs. Participants each received a cd with the test recordings as well as the worksheet for providing their answers. For each question, participants heard one English word pronounced two times, and they had to indicate which of two pictures best matched the word that they heard. Only words that could be pictured concretely and relatively unambiguously were included in the minimal pairs so that the worksheet contained only pictures and no orthographic word forms. This was done to reduce the influence of orthographic knowledge on test performance (e.g., Muneaux & Ziegler, 2004; Seidenberg & Tanenhaus, 1979; Taft & Hambly, 1985). An example item, testing the contrast between /s/ and /z/ via *ice* and *eyes* is given in Figure 30.

Figure 30. An example item from the sound discrimination task.

7.1.6 Reading Survey

Participants also completed a survey that asked them questions about how much difficulty they had learning to read and spell in their L1 and how much difficulty they have in English, their level of enjoyment reading in their L1 and in English, how often they read different types of materials for fun in their L1 and English, and how important it was for them to be a good reader and speller in their L1 and in English. On the post-test, participants were additionally asked to judge how much they had learned about English spelling and how much they felt that their reading, spelling, and word recognition abilities had improved during the semester. For all questions participants responded by choosing a number on a scale of 1 to 5 and they were provided with descriptions of the meaning of each number along the scale. Many of the questions were adapted from widely used surveys of student reading habits and motivations, including the Elementary Reading Attitude Survey (McKenna & Kear, 1990) and the Literacy Habits Survey (Paris et al., 2004). The exact questions used on the pretest and the post-test can be found in the appropriate tables in the Results below.

7.1.7 Intervention

All students enrolled in the high-intermediate reading courses during the period of data collection received the normal reading curriculum at the English Language Institute, which focused on academic vocabulary learning and reading skills such as skimming, scanning, and summarizing. To supplement this curriculum, during one of the three semesters of data collection all students enrolled in the high-intermediate reading courses additionally received a phonics-based instructional intervention. This intervention comprised four lessons, each of which was approximately 30-40 minutes in length (out of a 50-minute class period). The first two lessons were implemented once a week in consecutive weeks; the final two lessons were implemented once every other week following this. In addition to the four in-class lessons, students receiving the intervention also completed seven homework assignments that reinforced the material taught in the lessons and gave students additional practice with the skills and activities. These assignments generally required 30-45 minutes to complete, giving students a total of approximately 5.5 – 7.9 hours of phonics instruction and practice over the course of the semester.

The instructional intervention was adapted from the PHAST (Phonological and Strategy Training) program developed by Lovett and colleagues (Lovett, Lacerenza, & Borden, 2000; Lovett, Lacerenza, Borden, et al., 2000). This particular program was chosen as the basis for our intervention because most phonics-based interventions are designed and used with children who are learning to read for the first time in their native language. These interventions thus use very basic activities and materials, making them age-inappropriate for adult learners. In contrast, the PHAST program combines direct instruction with extensive metacognitive training, and the materials are largely adaptable to different ages, making them much more suited to adult learners. In addition, the PHAST program was the only phonics-based intervention we were aware of that had previous been used successfully in students up to age 20 who had reading delay. The PHAST program thus served as the basis of our intervention, which was designed in collaboration with the high-intermediate reading curriculum coordinator at the English Language Institute to ensure compatibility with the existing course schedule and materials.

The PHAST program teaches students five strategies to help them deal with reading unfamiliar words: ‘Sound It Out’, ‘Rhyming’, ‘Peeling Off’, ‘Vowel Alert’, and ‘I Spy’. The ‘Sound It Out’ strategy, presented at the beginning of the PHAST program, focuses on direct instruction of grapheme-phoneme correspondences, training students in their oral segmentation and blending skills, and applying these blending skills to sounding out unfamiliar text. The ‘Rhyming’ strategy emphasizes word identification by analogy to other words with the same spelling pattern. To use this strategy, students are introduced to the idea of keywords, a set of frequent words with common spelling patterns that can be used as ‘keys’ to unlock the ability to read a much larger set of words that also have the same spelling pattern. The ‘Peeling Off’ strategy is aimed particularly at longer, multimorphemic words. Students are taught about the concepts of prefixes and suffixes and how to identify and separate affixes from the root word, which can then be decoded and blended together with the affixes to read the whole word. The ‘Vowel Alert’ strategy takes advantage of the fact that (particularly L1) children generally have a high level of competence in their oral language and helps students connect unfamiliar written words with the oral words that they already know. Students are taught about the variability of single and especially double vowel letter pronunciations, and receive direct instruction about the frequency of the different pronunciations for each vowel type. When students encounter an unfamiliar word, they are encouraged to slow down at the vowels and try multiple possible pronunciations, starting with the most common, until they find one that forms a word that they know in their oral vocabulary (see also Cordts, 1965; Jones, 1996 for further discussion of the advantages of this type of strategy). Similar to the 'Peeling Off' strategy, the 'I Spy' strategy is designed for longer, more complex words, particularly compounds. For this strategy, students are told to look for familiar parts within a longer, unfamiliar word, and to then blend these familiar parts together. Throughout the program, students are taught to work through the problem of identifying unfamiliar words by talking themselves through the necessary steps with self-dialogue. This self-dialogue emphasizes the logic behind the various strategies and provides students with a feeling of agency in their attempts to deal with unfamiliar words. This focus on logic and self-direction, rather than rote memorization, is an additional aspect of the PHAST program that makes it particularly well suited for adaptation in an adult learning context.

The intervention used in the current study adapted and reorganized these five strategies to fit within the four class periods that were available for instruction. The first lesson focused on direct instruction of consonant sounds and common consonant grapheme-phoneme correspondences, including the consonant digraphs 'ck', 'ch', 'ph', 'sh', 'th', 'wr', 'qu', 'kn', 'ng', and 'wh'. The instruction also discussed the alternation between 'hard' and 'soft' 'c' and 'g' sounds (e.g., *cake* versus *city*) and how to determine the correct pronunciation. Finally, the first lesson introduced students to the concept of keywords for making analogies between unfamiliar and familiar words, and taught them the first of four groups of keywords, chosen from the list provided by Lovett and colleagues (Lovett, Lacerenza, & Borden, 2000; Lovett, Lacerenza, Borden, et al., 2000) to reflect the consonant patterns that were the focus of the lesson. This use of keywords and making analogies to other, known words takes advantage of students' natural tendency to use orthographic information and analogies to help them with unknown words (e.g., Goswami, 1986, 1988a, 1988b). The homework following this first lesson consisted of two worksheets on which students were asked to sort words on the basis of how their consonants were pronounced and a worksheet on which students were asked to find examples of words from their class reading with certain consonant digraphs. Students were also encouraged to practice reading aloud the keywords on a daily basis.

The second lesson focused on direct instruction of vowel sounds, a 'Double Vowel Alert' strategy very similar to the 'Vowel Alert' strategy in the original PHAST program, and a 'Rhyming' strategy again similar to the original program. The teacher began with reviewing single vowel letters, their most common short and long pronunciations, and the spelling patterns most commonly associated with those pronunciations. Teachers also introduced the 'silent –e' rule, in which vowels in a VCe pattern are generally pronounced as long vowels with a silent 'e' at the end. They then reviewed the most common vowel digraphs, their most common pronunciations, and any clues to pronunciation that might be available from the spelling pattern. The digraphs that were targeted were 'ea', 'ei/ey', 'ie', 'oo', 'ou/ow', 'ue', 'ui', 'ai/ay', 'au/aw' 'ee' and 'oa', with the most variable digraphs covered first. Students were given practice sounding out words with vowels and trying multiple pronunciations, starting with the most common, until they found one that sounded like a word they knew or had heard. Finally, students were again encouraged to make analogies with keywords and were introduced to a second set of keywords that reflected the vowel patterns emphasized in the lesson. The homework for the second lesson included a worksheet on which students were asked to sort words according to how their vowels were pronounced and a worksheet on which they identified words from their class reading that they found difficult to pronounce and identified keywords that would be helpful for getting the correct pronunciation. Students were again asked to read aloud the keywords on a daily basis. A third homework was assigned between the second and third lessons, in which students were again asked to identify difficult words from their normal class reading and indicate keywords that would be helpful for pronouncing them.

The third lesson focused on introducing morphology and the concept of word roots and affixes. Students reviewed common prefixes and suffixes and discussed their meanings and how they affected the pronunciation of the word they attached to, and both in a teacher-led context and in group work using a student-led discovery method. The affixes that were targeted were ‘em-/en-’, ‘-ic’, ‘in-/im-/il-/ir-’, ‘-able/-ible’, ‘dis-’, ‘-ion/-sion/-tion’, ‘-ive’, and ‘re-’. Students were given practice identifying affixes, removing them from multimorphemic words, and using strategies that were already familiar to them to decode the root. Students were also introduced to the third group of keywords and completed practice activities that combined separating roots from affixes, decoding, making analogies with keywords, and trying multiple vowel pronunciations. The homework assignment from the third lesson consisted of a worksheet on which students practiced identifying roots and affixes and also a worksheet on which they provided words from their class reading that had at least one affix and indicated the keywords or strategies that could be used to help them read those words. There was an additional homework assignment assigned between the third and fourth lessons. On this assignment, students were given a paragraph from their weekly reading that contained a number of words that could be read with the help of the strategies students had been practicing. They were asked to record themselves reading this paragraph aloud, and teachers provided general feedback on their pronunciation of the words.

The fourth and final lesson served primarily as a review and gave students additional practice with longer, more complex multimorphemic words. Students reviewed the strategies they had already learned and were introduced to a modified version of the 'I Spy' strategy, called 'SPY' or 'Seek the Part You know', that encouraged them to break down long, multimorphemic words and look for familiar parts inside of them (this type of strategy has also been suggested by other researchers, e.g., Sternberg, 1987). Student completed extra group work practice with complex words, were introduced to the final group of keywords for spelling analogies, and also focused on brainstorming sets of words with shared spelling patterns to emphasize analogies from a different perspective. The homework following this lesson consisted of a worksheet on which students practiced breaking long and complex words into recognizable parts and identifying keywords that would be helpful for pronouncing those parts, as well as a worksheet on which students were given particular spelling patterns and had to find examples of other words with those same spelling patterns from their class reading. The final homework assignment, completed two weeks after the fourth lesson, again asked students to record themselves reading aloud a paragraph from their class reading.

7.1.8 General Procedure

Data were collected for all tasks once in weeks 2 and 3 (pre-test) and once in weeks 12 and 13 (post-test) of a 14-week academic term, during all three semesters of data collection. The pre-test reading survey was conducted in class; the post-test reading survey was begun during a class period and finished as homework if students did not complete it before the end of class. Data for the pre-test and post-test phonological awareness tests were collected in a group administration setting, one class at a time, in a computer laboratory. Similar to the post-test reading survey, the spelling verification and sound discrimination tasks were begun during class and finished as homework if students did not complete them before the end of class.

During the intervention semester, each lesson generally occurred every other week, and students generally completed one homework assignment per week. The students' regular instructor delivered all lesson materials after undergoing a one-hours training workshop at the beginning of the semester. The instructor was in contact with the experimenter throughout the semester and the experimenter attended each class for one lesson to ensure lesson adherence. The weekly schedule for the intervention semester is given in Figure 31.

Week	Activities	Focus
Week 2	Introduction; pre-test reading survey	Pre-test
Week 3	Pre-test spelling verification, phonological awareness, sound discrimination	Pre-test
Week 4	Lesson 1, Homework 1	Consonants, keywords
Week 5	Lesson 2, Homework 2	Vowels, keywords
Week 6	Homework 3	Keyword practice, multisyllabic words
Week 7	Lesson 3, Homework 4	Affixes, "Peeling Away", keywords
Week 8	Homework 5	Reading aloud
Week 9	Lesson 4, Homework 6	"SPY", multisyllabic words, keywords
Week 10	None – Spring break	--
Week 11	Homework 7	Reading aloud, affixes, keyword practice
Weeks 12-13	All post-test activities	Post-test

Figure 31. Weekly schedule for the intervention semester.

7.2 RESULTS

In order to account for the nested structure of the classroom data (items nested within students, nested within classrooms, nested within semesters), hierarchical linear modeling was used to analyze the classroom data. The lme4 package in R was used, along with the general linear hypothesis testing (glht) function available within the multcomp package. For all post-test data the first predictor entered into the models was pre-test score in order to control for any differences among classes or L1 groups that had existed at pre-test and also because the auto-regressive effect of a variable is often one of the best predictors of later performance (Gollob & Reichardt, 1987).

7.2.1 Sound Discrimination

7.2.2 Pre-test – Sound Discrimination

Descriptive statistics for performance on the sound discrimination pre-test, organized by semester and L1 writing system type, can be found in Table 30. A binomial linear mixed-effects model was used to analyze accuracy on the sound discrimination pre-test. The first model that was estimated was a baseline random intercepts model, and included random intercepts for item, student, and class. The fixed effect of L1 writing system type (categorical, with 6 levels: English (native), L1 Roman alphabet, L1 non-Roman alphabet, L1 abjad/abugida, L1 syllabary, L1 morphosyllabary) was then added as a predictor of accuracy. The omnibus effect of L1 type was significant, $\chi^2(df = 5) = 43.28$, $p < .001$. Pairwise comparisons were calculated using general linear hypothesis testing with the Tukey correction for multiple comparisons. These analyses revealed that all five groups of non-native English speakers had a significantly lower log odds of accuracy than did the L1 English speakers; for the L1 Roman alphabet students, β = -3.07, z = -8.33, p < .001; for the L1 non-Roman alphabet students, β = -2.69, z = -7.39, p < .001; for the L1 abjad/abugida students, β = -2.83, z = 8.35, p < .001; for the L1 syllabary students, β = -2.83, z = -7.32, p < .001; and for the L1 morphosyllabary students, β = -3.27, z = -9.09, p < .001. In addition, the L1 morphosyllabary students had a significantly lower log odds of accuracy compared to the L1 abjad/abugida students, β = -.44, z = -2.84, p = .046; and compared to the L1 non-Roman alphabet students, β = -.58, z = -2.84, p = .045. The model of pre-test sound discrimination accuracy is summarized in Table 31 and the final pairwise comparisons among the L1 groups are given in Table 32.

Table 30. Proportion correct on the sound discrimination pre-test.

	Control semester 1	Intervention semester	Control semester 2
L1 Roman alphabet	85.2% (35.6)	85.1% (35.6)	78.0% (41.5)
L1 non-Roman alphabet	86.3% (34.4)	85.2% (35.6)	86.7% (34.1)
L1 abjad/abugida	83.7% (37.0)	85.3% (35.5)	84.1% (36.6)
L1 syllabary	83.3% (37.4)	88.0% (32.7)	83.8% (36.9)
L1 morphosyllabary	82.0% (38.4)	82.1% (38.5)	70.4% (45.7)
L1 English [a]	98.7% (11.4)		

Note. Standard deviations are given in parentheses. [a]Data from L1 English speakers were collected at one time point only and are listed in the control semester 1 column for convenience.

Table 31. Final model estimates of fixed effects and variance components for predicting proportion correct on the sound discrimination pre-test.

Fixed Effect	Coefficient	*SE*	*z*	*p*
Intercept	4.95	.37	13.54	< .001
L1 Roman alphabet[a]	-3.07	.37	-8.33	< .001
L1 non-Roman alphabet[a]	-2.69	.36	-7.39	< .001
L1 abjad/abugida[a]	-2.83	.34	-8.35	< .001
L1 syllabary[a]	-2.83	.39	-7.32	< .001
L1 morphosyllabary[a]	-3.27	.36	-9.09	< .001
Random Effect	**Variance Component**	***SD***		
Item	1.23	1.11		
Participant	.22	.47		
Class	.01	.11		

Note. [a]Baseline is L1 English speakers.

Table 32. Final pairwise comparisons among L1 writing system types for the model predicting proportion correct on the sound discrimination pre-test.

Comparison	Coefficient	*SE*	*z*	*p*
L1 English vs. L1 Roman alphabet	3.07	.37	8.33	< .001
L1 English vs. L1 non-Roman alphabet	2.69	.36	7.39	< .001
L1 English vs. L1 abjad/abugida	2.83	.34	8.35	< .001
L1 English vs. L1 syllabary	2.83	.39	7.32	< .001
L1 English vs. L1 morphosyllabary	3.27	.36	9.09	< .001
L1 Roman alphabet vs. L1 non-Roman alphabet	-.37	.22	-1.71	.50
L1 Roman alphabet vs. L1 abjad/abugida	-.23	.17	-1.35	.74
L1 Roman alphabet vs. L1 syllabary	-.24	.25	-.95	.93
L1 Roman alphabet vs. L1 morphosyllabary	.20	.21	.97	.92
L1 non-Roman alphabet vs. L1 abjad/abugida	.14	.16	.85	.95
L1 non-Roman alphabet vs. L1 syllabary	.13	.25	.55	.99
L1 non-Roman alphabet vs. L1 morphosyllabary	.58	.20	2.84	.045
L1 abjad/abugida vs. L1 syllabary	-.01	.21	-.03	> .99
L1 abjad/abugida vs. L1 morphosyllabary	.44	.15	2.84	.046
L1 syllabary vs. L1 morphosyllabary	.44	.24	1.85	.41

Note. Significance values reflect correction for multiple comparisons using the Tukey procedure.

7.2.3 Post-test – Sound Discrimination

Descriptive statistics for performance on the sound discrimination post-test, organized by semester and L1 writing system type, can be found in Table 33. A binomial linear mixed-effects model was again used to analyze accuracy on the sound discrimination post-test. Random intercepts for item, student, and class were again included in the baseline model. The first predictor added to the model was the fixed effect of an individual student's pre-test accuracy score; this variable was added first so that all subsequent comparisons could be made having already controlled for any differences that had existed at pre-test. The effect of pre-test accuracy was significant, β = 6.45, z = 8.90, p < .001. The fixed effect of semester was added to the model next. The omnibus effect of semester was not significant, $\chi^2(df = 2) = 1.59$, $p = .45$, thus, no comparisons among the semesters were examined.

Table 33. Proportion correct on the sound discrimination post-test.

	Control semester 1	Intervention semester	Control semester 2
L1 Roman alphabet	81.5% (39.0)	81.9% (38.6)	70.4% (45.8)
L1 non-Roman alphabet	88.6% (31.9)	86.7% (34.0)	86.1% (34.7)
L1 abjad/abugida	85.5% (35.2)	83.9% (36.8)	85.4% (35.3)
L1 syllabary	84.6% (36.2)	88.9% (31.6)	--[a]
L1 morphosyllabary	83.3% (37.3)	88.0% (32.7)	71.5% (45.2)

Note. Standard deviations are given in parentheses. [a]No data from L1 syllabary students were available for this task during this semester.

The next step was to include L1 writing system type as a predictor in the model to test its main effect. The omnibus main effect of L1 writing system type was significant, $\chi^2(df = 4) = 12.48$, $p = .01$. Pairwise comparisons corrected with the Tukey procedure indicated that the L1 Roman alphabet students had a significantly lower log odds of accuracy when compared to the L1 non-Roman alphabet students and also when compared to the L1 abjad/abugida students. The

L1 Roman alphabet students also had a marginally significant lower log odds of accuracy when compared to the L1 syllabary students and when compared to the L1 morphosyllabary students. Details of all the pairwise comparisons are given in Table 34. In this model the effect of pre-test accuracy was still significant, $\beta = 5.98$, $z = 8.34$, $p < .001$. The effect of the intervention was still not significant when compared to either control semester, $\beta = .14$, $z = .71$, $p = .48$ for fall, and $\beta = .01$, $z = .04$, $p = .97$ for summer.

The final step of our modeling was to add an interaction between L1 writing system type and semester. However, there were no available post-test data from the L1 syllabary students in the third semester of data collection, leaving an empty cell in the data set. To alleviate this problem, a new variable was created that combined the two semesters of control group data; the last model we tested thus included pre-test accuracy, instruction type (rather than semester per se), and L1 writing system type as predictors. In this final model the effect of pre-test accuracy was still significant, $\beta = 5.79$, $z = 8.22$, $p < .001$. The effect of receiving the instructional intervention, compared to the control instruction, remained non-significant, $\beta = .41$, $z = -1.23$, $p = .22$. General linear hypothesis testing with the Tukey correction for multiple comparisons was used to examine pairwise comparisons among the L1 writing system type x instruction type interactions. These comparisons revealed no significant differences among the L1 writing system type x instruction type interactions, thus the interaction term was not retained in our final model of post-test sound discrimination accuracy. The final model is summarized in Table 35, and the change across time for students receiving each type of instruction is displayed in Figure 32 and Figure 33.

Table 34. Final pairwise comparisons among L1 writing systems types for the model predicting proportion correct on the sound discrimination post-test.

Comparison	Coefficient	*SE*	*z*	*p*
L1 Roman alphabet vs. L1 non-Roman alphabet	-.72	.22	-3.26	.01
L1 Roman alphabet vs. L1 abjad/abugida	-.65	.17	-3.72	.002
L1 Roman alphabet vs. L1 syllabary	-.73	.29	-2.51	.08
L1 Roman alphabet vs. L1 morphosyllabary	-.55	.21	-2.56	.07
L1 non-Roman alphabet vs. L1 abjad/abugida	.07	.16	.56	.99
L1 non-Roman alphabet vs. L1 syllabary	-.01	.28	-.04	> .99
L1 non-Roman alphabet vs. L1 morphosyllabary	.17	.20	.86	.91
L1 abjad/abugida vs. L1 syllabary	-.08	.25	-.34	> .99
L1 abjad/abugida vs. L1 morphosyllabary	.10	.15	.65	.96
L1 syllabary vs. L1 morphosyllabary	.18	.27	.67	.96

Note. Reported significance values reflect correction for multiple comparisons using the Tukey procedure.

Table 35. Final model estimates of fixed effects and variance components for predicting proportion correct on the sound discrimination post-test.

Fixed Effect	Coefficient	*SE*	*z*	*p*
Intercept	-3.54	.64	-5.53	< .001
Sound discrimination pre-test score	5.98	.72	8.34	< .001
Control semester 1[a]	.14	.20	.71	.49
Control semester 2[a]	.01	.20	.04	.97
L1 Non-Roman alphabet[b]	.72	.22	3.26	.001
L1 abjad/abugida[b]	.65	.17	3.72	< .001
L1 syllabary[b]	.73	.29	2.51	.01
L1 morphosyllabary[b]	.55	.21	2.56	.01
Random Effect	**Variance Component**	***SD***		
Item	1.18	1.09		
Participant	.08	.29		
Class	.05	.22		

Note. [a]Baseline is intervention semester. [b]Baseline is L1 Roman alphabet.

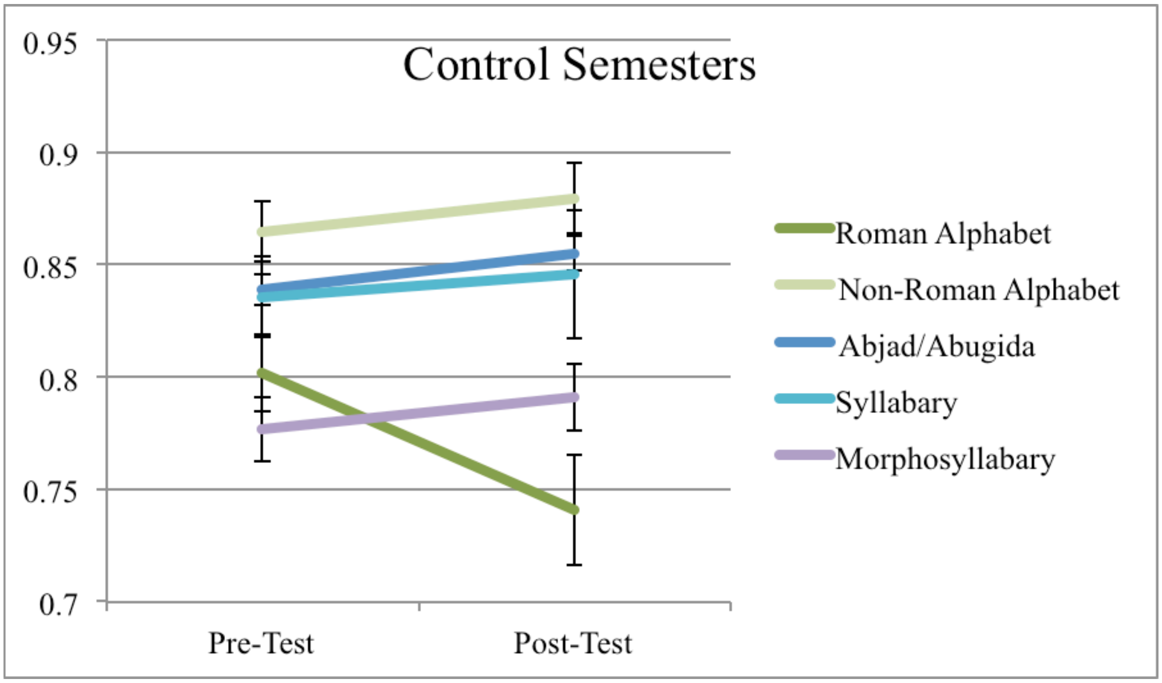

Figure 32. Proportion correct on the sound discrimination task at pre-test and post-test for each language group receiving the control instruction.

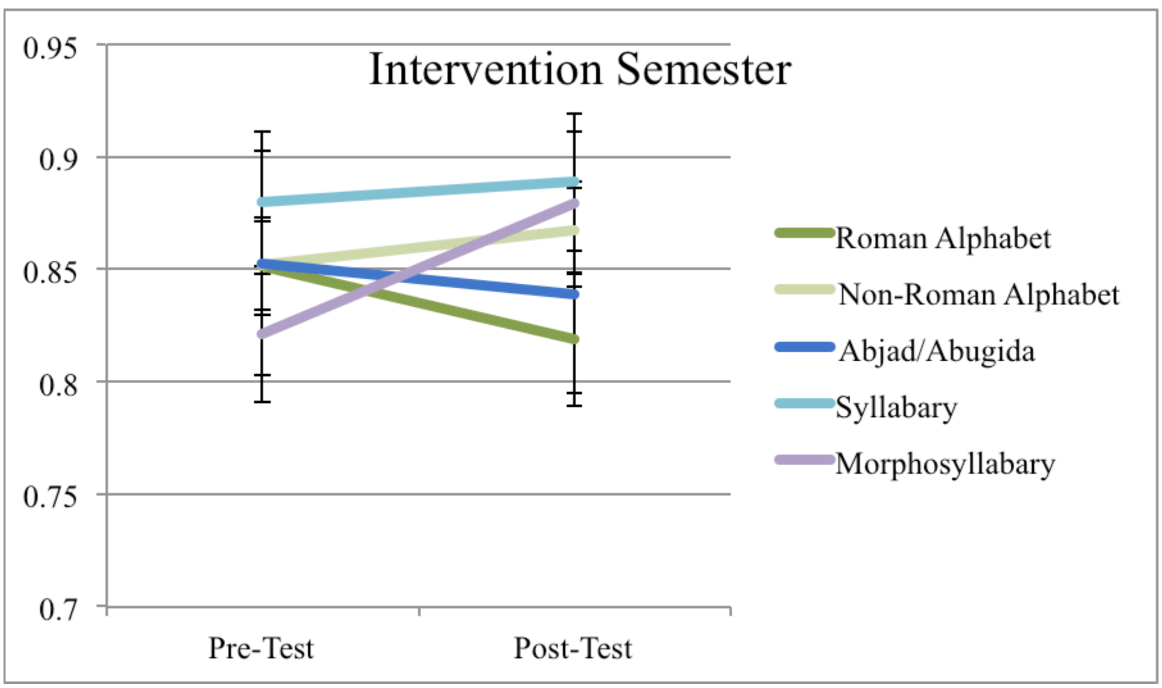

Figure 33. Proportion correct on the sound discrimination task at pre-test and post-test for each language group receiving the intervention instruction.

7.2.4 Phonological Awareness

7.2.5 Pre-test – Phonological Awareness

Descriptive statistics for performance on the phonological awareness pre-test, organized by task, semester, and L1 writing system type, can be found in Table 36. A binomial linear mixed-effects model was used to analyze accuracy on the two phonological awareness pre-test tasks. The first model that was estimated with a baseline random intercepts model, which included random intercepts for item, student, and class. Following this, the fixed effect of L1 writing system type (categorical, with 6 levels: English (native), L1 Roman alphabet, L1 non-Roman alphabet, L1 abjad/abugida, L1 syllabary, L1 morphosyllabary) was added as a predictor of accuracy. The omnibus effect of L1 was significant, $\chi^2(df = 5) = 37.51$, $p < .001$. Pairwise comparisons were made using general linear hypothesis testing with the Tukey correction for multiple comparisons. These comparisons showed that the L1 English speakers had a significantly higher log odds of responding accurately to a given item than all five non-native English groups. In addition, the L1 non-Roman alphabet and the L1 syllabary students had a significantly lower log odds of accuracy than the abjad/abugida L1 students, and the L1 Roman alphabet and the L1 morphosyllabary students had a marginally significant lower log odds than the abjad/abugida students. Details of all the pairwise comparisons are given in Table 37.

Table 36. Proportion correct on the phonological awareness pre-test.

	Elision		
	Control semester 1	Intervention semester	Control semester 2
L1 Roman alphabet	80.0% (40.3)	78.8% (41.2)	52.5% (50.1)
L1 non-Roman alphabet	58.6% (49.5)	62.5% (48.6)	64.0% (48.2)
L1 abjad/abugida	82.6% (37.9)	73.5% (44.2)	70.2% (45.8)
L1 syllabary	61.0% (49.2)	77.5% (42.3)	47.5% (50.3)
L1 morphosyllabary	69.8% (46.0)	71.7% (45.4)	56.8% (49.8)
L1 English [a]	92.8% (25.9)		
	Blending		
	Control semester 1	Intervention semester	Control semester 2
L1 Roman alphabet	71.7% (45.4)	68.8% (46.6)	48.4% (50.1)
L1 non-Roman alphabet	67.7% (47.0)	65.0% (47.9)	54.0% (50.1)
L1 abjad/abugida	75.7% (43.0)	69.4% (46.2)	63.8% (48.1)
L1 syllabary	50.8% (50.4)	72.5% (45.2)	48.8% (50.3)
L1 morphosyllabary	68.8% (46.4)	73.3% (44.6)	45.7% (50.1)
L1 English [a]	93.1% (25.3)		

Note. Standard deviations are given in parentheses. [a]Data from L1 English speakers were collected at one time point only and are listed in the control semester 1 column for convenience.

Table 37. Final pairwise comparisons among L1 writing system types for the model predicting proportion correct on the phonological awareness pre-test.

Comparison	Coefficient	*SE*	*z*	*p*
L1 English vs. L1 Roman alphabet	2.74	.51	5.33	< .001
L1 English vs. L1 non-Roman alphabet	2.84	.51	5.57	< .001
L1 English vs. L1 abjad/abugida	2.11	.47	4.48	< .001
L1 English vs. L1 syllabary	3.08	.54	5.65	< .001
L1 English vs. L1 morphosyllabary	2.70	.51	5.35	< .001
L1 Roman alphabet vs. L1 non-Roman alphabet	.10	.32	.32	> .99
L1 Roman alphabet vs. L1 abjad/abugida	-.63	.25	-2.57	.095
L1 Roman alphabet vs. L1 syllabary	-.34	.37	-.93	.93
L1 Roman alphabet vs. L1 morphosyllabary	-.03	.31	-.10	> .99
L1 non-Roman alphabet vs. L1 abjad/abugida	-.73	.24	-3.07	.02
L1 non-Roman alphabet vs. L1 syllabary	.24	.36	.66	.99
L1 non-Roman alphabet vs. L1 morphosyllabary	-.13	.30	-.45	> .99
L1 abjad/abugida vs. L1 syllabary	.97	.30	-3.20	.02
L1 abjad/abugida vs. L1 morphosyllabary	.60	.23	2.60	.09
L1 syllabary vs. L1 morphosyllabary	-.37	.36	-1.04	.89

Note. Reported significance values reflect correction for multiple comparisons using the Tukey procedure.

In the next step, task was added as a categorical predictor to determine whether there was a significant difference in performance on the elision versus the blending task. The effect of task was not significant, $\beta = .11$, $z = .26$, $p = .80$. An additional model was fit to determine whether task was a significant predictor of the log odds of accuracy prior to entering L1 writing system type as a predictor; however, task remained non-significant in this case as well. For the next step the interaction between task and L1 writing system type was entered, to determine whether the difference in task performance might have differed between students from different L1 writing systems. However, the addition of this interaction resulted in the model failing to converge; task and the task x L1 writing system type interaction were thus not retained as predictors.

In the final step, participants' sound discrimination pre-test score was added as a predictor to determine whether general sound discrimination ability could predict additional variance in pre-test phonological awareness accuracy. However, the addition of this variable also resulted in the model failing to converge. The final model for pre-test phonological awareness therefore included only L1 type as a predictor; this model is summarized in Table 38.

Table 38. Final model estimates of fixed effects and variance components for the model predicting proportion correct on the phonological awareness post-test.

Fixed Effect	Coefficient	*SE*	*z*	*p*
Intercept	3.48	.49	7.10	< .001
L1 Roman alphabet[a]	-2.66	.53	-5.06	< .001
L1 non-Roman alphabet[a]	-2.80	.52	-5.37	< .001
L1 abjad/abugida[a]	-2.07	.48	-4.28	< .001
L1 syllabary[a]	-3.03	.55	-5.47	< .001
L1 morphosyllabary[a]	-2.66	.52	-5.14	< .001

Random Effect	Variance Component	*SD*
Item	2.02	1.42
Participant	.58	.76
Class	.02	.13
Semester	.10	.32

Note. [a]Baseline is L1 English speakers.

7.2.6 Post-test – Phonological Awareness

Descriptive statistics for performance on the phonological awareness post-test, organized by task, semester, and L1 writing system type, can be found in Table 39. A binomial linear mixed-effects model was used to analyze the log odds of accuracy on the phonological awareness post-test tasks. The first model that was estimated was a baseline random intercepts model, which included random intercepts for item, student, class, and semester. The first predictor that was included was participants' overall score from the phonological awareness pre-test. This was a significant predictor of the log odds of accuracy on the post-test, $\beta = 4.51$, $z = 12.56$, $p < .001$. The next predictor that was added to the model was semester (categorical: control semester 1, control semester 2, intervention semester). Although the omnibus effect of semester was not significant, $\chi^2(df = 2) = 4.56$, $p = .10$, the planned comparisons between each control semester and the intervention semester indicated that control semester 1 was associated with a significantly higher log odds of accuracy on the phonological awareness post-test, $\beta = .44$, $z =$

2.24, p = .03, and control semester 2 was associated with a marginally significant higher log odds of accuracy on the phonological awareness post-test, β = .37, z = 1.93, p = .05. This factor was thus retained in the model.

Table 39. Proportion correct on the phonological awareness post-test.

	Elision		
	Control semester 1	Intervention semester	Control semester 2
L1 Roman alphabet	81.7% (39.0)	69.0% (46.5)	55.9% (49.9)
L1 non-Roman alphabet	82.0% (38.6)	57.5% (49.6)	71.0% (45.6)
L1 abjad/abugida	85.0% (35.8)	69.0% (46.3)	74.0% (43.9)
L1 syllabary	63.3% (48.6)	67.5% (47.4)	60.0% (49.6)
L1 morphosyllabary	67.3% (46.1)	85.0% (36.2)	63.0% (48.6)
	Blending		
	Control semester 1	Intervention semester	Control semester 2
L1 Roman alphabet	78.3% (41.5)	70.0% (46.1)	52.9% (50.1)
L1 non-Roman alphabet	74.0% (44.1)	65.8% (47.6)	56.1% (49.9)
L1 abjad/abugida	76.0% (42.7)	69.5% (46.1)	68.9% (46.4)
L1 syllabary	76.7% (42.7)	62.5% (49.0)	43.2% (50.2)
L1 morphosyllabary	71.8% (45.1)	55.0% (50.4)	55.7% (50.0)

Note. Standard deviations are given in parentheses.

The next step was to include L1 writing system type as a predictor to determine whether L1 type predicted additional variance on the post-test, beyond its ability to predict performance on the pre-test (the scores of which were already included in the post-test model). The omnibus test of L1 was not significant, $\chi^2(df = 4) = 1.52$, p = .82. The interaction between L1 writing system type and semester was also tested, to determine whether the change across a semester differed among students from different L1 backgrounds. The inclusion of the interaction effect did not improve the model, $\chi^2(df = 8) = 5.33$, p = .72. There was no evidence of a unique predictive value for L1 writing system type in post-test performance, and thus this variable was not retained in the model.

We then examined whether there was a significant effect of task (elision or blending) on post-test performance. The effect of task was not significant, $\beta = .21$, $z = .47$, $p = .64$. Before removing task from the model, however, we tested whether there was any evidence of an interaction between task and semester. The comparison between models with and without the task x semester interaction term revealed that the inclusion of this interaction resulted in a marginally better model fit, $\chi^2(df = 2) = 5.05$, $p = .08$. General linear hypothesis testing was used to examine the pairwise comparisons for this interaction. The results revealed that the effect of task was significant for the comparison between control semester 2 and the intervention semester, $\beta = .43$, $z = 2.19$, $p = .03$, and that the effect of task was marginally significant for the comparison between control semester 1 and the intervention semester, $\beta = .35$, z = 1.69, $p = .09$, but that there was no effect of task for the comparison between the two control semesters, $\beta = .08$, $z = .43$, $p = .67$. These means can be compared in Table 40.

Table 40. Proportion correct on the phonological awareness post-test, by semester and task.

	Control semester 1	Intervention semester	Control semester 2
Blending task	75.2% (43.2)	67.7% (46.8)	62.4% (48.5)
Elision task	79.5% (40.4)	67.9% (46.7)	69.1% (46.2)

Note. Standard deviations are given in parentheses.

For the final step in modeling post-test phonological awareness accuracy, students' pre-test sound discrimination score was added as the final predictor. However, the addition of this variable resulted in the model failing to converge, and this variable was thus kept out of the final model. The final model included pre-test phonological awareness score, semester, task, and the semester x task interaction as predictors and is summarized in Table 41. The change across time for students receiving each type of instruction is displayed in Figure 34 and Figure 35.

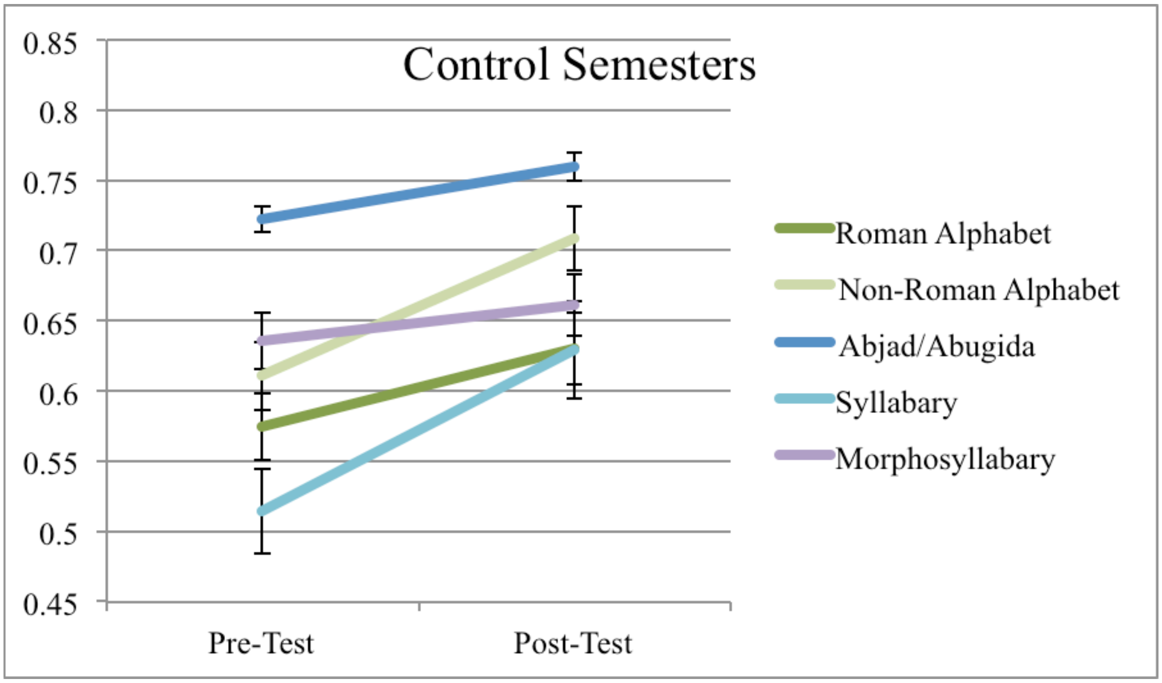

Figure 34. Proportion correct on the phonological awareness tasks at pre-test and post-test for each language group receiving the control instruction.

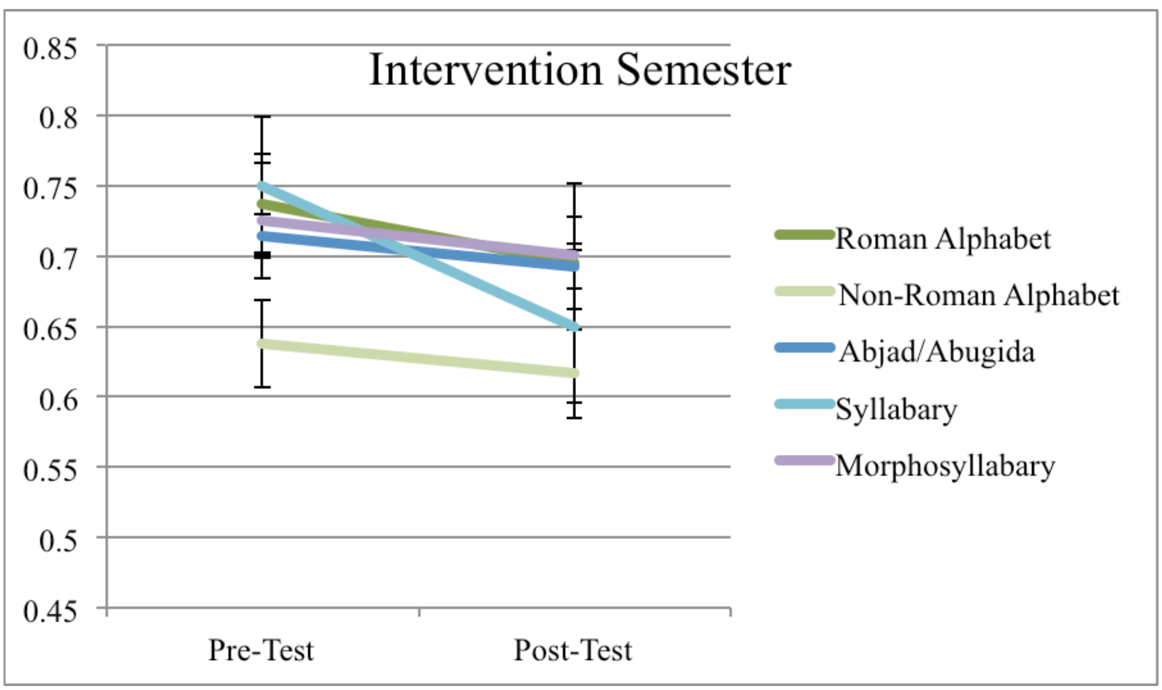

Figure 35. Proportion correct on the phonological awareness tasks at pre-test and post-test for each language group receiving the intervention instruction.

Table 41. Final model estimates of fixed effects and variance components for the model predicting proportion correct on the phonological awareness post-test.

Fixed Effect	Coefficient	*SE*	*z*	*p*
Intercept	-2.18	.43	-5.02	< .001
Phonological awareness pre-test score	4.54	.36	12.59	< .001
Control semester 1[a]	.26	.22	1.17	.24
Control semester 2[a]	.15	.22	.70	.48
Elision task[b]	-.07	.47	-.15	.88
Control semester 1[a] x elision task[b]	.35	.20	1.69	.09
Control semester 2[a] x elision task[b]	.43	.20	2.19	.03

Random Effect	Variance Component	*SD*
Item	1.99	1.41
Participant	.12	.34
Class	.04	.20

Note. [a]Baseline is intervention semester. [b]Baseline is blending task.

7.2.7 Spelling Verification

7.2.8 Pre-test – Spelling Verification

Descriptive statistics for performance on the spelling verification pre-test, organized by L1 writing system type, can be found in Table 42. A binomial linear mixed-effects model was also used to analyze accuracy on the spelling verification pre-test. The first model that was estimated was a baseline random intercepts model, and included random intercepts for item, student, and class. Following this, the four word-level characteristics were added step-by-step in order to determine whether they were unique predictors of performance on the pre-test spelling verification task. These variables were entered in order of expected influence, with the most influential variables added earlier. The first factor that was included was word frequency. This variable was included as a categorical predictor because continuous frequency estimates were not available on the same scale for all items. The effect of frequency was significant, with high frequency words being associated with an increase in the log odds of accuracy compared to low frequency words, β = 1.10, z = 4.78, p < .001. The number of syllables (monosyllabic vs. disyllabic) was added to the model next. Although it was not a significant predictor, β = -0.23, z = -0.99, p = .32, number of syllables was retained in the model for one additional step to examine whether its regression coefficient would change with more factors being considered.

Table 42. Proportion correct on the spelling verification pre-test.

	Control semester 1	Intervention semester	Control semester 2
L1 Roman alphabet	88.3% (32.1)	84.0% (36.7)	81.5% (30.6)
L1 non-Roman alphabet	88.0% (32.5)	88.3% (32.1)	90.2% (29.8)
L1 abjad/abugida	76.0% (42.7)	79.9% (40.1)	76.1% (42.7)
L1 syllabary	81.5% (38.9)	90.4% (29.5)	84.4% (36.3)
L1 morphosyllabary	78.4% (41.2)	68.6% (46.5)	77.8% (41.6)
L1 English [a]	98.1% (13.7)		

Note. Standard deviations are given in parentheses. [a]Data from L1 English speakers were collected at one time point only and are listed in the control semester 1 column for convenience.

In the next step spelling consistency (categorical: consistent, inconsistent, exception) was added as a predictor. The omnibus effect of consistency was significant, $\chi^2(df = 2) = 6.74$, p = .03. Pairwise comparisons among the three consistency types were calculated using general linear hypothesis testing with the Tukey correction for multiple comparisons. These comparisons revealed that the log odds of accuracy for exception words was significantly lower than that for a consistent word (β = -0.84, p = .03). In addition, the log odds of accuracy for exception words was marginally lower than that for an inconsistent word (β = .76, p = .05). Word frequency remained a significant predictor, β = 1.10, z = 4.91, p < .001, and in this model number of syllables was also a significant predictor of the log odds of accuracy, β = -.50, z = -2.01, p = .04. Number of syllables was thus retained as a predictor going forward.

The final word characteristic, answer type, was added to the model next. This factor indicated whether the stimulus was in fact correctly spelled (and students thus had to identify it as correctly spelled) or whether it was misspelled with either a pronunciation-preserving or a pronunciation-changing misspelling (see Method for details). The omnibus effect of answer type was not significant, $\chi^2(df = 2) = .73$, $p = .69$. No comparisons among the three answer types were thus made. As with the number of syllables, answer type was retained in the model for one additional step to evaluate whether it would become a significant predictor. However, it remained non-significant in the next model and was thus excluded as a predictor. Word frequency, number of syllables, and spelling consistency were the three word-level characteristics retained in the model.

L1 writing system type was included as the next predictor. As with the sound discrimination analyses, this was a categorical predictor with six levels: L1 English speaker, L1 Roman alphabet, L1 non-Roman alphabet, L1 abjad/abugida, L1 syllabary, and L1 morphosyllabary. The omnibus effect of L1 writing system type was significant, $\chi^2(df = 5) = 68.5$, $p < .001$. General linear hypothesis testing with the Tukey correction for multiple comparisons was again used to examine the pairwise comparisons among all L1 writing system types. All five non-native English groups had significantly lower log odds of responding correctly to a given item than the L1 English speakers. In addition, the L1 non-Roman alphabet and the Roman alphabet students each had a significantly higher log odds of responding to a given item correctly than the L1 abjad/abugida students, and the L1 Roman alphabet students also had a significantly higher log odds of accuracy than the L1 morphosyllabary students. Details of these pairwise comparisons are given in Table 43.

Table 43. Final pairwise comparisons among L1 writing system types for the model predicting proportion correct on the spelling verification pre-test.

Comparison	Coefficient	*SE*	*z*	*p*
L1 English vs. L1 Roman alphabet	2.32	.34	6.83	< .001
L1 English vs. L1 non-Roman alphabet	2.10	.34	6.19	< .001
L1 English vs. L1 abjad/abugida	3.23	.29	11.22	< .001
L1 English vs. L1 syllabary	2.66	.38	7.06	< .001
L1 English vs. L1 morphosyllabary	3.23	.33	9.74	< .001
L1 Roman alphabet vs. L1 non-Roman alphabet	-.23	.29	-.79	.97
L1 Roman alphabet vs. L1 abjad/abugida	.91	.22	4.09	< .001
L1 Roman alphabet vs. L1 syllabary	.34	.33	1.04	.90
L1 Roman alphabet vs. L1 morphosyllabary	.91	.28	3.28	.01
L1 non-Roman alphabet vs. L1 abjad/abugida	1.13	.22	5.15	< .001
L1 non-Roman alphabet vs. L1 syllabary	.57	.33	1.73	.50
L1 non-Roman alphabet vs. L1 morphosyllabary	1.13	.27	4.15	< .001
L1 abjad/abugida vs. L1 syllabary	-.56	.27	-2.05	.30
L1 abjad/abugida vs. L1 morphosyllabary	.002	.21	.01	> .99
L1 syllabary vs. L1 morphosyllabary	.57	.32	1.76	.47

Note. Reported significance values reflect correction for multiple comparisons using the Tukey procedure.

For the final step of the analysis, participants' overall phonological awareness pre-test score and their sound discrimination pre-test score were added one at a time to the model to determine whether they could predict further variability in participants' pre-test spelling verification accuracy. However, in each case the addition of this final variable resulted in the model failing to converge. Thus, these variables were not retained; the final model is summarized in Table 44.

Table 44. Final model estimates of fixed effects and variance components for the model predicting proportion correct on the spelling verification pre-test.

Fixed Effect	Coefficient	*SE*	*z*	*p*
Intercept	4.73	.36	12.98	< .001
High frequency[a]	1.10	.22	4.92	< .001
Disyllabic[b]	-.51	.25	-2.02	.04
Consistent spelling[c]	.08	.25	.33	.74
Exception spelling[c]	-.76	.33	-2.31	.02
L1 Roman alphabet[d]	-2.32	.34	-6.83	< .001
L1 non-Roman alphabet[d]	-2.10	.34	-6.19	< .001
L1 abjad/abugida[d]	-3.23	.29	-11.22	< .001
L1 syllabary[d]	-2.66	.38	-7.06	< .001
L1 morphosyllabary[d]	-3.23	.33	-9.74	< .001

Random Effect	Variance Component	*SD*
Item	1.41	1.19
Participant	.59	.77
Class	.01	.10

Note. [a]Baseline is low frequency words. [b]Baseline is monosyllabic words. [c]Baseline is inconsistent spelling words. [d]Baseline is L1 English speakers.

7.2.9 Post-test – Spelling Verification

Descriptive statistics for performance on the spelling verification post-test, organized by semester and L1 writing system type, can be found in Table 45. A binomial linear mixed-effects model was again used to analyze accuracy on the spelling verification post-test. Random intercepts for item, student, class, and semester were again included in the baseline model. The first predictor added to the model was the fixed effect of an individual student's pre-test accuracy score; this variable was added first so that all subsequent comparisons could be made having already controlled for any differences that had existed at pre-test. The effect of pre-test accuracy was significant, $\beta = 6.17$, $z = 10.50$, $p < .001$. In the next step the fixed effect of semester was added as a categorical predictor. There were three levels: control semester 1, control semester 2, and intervention semester. The omnibus effect of semester was marginally significant, $\chi^2(df = 2) = 5.51$, $p = .06$. Looking at the comparison of each control semester versus the intervention semester, students in control semester 1 had a marginally significant lower log odds of responding to a given item correctly, $\beta = -.29$, $z = -1.76$, $p = .08$. Students in control semester 2 had a significantly lower log odds of responding to a given item correctly, $\beta = -.46$, $z = -2.67$, $p = .008$.

Table 45. Proportion correct on the spelling verification post-test.

	Control semester 1	Intervention semester	Control semester 2
L1 Roman alphabet	94.6% (22.7)	87.8% (32.8)	87.7% (32.9)
L1 non-Roman alphabet	88.8% (31.5)	90.8% (28.9)	90.2% (29.8)
L1 abjad/abugida	78.7% (40.9)	83.1% (37.4)	75.3% (43.1)
L1 syllabary	89.2% (31.1)	92.1% (27.1)	--[a]
L1 morphosyllabary	77.5% (41.8)	87.9% (32.7)	87.0% (33.7)

Note. Standard deviations are given in parentheses. [a]No data from L1 syllabary students were available for this task during this semester.

The next predictor added to the model was L1 writing system type. The omnibus effect of this variable was significant, $\chi^2(df = 4) = 10.57$, $p = .03$. As before, general linear hypothesis testing with the Tukey correction for multiple comparisons was used to examine the pairwise comparisons among all L1 groups. However, with pre-test scores and instruction type already in the model, only one difference was marginally significant: the L1 Roman alphabet students had a marginally higher log odds of responding to a given item correctly than the L1 abjad/abugida students (difference = .58, $p = .053$).

The final step in modeling the spelling verification post-test data was to examine whether pre-test scores on the phonological awareness or sound discrimination tasks, or the four item characteristics, were unique predictors of post-test performance even after accounting for pre-test performance, instruction type, and L1 writing system type. Each of these six variables was entered one at a time into the model; however, in each case the addition of the new predictor resulted in a failure of the model to converge. Thus, the final model of post-test spelling verification accuracy included only pre-test score, instruction type, and L1 writing system type. This model is summarized in Table 46, and the final pairwise comparisons among L1 groups are given in Table 47. The change across time for students receiving each type of instruction is displayed in Figure 36 and Figure 37.

Table 46. Final model estimates of fixed effects and variance components for the model predicting proportion correct on the spelling verification post-test.

Fixed Effect	Coefficient	*SE*	*z*	*p*
Intercept	-1.40	.59	-2.38	.02
Spelling verification pre-test score	5.25	.63	8.34	< .001
Control semester 1[a]	-.27	.16	-1.72	.08
Control semester 2[a]	-.43	.16	-2.64	.01
L1 Non-Roman alphabet[b]	-.16	.26	-.61	.54
L1 abjad/abugida[b]	-.57	.21	-2.66	.008
L1 syllabary[b]	.02	.36	.06	.96
L1 morphosyllabary[b]	-.33	.26	-1.26	.21

Random Effect	Variance Component	*SD*
Item	1.86	1.36
Participant	.32	.57
Class	.001	.03

Note. [a]Baseline is intervention semester. [b]Baseline is L1 Roman alphabet.

Table 47. Final pairwise comparisons among L1 writing system types for the model predicting proportion correct on the spelling verification post-test.

Comparison	Coefficient	*SE*	*z*	*p*
L1 Roman alphabet vs. L1 non-Roman alphabet	.16	.26	.61	.97
L1 Roman alphabet vs. L1 abjad/abugida	.57	.21	2.66	.06
L1 Roman alphabet vs. L1 syllabary	-.02	.36	-.06	> .99
L1 Roman alphabet vs. L1 morphosyllabary	.33	.26	1.26	.70
L1 non-Roman alphabet vs. L1 abjad/abugida	.41	.22	1.90	.30
L1 non-Roman alphabet vs. L1 syllabary	-.18	.36	-.51	.99
L1 non-Roman alphabet vs. L1 morphosyllabary	.17	.25	.66	.96
L1 abjad/abugida vs. L1 syllabary	-.59	.32	-1.86	.33
L1 abjad/abugida vs. L1 morphosyllabary	-.24	.18	-1.32	.66
L1 syllabary vs. L1 morphosyllabary	.35	.35	1.00	.85

Note. Reported significance values reflect correction for multiple comparisons using the Tukey procedure.

7.2.10 Reading Survey

7.2.11 Pre-test – Reading Survey

In order to analyze the reading survey pre-test data, a separate linear mixed effects model was used to examine the effect of L1 writing system type on how participants responded to each question. For each question, the omnibus effect of L1 writing system type was examined, followed by pairwise comparisons among all L1 types if appropriate, with the Tukey correction for multiple comparisons. All answers were provided on a 5-point scale (see tables below for details). Means and standard deviations for each question, organize by L1 writing system type, are presented in Table 48.

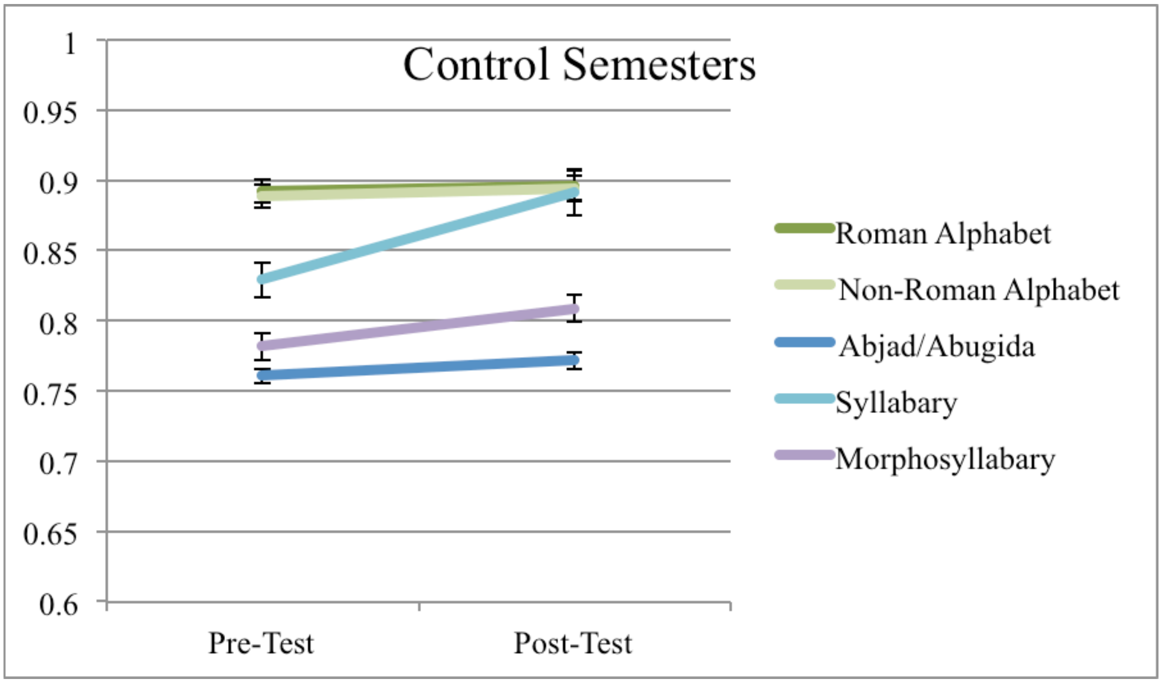

Figure 36. Proportion correct on the spelling verification task at pre-test and post-test for each language group receiving the control instruction.

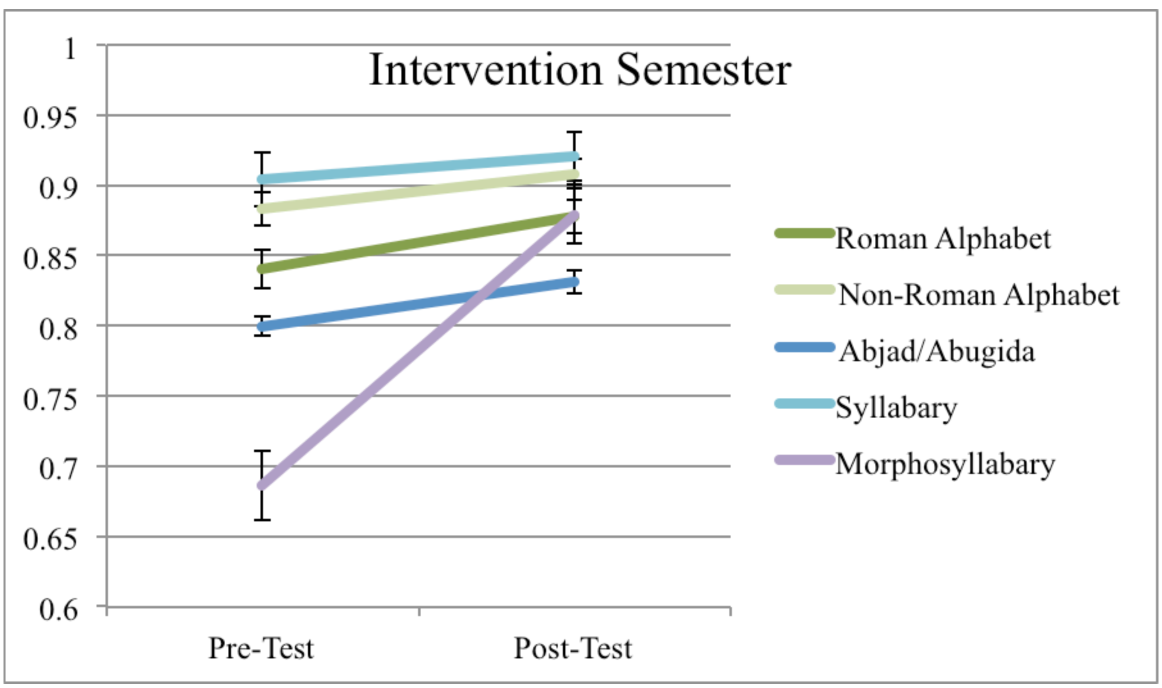

Figure 37. Proportion correct on the spelling verification task at pre-test and post-test for each language group receiving the intervention instruction.

Table 48. Descriptive statistics for survey responses on the pre-test.

		Control semester 1					Intervention semester					Control semester 2				
Question	Native speakers	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary
How much difficulty did you have learning to read your native language in elementary school? (1 = None, 5 = A large amount)	--[c]	1.67 (.58)	2.00 (1.10)	1.77 (1.07)	1.00 (.00)	1.20 (.42)	1.67 (.82)	1.00 (.00)	1.69 (.97)	2.50 (.71)	1.00 (.00)	1.57 (1.13)	1.33 (.58)	1.52 (.76)	2.33 (1.53)	1.67 (1.21)
How much difficulty do you have learning to read English? *(1 = None, 5 = A large amount)*	1.59 (.51)	3.00 (.00)	3.17 (.41)	3.15 (.77)	3.50 (1.29)	3.20 (.79)	2.83 (.75)	3.20 (.45)	2.92 (.56)	4.00 (1.41)	3.50 (.71)	2.86 (1.07)	3.67 (1.53)	3.18 (.68)	4.00 (1.00)	3.17 (.75)
Do you find it hard to pronounce new words in English? *(1 = No, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = All the time)*	1.88 (.70)	3.33 (.58)	3.17 (.98)	3.00 (.88)	4.25 (.96)	3.30 (1.06)	3.17 (.41)	3.20 (1.48)	3.31 (.55)	3.00 (1.41)	3.00 (.00)	3.43 (.98)	3.67 (1.15)	3.15 (.76)	4.67 (.58)	3.50 (.55)

		Control semester 1					Intervention semester					Control semester 2				
Question	Native speakers	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary
How would you compare your current reading speed in your native language to that of other people of the same age and education? (1 = Below average, 5 = Above average)	--[c]	4.00 (1.00)	3.50 (1.05)	4.11 (.97)	4.50 (.58)	3.70 (1.16)	4.00 (1.55)	3.20 (1.10)	3.96 (.92)	3.00 (1.51)	4.50 (.71)	3.43 (1.13)	4.67 (.58)	3.63 (1.18)	4.00 (.00)	4.17 (1.33)
How would you compare your current reading speed in English to that of other people of the same age and education? *(1 = Below average, 5 = Above average)*	3.82 (.73)	2.33 (.58)	2.33 (.52)	3.04 (1.04)	3.00 (.82)	2.80 (.63)	3.33 (.82)	2.60 (.55)	3.04 (.87)	3.00 (.00)	2.50 (.71)	2.86 (.38)	2.67 (.58)	2.84 (.77)	3.00 (1.00)	2.67 (1.21)

		Control semester 1					Intervention semester					Control semester 2				
Question	Native speakers	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary
How much difficulty did you have learning to spell your native language in elementary school? (1 = None, 5 = A large amount)	--[c]	2.00 (1.00)	2.00 (1.26)	1.89 (1.22)	1.25 (.50)	1.50 (.97)	1.50 (.55)	1.40 (.55)	1.62 (.80)	2.00 (.00)	2.00 (1.41)	1.57 (.79)	1.33 (.58)	1.67 (.85)	3.00 (2.00)	1.50 (.84)
How much difficulty do you have learning to spell English? *(1 = None, 5 = A large amount)*	1.82 (.88)	2.33 (.58)	3.00 (.63)	3.22 (.97)	3.75 (.96)	3.10 (.99)	3.17 (.98)	2.80 (1.30)	2.81 (.90)	3.00 (.00)	3.50 (.71)	3.43 (.96)	3.00 (1.00)	3.18 (1.04)	3.67 (1.15)	3.17 (.75)
How good are you at spelling in your native language now? (1 = Very bad, 5 = Excellent)	--[c]	4.67 (.58)	4.50 (.55)	4.70 (.54)	3.50 (.58)	4.50 (.53)	4.50 (.55)	4.60 (.55)	4.77 (.59)	3.50 (.71)	4.00 (.00)	4.29 (1.11)	5.00 (.00)	4.70 (.81)	3.33 (.58)	4.50 (.55)
How good are you at spelling English now? *(1 = Very bad, 5 = Excellent)*	4.06 (.83)	3.33 (.58)	2.50 (1.22)	3.11 (.89)	2.50 (.58)	2.80 (.79)	3.00 (1.10)	3.20 (1.30)	3.08 (.89)	3.00 (.00)	3.00 (.00)	3.14 (.69)	3.67 (.58)	2.94 (.70)	2.67 (.58)	2.83 (.41)

		Control semester 1					Intervention semester					Control semester 2				
Question	Native speakers	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary
How often do you use spell-check when you type in your native language? (1 = Never, 5 = For all words in everything I write)	--[c]	2.00 (.00)	2.33 (.103)	1.63 (1.11)	1.75 (.50)	2.40 (1.65)	2.0 (.63)	2.40 (1.52)	1.80 (1.19)	3.00 (1.41)	1.50 (.71)	2.50 (1.64)	1.67 (1.15)	1.85 (1.06)	3.67 (.58)	1.50 (.55)
How often do you use spell-check when you type in English? *(1 = Never, 5 – For all words in everything I write)*	3.12 (1.22)	3.33 (.58)	3.33 (.82)	3.48 (.75)	4.00 (.82)	3.00 (.82)	3.50 (.55)	3.20 (1.64)	3.15 (.83)	4.00 (.00)	4.00 (.00)	3.43 (.98)	3.00 (1.00)	3.42 (.83)	3.67 (.58)	3.33 (.52)
How would you compare your current spelling ability in your native language to that of other people of the same age and education? (1 = Below average, 5 = Above average)	--[c]	4.33 (.58)	3.83 (.75)	4.42 (.90)	3.50 (1.00)	4.60 (.70)	3.67 (1.03)	4.20 (.84)	4.50 (.71)	3.50 (.71)	4.50 (.71)	3.43 (1.13)	4.67 (.58)	4.16 (.95)	3.67 (1.15)	4.33 (.52)

Question	Native speakers	Control semester 1: Roman alphabet	Control semester 1: Non-Roman alphabet	Control semester 1: Abjad/abugida	Control semester 1: Syllabary	Control semester 1: Morphosyllabary	Intervention semester: Roman alphabet	Intervention semester: Non-Roman alphabet	Intervention semester: Abjad/abugida	Intervention semester: Syllabary	Intervention semester: Morphosyllabary	Control semester 2: Roman alphabet	Control semester 2: Non-Roman alphabet	Control semester 2: Abjad/abugida	Control semester 2: Syllabary	Control semester 2: Morphosyllabary
How would you compare your current spelling ability in English to that of other people of the same age and education? *(1 = Below average, 5 = Above average)*	3.82 (.81)	3.00 (.00)	2.83 (.41)	3.08 (1.02)	2.75 (.50)	2.60 (.70)	2.83 (.98)	3.20 (1.30)	3.19 (.94)	3.00 (1.41)	3.00 (.00)	2.57 (.53)	3.33 (1.15)	2.97 (.73)	2.67 (.58)	3.20 (.45)
How much do you like reading in your native language? (1 = None, 5 = A large amount)	--[c]	4.00 (.00)	4.50 (.55)	3.44 (1.09)	4.25 (.96)	4.50 (.97)	3.17 (1.17)	3.60 (1.52)	3.69 (1.16)	3.50 (.71)	3.50 (.71)	3.71 (1.60)	4.33 (.58)	3.48 (1.25)	3.67 (.58)	4.33 (.82)
How much do you like reading in English? *(1 = None, 5 = A large amount)*	3.53 (1.07)	2.67 (.58)	2.50 (.55)	2.81 (1.04)	3.75 (1.26)	2.80 (.79)	3.00 (1.10)	3.40 (.55)	2.81 (.80)	2.00 (.00)	3.00 (1.41)	3.14 (1.07)	2.67 (.58)	2.82 (.77)	3.00 (1.00)	2.83 (.75)
How much reading do you do for fun in your native language? (1 = None, 5 = A large amount)	--[c]	3.67 (.58)	4.33 (.82)	3.41 (1.15)	4.00 (.82)	4.30 (1.06)	2.83 (1.17)	4.00 (1.41)	3.88 (.77)	3.00 (1.41)	4.50 (.71)	3.00 (1.63)	4.67 (.58)	3.09 (1.35)	4.33 (1.15)	3.67 (1.21)

Question		Control semester 1					Intervention semester					Control semester 2				
		Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary
How much reading do you do for fun in English? *(1 = None, 5 = A large amount)*	2.53 (1.12)	2.33 (.58)	2.00 (.00)	2.30 (.82)	3.50 (.58)	2.70 (.67)	2.33 (.82)	2.80 (.84)	2.62 (.90)	2.00 (1.41)	3.50 (.71)	2.29 (.76)	2.33 (.58)	2.42 (.90)	3.00 (1.00)	2.17 (.75)
How many books do you read for fun each year in your native language? (1 = None, 2 = 1-2, 3 = 3-5, 4 = 6-10, 5 = More than 10)	--[c]	3.33 (1.15)	4.67 (.52)	2.67 (1.24)	4.00 (1.15)	3.80 (.92)	2.33 (1.03)	4.00 (1.00)	3.12 (1.24)	3.00 (1.41)	3.00 (1.41)	2.29 (.95)	3.33 (1.53)	2.48 (1.78)	3.67 (.58)	3.67 (1.51)
How many books do you read for fun each year in English? *(1 = None, 2 = 1-2, 3 = 3-5, 4 = 6-10, 5 = More than 10)*	2.76 (1.30)	1.33 (.58)	1.83 (.75)	1.74 (.90)	2.25 (1.50)	2.00 (.67)	2.33 (.151)	2.40 (1.14)	1.96 (.72)	1.00 (.00)	1.50 (.71)	1.86 (.90)	2.00 (.00)	2.00 (.97)	1.67 (.58)	1.83 (1.17)

		Control semester 1					Intervention semester					Control semester 2				
Question		Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary
How many newspapers and/or magazines do you read for fun each month in your native language? (1 = None, 2 = 1-2, not regularly, 3 = 1-2, regularly, 4 = 3-4, regularly, 5 = 5 or more)	--[c]	3.67 (1.15)	3.67 (1.37)	3.85 (1.10)	4.00 (.82)	2.90 (1.20)	2.83 (1.17)	3.80 (1.30)	3.38 (1.10)	4.00 (1.41)	4.00 (1.41)	3.43 (1.62)	3.33 (1.53)	2.48 (1.78)	3.00 (1.73)	4.00 (1.26)
How many newspapers and/or magazines do you read for fun each month in English? *(1 = None, 2 = 1-2, not regularly, 3 = 1-2, regularly, 4 = 3-4, regularly, 5 = 5 or more)*	2.53 (.72)	1.67 (.58)	1.50 (.84)	2.04 (.81)	2.25 (.96)	1.80 (.63)	2.50 (1.64)	2.00 (.71)	2.08 (.80)	2.00 (.00)	2.00 (.00)	2.14 (.69)	2.67 (.58)	1.76 (.75)	1.67 (.58)	2.00 (1.55)

		Control semester 1					Intervention semester					Control semester 2				
Question		Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary
Do you read articles on the internet in your native language? (1 = Never, 2 = Rarely, 3 = Once in a while, 4 = Once a week, 5 = Every day)	--[c]	5.00 (.00)	4.50 (1.22)	3.78 (1.31)	4.50 (1.00)	4.20 (1.14)	3.67 (1.03)	4.20 (1.10)	3.81 (1.17)	5.00 (.00)	4.50 (.71)	3.29 (1.38)	5.00 (.00)	3.58 (1.28)	4.67 (.58)	4.67 (.82)
Do you read articles on the internet in English? *(1 = Never, 2 = Rarely, 3 = Once in a while, 4 = Once a week, 5 = Every day)*	4.12 (.93)	3.67 (1.15)	2.00 (1.10)	3.07 (1.30)	3.00 (1.41)	3.10 (1.10)	3.17 (.75)	3.20 (1.30)	3.12 (1.03)	3.50 (.71)	3.00 (1.41)	2.71 (1.11)	2.67 (1.15)	2.82 (1.31)	3.67 (1.53)	2.67 (1.51)
It is very important to me to be a good reader in my native language. (1 = Disagree strongly, 5 = Agree strongly)	--[c]	5.00 (.00)	4.67 (.52)	4.67 (.73)	3.75 (.96)	4.50 (.85)	4.83 (.41)	4.20 (.84)	4.81 (.57)	5.00 (.00)	4.00 (1.41)	4.29 (1.50)	3.67 (.58)	4.76 (.61)	3.67 (1.53)	4.20 (.84)

Question		Control semester 1					Intervention semester					Control semester 2				
		Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary
It is very important to me to be a good speller in my native language. (1 = Disagree strongly, 5 = Agree strongly)	--[c]	5.00 (.00)	4.17 (.98)	4.74 (.59)	3.50 (.58)	4.30 (.67)	4.83 (.41)	3.40 (1.14)	4.77 (.59)	5.00 (.00)	3.50 (2.12)	4.29 (1.25)	4.33 (.58)	4.58 (.97)	4.67 (.58)	4.40 (.89)
I am a good reader in my native language. (1 = Disagree strongly, 5 = Agree strongly)	--[c]	4.67 (.58)	3.83 (.75)	4.59 (.64)	3.75 (.96)	4.20 (1.32)	4.17 (.75)	4.00 (1.00)	4.65 (.63)	3.00 (1.41)	4.50 (.71)	4.00 (1.41)	3.33 (.58)	4.39 (.86)	2.67 (1.15)	4.40 (.89)
It is very important to me to be a good reader in English. *(1 = Disagree strongly, 5 = Agree strongly)*	4.24 (.66)	5.00 (.00)	4.67 (.52)	4.59 (.84)	4.00 (.82)	4.30 (1.06)	4.83 (.41)	3.60 (1.52)	4.58 (.86)	5.00 (.00)	4.00 (.00)	4.71 (.76)	3.33 (1.15)	4.67 (.74)	4.00 (1.73)	4.40 (.89)
It is very important to me to be a good speller in English. *(1 = Disagree strongly, 5 = Agree strongly)*	3.94 (.97)	5.00 (.00)	4.50 (.84)	4.37 (.97)	3.75 (.50)	4.70 (.48)	4.83 (.41)	4.00 (1.00)	4.73 (.67)	5.00 (.00)	4.50 (.71)	4.71 (.76)	4.33 (.58)	4.73 (.57)	4.00 (1.73)	4.60 (.55)

		Control semester 1					Intervention semester					Control semester 2				
Question		Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary
I am a good reader in English. *(1 = Disagree strongly, 5 = Agree strongly)*	4.12 (.78)	2.67 (.58)	2.0 (.63)	2.93 (.83)	3.00 (.82)	2.80 (.79)	3.00 (.63)	2.80 (.84)	3.15 (.83)	2.00 (.00)	2.50 (.71)	2.86 (1.07)	2.00 (.00)	3.12 (.89)	2.00 (1.00)	2.60 (1.14)
If the teacher talks about something interesting, I might read more about it. (1 = Disagree strongly, 5 = Agree strongly)	3.65 (.93)	3.00 (1.00)	3.83 (.98)	3.67 (1.11)	3.75 (.96)	4.30 (.67)	4.17 (.41)	4.00 (1.00)	3.77 (.91)	3.00 (1.41)	4.00 (.00)	4.29 (.95)	4.00 (1.00)	3.67 (1.14)	4.33 (.58)	4.40 (.89)
I read to learn more about new topics that are interesting to me. (1 = Disagree strongly, 5 = Agree strongly)	4.06 (.90)	3.33 (.68)	4.0 (.63)	3.70 (1.10)	4.00 (1.15)	4.00 (.67)	4.33 (.52)	3.40 (1.14)	4.04 (.96)	4.00 (.00)	5.00 (.00)	4.00 (1.15)	4.00 (1.00)	3.85 (1.23)	4.67 (.58)	4.60 (.55)

Note. [a]Sufficient data were not available from these students for this question during this semester to calculate a standard deviation.
[b]Data were not available from these students for this question during this semester.

The results from the questions asking about L1 reading and spelling behaviors were examined first. There was no effect of L1 type on students' reported difficult learning to read in their L1 in elementary school, or on their reported difficulty learning to spell in their L1 in elementary school. There was no effect of L1 type on students' reported reading speed in their L1 compared to other people of the same age and education. However, there was a significant effect of L1 writing system type on students' reported current spelling ability in their L1, $\chi^2(df = 4) = 31.34$, $p < .001$. Follow-up pairwise comparisons showed that the L1 syllabary students reported a significantly lower current L1 spelling ability compared to all other L1 groups: L1 Roman alphabet students, $\beta = -.99$, $z = -3.77$, $p < .01$; L1 non-Roman alphabet students, $\beta = -1.20$, $z = -4.44$, $p < .001$; L1 abjad/abugida students, $\beta = -1.28$, $z = -5.77$, $p < .001$; L1 morphosyllabary students, $\beta = -1.00$, $z = -3.88$, $p < .01$. There was also a significant effect of L1 writing system type on students' reported L1 spelling ability compared to other people of the same age and education, $\chi^2(df = 4) = 15.29$, $p < .01$. Pairwise comparisons showed that the L1 syllabary students reported a significantly lower comparative L1 spelling ability than the morphosyllabary students, $\beta = -.94$, $z = -2.76$, $p < .05$, and also a marginally lower comparative L1 spelling ability than the abjad/abugida L1 students, $\beta = -.79$, $z = -2.69$, $p = .05$. The L1 Roman alphabet speakers also reported a significantly lower comparative L1 spelling ability than the L1 morphosyllabary speakers, $\beta = -.81$, $z = -2.82$, $p < .05$, and the L1 abjad/abugida L1 students, $\beta = -.66$, $z = -2.88$, $p < .05$. Despite this, there was no difference in among students' self-reported frequency using spell-check when typing in their L1.

There was a significant effect of L1 writing system type on students' self-reported enjoyment of reading in their L1, $\chi^2(df = 4) = 10.33$, $p < .05$. Pairwise comparisons revealed that the only significant difference was between L1 morphosyllabary students and L1 abjad/abugida students: the L1 morphosyllabary students reported a significantly higher level of enjoyment of reading in their L1 than the L1 abjad/abugida students, $\beta = .80$, $z = 2.76$, $p < .05$. There was a significant effect of L1 writing system type on the quantity of reading that students reported doing for fun in their L1, $\chi^2(df = 4) = 14.16$, $p < .01$. Pairwise comparisons revealed that the L1 Roman alphabet students reported a significantly lower quantity of reading for fun in their L1 compared to the L1 non-Roman alphabet students, $\beta = -1.25$, $z = -2.90$, $p < .05$, and also a marginally lower quantity of reading for fun compared to the L1 morphosyllabary students, $\beta = -1.05$, $z = -2.65$, $p = .06$. The L1 non-Roman alphabet students reported a marginally higher quantity of reading for fun in their L1 compared to the abjad/abugida L1students, $\beta = .88$, $p = .07$.

There was a significant effect of L1 writing system type on the number of books students reported reading per year in their L1, $\chi^2(df = 4) = 26.71$, $p < .001$. Pairwise comparisons revealed that the L1 Roman alphabet students reported reading significantly fewer books per year in their L1 than the L1 non-Roman alphabet students, $\beta = -1.64$, $z = -3.76$, $p < .01$, and also than the L1 morphosyllabary students, $\beta = -1.17$, $z = -2.85$, $p < .05$. Both the L1 non-Roman alphabet students and the L1 morphosyllabary students reported reading significantly more books per year in their L1 than the L1 abjad/abugida students, $\beta = 1.41$, $z = 4.10$, $p < .001$, and $\beta = .94$, $z = 3.02$, $p < .05$. There was also a significant effect of L1 writing system type on how often students reported reading articles online in their L1, $\chi^2(df = 4) = 13.02$, $p = .01$. However, follow-up pairwise comparisons did not reveal any significant differences among L1 groups. There was no effect of L1 writing system type on the number of newspapers and magazines that students reported reading per month in their L1.

Students were also asked whether it was important to them to be a good reader and to be a good speller in their L1. There was a significant effect of L1 writing system type on how important students reported it was for them to be a good reader in their L1, $\chi^2(df = 4) = 12.82$, $p = .01$. Pairwise comparisons revealed that the only significant difference was between the L1 syllabary and the L1 abjad/abugida students: the L1 syllabary students reported that it was significantly less important for them to be a good reader in their L1 than the L1 abjad/abugida students, $\beta = -.74$, $z = -2.80$, $p < .05$. There was also a significant effect of L1 writing system type on how important students reported it was for them to be a good speller in their L1, $\chi^2(df = 4) = 14.20$, $p < .01$. Pairwise comparisons again revealed only one significant difference: the L1 non-Roman alphabet students reported that it was significantly less important for them to be a good speller in their L1 than the L1 abjad/abugida students, $\beta = -.76$, $z = -3.22$, $p = .01$. Students were also asked how strongly they agreed with the statement "I am a good reader in my native language". There was a significant effect of L1 writing system type on this variable, $\chi^2(df = 4) = 24.98$, $p < .001$. The L1 syllabary students disagreed with this statement significantly more strongly than the L1 Roman alphabet students, $\beta = -.97$, $z = -2.72$, $p < .05$, the L1 abjad/abugida students, $\beta = -1.31$, $z = -4.40$, $p < .001$, and the L1 morphosyllabary students, $\beta = - 1.07$, $z = -3.06$, $p < .05$. The L1 non-Roman alphabet students also disagreed with this statement significantly more strongly than the L1 abjad/abugida students, $\beta = -.75$, $z = -3.05$, $p < .05$.

Finally, students were asked whether they were likely to read more about something interesting if their teacher talked about it, and whether they read to learn about new topics that were interesting to them. There was no effect of L1 writing system type for either of these variables.

The results from the questions asking about English reading and spelling behaviors were examined next. These questions were asked both of the five groups of non-native ESL students and also the comparison group of native English speakers. There was a significant effect of L1 writing system type on the level of difficulty that students reported with learning to read English, $\chi^2(df = 5) = 31.66$, $p < .001$. Pairwise comparisons revealed that all five ESL student groups reported a significantly higher level of difficulty learning to read English than did the L1 English speakers: L1 Roman alphabet students, $\beta = 1.30$, $z = 4.82$, $p < .001$; L1 non-Roman alphabet students, $\beta = 1.70$, $z = 6.09$, $p < .001$; L1 abjad/abugida students, $\beta = 1.51$, $z = 6.94$, $p < .001$; L1 syllabary students, $\beta = 2.21$, $z = 7.07$, $p < .001$; and L1 morphosyllabary students, $\beta = 1.62$, $z = 6.12$, $p < .001$. In addition, the L1 syllabary students reported a significantly higher level of difficulty than the L1 Roman alphabet students, $\beta = .91$, $z = 3.05$, $p < .05$, and a marginally higher level of difficulty than the L1 abjad/abugida students, $\beta = .70$, $z = 2.82$, $p = .05$. There was a significant effect of L1 writing system type on the level of difficulty that students reported with learning to spell English, $\chi^2(df = 5) = 19.43$, $p < .01$. Again all five ESL student groups reported a significantly higher level of difficulty learning to spell English than did the L1 English speakers: L1 Roman alphabet students, $\beta = 1.30$, $z = 3.94$, $p < .01$; L1 non-Roman alphabet students, $\beta = 1.11$, $z = 3.23$, $p < .05$; L1 abjad/abugida students, $\beta = 1.26$, $z = 5.00$, $p < .001$; L1 syllabary students, $\beta = 1.73$, $z = 4.43$, $p < .001$; and L1 morphosyllabary students, $\beta = 1.34$, $z = 4.19$, $p < .001$. There were no significant differences among any of the ESL student groups. There was also a significant effect of L1 writing system type on students' self-reported difficulty with pronouncing new words in English, $\chi^2(df = 5) = 29.33$, $p < .001$. Pairwise comparisons revealed that all five ESL student groups reported a significantly higher level of difficulty with pronouncing new words in English: L1 Roman alphabet students, $\beta = 1.44$, $z = 4.59$, $p < .001$;

L1 non-Roman alphabet students, $\beta = 1.41$, $z = 4.37$, $p < .001$; L1 abjad/abugida students, $\beta = 1.26$, $z = -4.93$, $p < .001$; L1 syllabary students, $\beta = 2.25$, $z = 6.35$, $p < .001$; and L1 morphosyllabary students, $\beta = 1.43$, $z = 4.66$, $p < .001$. There were no significant differences among any of the ESL student groups.

Students were also asked to self-rate their current reading speed in English compared to other people of the same age and education. There was again a significant effect of L1, $\chi^2(df = 5) = 18.81$, $p < .01$. For this variable, four of the five ESL student groups reported that their current comparative English reading speed was significantly lower than that reported by the L1 English speakers: L1 Roman alphabet students, $\beta = -.89$, $z = 3.15$, $p < .52$; L1 non-Roman alphabet students, $\beta = -1.32$, $z = -4.54$, $p < .001$; L1 abjad/abugida students, $\beta = -.86$, $z = 4.00$, $p < .001$; and L1 morphosyllabary students, $\beta = -1.10$, $z = -4.03$, $p < .001$. However, the L1 syllabary students did not differ from the L1 English students.

Regarding spelling abilities, there was a significant effect of L1 writing system type on students' reported current spelling ability in English, $\chi^2(df = 5) = 19.92$, $p < .01$. All five ESL student groups reported a significantly lower current English spelling ability: L1 Roman alphabet students, $\beta = -.93$, $z = -3.26$, $p = .01$; L1 non-Roman alphabet students, $\beta = -1.06$, $z = -3.57$, $p < .001$; L1 abjad/abugida L1 students, $\beta = -1.02$, $z = -4.69$, $p < .001$; L1 syllabary students, $\beta = -1.39$, $z = -4.11$, $p < .001$; and L1 morphosyllabary students, $\beta = -1.23$, $z = -4.41$, $p < .001$. There were no significant differences among any of the ESL student groups. There was also a significant effect of L1 on students' reported English spelling ability compared to other people of the same age and education, $\chi^2(df = 5) = 15.05$, $p = .01$. Pairwise comparisons showed that four of the five ESL student groups reported a significantly lower comparative English spelling ability than did the L1 English speakers: L1 Roman alphabet students, $\beta = -1.07$, $z = -$

3.73, $p < .01$; L1 abjad/abugida students, $\beta = -.75$, $z = -3.43$, $p < .01$; L1 syllabary students, $\beta = -1.05$, $z = -3.07$, $p < .05$; and L1 morphosyllabary students, $\beta = -1.00$, $z = -3.53$, $p < .01$. However, there was no difference between the L1 non-Roman alphabet students and the L1 English speakers. Despite these differences in reported spelling ability, there was no significant effect of L1 writing system type on the frequency that participants reported using spell-check when the typed in English.

There was no significant effect of L1 writing system type on how much participants reported liking to read in English. There was also no significant effect of L1 on how much participants reported reading for fun in English per year. This included no L1 effect on the number of books per year that participants reported reading for fun in English and no L1 effect on the number of newspapers and magazines per month that participants reported reading for fun in English. There was a marginal effect of L1 type on how often participants reported reading articles online in English, $\chi^2(df = 5) = 9.98$, $p = .08$. Pairwise comparisons revealed that the only difference was between L1 non-Roman alphabet students and L1 English speakers; the L1 non-Roman alphabet students reported reading articles online in English significantly less than did the L1 English speakers, $\beta = -1.60$, $z = -3.00$, $p < .05$.

Finally, there was a significant effect of L1 writing system type on how important participants said it was to them to be a good reader in English, $\chi^2(df = 5) = 11.75$, $p < .05$. However, pairwise comparisons revealed only one marginally significant difference: the L1 Roman alphabet students reported that it was marginally more important to them to be a good reader in English than did the L1 non-Roman alphabet students, $\beta = .81$, $z = 2.65$, $p = .08$. There was also a significant effect of L1 on how important participants said it was to them to be a good speller in English, $\chi^2(df = 5) = 15.96$, $p < .01$. The pairwise comparisons showed that the L1

Roman alphabet students and the L1 abjad/abugida students each reported that it was significantly more important to them to be a good speller in English compared to the L1 English speakers, $\beta = .87$, $z = 3.27$, $p = .01$ for the L1 Roman alphabet students and $\beta = .68$, $z = -3.33$, $p = .01$ for the L1 abjad/abugida students. In addition, the L1 morphosyllabary students said that it was marginally more important to them to be a good speller in English compared to the L1 English speakers, $\beta = .71$, $z = 2.69$, $p = .07$. For the last question, participants were asked to say how strongly they agreed with the statement, "I am a good reader in English". There was a significant effect of L1 writing system type on how strongly participants agreed with this statement, $\chi^2(df = 5) = 28.75$, $p < .001$. All five ESL student groups reported agreeing with this statement significantly less strongly than did the L1 English speakers: L1 Roman alphabet students, $\beta = -1.24$, $z = -4.14$, $p < .001$; L1 non-Roman alphabet students, $\beta = -1.83$, $z = -5.94$, $p < .001$; L1 abjad/abugida students, $\beta = -1.05$, $z = -4.51$, $p < .001$; L1 syllabary students, $\beta = -1.67$, $z = -4.78$, $p < .001$; and L1 morphosyllabary students, $\beta = -1.41$, $z = -4.78$, $p < .001$. In addition, the non-Roman alphabet students reported agreeing with this statement significantly less strongly than the L1 abjad/abugida students, $\beta = -.78$, $z = -3.27$, $p = .01$.

7.2.12 Post-test – Reading Survey

Participants were asked a subset of the same questions, focusing mainly on English, on the reading survey post-test (ESL students only). They were also asked four new questions, in which they were asked to evaluate how much they felt they had learned about various aspects of English literacy skills during that semester. For each question, pre-test responses were entered as the first predictor of their post-test responses, and the regression included a random effect for class to account for the nested structure of the data. Following this, semester was entered as the second predictor, to determine whether there were any differences among groups that received different types of instruction after accounting for existing differences at pre-test. L1 writing system type was also tested as a predictor following semester, again to determine whether there were any remaining differences among L1 groups after account for any existing differences at pre-test. Means and standard deviations for each question, organized by semester and L1 writing system type, are given in Table 49.

Pre-test responses were the only significant predictor of post-test responses for most questions that were asked at both pre-test and post-test: the amount of difficulty that students reported learning to read English, $\beta = .44$, $t = 4.55$, $p < .001$; the amount of difficulty that students reported learning to spell English, $\beta = .46$, $t = 5.69$, $p < .001$; how hard students found it to pronounce new words in English, $\beta = .66$, $t = 10.56$, $p < .001$; how students compared their current reading speed in English to that of other people of the same age and education, $\beta = .34$, $t = 4.12$, $p < .001$; how students compared their current spelling ability in English to that of other people of the same age and education, $\beta = .63$, $t = 9.02$, $p < .001$; students' evaluation of the current spelling abilities in English, $\beta = .61$, $t = 8.95$, $p < .001$; the amount of reading that students did for fun in English, $\beta = .52$, $t = 5.46$, $p < .001$; the number of books per year that students read for fun in English, $\beta = .51$, $t = 6.45$, $p < .001$; how often students reported reading articles online in English, $\beta = .41$, $t = 5.81$, $p < .001$; and how strongly students agreed with the statement "I am a good reader in English", $\beta = .50$, $t = 6.35$, $p < .001$.

Table 49. Descriptive statistics for survey responses on the post-test.

	Control semester 1					Intervention semester					Control semester 2				
Question	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary
How much difficulty do you have learning to read English? *(1 = None, 5 = A large amount)*	2.67 (1.15)	3.33 (.52)	3.04 (.84)	2.33 (.58)	2.60 (.52)	2.20 (.84)	2.67 (1.37)	2.88 (.88)	4.00 (.00)	3.00 (1.41)	2.17 (.75)	3.20 (.45)	2.77 (.86)	3.00 --[a]	2.40 (.55)
Do you find it hard to pronounce new words in English? *(1 = No, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = All the time)*	3.33 (.58)	3.17 (.75)	3.04 (.89)	3.33 (.58)	3.40 (.70)	3.00 (.00)	3.17 (1.17)	3.00 (.58)	3.50 (.71)	2.50 (.71)	3.17 (.98)	3.40 (.89)	2.92 (.69)	5.00 -- [a]	3.60 (.55)
How would you compare your current reading speed in English to that of other people of the same age and education? *(1 = Below average, 5 = Above average)*	3.00 (1.73)	3.50 (1.05)	3.16 (.90)	3.00 (1.00)	3.60 (.84)	3.40 (1.14)	-- [b] -- [a]	3.24 (.78)	3.50 (.71)	3.50 (.71)	3.00 (.63)	2.80 (.84)	2.96 (.82)	4.00 -- [a]	3.20 (.84)
How much difficulty do you have learning to spell English? *(1 = None, 5 = A large amount)*	2.33 (1.15	2.70 (.82)	2.96 (.98)	3.33 (1.15)	2.90 (1.20)	2.80 (.45)	2.00 (1.10)	2.88 (.73)	3.00 (.00)	2.00 (.00)	2.83 (.98)	1.60 (.55)	2.96 (.96)	2.00 -- [a]	3.00 (.00)
How good are you at spelling English now? *(1 = Very bad, 5 = Excellent)*	3.33 (.58)	3.17 (.98)	3.16 (.90)	2.67 (1.15)	3.00 (.67)	3.60 (.55)	3.50 (1.05)	3.32 (.80)	4.00 (.00)	3.00 (.00)	3.17 (.98)	3.40 (.89)	3.19 (.80)	4.00 -- [a]	3.00 (.00)

	Control semester 1					Intervention semester					Control semester 2				
Question	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary
How often do you use spell-check when you type in English? *(1 = Never, 5 – For all words in everything I write)*	2.33 (.58)	2.83 (1.17)	3.08 (.81)	3.33 (1.15)	3.10 (.99)	2.80 (.84)	2.17 (.75)	2.76 (.93)	2.50 (.71)	3.00 (.00)	3.17 (.75)	3.20 (.84)	3.12 (.86)	3.00 -- [a]	3.40 (.89)
How would you compare your current spelling ability in English to that of other people of the same age and education? *(1 = Below average, 5 = Above average)*	4.00 (1.00)	3.17 (.75)	3.20 (1.04)	2.67 (1.15)	3.20 (.92)	3.00 (.71)	3.50 (.84)	3.36 (.76)	3.00 (1.41)	3.00 (.00)	3.33 (.52)	3.40 (.55)	3.15 (.92)	3.00 -- [a]	3.20 .45
How much do you like reading in English? *(1 = None, 5 = A large amount)*	3.00 (.00)	2.82 (.75)	2.88 (.97)	4.33 (.58)	3.80 (.79)	3.40 (.55)	3.67 (.52)	3.36 (.81)	3.00 (1.41)	4.00 (1.41)	3.67 (1.03)	3.20 (.71)	3.27 (.83)	4.00 -- [a]	3.20 (.84)
How much reading do you do for fun in English? *(1 = None, 5 = A large amount)*	3.00 (.00)	2.67 (1.21)	2.52 (.92)	3.67 (1.53)	3.20 (.92)	2.80 (.45)	3.50 (.55)	3.04 (.89)	3.00 (1.41)	3.50 (.71)	3.00 (1.41)	2.80 (.84)	2.77 (.99)	3.00 -- [a]	3.20 (.84)
How many books do you read for fun each year in English? *(1 = None, 2 = 1-2, 3 = 3-5, 4 = 6-10, 5 = More than 10)*	1.67 (1.15)	1.33 (.52)	1.96 (.73)	2.33 (1.53)	2.10 (.74)	2.20 (.45)	3.00 (.89)	2.36 (.95)	1.00 (.00)	2.00 (.00)	1.67 (.82)	1.60 (.55)	2.31 (1.16)	2.00 -- [a]	2.00 (.71)

	Control semester 1					Intervention semester					Control semester 2				
Question	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary
How many newspapers and/or magazines do you read for fun each month in English? *(1 = None, 2 = 1-2, not regularly, 3 = 1-2, regularly, 4 = 3-4, regularly, 5 = 5 or more)*	3.33 (1.53)	3.00 (1.10)	2.56 (1.12)	3.33 (1.15)	2.40 (.70)	2.80 (.84)	2.67 (1.21)	2.20 (1.08)	2.00 (.00)	2.50 (.71)	2.33 (1.03)	2.20 (.45)	2.08 (.80)	3.00 -- [a]	2.80 (.84)
Do you read articles on the internet in English? *(1 = Never, 2 = Rarely, 3 = Once in a while, 4 = Once a week, 5 = Every day)*	3.67 (1.15)	2.67 (1.03)	2.80 (1.08)	4.33 (1.15)	3.50 (.53)	3.20 (.84)	3.33 (1.21)	3.24 (1.01)	3.50 (.71)	4.50 (.71)	3.50 (1.22)	2.80 (1.30)	3.12 (1.03)	4.00 -- [a]	2.40 (.55)
It is very important to me to be a good reader in English. *(1 = Disagree strongly, 5 = Agree strongly)*	5.00 (.00)	4.50 (.55)	4.68 (.63)	5.00 (.00)	4.10 (.57)	4.60 (.55)	4.33 (.82)	4.60 (.76)	5.00 (.00)	4.50 (.71)	4.67 (.82)	3.80 (1.10)	4.85 (.46)	5.00 -- [a]	3.80 (1.30)
It is very important to me to be a good speller in English. *(1 = Disagree strongly, 5 = Agree strongly)*	-- [b]	-- [b]	-- [b]	-- [b]	-- [b]	4.60 (.55)	4.33 (.82)	4.44 (.87)	4.00 (1.41)	3.50 (2.12)	4.00 (1.26)	3.40 (.89)	4.73 (.72)	4.00 -- [a]	4.00 (1.22)
I am a good reader in English. *(1 = Disagree strongly, 5 = Agree strongly)*	3.33 (.58)	2.00 (.63)	3.04 (.89)	3.67 (1.53)	3.60 (.97)	3.60 (.55)	3.33 (.52)	3.12 (.97)	2.00 (.00)	3.00 (.00)	3.67 (.82)	2.60 (.55)	3.15 (.67)	3.00 -- [a]	3.00 (.00)

	Control semester 1					Intervention semester					Control semester 2				
Question	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary	Roman alphabet	Non-Roman alphabet	Abjad/abugida	Syllabary	Morphosyllabary
How much did you learn about English spelling this semester? *(1 = Nothing, 5 = A large amount)*	3.00 (1.00)	3.00 (.89)	3.44 (.87)	2.67 (.58)	3.60 (.70)	4.00 (.00)	2.83 (.75)	3.48 (.92)	3.00 (1.41)	2.50 (.71)	3.33 (.82)	3.00 (.71)	3.58 (.70)	3.00 -- [a]	3.40 (.55)
How much did your spelling improve this semester? *(1 = None, 5 = A large amount)*	3.00 (1.00)	3.00 (.89)	3.44 (.87)	2.67 (.58)	3.60 (.70)	4.00 (.00)	2.83 (.75)	3.48 (.92)	3.00 (1.41)	2.50 (.71)	3.33 (.82)	3.00 (.71)	3.58 (.70)	3.00 -- [a]	3.40 (.55)
How much did your ability to read in English improve this semester? *(1 = None, 5 = A large amount)*	3.33 (1.53)	3.83 (.75)	3.68 (.80)	3.67 (1.53)	3.80 (.79)	4.20 (.45)	4.00 (.63)	3.60 (.82)	3.50 (.71)	3.50 (.71)	4.00 (.63)	3.20 (.45)	3.65 (.75)	4.00 -- [a]	3.80 (.45)
How much did your ability to recognize words in English improve this semester? *(1 = None, 5 = A large amount)*	3.67 (1.53)	3.50 (.55)	3.60 (.71)	3.67 (1.53)	3.50 (.71)	4.80 (.45)	4.00 (.63)	3.84 (.55)	3.00 (1.41)	3.50 (.71)	4.00 (.63)	3.20 (.45)	3.69 (.79)	4.00 -- [a]	3.40 (.55)

Note. [a]Sufficient data were not available from these students for this question during this semester to calculate a standard deviation. [b]Data were not available from these students for this question during this semester.

However, there were a small number of questions where there was also evidence for an effect of semester or L1 writing system type. When asked how often they used spell-check when typing in English, the students in control semester 2 indicated that they used spell-check significantly more than the students in the intervention semester, $\beta = .40$, $t = 2.15$, $p < .05$ (pre-test response was also a significant predictor, $\beta = .33$, $t = 5.00$, $p < .001$). There was a marginally significant omnibus effect of L1 writing system type on how much students said that they liked reading in English, $\chi^2(df = 4) = 8.13$, $p = .09$ (pre-test response was also a significant predictor, $\beta = .51$, $t = 6.56$, $p < .001$). However, pairwise comparisons revealed no significant differences among the ESL student groups. There was an omnibus main effect of semester on how many newspapers and magazines students reported reading for fun each month in English, $\chi^2(df = 2) = 9.45$, $p < .01$ (pre-test response was also a significant predictor, $\beta = .54$, $t = 5.49$, $p < .001$). Pairwise comparisons showed that the students in the intervention semester and the students in control semester 2 both read significantly fewer newspapers and magazines for fun each month in English than the students in control semester 1, $\beta = -.54$, $z = -2.70$, $p < .05$, and $\beta = -.56$, $z = -2.86$, $p = .01$, respectively. There was a significant omnibus main effect of L1 type on how important students said that it was for them to be a good reader in English, $\chi^2(df = 4) = 18.54$, $p = .001$ (pre-test response was also a significant predictor, $\beta = .25$, $t = 3.76$, $p < .001$). Pairwise comparisons among the ESL groups indicated that both the L1 abjad/abugida students and the L1 syllabary students said that being a good reader in English was more important to them than did the L1 morphosyllabary students, $\beta = .61$, $z = 3.58$, $p < .01$, and $\beta = .94$, $z = 2.93$, $p < .05$, respectively. Finally, there was a marginally significant omnibus main effect of L1 writing system type on how important students said that it was for them to be a good speller in English, $\chi^2(df = 4) = 8.15$, $p = .09$ (pre-test response was also a significant predictor, $\beta = .34$, $t =$

2.18, $p < .05$). However, pairwise comparisons revealed no significant differences among the ESL student groups.

The final analyses of the post-test survey data examined students' responses to the four questions that were only asked on the post-test version of the survey, and which asked them to indicate how much they felt they learned about English spelling during the semester, as well as to indicate how much they felt that their spelling, ability to read in English, and ability to recognize words in English improved. There was no significant main effect of semester on students' ratings of how much they had learned about English spelling that semester. However, there was a significant omnibus main effect of L1 writing system type, $\chi^2(df = 4) = 10.48$, $p < .05$. Pairwise comparisons revealed that the L1 non-Roman alphabet students felt they had learned marginally less than the L1 abjad/abugida students, $\beta = -.54$, $z = -2.54$, $p = .08$. There was no main effect of semester and no main effect of L1 writing system type on students' ratings of how much their spelling and their ability to read in English had improved during the semester. However, there was a marginal omnibus main effect of semester on students' ratings of much their ability to recognize words in English had improved, $\chi^2(df = 4) = 5.02$, $p = .08$. Pairwise comparisons showed that the students in the intervention semester reported that their ability to recognize words in English had improved marginally more than the students in control semester 1, $\beta = .35$, $z = 2.11$, $p = .09$. There was also a marginal omnibus main effect of L1 writing system type, $\chi^2(df = 4) = 8.67$, $p = .07$. The L1 Roman alphabet students reported a marginally larger improvement than the L1 non-Roman alphabet students, $\beta = .64$, $z = 2.44$, $p = .097$; than the L1 abjad/abugida students, $\beta = .51$, $z = 2.44$, $p = .098$; and than the L1 morphosyllabary students, $\beta = .68$, $z = 2.57$, $p = .07$.

7.3 STUDY 3 SUMMARY AND DISCUSSION

The goal of Study 3 was to extend the cross-linguistic investigation of sub-lexical literacy skills to classroom ESL learners. Data were collected in a group administration environment so that the efficacy of this type of data collection could be evaluated. In addition, data were collected at the beginning and again at the end of a semester of intensive English instruction, allowing for an exploration of whether, and how, literacy-related related skills change over time in adult (as opposed to younger) ESL learners. Finally, given the importance of good literacy skills for successful reading outcomes in L1 English speakers, we also developed and implemented an instructional intervention that used a combination of phonics-based direct instruction and metacognitive training. This intervention was implemented during one of three semesters of data collection, allowing for a comparison of student development with and without the additional instruction.

The data collected in Study 3 were extremely rich and provide a number of insights regarding real classroom learners' literacy skills, their development with and without specialized instruction, and also differences among L1 groups. Looking first at the results from the sound discrimination task, there were some unexpected patterns of L1 differences both at pre-test and at post-test. At pre-test, the L1 morphosyllabic students were less accurate at English sound discrimination than the L1 abjad/abugida and the L1 non-Roman alphabet students. Although somewhat speculative, one possible explanation for this finding relates to the pattern of results found for L1 Chinese speakers in Study 1 and Study 2 of relatively weak phonological skills across multiple tasks. At post-test, the L1 Roman alphabet students were either significantly or marginally less accurate than all the other non-native speaker groups. In fact, examination of the pattern of their development from pre-test to post-test shows that regardless of the type of instruction they received, the L1 Roman alphabet students had a decrease in their accuracy on sound discrimination from pre-test to post-test. This is a surprising result, and is not expected to occur as a result of *increased* exposure to English. Thus, the underlying reason for this pattern is not clear, and further work is needed to see if this result replicates in other samples. If this finding does replicate, then it suggests an important area of focus for targeted instruction by ESL teachers. It should be noted that although there was no evidence of greater improvement on English sound discrimination during the intervention semester than the two control semesters, this was not a skill targeted by the instructional intervention and it is thus not a surprising finding.

The results from the phonological awareness data were mixed. At pretest, we found that the L1 abjad/abugida students did significantly or marginally better than all other non-native English groups. This is somewhat surprising given the general finding of relatively weak phonological skills, particularly in L1 Arabic speakers, who formed a large portion of this group (e.g., Abbott, 2006; Alsulaimani, 1990; Fender, 2008; Ryan & Meara, 1991; M. Taylor, 2008). There may be a number of explanations for this finding. First, as was noted above in relation to the high performance on the oddity task by L1 Chinese speakers, some studies have shown that ESL students may be able to perform successfully on measures of sub-lexical literacy without being able to use those skills productively to support other literacy tasks. The unexpectedly high performance by the L1 abjad/abugida students in this sample may indicate a similar type of ability: these learners are able to develop some fundamental phonological awareness skills, but struggle to use them effectively within literacy more broadly defined. If this is indeed the case, it again highlights an area where ESL instructors may be able to effectively target their classroom activities to improve students' literacy skills.

Another key consideration is that relatively few studies have measured phonological awareness in L1 abjad/abugida speakers, and so our understanding of their skills is quite limited. Although previous research has suggested that speakers of these languages have relatively weak phonological skills, this work generally shows that their problem is most specifically for vowels rather than consonants (Alsulaimani, 1990; Fender, 2003, 2008; Hayes-Harb, 2006; K. I. Martin & Juffs, In revision; Ryan & Meara, 1991, 1996; Saigh & Schmitt, 2012). However, in the phonological awareness tasks used in this study primarily consonants were targeted for manipulation. This was most obvious in the Elision task, which only ever required the deletion of a consonant. The consonantal root-based structure of L1 Arabic is likely a contributing factor, and the relatively high accuracy of the L1 abjad/abugida speakers on the phonological awareness tasks may therefore simply be a typical reflection of their processing advantage for consonants.

The post-test data for this task showed an unexpected finding. There was a general trend toward improvement in phonological awareness scores across a semester for the students receiving the control instruction, which is a positive result as it suggests that phonological awareness is still able to improve in adult ESL learners (e.g., Durgunoğlu & Öney, 2002; Morais et al., 1986). However, the pattern changed drastically for the intervention semester, so that there was a general trend toward lower scores across a semester for students receiving the intervention instruction. This result may initially appear alarming, as it suggests a decrease in performance on a crucial underlying literacy skill as a result of the instructional intervention. However, there are a number of possible explanations that suggest a high level of concern is not warranted. First, it is important to recognize that the results show an impact of the intervention on participants' phonological awareness task performance. Thus, the intervention had an observable effect on the students, in contrast to other work with adult phonics-based instructional interventions, which have shown no effect of intervention even with many more hours of instruction and training (e.g., M. Taylor, 2008). Second, data from only two time points are available and it is thus not clear what the long-term developmental trajectory looks like for these students. It is possible that the decrease in phonological awareness performance in students receiving the intervention is a result of the students beginning to restructure their mental lexicon and think about English words and their component parts in a different way than before. If this is what is occurring, this process may lead to short-term decreases in performance, as observed in this study, but long-term gains in students' underlying abilities. We note that this explanation is speculative at this point, and much more longitudinal work is needed to rule out alternative explanations. Given the differing levels of pre-test performance on the phonological awareness task in each semester of the current study, a simple explanation such as regression to the mean is a possibility that needs to be ruled out with future work.

It is also important to recognize that the phonological awareness tasks themselves are not a literacy task that ever needs to be performed directly by students. Although the Elision and Blending tasks are measures of an important underlying skill, students will not need to delete single phonemes from inside an English word in the course of their normal literacy activities. However, spelling knowledge and spelling abilities are an important and directly observable literacy skill, and one that will actually be used and evaluated in students' lives. It is therefore important to note that although the instructional intervention apparently led to a decrease in phonological awareness performance per se in the ESL students, it also led to an increase in their spelling knowledge compared to students who only received the control instruction.

The improvement in spelling verification performance that results from the instructional intervention is one of the most promising findings from this study. However, it is also interesting to note the influences of word characteristics and L1 background on performance. Considering the pre-test data, three of the four word characteristics we tested (frequency, number of syllables, and grapheme-phoneme correspondence consistency) had a significant impact on spelling knowledge performance. These results are not surprising given the effect of these same variables for L1 English speakers, but they do confirm that L2 speakers of a language also show similar sensitivity to (at least these) word characteristics. There were also differences among some of the L1 groups that are consistent with previous research (Fender, 2008; Saigh & Schmitt, 2012). Specifically, the L1 non-Roman alphabet and the L1 Roman alphabet students were more accurate on the spelling test than the L1 abjad/abugida students at pre-test, and this difference persisted for the L1 Roman alphabet students at post-test, even after controlling for the differences that already existed at pre-test. The relatively poor performance by the L1 abjad/abugida students is particularly surprising, given their relatively high performance on the phonological awareness tasks. However, this result is consistent with the idea that these students may have developed good underlying phonological awareness skills, but are unable to apply them to performance on other literacy tasks. It is also important to note that other studies have shown reduced or non-existent relationships among phonological and literacy skills in readers of abjad languages such as Arabic, a finding which is also consistent with the current results (Arab-Moghaddam & Sénéchal, 2001; Nassaji & Geva, 1999; Saiegh-Haddad, 2003).

One final result from the spelling verification data bears mentioning. Inspection of the change across a semester for the different L1 groups shows that the L1 morphosyllabary students showed a large amount of improvement in response to the intervention instruction, and that their improvement was much larger than that of the other L1 groups. This result is consistent with previous research demonstrating the importance of cross-cultural differences in attitudes toward texts, literacy, and study skills for how students respond to instructional activities (Juffs & Friedline, 2014), and the generally more positive attitude than students from East Asian cultures have towards texts, reading, and related instruction.

The results from the reading survey are also consistent with the findings from Juffs and Friedline (2014) regarding attitudes toward literacy and classroom activities in ESL students from different cultural backgrounds. Specifically, the L1 morphosyllabary students reported higher levels of enjoyment reading in their L1 than the L1 abjad/abugida students, and the L1 Roman alphabet and the L1 abjad/abugida students reported reading less material in their L1 than the L1 non-Roman alphabet and the L1 morphosyllabary students. Despite this, the L1 abjad/abugida students indicated that it was highly important for them to be good readers in their L1 and in English, that it was also highly important for them to be good spellers in English, and that they were in fact good readers in their L1. These responses suggest that high levels of literacy may be objectively important as a goal for the L1 abjad/abugida students, but that these learners may be less willing to invest the time and effect necessary to develop these skills through activities such as extensive reading.

In general, there were few differences among the L1 groups with regard to their experiences learning to read in L1 and their self-reported L1 abilities, but unsurprisingly all ESL groups reported experiencing significantly more difficulty with English than the L1 English speakers. In contrast, there were generally no significant differences between the native and non-native English speakers with regard to how much and how often they reported reading for fun in English. This result may be a combination of ESL students reading somewhat more than expected in their L2, perhaps as a result of their immersion experience, and also L1 English speakers reading relatively little for enjoyment. Given the importance of print exposure for developing literacy skills and vocabulary growth (e.g., Berninger & Abbott, 1994; Cunningham & Stanovich, 1990; Juel, Griffith, & Gough, 1986; Sternberg, 1987), this has implications for students' continuing language and literacy growth.

From a pedagogical perspective, it is useful to consider the tasks for which effects of the intervention were or were not found. As mentioned above, it was unsurprising to find no effect of instruction type on the sound discrimination task because sound discrimination was not a skill targeted by the intervention activities. We did, however, hope to find effects of the intervention on phonological awareness and spelling verification (orthographic knowledge) because both of these skills are associated with the development of alphabetic literacy and have been successfully developed in training studies with children learning to read in their L1 (e.g., Ball & Blachman, 1991; Bradley & Bryant, 1983, 1985, 1991; Content et al., 1982; Gaskins et al., 1988; Kyle et al., 2013; Lovett, Lacerenza, & Borden, 2000; Lovett, Lacerenza, Borden, et al., 2000; Lundberg et al., 1988; Morais et al., 1988; Tunmer & Hoover, 1993; Wallach & Wallach, 1976; Williams, 1980). However, of the two tasks we only saw evidence for improvement on the spelling verification task. One factor that may have contributed to this result is the focus of the instruction and practice activities that students experienced as part of the instructional intervention. Specifically, although phonological awareness is highly correlated with decoding and spelling abilities (de Manrique & Signorini, 1994; Goswami & Bryant, 1990; Lundberg et al., 1988; Torgesen, Wagner, Rashotte, Burgess, & Hecht, 1997) we did not specifically train phonological awareness as part of the intervention. Rather, the intervention focused primarily on grapheme-phoneme correspondences, the most frequent pronunciations associated with various spelling patterns, and making analogies between words with shared spelling patterns (see above for details). All of these skills are directly related to spelling abilities, but not necessarily phonological awareness abilities. As others have observed, phonics-based instruction may indeed be as important or more important for spelling development than reading development (Jones, 1996), and we appear to have further evidence for this hypothesis with the current results.

Taken as a whole, the results from Study 3 suggest that ESL students develop differing levels of ability on sub-lexical literacy skills, and that many of these differences may be at least partially driven by the characteristics of their L1. However, just as in Study 1 and Study 2, the patterns of performance that we saw were not always directly predictable just from knowing participants' L1. As mentioned before, this is unsurprising given the complex competition and interactions that occur among the multiple languages an individual speaks (Li & MacWhinney, 2012; MacWhinney, 2005, 2008, 2012). The results also hint at the importance of students' cultural backgrounds and attitudes in the classroom for their literacy performance. However, the results show promise for the development of students' skills, particularly in response to targeted, phonics-based classroom interventions. The intervention instruction used in this study will provide an important starting point for the development of more targeted and more effective intervention activities going forward.

8.0 PEDAGOGICAL SUGGESTIONS

The results of this research allow for a number of specific, if tentative, pedagogical suggestions to be made. First, it is crucial for ESL instructors to understand that the difficulties their students face may differ substantially as a result of their L1 background, both because of linguistic influence and because of cultural differences in attitudes toward reading and literacy. As a result, effective instruction in a mixed-L1 classroom cannot be one-size-fits all. Rather, instructors need to be prepared to use a wide variety of activities to target specific skills for each L1 group.

Looking at some specific examples, the results from this study suggest that L1 Chinese speakers may do relatively well on spelling tasks compared to speakers from segmental L1s. In addition, L1 Chinese speakers may have developed some phonological awareness skills if they have experience with systems such as *pinyin* or *zhuyin fuhao*. However, these skills are likely to be relatively weak and in need of targeted instruction and practice to strengthen them. L1 Chinese speakers may also need instruction to help them learn how to apply their developing phonological skills to other literacy abilities, such as word identification, rather than relying on their preferred L1 strategy of using whole word form recognition.

The results from the orthographic knowledge tasks in particular emphasize the need for vowels and vowel spellings to receive extra attention in the ESL classroom. Due to the large number of vowel phonemes but the small number of vowel letters, the mappings between

English written and spoken vowels are extremely complex (Share, 2008). Although vowels are likely to present a challenge to all ESL learners, the current study showed that learners coming from an L1 that does not normally include written vowels as part of their writing system have significantly more difficulty with vowel spellings than with consonant spellings. This is therefore an excellent area of focus and emphasis for speakers from these types of L1.

The relative success of the instructional intervention provides a starting point for making effective pedagogical adjustments. Many of the skills and strategies used in the intervention activities overlapped with other pedagogical research, providing further evidence for their effectiveness. These included encouraging students to sound out (rather than blindly guess) an unknown word in a text, making analogies to other words with similar spelling patterns, and trying multiple possible pronunciations for an unknown word to find one that sounds familiar. In addition, as part of the intervention we encouraged teachers to emphasize writing words on the board when discussing unfamiliar vocabulary to help students build a connection between the written and the spoken forms, and encouraging students to write and pronounce the word as well. Instructors can make all of these adjustments quickly and easily without needing to add new material to their lessons, thus requiring a minimum level of preparation yet still providing the opportunity for significant gains by students.

9.0 GENERAL CONCLUSIONS

Obtaining literacy is crucial for success in both educational and employment contexts. However, adult learners of an L2 face a number of challenges when learning to read in their new language. In contrast to children, adults generally have limited oral proficiency in the target language, and they may have already developed fluent reading skills in their L1. These L1 reading experiences can impact the development of L2 literacy skills in complex ways. For example, previous research has demonstrated that the characteristics of a speaker's L1 can impact the level of skill that they develop on crucial sub-lexical abilities and the extent to which they rely on these skills to support performance on broader literacy tasks such as word recognition and reading comprehension (see Frost, 2012; Goswami & Bryant, 1990; Koda, 2004; Koda & Zehler, 2008; Ziegler & Goswami, 2005 for reviews).

One of the most useful frameworks for understanding the impacts of L1 experiences on L2 reading skills is the component skills approach (Carr & Levy, 1990; Koda, 2004). This framework emphasizes the contributions of separable cognitive and linguistic skills to overall reading success and thus allows for an examination of individual skills, their contributions to reading, and the extent to which they are independently affected by L1 background and other individual characteristics. Within this approach, two of the most important sub-lexical skills that support reading are phonological awareness and orthographic knowledge. Phonological awareness is the ability to manipulate and segment words into their component phonological units (McBride-Chang, 1995b), and orthographic knowledge consists of conscious and unconscious knowledge of the conventions of the writing system in a particular language (Berninger, 1994). Each of these skills develops in specific ways based on the characteristics of a reader's L1, and these different patterns of development can then transfer into an L2 and impact reading behaviors and performance in the L2. For example, learners from an alphabetic L1 tend to rely relatively more on phonological information for word recognition, and learners from a non-alphabetic L1 tend to rely relatively more on orthographic information for word recognition (e.g., Arab-Moghaddam & Sénéchal, 2001; McBride-Chang & Ho, 2005; Nassaji & Geva, 1999; Tong et al., 2009; Wade-Woolley, 1999).

Although the importance of these skills has been clearly demonstrated in both L1 and L2 contexts, there is still a relatively limited understanding of the fine-grained details of these skills in a cross-linguistic context. The majority of research on phonological awareness and orthographic knowledge has been conducted with monolingual children learning to read in their L1, typically English (Frost, 2012; Share, 2008). Many of the bilingual and second language studies that have been conducted involve speakers of alphabetic L1s or comparisons of speakers from an alphabetic L1 (e.g., Korean) and a morphosyllabic L1 (e.g., Chinese). Other types of languages, such as abjads and abugidas, have been largely left out of this research (see Tolchinsky et al., 2012 for a recent exception). In addition, there has been little consideration of the effect of task (e.g., response type, task demands) on performance. As a result, our understanding of the development of L2 literacy skills across multiple languages is relatively constrained, and little is definitely known regarding the impact of task choice on measured outcomes.

Another important gap in the existing literature is a lack of classroom-based studies of literacy skills development in ESL learners. Although the L1 literature demonstrates the trainability of phonological awareness and the success of phonics-based approaches for improving literacy skills (see NRP, 2000 for a review), relatively few studies have been conducted with ESL children and even fewer have included adult ESL students as the target population. Despite the lack of research in this area, the metacognitive skills of adult learners and the success of phonics training in L1 suggest that this is a fruitful area for future work, both for measuring the natural development of literacy skills and for evaluating the effectiveness of different types of instruction for improving these skills.

The goal of the present research was to begin addressing some of these gaps in the literature. Specifically, we conducted cross-linguistic comparisons of phonological awareness, orthographic knowledge, and the extent to which learners from different L1s rely on these skills for broader literacy tasks. We also measured the development of ESL literacy skills in classroom learners who received different types of reading instruction, either traditional vocabulary- and skills-based instruction or a supplemental phonics-based intervention. For Study 1 and Study 2 we targeted learners from three representative L1s: French (an alphabet), Hebrew (an abjad), and Chinese (a morphosyllabary). For Study 3 adult ESL students from a wide variety of L1 backgrounds provided data for the study.

Across all three studies there were pervasive effects of L1 influence on participants' relative level of skill on literacy tasks and on the relationships among performance on various literacy tasks. There was also evidence of phonological influence from the L1, separate from the confounded impacts of orthographic-phonological-morphological structure. However, in many cases our predictions were only partially confirmed, revealing the complexity of interacting factors that influence literacy acquisition. The work done here is particularly valuable in contributing to the limited body of cross-linguistic literacy research, an important step toward developing a better understanding of the universal and the language-specific aspects of reading (Frost, 2012).

This work also has important practical applications. Gaining a high level of English literacy is increasingly important for individuals around the world, and yet learners face very different challenges based on their previous language and literacy experiences. Specific suggestions for pedagogical changes can be made based on the results of the current research; this include emphasizing decoding skills, teaching common pronunciations for complex graphemes such as vowels, and encouraging students to make analogies to words they already know. However, much additional work in both laboratory and classroom contexts is needed to help clarify when and how L1 and task demands influence performance, as well as how instructors can best target their instruction to learners from particular L1s. Success in literacy skills is an important goal for learners of all backgrounds, and a careful cross-linguistic comparisons of literacy skills, similar to the current research, will continue to make valuable contributions in this area.

APPENDIX A

ODDITY STIMULI

Note: Within each triplet, stimuli are provided in the order in which participants heard them. The oddity item appears in *italic* type. Triplets were randomized within word location (beginning, middle, or end) for each participant.

<u>First Syllable</u>

office, *listen*, often
setting, settle, *handle*
deeply, closely, clothing
rider, *close*, rifle

weather, former, forward
funding, *sorry*, funny
handle, handsome, *grateful*
endless, homeless, homework

<u>Second Syllable</u>

water, after, *angry*
easy, even, happen
garlic, *travel*, topic
garbage, mortgage, *gifted*

really, *over*, only
market, target, *manage*
warning, darkness, weakness
rental, *carpet*, coastal

<u>Onset</u>

speak, break, bring
steal, *trail*, stare
smile, smooth, *pride*
cute, spite, spoon

still, stuff, *quick*
scale, grab, grade
trap, *spill*, trick
dried, drown, *queen*

Rime

his, hot, not
song, *rate*, wrong
shape, tape, *shoot*
leap, knock, lock

right, white, *reach*
cup, rule, cool
meal, *match*, wheel
cheat, neat, *chill*

Body

big, bit, *get*
call, fill, fish
beach, *coach*, beat
fame, fair, *sheer*

deal, *fail*, deep
like, live (/laɪv/), *shake*
beam, shade, shame
soak, *tuck*, soap

Coda

breast, last, *health*
search, shift, soft
cast, *depth*, dust
drunk, shrink, *drift*

drink, *stand*, thank
gift, craft, *glance*
shelf, fence, bounce
dump, *wrist*, shrimp

Initial Phoneme (C)

brain, breath, *dress*
ground, friend, from
sleep, *flood*, slide
swear, sweat, *twelve*

dream, *bright*, drive
blast, blood, *class*
blind, slight, slope
cruise, *drum*, crush

Medial Phoneme (C)

most, find, month
grant, *length*, trend
black, plain, *bread*
boast, harm, bird

burn, forth, *fast*
stiff, thrive, trash
drop, *steal*, trade
barn, purse, *beast*

Medial Phoneme (V)

street, *skin*, clean
frame, steak, *swim*
stick, sweet, freeze
glove, *stove*, shrug

track, climb, twice
slip, *stone*, twin
closed, scope, *crop*
cliff, scheme, grief

Final Phoneme (C)

build, *help*, hold
pound, bind, *bench*
skirt, hers, yours
shared, *spark*, sword

bunch, bend, found
scared, *storm*, tired
chest, blast, *grasp*
skilled, silk, sulk

APPENDIX B

DELETION STIMULI

Note: Stimuli were presented in groups according to the length of the phonological unit: first and second syllable, body and rime, onset and coda, and single phoneme. Within these groups stimuli were randomized for each participant.

Word Stimuli

First Syllable

wildlife without *wild*
birthday without *birth*
outside without *out*
sunlight without *sun*

Second Syllable

nightmare without *mare*
highway without *way*
meanwhile without *while*
peanut without *nut*

Body

mike without *my*
rode without *row*
state without *stay*
lane without *lay*

Pseudoword Stimuli

First Syllable

/dædʒɪt/ without /dæ/
/favuk/ without /fa/
/lɛtup/ without /lɛ/
/teɪnoʊs/ without /teɪ/

Second Syllable

/beɪmɛt/ without /ɛt/
/kælum/ without /um/
/doʊbas/ without /as/
/teɪgɔɪ/ without /ɔɪ/

Body

/zaɪg/ without /zaɪ/
/teɪv/ without /teɪ/
/fun/ without /fu/
/hoɪb/ without /hoɪ/

Rime

hand without *and*
send without *end*
fact without *act*
buy without *I*

Onset

bold without /b/
land without /l/
shout without /ʃ/
sit without /s/

Coda

same without /m/
wrote without /t/
light without /t/
bean without /n/

Initial Phoneme

play without /p/
black without /b/
thread without /θ/
slow without /s/

Medial Phoneme

stream without /ɹ/
sand without /n/
smell without /m/
snail without /n/

Final Phoneme

milk without /k/
belt without /t/
cold without /d/
wind without /d/

Rime

/zid/ without /id/
/joʊl/ without /oʊl/
/mug/ without /ug/
/paɪt/ without /aɪt/

Onset

/toʊb/ without /t/
/heɪn/ without /h/
/vaɪt/ without /v/
/θus/ without /θ/

Coda

/git/ without /t/
/feɪb/ without /b/
/zoʊʃ/ without /ʃ/
/dʒaɪn/ without /n/

Initial Phoneme

/plɛm/ without /p/
/spoʊn/ without /s/
/bluf/ without /b/
/snʌp/ without /s/

Medial Phoneme

/blaɪs/ without /l/
/mivz/ without /v/
/beɪmp/ without /m/
/skæf/ without /k/

Final Phoneme

/hɛts/ without /s/
/ɹɪlz/ without /z/
/hikt/ without /t/
/lʌsk/ without /k/

APPENDIX C

WORD/PSEUDOHOMOPHONE DISCRIMINATION STIMULI

Misspelled Items Involving Consonants

Word	Pseudohomophone
bite	bight
blow	bloe
church	churtch
claim	klame
clay	klay
clock	klock
cloud	kloud
clue	klue
cop	kop
crew	krew
cut	kut
dog	dawg
eight	ait
fade	phaid
fight	fite
fly	flight
grow	groe
horse	horce
main	mayn
more	mohr
nod	nodd
noise	noize
once	wunce
quote	kwote
raise	raize
roof	rooph
scare	skair

screw	skroo
should	shud
show	sho
sick	sikk
sigh	sye
stop	stawp
through	thrue
thumb	thum
tight	tite
toll	toal
wall	whall
wheat	weet
whole	hoal

Misspelled Items Involving Vowels

Word	Pseudohomophone
blade	blaid
born	boarn
cake	caik
cheek	cheke
cope	coap
deem	deam
dirt	durt
doll	daul
field	feeld
first	furst
floor	flore
free	frea
gaze	gaize
girl	gurl
goal	gole
gold	goald
great	grait
home	hoam
jail	jale
joke	joak
late	layt
load	lode
loop	lupe
mate	mait
pie	pye
rape	raip
shirt	shurt
small	smol
smoke	smoak

sneak	sneek
snow	snoe
soon	sune
suit	sute
sweep	sweap
tear	tair
third	thurd
threat	thret
word	wurd
works	werks
wound	woond

APPENDIX D

WORDLIKENESS JUDGMENTS STIMULI

Legal, High Positional Frequency	Illegal, High Positional Frequency
remond	mnteiu
siflet	mrates
sinald	pahpne
gonurd	olnged
nagred	irfned
firgen	afmrer
rigund	lbuoed
tecird	beoknr
golben	hnbied
naroud	uaotnm
paccet	pitisr
troper	lrlaey
cofife	tnruer
sufeer	aerrgd
asones	hcools
sirset	srttee
desund	srmmeu
valely	saocrs
lamian	aappre
vairer	bealtl

Legal, High Positional Frequency	Legal, Low Positional Frequency
samolt	incato
lewoly	edwron
werint	trelva
watord	odelub
sardep	engsil
limpes	evsril

socend	ineram
surtel	oseran
cipner	edirop
murben	egdrib
ceabem	ebecom
boreef	edesib
betret	ontcan
cachen	ededic
gedeer	eddman
fofret	hteeri
caseep	ellfow
rufuet	evehan
pirist	trerop
leraly	eficof

Legal, High Positional Frequency	Illegal, Low Positional Frequency
muinet	eflsti
matser	inlsda
phanep	ondrgu
genold	endnrga
firden	inerfg
ramfer	undrgi
boudel	etrcdi
ronkeb	onelgb
hibned	ndorau
acoint	eylwol
rutner	esrefu
soloch	esrtsi
rettes	dnesdu
merums	ylelav
casors	rdfaai
faraif	ilnmaa
papare	rreaiv
tabtel	ebeamc
boceem	oreefb
bidees	ettrbe

Illegal, High Positional Frequency	Legal, Low Positional Frequency
coaitn	elylow
wnroed	etwrin
avrlet	otdraw
mrtaes	edrsap
sgniel	empsil
serilv	endsoc
aeminr	estrul
peoird	embrun

rmnoed	itenum
tlfies	estram
afaifr	inalma
mbcoee	ebefor
bsieed	erbett
aoctnn	echcan
cdeied	eredge
anmded	edesir
aorlld	etorff
htriee	ecesap
llfoew	uferut
tproer	lyeral

Illegal, High Positional Frequency	Illegal, Low Positional Frequency
lanisd	ehpnpa
ruognd	eldngo
arnged	efnrdi
frnieg	erarfm
uirgnd	ebdluo
tericd	erknbo
benolg	inedhb
uoarnd	nmutao
maotls	ntiaco
olwley	edwnro
ccapet	tpsrii
fcfieo	rrentu
srfuee	rdgera
seoans	lsocho
srties	rteste
dsnued	rsemmu
llvaey	ocssra
araifd	rfaafi
aaimln	pperaa
rraiev	etlatb

Legal, Low Positional Frequency	Illegal, Low Positional Frequency
endrom	inemtu
estfil	entrsa
indsal	osltma
enpaph	etwnri
odgrun	tlerav
engdol	rtdwao
edgran	mtersa
endrif	esrdpa
efgrin	ngelsi

ermraf
ectcap
ertrun
eseruf
edgrar
esosan
oschol
eterts
esddun
emsrum
vyelal

esplmi
erptro
nsseoa
emeobd
iseedb
ncntao
enhcca
ollrda
eeerdg
ednadm
iseedr

APPENDIX E

OPERATION SPAN WORD STIMULI

English	French	Hebrew	Chinese
father	père	אבא	爸爸
uncle	oncle	דוד	叔叔
money	argent	כסף	錢
degree	degré	תואר	學位
dollar	dollar	דולר	美元
wood	bois	עץ	木材
metal	métal	מתכת	金屬
dog	chien	כלב	狗
chest	poitrine	חזה	胸
blood	sang	דם	血
hotel	hôtel	מלון	旅館
author	écrivain	סופר	作者
poem	poème	שיר	詩
mouth	bouche	פה	嘴
piano	piano	פסנתר	鋼琴
tree	arbre	אילן	樹
foot	pied	רגל	腳
rain	pluie	גשם	雨
group	groupe	קבוצה	組
clock	horloge	שעון	時鐘
dust	poussière	אבק	灰塵

English	French	Hebrew	Chinese
island	île	אי	島嶼
dinner	dîner	ארוחה	晚餐
bottle	bouteille	בקבוק	瓶子
hill	colline	גבעה	山丘
lake	lac	אגם	湖
king	roi	מלך	國王
girl	fille	ילדה	女孩
bank	banque	בנק	銀行
moon	lune	ירח	月亮
sign	panneau	סימן	記號
guide	guide	מדריך	導遊
bridge	pont	גשר	橋
chain	chaîne	שרשרת	鍊子
knife	couteau	סכין	刀子
world	monde	עולם	世界
pipe	tuyau	צינור	管子
leaf	feuille	עלה	葉子
site	site	מקום	地點
train	train	רכבת	火車
band	bande	להקה	樂團
plan	plan	תוכנית	計劃
rifle	fusil	רובה	步槍
nail	clou	מסמר	指甲
black	noir	שחור	黑色
paper	papier	נייר	紙
lion	lion	אריה	獅子
radio	radio	רדיו	收音機
finger	doigt	אצבע	手指
street	rue	רחוב	街
team	équipe	נבחרת	團隊
hand	main	יד	手指
boat	bateau	סירה	船
face	visage	פנים	臉

English	French	Hebrew	Chinese
valley	vallée	עמק	山谷
wine	vin	יין	酒
pear	poire	אגס	梨子
line	ligne	שורה	線
wall	mur	קיר	牆
floor	sol	קומה	地板
tooth	dent	שן	牙齒
rock	pierre	סלע	石頭
cloud	nuage	ענן	雲
month	mois	חודש	月份
beach	plage	ים	海灘
oven	four	תנור	烤箱
rule	règle	חוק	規則
flower	fleur	פרח	花
skirt	jupe	חצאית	裙子
coast	côte	חוף	海岸

APPENDIX F

LEXICAL DECISION STIMULI

Table 50. Lexical decision stimuli.

Consistent Words	Inconsistent Words	Exception Words	High Bigram Pseudo-words	Low Bigram Pseudo-words	Pseudo-homophones	Unpronounceable Non-words
lung	hat	door	kest	skows	eet	bnad
nice	has	egg	jines	skops	leen	drne
red	boot	faith	reen	swaft	eer	stagv
wise	leaf	inn	sind	vuffs	stoar	rakv
cheap	full	width	dind	smub	lern	twse
leave	four	huge	hest	gaft	bie	scolr
pace	watch	guess	dites	vops	baik	dlun
set	tough	odd	rones	vays	leest	ebnt
rag	shave	heart	zate	jods	fiew	tisf
safe	path	burst	dest	safs	feer	cdil
feel	tap	weigh	lind	gluff	steap	cnif
meet	wool	clothe	linds	puds	nale	stagb
hill	bear	world	lert	zake	taik	latv
wait	fall	yes	leat	krug	rute	safv
loan	tool	midst	trin	yags	lede	grunw
leg	cave	board	tind	twip	hert	tritv
feed	does	void	reats	slub	seak	tlnu
seed	worth	worse	sest	vulk	ded	stuml
hope	house	dealt	flate	jeeks	boan	aglf
roar	wash	off	stin	yake	saik	gtne
hole	mat	haul				
weed	jar	truth				
coin	fog	twelve				
dish	war	staff				
let	come	desk				
well	wood	else				

Consistent Words	Inconsistent Words	Exception Words
wire	boss	solve
mail	down	glimpse
mud	fat	curb
rug	loss	young
men	log	
need	cheese	
wet	rear	
file	good	
yet	heat	
rude	rise	
gap	deaf	
make	dive	
hire	sat	
thief	work	

APPENDIX G

NAMING STIMULI

Consistent Words	Inconsistent Words	Exception Words
bell	bare	board
cheap	bath	breathe
chore	both	crowd
coat	care	door
duck	catch	eye
feet	cave	faith
fell	comb	guess
foam	dare	guest
game	deaf	haul
gate	dear	heart
goat	dive	huge
him	food	lamb
hope	fool	laugh
keep	foot	lens
kneel	gear	loaf
lack	gone	meant
lawn	have	owe
leave	head	sleeve
less	hear	staff
much	leak	straight
neck	limb	teeth
room	loose	these
rope	love	tongue
safe	meat	truth
sell	mere	waist
shell	move	weigh
such	near	weird
tame	none	worse
teach	pour	yes

Consistent Words	Inconsistent Words	Exception Words
team	rough	young
tide	shall	
wage	tall	
wake	town	
week	wash	
wife	wave	
wing	weak	
wipe	wear	
wish	whose	
wreck	wool	
write	year	

APPENDIX H

PSEUDOWORD DECODING STIMULI

Note: Stimuli were taken from the non-word reading test in the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999). Participants read down the list of words in columns.

mo	mest	flimp
ik	stree	girtus
pu	weaf	strale
bi	barch	debmer
ib	glack	happon
ku	prot	framble
eb	runk	progus
pog	loast	supken
dat	mact	jeltic
mip	blork	tegwop
ral	phet	slinperk
nas	wogger	plinders
mib	klup	thundelp
faw	skad	bramtich
shum	keast	chimdruff
bice	churt	darlankert
nade	glamp	stremfick
teap	prait	morlingdon
derl	flact	revignuf
marl	throbe	obsorfelm
berk	creft	pitocrant

APPENDIX I

SPELLING VERIFICATION STIMULI

Table 51. Spelling verification stimuli.

Base Word	Actual Stimulus	Syllables	Frequency	Grapheme-Phoneme Correspondence Consistency	Answer Type[a]
came	came	Monosyllabic	High	Consistent	Correct
hair	hair	Monosyllabic	High	Consistent	Correct
brain	brain	Monosyllabic	High	Consistent	Correct
life	life	Monosyllabic	High	Consistent	Correct
health	heelth	Monosyllabic	High	Consistent	Pron-Alt
real	ril	Monosyllabic	High	Consistent	Pron-Alt
flight	fleght	Monosyllabic	High	Consistent	Pron-Alt
church	chorch	Monosyllabic	High	Consistent	Pron-Alt
learn	lern	Monosyllabic	High	Consistent	Pron-Pres
fact	fakt	Monosyllabic	High	Consistent	Pron-Pres
fate	fait	Monosyllabic	High	Consistent	Pron-Pres
night	niight	Monosyllabic	High	Consistent	Pron-Pres
deed	deed	Monosyllabic	Low	Consistent	Correct
starve	starve	Monosyllabic	Low	Consistent	Correct
bean	bean	Monosyllabic	Low	Consistent	Correct
bold	bold	Monosyllabic	Low	Consistent	Correct
verb	veerb	Monosyllabic	Low	Consistent	Pron-Alt
melt	meelt	Monosyllabic	Low	Consistent	Pron-Alt
bride	brid	Monosyllabic	Low	Consistent	Pron-Alt
noon	noom	Monosyllabic	Low	Consistent	Pron-Alt
shook	shookk	Monosyllabic	Low	Consistent	Pron-Pres
porch	portch	Monosyllabic	Low	Consistent	Pron-Pres
lend	lennd	Monosyllabic	Low	Consistent	Pron-Pres
true	troo	Monosyllabic	Low	Consistent	Pron-Pres
tongue	tongue	Monosyllabic	High	Exception	Correct
heart	heart	Monosyllabic	High	Exception	Correct
poor	poor	Monosyllabic	High	Exception	Correct

Base Word	Actual Stimulus	Syllables	Frequency	Grapheme-Phoneme Correspondence Consistency	Answer Type[a]
truth	truth	Monosyllabic	High	Exception	Correct
huge	huj	Monosyllabic	High	Exception	Pron-Alt
depth	deepth	Monosyllabic	High	Exception	Pron-Alt
meant	mant	Monosyllabic	High	Exception	Pron-Alt
worse	worze	Monosyllabic	High	Exception	Pron-Alt
peace	peece	Monosyllabic	High	Exception	Pron-Pres
solve	solv	Monosyllabic	High	Exception	Pron-Pres
source	sorce	Monosyllabic	High	Exception	Pron-Pres
desk	desc	Monosyllabic	High	Exception	Pron-Pres
pier	pier	Monosyllabic	Low	Exception	Correct
soothe	soothe	Monosyllabic	Low	Exception	Correct
sauce	sauce	Monosyllabic	Low	Exception	Correct
bulb	bulb	Monosyllabic	Low	Exception	Correct
garb	garp	Monosyllabic	Low	Exception	Pron-Alt
false	folse	Monosyllabic	Low	Exception	Pron-Alt
tempt	timpt	Monosyllabic	Low	Exception	Pron-Alt
sparse	sperse	Monosyllabic	Low	Exception	Pron-Alt
mourn	morne	Monosyllabic	Low	Exception	Pron-Pres
chef	shef	Monosyllabic	Low	Exception	Pron-Pres
watt	wat	Monosyllabic	Low	Exception	Pron-Pres
realm	relm	Monosyllabic	Low	Exception	Pron-Pres
child	child	Monosyllabic	High	Inconsistent	Correct
rare	rare	Monosyllabic	High	Inconsistent	Correct
meat	meat	Monosyllabic	High	Inconsistent	Correct
choose	choose	Monosyllabic	High	Inconsistent	Correct
could	culd	Monosyllabic	High	Inconsistent	Pron-Alt
lose	loze	Monosyllabic	High	Inconsistent	Pron-Alt
gave	gav	Monosyllabic	High	Inconsistent	Pron-Alt
year	yer	Monosyllabic	High	Inconsistent	Pron-Alt
wood	wud	Monosyllabic	High	Inconsistent	Pron-Pres
bread	bredd	Monosyllabic	High	Inconsistent	Pron-Pres
height	hight	Monosyllabic	High	Inconsistent	Pron-Pres
death	deth	Monosyllabic	High	Inconsistent	Pron-Pres
nut	nut	Monosyllabic	Low	Inconsistent	Correct
tomb	tomb	Monosyllabic	Low	Inconsistent	Correct
scarce	scarce	Monosyllabic	Low	Inconsistent	Correct
cone	cone	Monosyllabic	Low	Inconsistent	Correct
stove	stov	Monosyllabic	Low	Inconsistent	Pron-Alt
gap	gep	Monosyllabic	Low	Inconsistent	Pron-Alt
warn	wern	Monosyllabic	Low	Inconsistent	Pron-Alt
drown	dron	Monosyllabic	Low	Inconsistent	Pron-Alt
mall	mawl	Monosyllabic	Low	Inconsistent	Pron-Pres
stool	stule	Monosyllabic	Low	Inconsistent	Pron-Pres
letter	letter	Disyllabic	High	Consistent	Correct

Base Word	Actual Stimulus	Syllables	Frequency	Grapheme-Phoneme Correspondence Consistency	Answer Type[a]
little	little	Disyllabic	High	Consistent	Correct
budget	budget	Disyllabic	High	Consistent	Correct
number	number	Disyllabic	High	Consistent	Correct
silent	selent	Disyllabic	High	Consistent	Pron-Alt
people	peoble	Disyllabic	High	Consistent	Pron-Alt
garden	gairden	Disyllabic	High	Consistent	Pron-Alt
market	mairket	Disyllabic	High	Consistent	Pron-Alt
pocket	pockit	Disyllabic	High	Consistent	Pron-Pres
coffee	cofee	Disyllabic	High	Consistent	Pron-Pres
silver	silvur	Disyllabic	High	Consistent	Pron-Pres
permit	purmit	Disyllabic	High	Consistent	Pron-Pres
burden	burden	Disyllabic	Low	Consistent	Correct
coffin	coffin	Disyllabic	Low	Consistent	Correct
beetle	beetle	Disyllabic	Low	Consistent	Correct
bitter	bitter	Disyllabic	Low	Consistent	Correct
toilet	tolet	Disyllabic	Low	Consistent	Pron-Alt
gentle	gintle	Disyllabic	Low	Consistent	Pron-Alt
tickle	tickile	Disyllabic	Low	Consistent	Pron-Alt
bucket	bocket	Disyllabic	Low	Consistent	Pron-Alt
helmet	helmit	Disyllabic	Low	Consistent	Pron-Pres
cherry	cherrie	Disyllabic	Low	Consistent	Pron-Pres
muffin	mufin	Disyllabic	Low	Consistent	Pron-Pres
bundle	bundel	Disyllabic	Low	Consistent	Pron-Pres
create	create	Disyllabic	High	Inconsistent	Correct
senior	senior	Disyllabic	High	Inconsistent	Correct
native	native	Disyllabic	High	Inconsistent	Correct
volume	volume	Disyllabic	High	Inconsistent	Correct
prison	preson	Disyllabic	High	Inconsistent	Pron-Alt
damage	demage	Disyllabic	High	Inconsistent	Pron-Alt
reason	rison	Disyllabic	High	Inconsistent	Pron-Alt
future	foture	Disyllabic	High	Inconsistent	Pron-Alt
window	windo	Disyllabic	High	Inconsistent	Pron-Pres
direct	direkt	Disyllabic	High	Inconsistent	Pron-Pres
column	colomn	Disyllabic	High	Inconsistent	Pron-Pres
person	purson	Disyllabic	High	Inconsistent	Pron-Pres
petrol	petrol	Disyllabic	Low	Inconsistent	Correct
borrow	borrow	Disyllabic	Low	Inconsistent	Correct
versus	versus	Disyllabic	Low	Inconsistent	Correct
coupon	coupon	Disyllabic	Low	Inconsistent	Correct
spiral	speiral	Disyllabic	Low	Inconsistent	Pron-Alt
chorus	cherus	Disyllabic	Low	Inconsistent	Pron-Alt
cousin	coasin	Disyllabic	Low	Inconsistent	Pron-Alt
garage	garege	Disyllabic	Low	Inconsistent	Pron-Alt
climax	climaks	Disyllabic	Low	Inconsistent	Pron-Pres

Base Word	Actual Stimulus	Syllables	Frequency	Grapheme-Phoneme Correspondence Consistency	Answer Type[a]
gallon	galon	Disyllabic	Low	Inconsistent	Pron-Pres
collar	kollar	Disyllabic	Low	Inconsistent	Pron-Pres
hazard	hasard	Disyllabic	Low	Inconsistent	Pron-Pres
letter	letter	Disyllabic	High	Consistent	Correct
little	little	Disyllabic	High	Consistent	Correct

Note. [a]Phon-Alt indicates items whose misspelled form was designed to change the pronunciation of the base word; Phon-Pres indicates items whose misspelled form was design to preserve the pronunciation of the base word.

APPENDIX J

SOUND DISCRIMINATION STIMULI

Table 52. Sound discrimination stimuli.

Sound Contrast	Word 1	Word 2
/a/ vs. /o/	cot	coat
	clock	cloak
/æ/ vs. /aɪ/	back	bike
	cat	kite
/æ/ vs. /ɑ/ or /a/	map	mop
	sack	sock
/a/ vs. /ʌ/	paddle	puddle
	bag	bug
/æ/ vs. /ɛ/	man	men
	pan	pen
/æ/ vs. /eɪ/	snack	snake
	pants	paints
/ɑ/ vs. /eɪ/	fall	fail
	hall	hail
/ɑ/ vs. /ʌ/	boss	bus
	caught	cut
/ɑ/ or /ɔ/ vs. /o/	fawn	phone
	hall	hole
/eɪ/ vs. /aɪ/	tail	tile
	pail	pile
/eɪ/ vs. /i/	mail	meal
	tray	tree
/eɪ/ vs. /ɛ/	braid	bread
	pain	pen
/ɛ/ vs. /i/	check	cheek
	sweat	sweet
/ɛ/ vs. /ʌ/	net	nut
	desk	dusk

/ɛ/ vs. /a/	pet	pot
	step	stop
/ɤ/ vs. /o/	bull	bowl
	pull	pole
/i/ vs /ɪ/	peel	pill
	heel	hill
/ɪ/ vs. /aɪ/	win	wine
	hit	height
/o/ vs. /ɔʊ/	hole	howl
	coach	couch
/o/ vs. /ʌ/	coat	cut
	note	nut
/u/ vs. /aɪ/	spoon	spine
	noon	nine
/u/ or /ʊ/ vs. /ʌ/ or /ɤ/	fool	full
	pool	pull
/ʌ/ vs. /ɔ/	duck	dock
	luck	lock
/p/ vs. /b/	rope	robe
	pear	bear
/ɹ/ vs. /l/	list	wrist
	tile	tire
/s/ vs. /z/	ice	eyes
	peace	peas
/θ/ vs. /ð/	bath	bathe
	teeth	teethe

BIBLIOGRAPHY

Abbott, M. L. (2006). ESL reading strategies: Differences in Arabic and Mandarin speaker test performance. *Language Learning, 56*(4), 633-670. doi: 10.1111/j.1467-9922.2006.00391.x

Abu-Rabia, S. (1997a). Reading in Arabic orthography: The effect of vowels and context on reading accuracy of poor and skilled native Arabic readers in reading paragraphs, sentences, and isolated words. *Journal of Psycholinguistic Research, 26*(4), 465-482. doi: 10.1023/a:1025034220924

Abu-Rabia, S. (1997b). Verbal and working-memory skills of bilingual Hebrew-English speaking children. *International Journal of Psycholinguistics, 13*(1), 25-40.

Abu-Rabia, S. (1998). Reading Arabic texts: Effects of text type, reader type and vowelization. *Reading and Writing, 10*(2), 105-119. doi: 10.1023/a:1007906222227

Abu-Rabia, S. (2001). The role of vowels in reading Semitic scripts: Data from Arabic and Hebrew. *Reading and Writing, 14*(1), 39-59. doi: 10.1023/a:1008147606320

Abu-Rabia, S. (2002). Reading in a root–based–morphology language: The case of Arabic. *Journal of Research in Reading, 25*(3), 299-309. doi: 10.1111/1467-9817.00177

Abu-Rabia, S., & Awwad, J. (2004). Morphological structures in visual word recognition: the case of Arabic. *Journal of Research in Reading, 27*(3), 321-336. doi: 10.1111/j.1467-9817.2004.00235.x

Adams, M. J. (1990). *Begnning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2011). Pedagogical strategies for teaching literacy to ESL immigrant students: A meta-analysis. *British Journal of Educational Psychology, 81*(4), 629-653. doi: 10.1111/j.2044-8279.2010.02015.x

Akamatsu, N. (1999). The effects of first language orthographic features on word recognition processing in English as a second language. *Reading and Writing, 11*(4), 381-403. doi: 10.1023/a:1008053520326

Akamatsu, N. (2002). A similarity in word-recognition procedures among second language readers with different first language backgrounds. *Applied Psycholinguistics, 23*(01), 117-133. doi: 10.1017/S0142716402000061

Akamatsu, N. (2003). The effects of first language orthographic features on second language reading in text. *Language Learning, 53*(2), 207-231. doi: 10.1111/1467-9922.00216

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. doi: http://dx.doi.org/10.1016/j.jecp.2007.02.001

Alegria, J., Pignot, E., & Morais, J. (1982). Phonetic analysis of speech and memory codes in beginning readers. *Memory & Cognition, 10*(5), 451-456. doi: 10.3758/BF03197647

Alsulaimani, A. (1990). *Reading problems in Arabic learners of English.* (Unpublished Doctoral dissertation), London University, London, UK.

Alvarez, C. J., Carreiras, M., & Taft, M. (2001). Syllables and morphemes: Contrasting frequency effects in Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(2), 545-555. doi: 10.1037/0278-7393.27.2.545

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson, R. Barr, M. L. Kamil & P. Mosenthal (Eds.), *Handbook of reading research* (Vol. 1, pp. 255-291). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition, 10*(6), 565-575. doi: 10.3758/BF03202439

Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly, 38*(4), 470-487. doi: 10.1598/RRQ.38.4.3

Anthony, J. L., Williams, J. M., Durán, L. K., Gillam, S. L., Liang, L., Aghara, R., . . . Landry, S. H. (2011). Spanish phonological awareness: Dimensionality and sequence of development during the preschool and kindergarten years. *Journal of Educational Psychology, 103*(4), 857-876. doi: 10.1037/a0025024

Arab-Moghaddam, N., & Sénéchal, M. (2001). Orthographic and phonological processing skills in reading and spelling in Persian/English bilinguals. *International Journal of Behavioral Development, 25*(2), 140-147. doi: 10.1080/01650250042000320

Baddeley, A. D., Ellis, N. C., Miles, T. R., & Lewis, V. J. (1982). Developmental and acquired dyslexia: A comparison. *Cognition, 11*(2), 185-199. doi: http://dx.doi.org/10.1016/0010-0277(82)90025-7

Baker, C., & Jones, S. P. (1998). *Encyclopedia of bilingualism and bilingual education*: Multilingual Matters.

Ball, E. W., & Blachman, B. A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26*(1), 49-66.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*(3), 340-357. doi: 10.1037/0096-1523.10.3.340

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459.

Baron, J. (1979). Orthographic and word-specific mechanisms in children's reading of words. *Child Development, 50*(1), 60-72. doi: 10.2307/1129042

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

Barron, R. W. (1991). Proto-literacy, literacy and the acquisition of phonological awareness. *Learning and Individual Differences, 3*(3), 243-255. doi: http://dx.doi.org/10.1016/1041-6080(91)90010-X

Bear, D. R. (1991). “Learning to fasten the seat of my union suit without looking around”: The synchrony of literacy development. *Theory Into Practice, 30*(3), 149-157. doi: 10.1080/00405849109543495

Ben-Dror, I., Frost, R., & Bentin, S. (1995). Orthographic representation and phonemic segmentation in skilled readers: A cross-language comparison. *Psychological Science, 6*(3), 176-181.

Berninger, V. W. (1994). Introduction to the varieties of orthographic knowledge I: Theoretical and developmental issues. In V. W. Berninger (Ed.), *The Varieties of Orthographic Knowledge I: Theoretical and Developmental Issues*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Berninger, V. W., & Abbott, R. D. (1994). Multiple orthographic and phonological codes in literacy acquisition: An evolving research program. In V. Berninger (Ed.), *The Varieties of Orthographic Knowledge* (Vol. 8, pp. 277-319): Springer Netherlands.

Berninger, V. W., & Alsdorf, B. (1989). Are there errors in error analysis? *Journal of Psychoeducational Assessment, 7*(3), 209-222. doi: 10.1177/073428298900700303

Bertelson, P., de Gelder, B., & van Zon, M. (1997). Explicit speech segmentation and syllabic onset structure: Developmental trends. *Psychological Research, 60*(3), 183-191. doi: 10.1007/BF00419766

Besner, D., & Davelaar, E. (1983). Suedohomofoan effects in visual word recognition: Evidence for phonological processing. *Canadian Journal of Psychology/Revue canadienne de psychologie, 37*(2), 300-305. doi: 10.1037/h0080719

Bhide, A., Gadgil, S., Zelinsky, C. M., & Perfetti, C. A. (2013). Does reading in an alphasyllabary affect phonemic awareness? Inherent schwa effects in Marathi-English bilinguals. *Writing Systems Research*, 1-21. doi: 10.1080/17586801.2013.855619

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*(2), 290-303.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Developmental Science, 7*(3), 325-339. doi: 10.1111/j.1467-7687.2004.00351.x

Bialystok, E., McBride-Chang, C., & Luk, G. (2005). Bilingualism, language proficiency, and learning to read in two writing systems. *Journal of Educational Psychology, 97*(4), 580-590. doi: 10.1037/0022-0663.97.4.580

Birch, B. M. (2015). *English L2 reading: Getting to the bottom*. New York, NY: Routledge.

Borowsky, R., Owen, W., & Masson, M. J. (2002). Diagnostics of phonological lexical processing: Pseudohomophone naming advantages, disadvantages, and base-word frequency effects. *Memory & Cognition, 30*(6), 969-987. doi: 10.3758/BF03195781

Bosman, A. M. T. (1996). Phonologic mediation is fundamental to reading: Evidence from beginning readers. *The Quarterly Journal of Experimental Psychology Section A, 49*(3), 715-744. doi: 10.1080/713755641

Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organisation as a possible cause of reading backwardness. *Nature, 271*(5647), 746-747. doi: 10.1038/271746a0

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—A causal connection. *Nature, 301*(5899), 419-421. doi: 10.1038/301419a0

Bradley, L., & Bryant, P. E. (1985). Rhyme and reason in reading and spelling (International Academy for Research in Learning Disabilities Monograph Series No. 1). *Ann Arbor: University of Michigan Press*.

Bradley, L., & Bryant, P. E. (1991). Phonological skills before and after learning to read. In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Lisabelle Y. Liberman* (Vol. 37-46): Routledge.

Brady, S. A., & Shankweiler, D. P. (1991). *Phonological processes in literacy: A tribute to Isabelle Y. Liberman*: Routledge.

Brown, G. D. A., & Deavers, R. P. (1999). Units of analysis in nonword reading: Evidence from children and adults. *Journal of Experimental Child Psychology, 73*(3), 208-242. doi: http://dx.doi.org/10.1006/jecp.1999.2502

Brown, T. L., & Haynes, M. (1985). Literacy background and reading development in a second language. In H. Carr (Ed.), *The development of reading skills* (pp. 19-34). San Francisco, CA.

Bruce, D. J. (1964). The analysis of word sounds by young children. *British Journal of Educational Psychology, 34*(2), 158-170. doi: 10.1111/j.2044-8279.1964.tb00620.x

Bruck, M., Genesee, F., & Caravolas, M. (1997). A cross-linguistic study of early literacy acquisition. In B. A. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 145-162). Mahwah, NJ: Lawrence Erlbaum Associates.

Bruck, M., & Treiman, R. (1990). Phonological awareness and spelling in normal children and dyslexics: The case of initial consonant clusters. *Journal of Experimental Child Psychology, 50*(1), 156-178. doi: http://dx.doi.org/10.1016/0022-0965(90)90037-9

Bryant, P. E., & Bradley, L. (1983). Psychological strategies and the development of reading and writing. In M. Martlew (Ed.), *The psychology of written language: Developmental and educational perspectives* (pp. 163-178). Chichester: Wiley.

Bryant, P. E., Bradley, L., Maclean, M., & Crossland, J. (1989). Nursery rhymes, phonological skills and reading. *Journal of Child Language*, *16*(02), 407-428. doi: doi:10.1017/S0305000900010485

Bryant, P. E., & Impey, L. (1986). The similarities between normal readers and developmental and acquired dyslexics. *Cognition, 24*(1–2), 121-137. doi: http://dx.doi.org/10.1016/0010-0277(86)90007-7

Bryant, P. E., Maclean, M., & Bradley, L. (1990). Rhyme, language, and children's reading. *Applied Psycholinguistics*, *11*(03), 237-252. doi: doi:10.1017/S0142716400008870

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977-990. doi: 10.3758/BRM.41.4.977

Burgess, S. R., & Lonigan, C. J. (1998). Bidirectional relations of phonological sensitivity and prereading abilities: Evidence from a preschool sample. *Journal of Experimental Child Psychology*, *70*(2), 117-141. doi: http://dx.doi.org/10.1006/jecp.1998.2450

Calfee, R. C. (1977). Assessment of individual reading skills: Basic research and practical applications. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading*. New York: Lawrence Erlbaum Associates Inc.

Calfee, R. C., Lindamood, P., & Lindamood, C. (1973). Acoustic-phonetic skills and reading: Kindergarten through twelfth grade. *Journal of Educational Psychology, 64*(3), 293-298. doi: 10.1037/h0034586

Campbell, R. (1985). When children write nonwords to dictation. *Journal of Experimental Child Psychology*, *40*(1), 133-151. doi: http://dx.doi.org/10.1016/0022-0965(85)90069-4

Caravolas, M., & Bruck, M. (1993). The effect of oral and written language input on children′s phonological awareness: A cross-linguistic study. *Journal of Experimental Child Psychology, 55*(1), 1-30. doi: http://dx.doi.org/10.1006/jecp.1993.1001

Carr, T. H., & Levy, B. A. (1990). *Reading and its development: Component skills approaches*. San Diego, CA, US: Academic Press.

Carreiras, M., & Perea, M. (2002). Masked priming effects with syllabic neighbors in a lexical decision task. *Journal of Experimental Psychology: Human Perception and Performance, 28*(5), 1228-1242. doi: 10.1037/0096-1523.28.5.1228

Carrillo, M. (1994). Development of phonological awareness and reading acquisition. *Reading and Writing, 6*(3), 279-298. doi: 10.1007/BF01027086

Carroll, J. M., & Snowling, M. J. (2001). The effects of global similarity between stimuli on children's judgment of rime and alliteration. *Applied Psycholinguistics, 22*(03), 327-342. doi: doi:null

Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech, Language, and Hearing Research, 36*(5), 948-958. doi: 10.1044/jshr.3605.948

Chéreau, C., Gaskell, M. G., & Dumay, N. (2007). Reading spoken words: Orthographic effects in auditory priming. *Cognition, 102*(3), 341-360. doi: http://dx.doi.org/10.1016/j.cognition.2006.01.001

Cheung, H., Chen, H.-C., Lai, C. Y., Wong, O. C., & Hills, M. (2001). The development of phonological awareness: effects of spoken language experience and orthography. *Cognition, 81*(3), 227-241. doi: http://dx.doi.org/10.1016/S0010-0277(01)00136-6

Chikamatsu, N. (1996). The effects of L1 orthography on L2 word recognition: A study of American and Chinese learners of Japanese. *Studies in Second Language Acquisition, 18*(04), 403-432. doi: doi:10.1017/S0272263100015369

Chitiri, H.-F., Sun, Y., Willows, D. M., & Taylor, I. (1992). Word recognition in second-language reading. In H. Richard Jackson (Ed.), *Advances in Psychology* (Vol. Volume 83, pp. 283-297): North-Holland.

Chitiri, H.-F., & Willows, D. (1994). Word recognition in two languages and orthographies: English and Greek. *Memory & Cognition, 22*(3), 313-325. doi: 10.3758/BF03200859

Chomsky, C. (1970). Reading, writing, and phonology. *Harvard Educational Review, 40*(2), 287-309.

Chomsky, C. (1971). Write first, read later. *Childhood education, 47*(6), 296-299.

Chomsky, C. (1976). After decoding: what? *Language Arts*, 288-314.

Chukovsky, K. (1963). *From two to five*: University of California Press Berkeley.

Clay, M. M. (1979). *What did I write? Beginning writing behaviour*. Portsmouth, NH: Heinemann Educational Books, Inc.

Cobb, T. (2000). One size fits all? *Canadian Modern Language Review, 57*(2), 295-324.

Coltheart, M., Besner, D., Jonasson, J. T., & Davelaar, E. (1979). Phonological encoding in the lexical decision task. *Quarterly Journal of Experimental Psychology, 31*(3), 489-507. doi: 10.1080/14640747908400741

Coltheart, V., Patterson, K., & Leahy, J. (1994). When a ROWS is a ROSE: Phonological effects in written word comprehension. *The Quarterly Journal of Experimental Psychology Section A, 47*(4), 917-955. doi: 10.1080/14640749408401102

Comeau, L., Cormier, P., Grandmaison, É., & Lacroix, D. (1999). A longitudinal study of phonological processing skills in children learning to read in a second language. *Journal of Educational Psychology, 91*(1), 29-43. doi: 10.1037/0022-0663.91.1.29

Content, A., Morais, J., Alegria, J., & Bertelson, P. (1982). Accelerating the development of phonetic segmentation skills in kindergartners. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 2*(3), 259-269.

Cordts, A. D. (1965). *Phonics for the reading teacher*: Holt, Rinehart and Winston.

Cossu, G., Shankweiler, D., Liberman, I. Y., Katz, L., & Tola, G. (1988). Awareness of phonological segments and reading ability in Italian children. *Applied Psycholinguistics, 9*(01), 1-16. doi: doi:10.1017/S0142716400000424

Cossu, G., Shankweiler, D., Liberman, I. Y., Tola, G., & Katz, L. (1987). Awareness of phonological segments and reading ability in Italian children. *Haskins Labs Status Report on Speech Research, No. SR-91*.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition, 113*(2), 135-149. doi: http://dx.doi.org/10.1016/j.cognition.2009.08.001

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition, 106*(1), 59-86. doi: 10.1016/j.cognition.2006.12.013

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222-251. doi: 10.3102/00346543049002222

Cummins, J. (1984). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. In C. Rivera (Ed.), *Language proficiency and academic achievement: Multilingual matters 10* (pp. 2-19). Clevedon, England: Multilingual Matters Ltd.

Cunningham, A. E., Perry, K. E., & Stanovich, K. E. (2001). Converging evidence for the concept of orthographic processing. *Reading and Writing, 14*(5-6), 549-568. doi: 10.1023/A:1011100226798

Cunningham, A. E., & Stanovich, K. E. (1990). Assessing print exposure and orthographic processing skill in children: A quick measure of reading experience. *Journal of Educational Psychology, 82*(4), 733-740. doi: 10.1037/0022-0663.82.4.733

Cunningham, A. E., & Stanovich, K. E. (1993). Children's literacy environments and early word recognition subskills. *Reading and Writing, 5*(2), 193-204. doi: 10.1007/bf01027484

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*(4), 385-400. doi: http://dx.doi.org/10.1016/0749-596X(86)90033-1

Damian, M. F., & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language, 49*(1), 119-132. doi: http://dx.doi.org/10.1016/S0749-596X(03)00008-1

Daniels, P. T., & Bright, W. (1996). *The world's languages*. New York: Oxford University Press.

Davies, M. (2008-). The Corpus of Contemporary American English: 450 million words, 1990-present.

Davis, C. (1998). Masked homophone and pseudohomophone priming in children and adults. *Language and Cognitive Processes, 13*(6), 625-651. doi: 10.1080/016909698386401

De Cara, B., & Goswami, U. (2003). Phonological neighbourhood density: Effects in a rhyme awareness task in five-year-old children. *Journal of Child Language, 30*(03), 695-710. doi: 10.1017/S0305000903005725

de Jong, P. F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology, 95*(1), 22-40. doi: 10.1037/0022-0663.95.1.22

de Manrique, A. M. B., & Signorini, A. (1994). Phonological awareness, spelling and reading abilities in Spanish-speaking children. *British Journal of Educational Psychology, 64*(3), 429-439. doi: 10.1111/j.2044-8279.1994.tb01114.x

de Santos Loureiro, C., Willadino Braga, L., do Nascimento Souza, L., Nunes Filho, G., Queiroz, E., & Dellatolas, G. (2004). Degree of illiteracy and phonological and metaphonological skills in unschooled adults. *Brain and Language, 89*(3), 499-502. doi: http://dx.doi.org/10.1016/j.bandl.2003.12.008

Deacon, S. H., Wade-Woolley, L., & Kirby, J. R. (2009). Flexibility in young second-language learners: Examining the language specificity of orthographic processing. *Journal of Research in Reading, 32*(2), 215-229. doi: 10.1111/j.1467-9817.2009.01392.x

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., . . . Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science, 330*(6009), 1359-1364. doi: 10.1126/science.1194140

Demont, E., & Gombert, J. E. (1996). Phonological awareness as a predictor of recoding skills and syntactic awareness as a predictor of comprehension skills. *British Journal of Educational Psychology, 66*(3), 315-332. doi: 10.1111/j.2044-8279.1996.tb01200.x

Domínguez, A. B. (1996). Evaluating of long-term effects of teaching phonological analysis abilities on reading and writing acquisition. *Infancia y Aprendizaje, 19*(76), 83-96. doi: 10.1174/021037096762905571

Dowker, A. (1989). Rhyme and alliteration in poems elicited from young children. *Journal of Child Language, 16*(01), 181-202. doi: doi:10.1017/S0305000900013507

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2007). Do transposed-letter similarity effects occur at a morpheme level? Evidence for morpho-orthographic decomposition. *Cognition, 105*(3), 691-703. doi: http://dx.doi.org/10.1016/j.cognition.2006.12.001

Durgunoğlu, A. Y., & Hancin, B. J. (1992). An overview of cross-language transfer in bilingual reading. In H. Richard Jackson (Ed.), *Advances in Psychology* (Vol. Volume 83, pp. 391-411): North-Holland.

Durgunoğlu, A. Y., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology, 85*(3), 453-465. doi: 10.1037/0022-0663.85.3.453

Durgunoğlu, A. Y., & Öney, B. (1999). A cross-linguistic comparison of phonological awareness and word recognition. *Reading and Writing, 11*(4), 281-299. doi: 10.1023/A:1008093232622

Durgunoğlu, A. Y., & Öney, B. (2002). Phonological awareness in literacy acquisition: It's not only for children. *Scientific Studies of Reading, 6*(3), 245-266. doi: 10.1207/S1532799XSSR0603_3

Ehri, L. C. (1980a). Reading and spelling in beginners: The development of orthographic images as word symbols in lexical memory *Cognitive processes in spelling*: Academic Press New York/London.

Ehri, L. C. (1980b). The role of orthographic images in learning printed words. *Orthography, reading and dyslexia*, 155-170.
Ehri, L. C. (1986). Sources of difficulty in learning to spell and read. *Advances in Developmental & Behavioral Pediatrics, 7*, 121-195.
Ehri, L. C. (1989). The development of spelling knowledge and its role in reading acquisition and reading disability. *Journal of Learning Disabilities, 22*(6), 356-365. doi: 10.1177/002221948902200606
Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. B. Gough, L. C. Ehri & R. Treiman (Eds.), *Reading acquisition* (pp. 107-143). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
Ehri, L. C. (1994). Development of the ability to read words: Update. In R. Ruddell, M. R. Ruddell & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed. ed., pp. 323-358). Newark, DE: International Reading Association.
Ehri, L. C. (1997). *Sight word learning in normal readers and dyslexics*. Mahwah, NJ: Erlbaum.
Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential to learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3-40). Mahwah, NJ: Erlbaum.
Ehri, L. C. (1999). Phases of development in learning to read words. In J. O. R. Beard (Ed.), *Reading development and the teaching of reading: A psychological perspective* (pp. 79-108). Oxford, England: Blackwell Science.
Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*(3), 250-287. doi: 10.1598/RRQ.36.3.2
Ehri, L. C., & Wilce, L. S. (1980). The influence of orthography on readers' conceptualization of the phonemic structure of words. *Applied Psycholinguistics, 1*(04), 371-385. doi: doi:10.1017/S0142716400009802
Ellis, N., & Large, B. (1987). The development of reading: As you seek so shall you find. *British Journal of Psychology, 78*(1), 1-28. doi: 10.1111/j.2044-8295.1987.tb02222.x
Eriksen, C. W. (1995). The Flankers Task and response competition: A useful tool for investigating a variety of cognitive problems. In C. Bundesen & H. Shibuya (Eds.), *Visual Selective Attention* (pp. 101-118). East Sussex, UK: Lawrence Erlbaum.
Fender, M. (2003). English word recognition and word integration skills of native Arabic- and Japanese-speaking learners of English as a second language. *Applied Psycholinguistics, 24*(02), 289-315. doi: 10.1017/S014271640300016X
Fender, M. (2008). Spelling knowledge and reading development: Insights from Arab ESL learners. *Reading in a Foreign Language, 20*(1), 19-42.
Ferrand, L., & Grainger, J. (1992). Phonology and orthography in visual word recognition: Evidence from masked non-word priming. *The Quarterly Journal of Experimental Psychology Section A, 45*(3), 353-372. doi: 10.1080/02724989208250619
Fitts, P. M., & Posner, M. I. (1967). *Human Performance*. Belmont, CA: Brooks/Cole.
Fleming, J. T. (1976). Alternative interpretations of evidence for phonemic recoding in visual word recognition. *Journal of Literacy Research, 8*(1), 7-18. doi: 10.1080/10862967609547161

Folk, J. R. (1999). Phonological codes are used to access the lexicon during silent reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 892-906. doi: 10.1037/0278-7393.25.4.892

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*(6), 627-635. doi: http://dx.doi.org/10.1016/S0022-5371(73)80042-8

Forster, K. I., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation? *The Quarterly Journal of Experimental Psychology Section A, 39*(2), 211-251. doi: 10.1080/14640748708401785

Fox, B., & Routh, D. K. (1975). Analyzing spoken language into words, syllables, and phonomes: A developmental study. *Journal of Psycholinguistic Research, 4*(4), 331-342. doi: 10.1007/bf01067062

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(3), 361-379. doi: 10.1037/0096-1523.2.3.361

Frith, U. (1980). Unexpected spelling problems. In U. Frith (Ed.), *Cognitive processes in spelling*. London: Academic Press.

Frith, U., & Snowling, M. (1983). Reading for meaning and reading for sound in autistic and dyslexic children. *British Journal of Developmental Psychology, 1*(4), 329-342. doi: 10.1111/j.2044-835X.1983.tb00906.x

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123*(1), 71-99. doi: 10.1037/0033-2909.123.1.71

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences, 35*(05), 263-279. doi: doi:10.1017/S0140525X12001574

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. doi: 10.1037/0096-1523.13.1.104

Gaskins, I. W., Downer, M. A., Anderson, R. C., Cunningham, P. M., Gaskins, R. W., & Schommer, M. (1988). A metacognitive approach to phonics: Using what you know to decode what you don't know. *Remedial and Special Education*, *9*(1), 36-41. doi: 10.1177/074193258800900107

Gaskins, I. W., Ehri, L. C., Cress, C., O'Hara, C., & Donnelly, K. (1996). Procedures for word learning: Making discoveries about words. *The Reading Teacher, 50*(4), 312-327. doi: 10.2307/20201768

Georgiou, G. K., Parrila, R., & Papadopoulos, T. C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology, 100*(3), 566-580. doi: 10.1037/0022-0663.100.3.566

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*(4), 674-691. doi: 10.1037/0096-1523.5.4.674

Goikoetxea, E. (2005). Levels of phonological awareness in preliterate and literate Spanish-speaking children. *Reading and Writing*, *18*(1), 51-79. doi: 10.1007/s11145-004-1955-7

Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *The Journal of Neuroscience, 33*(2), 387-396. doi: 10.1523/jneurosci.3837-12.2013

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development, 58*(1), 80-92. doi: 10.2307/1130293

Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology, 42*(1), 73-83. doi: http://dx.doi.org/10.1016/0022-0965(86)90016-0

Goswami, U. (1988a). Children's use of analogy in learning to spell. *British Journal of Developmental Psychology, 6*(1), 21-33. doi: 10.1111/j.2044-835X.1988.tb01077.x

Goswami, U. (1988b). Orthographic analogies and reading development. *The Quarterly Journal of Experimental Psychology Section A, 40*(2), 239-268. doi: 10.1080/02724988843000113

Goswami, U. (1990). Phonological priming and orthographic analogies in reading. *Journal of Experimental Child Psychology, 49*(2), 323-340. doi: http://dx.doi.org/10.1016/0022-0965(90)90062-D

Goswami, U. (1999). The relationship between phonological awareness and orthographic representation in different orthographies. In M. Harris & G. Hatano (Eds.), *Learning to read and write: A cross-linguistic perspective* (pp. 134-156). New York: Cambridge University Press.

Goswami, U. (2012). Universals of reading: Developmental evidence for linguistic plausibility. *Behavioral and Brain Sciences, 35*(05), 287-288. doi: doi:10.1017/S0140525X12000052

Goswami, U., & Bryant, P. E. (1990). *Phonological skills and learning to read*. Hove, UK: Erlbaum.

Goswami, U., & East, M. (2000). Rhyme and analogy in beginning reading: Conceptual and methodological issues. *Applied Psycholinguistics, 21*(01), 63-93.

Goswami, U., Gombert, J. E., & de Barrera, L. F. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics, 19*(01), 19-52. doi: doi:10.1017/S0142716400010560

Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2001). Pseudohomophone effects and phonological recoding procedures in reading development in English and German. *Journal of Memory and Language, 45*(4), 648-664. doi: http://dx.doi.org/10.1006/jmla.2001.2790

Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2003). Nonword reading across orthographies: How flexible is the choice of reading units? *Applied Psycholinguistics, 24*(02), 235-247. doi: doi:10.1017/S0142716403000134

Gottardo, A., Chiappe, P., Yan, B., Siegel, L., & Gu, Y. (2006). Relationships between first and second language phonological processing skills and reading in Chinese-English speakers living in English-speaking contexts. *Educational Psychology, 26*(3), 367-393. doi: 10.1080/01443410500341098

Gottardo, A., Yan, B., Siegel, L. S., & Wade-Woolley, L. (2001). Factors related to English reading performance in children with Chinese as a first language: More evidence of cross-language transfer of phonological processing. *Journal of Educational Psychology, 93*(3), 530-542. doi: 10.1037/0022-0663.93.3.530

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge: Cambridge University Press.

Graddol, D. (1997). *The future of English: A guide to forecasting the popularity of the English language in the 21st century*: The British Council.

Greaney, K. T., Tunmer, W. E., & Chapman, J. W. (1997). Effects of rime-based orthographic analogy training on the word recognition skills of children with reading disability. *Journal of Educational Psychology, 89*(4), 645-651. doi: 10.1037/0022-0663.89.4.645

Green, D. W., Liow, S. J. R., Tng, S. K., & Zielinski, S. (1996). Are visual search procedures adapted to the nature of the script? *British Journal of Psychology, 87*(2), 311-326. doi: 10.1111/j.2044-8295.1996.tb02592.x

Green, D. W., & Meara, P. (1987). The effects of script on visual search. *Second Language Research, 3*(2), 102-113. doi: 10.1177/026765838700300202

Hallé, P. A., Chéreau, C., & Segui, J. (2000). Where Is the /b/ in “absurde” [apsyrd]? It Is in French Listeners' Minds. *Journal of Memory and Language, 43*(4), 618-639. doi: http://dx.doi.org/10.1006/jmla.2000.2718

Hamada, M., & Koda, K. (2008). Influence of first language orthographic experience on second language decoding and word learning. *Language Learning, 58*(1), 1-31. doi: 10.1111/j.1467-9922.2007.00433.x

Harris, L. N., Perfetti, C. A., & Rickles, B. (2014). Error-related negativities during spelling judgments expose orthographic knowledge. *Neuropsychologia, 54*(0), 112-128. doi: http://dx.doi.org/10.1016/j.neuropsychologia.2013.12.007

Harris, M., & Giannouli, V. (1999). Learning to read and spell in Greek: The importance of letter knowledge and morphological awareness. In M. Harris & G. Hatano (Eds.), *Learning to read and write: A cross-linguistic perspective*. New York, NY: Cambridge University Press.

Hayes-Harb, R. (2006). Native speakers of Arabic and ESL texts: Evidence for the transfer of written word identification processes. *TESOL Quarterly, 40*(2), 321-339.

Haynes, M., & Carr, T. H. (1990). Writing system background and second language reading: A component skills analysis of English reading by native speaker-readers of Chinese. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 375-422). San Diego, CA: Academic Press, Inc.

Heywood, J. A., & Nahmad, H. M. (1965). *The Arabic language: A new Arabic grammar of the written language*. London: Percy Lund, Humphries.

Ho, C. S.-H., & Bryant, P. E. (1997). Phonological skills are important in learning to read Chinese. *Developmental Psychology, 33*(6), 946-951. doi: 10.1037/0012-1649.33.6.946

Ho, C. S.-H., & Lai, D. N.-C. (1999). Naming-speed deficits and phonological memory deficits in Chinese developmental dyslexia. *Learning and Individual Differences, 11*(2), 173-186. doi: http://dx.doi.org/10.1016/S1041-6080(00)80004-7

Høien, T., Lundberg, I., Stanovich, K., & Bjaalid, I.-K. (1995). Components of phonological awareness. *Reading and Writing, 7*(2), 171-188. doi: 10.1007/bf01027184

Holligan, C., & Johnston, R. (1988). The use of phonological information by good and poor readers in memory and reading tasks. *Memory & Cognition, 16*(6), 522-532. doi: 10.3758/BF03197053

Holm, A., & Dodd, B. (1996). The effect of first written language on the acquisition of English literacy. *Cognition, 59*(2), 119-147. doi: 10.1016/0010-0277(95)00691-5

Homer, B. D. (2009). Literacy and metalinguistic development. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 487-500). Cambridge, UK: Cambridge University Press.

Hu, C.-F., & Catts, H. W. (1998). The role of phonological processing in early reading ability: What we can learn from Chinese. *Scientific Studies of Reading, 2*(1), 55-79. doi: 10.1207/s1532799xssr0201_3

Huang, H. S., & Hanley, J. R. (1995). Phonological awareness and visual skills in learning to read Chinese and English. *Cognition, 54*(1), 73-98. doi: http://dx.doi.org/10.1016/0010-0277(94)00641-W

Huang, H. S., & Hanley, J. R. (1997). A longitudinal study of phonological awareness, visual skills, and Chinese reading acquisition among first-graders in Taiwan. *International Journal of Behavioral Development, 20*(2), 249-268. doi: 10.1080/016502597385324

Hung, D. L., & Tzeng, O. J. (1981). Orthographic variations and visual information processing. *Psychological Bulletin, 90*(3), 377-414. doi: 10.1037/0033-2909.90.3.377

Inagaki, K., Hatano, G., & Otake, T. (2000). The effect of kana literacy acquisition on the speech segmentation unit used by Japanese young children. *Journal of Experimental Child Psychology*, *75*(1), 70-91. doi: http://dx.doi.org/10.1006/jecp.1999.2523

IPA. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Melbourne, Australia: Cambridge University Press.

Jackson, N., Lu, W.-H., & Ju, D. (1994). Reading Chinese and reading English: Similarities, differences, and second-language reading. In V. Berninger (Ed.), *The Varieties of Orthographic Knowledge* (Vol. 8, pp. 73-109): Springer Netherlands.

Jakimik, J., Cole, R. A., & Rudnicky, A. I. (1985). Sound and spelling in spoken word recognition. *Journal of Memory and Language, 24*(2), 165-178. doi: http://dx.doi.org/10.1016/0749-596X(85)90022-1

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*(1), 160-212. doi: 10.1111/lang.12034

Jiménez González, J. E., & Ortiz González, M. d. R. (2000). Metalinguistic awareness and reading acquisition in the Spanish language. *The Spanish Journal of Psychology, 3*, 37-46. doi: doi:10.1017/S1138741600005527

Johnston, T. C., & Kirby, J. R. (2006). The contribution of naming speed to the Simple View of Reading. *Reading and Writing, 19*(4), 339-361. doi: 10.1007/s11145-005-4644-2

Jones, M. L. (1996). *Phonics in ESL literacy instruction: Functional or not?* Paper presented at the World Conference on Literacy, Philadelphia, PA.

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78*(4), 243-255. doi: 10.1037/0022-0663.78.4.243

Juffs, A. (2004-2015). *University of Pittsburgh English Language Institute Database*.

Juffs, A., & Friedline, B. E. (2014). Sociocultural influences on the use of a web-based tool for learning English vocabulary. *System, 42*(0), 48-59. doi: http://dx.doi.org/10.1016/j.system.2013.10.015

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189-217. doi: 10.1037/0096-3445.133.2.189

Kang, H., & Simpson, G. B. (1996). Development of semantic and phonological priming in a shallow orthography. *Developmental Psychology, 32*(5), 860-866. doi: 10.1037/0012-1649.32.5.860

Katz, L., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(1), 157-166. doi: 10.1037/0278-7393.9.1.157

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The Orthographic Depth Hypothesis. In R. Frost & M. Katz (Eds.), *Orthography, Phonology, Morphology, and Meaning*. Amsterdam: Elsevier.

Katzir, T., Kim, Y., Wolf, M., O'Brien, B., Kennedy, B., Lovett, M., & Morris, R. (2006). Reading fluency: The whole is more than the parts. *Annals of Dyslexia, 56*(1), 51-82. doi: 10.1007/s11881-006-0003-5

Kim, Y.-S. (2007). Phonological awareness and literacy skills in Korean: An examination of the unique role of body-coda units. *Applied Psycholinguistics, 28*(01), 69-94. doi: doi:10.1017/S014271640707004X

Kim, Y.-S. (2009). Crosslinguistic influence on phonological awareness for Korean–English bilingual children. *Reading and Writing, 22*(7), 843-861. doi: 10.1007/s11145-008-9132-z

Kimura, Y., & Bryant, P. E. (1983). Reading and writing in English and Japanese: A cross-cultural study of young children*. *British Journal of Developmental Psychology, 1*(2), 143-153. doi: 10.1111/j.2044-835X.1983.tb00552.x

Kirsch, I. S. (1993). *Adult literacy in America: A first look at the results of the National Adult Literacy Survey*: ERIC.

Kochnower, J., Richardson, E., & DiBenedetto, B. (1983). A comparison of the phonic decoding ability of normal and learning disabled children. *Journal of Learning Disabilities, 16*(6), 348-351. doi: 10.1177/002221948301600609

Koda, K. (1989a). Effects of L1 orthographic representation on L2 phonological coding strategies. *Journal of Psycholinguistic Research, 18*(2), 201-222. doi: 10.1007/bf01067782

Koda, K. (1989b). The effects of transferred vocabulary knowledge on the development of L2 reading proficiency. *Foreign Language Annals, 22*(6), 529-540. doi: 10.1111/j.1944-9720.1989.tb02780.x

Koda, K. (1990). The use of L1 reading strategies in L2 reading: Effects of L1 orthographic structures on L2 phonological recoding strategies. *Studies in Second Language Acquisition, 12*(04), 393-410. doi: 10.1017/S0272263100009499

Koda, K. (1998). The role of phonemic awareness in second language reading. *Second Language Research, 14*(2), 194-215. doi: 10.1191/026765898676398460

Koda, K. (1999). Development of L2 intraword orthographic sensitivity and decoding skills. *The Modern Language Journal, 83*(1), 51-64. doi: 10.1111/0026-7902.00005

Koda, K. (2004). *Insights into second language reading: A cross-linguistic approach*. New York: Cambridge University Press.

Koda, K., & Zehler, A. M. (2008). *Learning to read across languages: Cross-linguistic relationships in first- and second-language literacy development*. New York: Routledge.

Ktori, M., & Pitchford, N. J. (2008). Effect of orthographic transparency on letter position encoding: A comparison of Greek and English monoscriptal and biscriptal readers. *Language and Cognitive Processes, 23*(2), 258-281. doi: 10.1080/01690960701536797

Kubozono, H. (2002). Mora and syllable. In N. Tsujimura (Ed.), *The handbook of Japanese linguistics* (pp. 31-61). Malden, MA: Blackwell Publishers, Inc.

Kurvers, J., Stockmann, W., & van de Craats, I. (2010). Predictors of success in adult L2 literacy acquisition. In T. Wall & M. Leong (Eds.), *Low Educated Second Language and Literacy Acquisition* (pp. 64-79). Calgary: Bow Valley College.

Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG Rime and GG Phoneme. *Reading Research Quarterly, 48*(1), 61-76. doi: 10.1002/rrq.038

Lam, A. S. L., Perfetti, C. A., & Bell, L. (1991). Automatic phonetic transfer in bidialectal reading. *Applied Psycholinguistics, 12*(03), 299-311. doi: doi:10.1017/S0142716400009243

Laufer, B., & Nation, I. S. P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33-51. doi: 10.1177/026553229901600103

Laxon, V. J., Masterson, J., Pool, M., & Keating, C. (1992). Nonword naming: Further exploration of the pseudohomophone effect in terms of orthographic neighborhood size, graphemic changes, spelling-sound consistency, and reader accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(4), 730-748. doi: 10.1037/0278-7393.18.4.730

Lenel, J., & Cantor, J. (1981). Rhyme recognition and phonemic perception in young children. *Journal of Psycholinguistic Research, 10*(1), 57-67. doi: 10.1007/BF01067361

Li, P., & MacWhinney, B. (2012). Competition model. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*: Blackwell Publishing Ltd.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*(2), 201-212. doi: http://dx.doi.org/10.1016/0022-0965(74)90101-5

Liow, S. J. R., & Poon, K. K. L. (1998). Phonological awareness in multilingual Chinese children. *Applied Psycholinguistics, 19*(03), 339-362. doi: 10.1017/S0142716400010213

Liu, Y., Dunlap, S., Fiez, J., & Perfetti, C. A. (2007). Evidence for neural accommodation to a writing system following learning. *Human Brain Mapping, 28*(11), 1223-1234. doi: 10.1002/hbm.20356

Liu, Y., Perfetti, C. A., & Wang, M. (2006). Visual analysis and lexical access of Chinese characters by Chinese as second language readers. *Language and Linguistics, 7*(3), 637-357.

Lovett, M. W., Lacerenza, L., & Borden, S. L. (2000). Putting struggling readers on the PHAST track: A program to integrate phonological and strategy-based remedial reading instruction and maximize outcomes. *Journal of Learning Disabilities, 33*(5), 458-476. doi: 10.1177/002221940003300507

Lovett, M. W., Lacerenza, L., Borden, S. L., Frijters, J. C., Steinbach, K. A., & De Palma, M. (2000). Components of effective remediation for developmental reading disabilities: Combining phonological and strategy-based instruction to improve outcomes. *Journal of Educational Psychology, 92*(2), 263-283. doi: 10.1037/0022-0663.92.2.263

Lukatela, G., Eaton, T., Lee, C. H., Carello, C., & Turvey, M. T. (2002). Equal homophonic priming with words and pseudohomophones. *Journal of Experimental Psychology: Human Perception and Performance, 28*(1), 3-21. doi: 10.1037/0096-1523.28.1.3

Lukatela, G., & Turvey, M. T. (1991). Phonological access of the lexicon: Evidence from associative priming with pseudohomophones. *Journal of Experimental Psychology: Human Perception and Performance, 17*(4), 951-966. doi: 10.1037/0096-1523.17.4.951

Lukatela, G., & Turvey, M. T. (1994a). Visual lexical access is initially phonological: 2. Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General, 123*(4), 331-353. doi: 10.1037/0096-3445.123.4.331

Lukatela, G., & Turvey, M. T. (1994b). Visual lexical access is initially phonological: I. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General, 123*(2), 107-128. doi: 10.1037/0096-3445.123.2.107

Lukatela, G., & Turvey, M. T. (2000). Do spelling variations affect associative and phonological priming by pseudohomophones? *Perception & Psychophysics, 62*(1), 196-217. doi: 10.3758/BF03212072

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28*(2), 203-208. doi: 10.3758/BF03204766

Lundberg, I., Frost, J., & Petersen, O.-P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23*(3), 263-284. doi: 10.2307/748042

Lundberg, I., Olofsson, Å., & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21*(1), 159-173. doi: 10.1111/j.1467-9450.1980.tb00356.x

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. doi: 10.3758/bf03211402

Lupker, S. J., & Pexman, P. M. (2010). Making things difficult in lexical decision: The impact of pseudohomophones and transposed-letter nonwords on frequency and semantic priming effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1267-1289. doi: 10.1037/a0020125

Maclean, M., Bryant, P. E., & Bradley, L. (1987). Rhymes, nursery rhymes, and reading in early childhood. *Merrill-Palmer Quarterly (1982-)*, 255-281.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide*: Taylor & Francis.

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49-67). New York: Oxford University Press.

MacWhinney, B. (2008). A unified model. In N. C. Ellis & P. Robinson (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 341-371). Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (2012). The logic of the unified model. In S. M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition*. New York, NY: Routledge.

Manis, F. R., Doi, L. M., & Bhadha, B. (2000). Naming speed, phonological awareness, and orthographic knowledge in second graders. *Journal of Learning Disabilities, 33*(4), 325-333. doi: 10.1177/002221940003300405

Manis, F. R., Seidenberg, M. S., Doi, L. M., McBride-Chang, C., & Petersen, A. (1996). On the bases of two subtypes of development dyslexia. *Cognition, 58*(2), 157-195. doi: http://dx.doi.org/10.1016/0010-0277(95)00679-6

Mann, V. A. (1986). Phonological awareness: The role of reading experience. *Cognition, 24*(1–2), 65-92. doi: http://dx.doi.org/10.1016/0010-0277(86)90005-3

Marsh, G., Friedman, M., Desberg, P., & Saterdahl, K. (1981). Comparison of reading and spelling strategies in normal and reading disabled children. In M. Friedman, J. P. Das & N. O'Connor (Eds.), *Intelligence and Learning* (Vol. 14, pp. 363-367): Springer US.

Martin, K. I., & Juffs, A. (In revision). *L1 affects sensitivity to vowel information while reading in L2: Evidence from eye-tracking*. Pittsburgh, PA.

Martin, R. C. (1982). The pseudohomophone effect: The role of visual similarity in non-word decisions. *The Quarterly Journal of Experimental Psychology Section A, 34*(3), 395-409. doi: 10.1080/14640748208400851

Martinet, C., Valdois, S., & Fayol, M. (2004). Lexical orthographic knowledge develops from the beginning of literacy acquisition. *Cognition, 91*(2), B11-B22. doi: http://dx.doi.org/10.1016/j.cognition.2003.09.002

Massaro, D. W., Venezky, R. L., & Taylor, G. A. (1979). Orthographic regularity, positional frequency, and visual processing of letter strings. *Journal of Experimental Psychology: General, 108*(1), 107-124. doi: 10.1037/0096-3445.108.1.107

Matthews, P. H. (2007). *Oxford concise dictionary of linguistics* (Second ed.). Oxford: Oxford University Press.

McBride-Chang, C. (1995a). Phonological processing, speech perception, and reading disability: An integrative review. *Educational Psychologist, 30*(3), 109-121. doi: 10.1207/s15326985ep3003_2

McBride-Chang, C. (1995b). What is phonological awareness? *Journal of Educational Psychology, 87*(2), 179-192. doi: 10.1037/0022-0663.87.2.179

McBride-Chang, C. (1996). Models of speech perception and phonological processing in reading. *Child Development, 67*(4), 1836-1856. doi: 10.1111/j.1467-8624.1996.tb01831.x

McBride-Chang, C., Bialystok, E., Chong, K. K. Y., & Li, Y. (2004). Levels of phonological awareness in three cultures. *Journal of Experimental Child Psychology, 89*(2), 93-111. doi: http://dx.doi.org/10.1016/j.jecp.2004.05.001

McBride-Chang, C., Cho, J.-R., Liu, H., Wagner, R. K., Shu, H., Zhou, A., . . . Muse, A. (2005). Changing models across cultures: Associations of phonological awareness and morphological structure awareness with vocabulary and word recognition in second graders from Beijing, Hong Kong, Korea, and the United States. *Journal of Experimental Child Psychology, 92*(2), 140-160. doi: 10.1016/j.jecp.2005.03.009

McBride-Chang, C., & Ho, C. S.-H. (2000). Developmental issues in Chinese children's character acquisition. *Journal of Educational Psychology, 92*(1), 50-55. doi: 10.1037/0022-0663.92.1.50

McBride-Chang, C., & Ho, C. S.-H. (2005). Predictors of beginning reading in Chinese and English: A 2-year longitudinal study of Chinese kindergartners. *Scientific Studies of Reading, 9*(2), 117-144. doi: 10.1207/s1532799xssr0902_2

McBride-Chang, C., Tong, X., Shu, H., Wong, A. M. Y., Leung, K.-w., & Tardif, T. (2008). Syllable, phoneme, and tone: Psycholinguistic units in early Chinese and English word recognition. *Scientific Studies of Reading, 12*(2), 171-194. doi: 10.1080/10888430801917290

McBride–Chang, C., & Kail, R. V. (2002). Cross–cultural similarities in the predictors of reading acquisition. *Child Development, 73*(5), 1392-1407. doi: 10.1111/1467-8624.00479

McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of*

Experimental Psychology: Human Perception and Performance, 13(1), 14-24. doi: 10.1037/0096-1523.13.1.14

McCarthy, R., & Warrington, E. K. (1990). *Cognitive neuropsychology: A clinical introduction*. New York, NY: Academic Press.

McClure, K. K., Ferreira, F., & Bisanz, G. L. (1996). Effects of grade, syllable segmentation, and speed of presentation on children's word-blending ability. *Journal of Educational Psychology*, *88*(4), 670-681. doi: 10.1037/0022-0663.88.4.670

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. doi: http://dx.doi.org/10.1016/S0022-5371(82)90870-2

McKenna, M. C., & Kear, D. J. (1990). Measuring attitude toward reading: A new tool for teachers. *The Reading Teacher, 43*(9), 626-639. doi: 10.2307/20200500

Meisel, J. M., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition, 3*(02), 109-135. doi: doi:10.1017/S0272263100004137

Morais, J., Alegria, J., & Content, A. (1987). The relationships between segmental analysis and alphabetic literacy: An interactive view. *Current Psychology of Cognition*, *7*(5), 415-438.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, *24*(1–2), 45-64. doi: http://dx.doi.org/10.1016/0010-0277(86)90004-1

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*(4), 323-331. doi: http://dx.doi.org/10.1016/0010-0277(79)90020-9

Morais, J., Content, A., Bertelson, P., Cary, L., & Kolinsky, R. (1988). Is there a critical period for the acquisition of segmental analysis? *Cognitive Neuropsychology, 5*(3), 347-352. doi: 10.1080/02643298808252940

Muneaux, M., & Ziegler, J. C. (2004). Locus of orthographic effects in spoken word recognition: Novel insights from the neighbour generation task. *Language and Cognitive Processes, 19*(5), 641-660. doi: 10.1080/01690960444000052

Nagy, W., Berninger, V. W., Abbott, R., Vaughan, K., & Vermeulen, K. (2003). Relationship of morphology and other language skills to literacy skills in at-risk second-grade readers and at-risk fourth-grade writers. *Journal of Educational Psychology, 95*(4), 730-742. doi: 10.1037/0022-0663.95.4.730

Näslund, J. C., & Schneider, W. (1996). Kindergarten letter knowledge, phonological skills, and memory processes: Relative effects on early literacy. *Journal of Experimental Child Psychology*, *62*(1), 30-59. doi: http://dx.doi.org/10.1006/jecp.1996.0021

Nassaji, H., & Geva, E. (1999). The contribution of phonological and orthographic processing skills to adult ESL reading: Evidence from native speakers of Farsi. *Applied Psycholinguistics*, *20*(02), 241-267. doi: doi:null

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The language teacher, 31*(7), 9-13.

Nation, K., & Hulme, C. (1997). Phonemic Segmentation, Not Onset-Rime Segmentation, Predicts Early Reading and Spelling Skills. *Reading Research Quarterly, 32*(2), 154-167. doi: 10.1598/RRQ.32.2.2

Neufeld, S., & Billuroğlu, A. (2005). In search of the critical lexical mass: How 'general' is the GSL? How 'academic' is the AWL?

Nilsen, D. L. F., & Nilsen, A. P. (2010). *Pronunciation contrasts in English* (2nd ed. ed.). Long Grove, IL: Waveland Press, Inc.

Nishanimut, S. P., Johnston, R. S., Joshi, R. M., Thomas, P. J., & Padakannaya, P. (2013). Effect of synthetic phonics instruction on literacy skills in an ESL setting. *Learning and Individual Differences, 27*(0), 47-53. doi: http://dx.doi.org/10.1016/j.lindif.2013.06.007

NRP. (2000). *Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*: National Institute of Child Health and Human Development, National Institutes of Health.

Olson, R. K., Davidson, B. J., Kliegl, R., & Foltz, G. (1985). Individual and developmental differences in reading disability. . In G. E. MacKinnon & T. G. Waller (Eds.), *Reading research: Advances in theory and practice* (Vol. Vol. 4, pp. 1-64). New York: Academic Press.

Olson, R. K., Forsberg, H., & Wise, B. (1994). Genes, environment, and the development of orthographic skills. In V. W. Berninger (Ed.), *The varieties of orthographic knowledge I: Theoretical and developmental issues* (pp. 27-71). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Paris, S. G., Pearson, P. D., Cervetti, G., Carpenter, R., Paris, A. H., DeGroot, J., . . . Papanastasiou, E. (2004). Assessing the effectiveness of summer reading programs. *Summer learning: Research, policies, and programs*, 121-161.

Parkin, A. (1982). Phonological recoding in lexical decision: Effects of spelling-to-sound regularity depend on how regularity is defined. *Memory & Cognition, 10*(1), 43-53. doi: 10.3758/bf03197624

Perea, M., & Carreiras, M. (2006a). Do transposed-letter effects occur across lexeme boundaries? *Psychonomic Bulletin & Review, 13*(3), 418-422. doi: 10.3758/BF03193863

Perea, M., & Carreiras, M. (2006b). Do transposed-letter similarity effects occur at a prelexical phonological level? *The Quarterly Journal of Experimental Psychology, 59*(9), 1600-1613. doi: 10.1080/17470210500298880

Perea, M., & Carreiras, M. (2006c). Do transposed-letter similarity effects occur at a syllable level? *Experimental Psychology, 53*(4), 308-315. doi: 10.1027/1618-3169.53.4.308

Perea, M., & Lupker, S. J. (2003). Transposed-letter confusability effects in masked form priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: The state of the art* (pp. 97-120). New York, NY: Psychology Press.

Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language, 51*(2), 231-246. doi: http://dx.doi.org/10.1016/j.jml.2004.05.005

Perea, M., Mallouh, R., & Carreiras, M. (2010). The search for an input-coding scheme: Transposed-letter priming in Arabic. *Psychonomic Bulletin & Review, 17*(3), 375-380. doi: 10.3758/PBR.17.3.375

Perea, M., & Pérez, E. (2008). Beyond alphabetic orthographies: The role of form and phonology in transposition effects in Katakana. *Language and Cognitive Processes, 24*(1), 67-88. doi: 10.1080/01690960802053924

Perfetti, C. A. (1988). Verbal efficiency in reading ability. In G. E. MacKinnon, T. G. Waller & M. Daneman (Eds.), *Reading research: Advances in theory and practice, Vol. 6* (pp. 109-143). New York: Academic Press, Inc.

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri & R. Treiman (Eds.), *Reading acquisition* (pp. 145-174). Hillsdale, NJ: Erlbaum.
Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357-383. doi: 10.1080/10888430701530730
Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33*(3), 283-319.
Perfetti, C. A., & Bell, L. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and priming. *Journal of Memory and Language, 30*(4), 473-485. doi: http://dx.doi.org/10.1016/0749-596X(91)90017-E
Perfetti, C. A., & Dunlap, S. (2008). Learning to read: General principles and writing system variations. In K. Koda & A. M. Zehler (Eds.), *Learning to read across languages: Cross-linguistic relationships in first- and second-language literacy development* (pp. 13-38). New York, NY: Routledge.
Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67-86). Washington, DC: American Psychological Association.
Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Philadelphia, PA: John Benjamins.
Perfetti, C. A., & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 141-183). Hillsdale, NJ.
Perfetti, C. A., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., & Tan, L. H. (2007). Reading in two writing systems: Accommodation and assimilation of the brain's reading network. *Bilingualism: Language and Cognition, 10*(02), 131-146. doi: doi:10.1017/S1366728907002891
Perfetti, C. A., & McCutchen, D. (1982). Speech processes in reading. In N. Lass (Ed.), *Advances in speech and language* (Vol. Vol. 7). New York, NY: Academic Press.
Perfetti, C. A., & Stafura, J. (2013). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*(1), 22-37. doi: 10.1080/10888438.2013.827687
Perfetti, C. A., & Zhang, S. (1995). Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 24-33. doi: 10.1037/0278-7393.21.1.24
Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a "universal" phonological principle. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 227-248). Amsterdam: North-Holland.
Perin, D. (1983). Phonemic segmentation and spelling. *British Journal of Psychology, 74*(1), 129-144. doi: 10.1111/j.2044-8295.1983.tb01849.x
Perre, L., Pattamadilok, C., Montant, M., & Ziegler, J. C. (2009). Orthographic effects in spoken language: On-line activation or phonological restructuring? *Brain Research, 1275*(0), 73-80. doi: http://dx.doi.org/10.1016/j.brainres.2009.04.018
Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher, 58*(6), 510-519. doi: 10.1598/rt.58.6.2
Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition, 13*(02), 253-262. doi: 10.1017/S1366728909990526

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A reveiw. *Reading Research Quarterly, 27*(1), 28-53.
Randall, M., & Meara, P. (1988). How Arabs read Roman letters. *Reading in a Foreign Language, 4*(2), 133-145.
Rastle, K., & Coltheart, M. (1999). Lexical and nonlexical phonological priming in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 25*(2), 461-481. doi: http://dx.doi.org/10.1037/0096-1523.25.2.461
Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, *2*(2), 31-74. doi: 10.1111/1529-1006.00004
Rayner, K., Sereno, S. C., Lesch, M. F., & Pollatsek, A. (1995). Phonological codes are automatically activated during reading: Evidence from an eye movement priming paradigm. *Psychological Science, 6*(1), 26-32. doi: 10.1111/j.1467-9280.1995.tb00300.x
Read, C., Zhang, Y.-F., Nie, H.-Y., & Ding, B.-Q. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition*, *24*(1–2), 31-44. doi: http://dx.doi.org/10.1016/0010-0277(86)90003-X
Reynolds, M., & Besner, D. (2005). Basic processes in reading: A critical review of pseudohomophone effects in reading aloud and a new computational account. *Psychonomic Bulletin & Review, 12*(4), 622-646. doi: 10.3758/BF03196752
Richards, J. C., Hull, J., & Proctor, S. (2004a). *Interchange student's book 1 with audio CD*. Cambridge: Cambridge University Press.
Richards, J. C., Hull, J., & Proctor, S. (2004b). *Interchange student's book 2 with audio CD*. Cambridge: Cambridge University Press
Richards, J. C., Hull, J., & Proctor, S. (2004c). *Interchange student's book 3 with audio CD*. Cambridge: Cambridge University Press.
Roelofs, A. (2006). The influence of spelling on phonological encoding in word reading, object naming, and word generation. *Psychonomic Bulletin & Review, 13*(1), 33-37. doi: 10.3758/BF03193809
Rogerson, H. D., Davis, B., Hershelman, S. T., & Jasnow, C. (1992). *Words for students of English: A vocabulary series for ESL, Vol. 1*. Ann Arbor, MI: University of Michigan Press.
Rogerson, H. D., Esarey, G., Hershelman, S. T., Jasnow, C., Moltz, C., Schmandt, L. M., & Smith, D. A. (1992). *Words for students of English: A vocabulary series for ESL, Vol. 4*. Ann Arbor, MI: University of Michigan Press.
Rogerson, H. D., Esarey, G., Hershelman, S. T., Jasnow, C., Schmandt, L. M., Smith, D. A., & Snellings, C. M. (1992). *Words for students of English: A vocabulary series for ESL, Vol. 6*. Ann Arbor, MI: University of Michigan Press.
Rogerson, H. D., Esarey, G., Schmandt, L. M., & Smith, D. A. (1992). *Words for students of English: A vocabulary series for ESL, Vol. 2*. Ann Arbor, MI: University of Michigan Press.
Rogerson, H. D., Hershelman, S. T., & Jasnow, C. (1992). *Words for students of English: A vocabulary series for ESL, Vol. 5*. Ann Arbor, MI: University of Michigan Press.
Rogerson, H. D., Hershelman, S. T., Jasnow, C., & Moltz, C. (1992). *Words for students of English: A vocabulary series for ESL, Vol. 3*. Ann Arbor, MI: University of Michigan Press.

Rohl, M., & Tunmer, W. E. (1988). Phonemic segmentation skill and spelling acquisition. *Applied Psycholinguistics, 9*(04), 335-350. doi: doi:10.1017/S0142716400008043

Rosner, J., & Simon, D. P. (1971). The Auditory Analysis Test: An initial report. Learning Research and Development Center, University of Pittsburgh.

Ryan, A., & Meara, P. (1991). The case of the invisible vowels: Arabic speakers reading English words. *Reading in a Foreign Language, 7*(2), 531-540.

Ryan, A., & Meara, P. (1996). *A diagnostic test for vowel blindness in Arabic speaking learners of English*. The _lognostics Virtual Library, Retrieved from The _lognostics Virtual Library database. Swansea University.

Saiegh-Haddad, E. (2003). Bilingual oral reading fluency and reading comprehension: The case of Arabic/Hebrew (L1) - English (L2) readers. *Reading and Writing, 16*(8), 717-736. doi: 10.1023/a:1027310220036

Saigh, K., & Schmitt, N. (2012). Difficulties with vocabulary word form: The case of Arabic ESL learners. *System, 40*(1), 24-36. doi: 10.1016/j.system.2012.01.005

Sampson, G. (2015). *Writing systems* (2nd ed.). Bristol, CT: Equinox

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3*(1), 1-17. doi: 10.1037/0096-1523.3.1.1

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. *Specific reading disability: A view of the spectrum*, 75-119.

Schatschneider, C., Francis, D. J., Foorman, B. R., Fletcher, J. M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology, 91*(3), 439-449. doi: http://dx.doi.org/10.1037/0022-0663.91.3.439

Schneider, W., Küspert, P., Roth, E., Visé, M., & Marx, H. (1997). Short- and Long-Term Effects of Training Phonological Awareness in Kindergarten: Evidence from Two German Studies. *Journal of Experimental Child Psychology, 66*(3), 311-340. doi: http://dx.doi.org/10.1006/jecp.1997.2384

Schneider, W., Schman, A., & Zuccolutto, A. (2002). E-Prime reference guide. Pittsburgh, PA: Psychology Software Tools, Inc.

Scholes, R. (1998). The case against phonemic awareness. *Journal of Research in Reading, 21*(3), 177-188. doi: 10.1111/1467-9817.00054

Schoonbaert, S., & Grainger, J. (2004). Letter position coding in printed word perception: Effects of repeated and transposed letters. *Language and Cognitive Processes, 19*(3), 333-367. doi: 10.1080/01690960344000198

Seidenberg, M. S. (1992). Beyond orthographic depth in reading: Equitable division of labor. In F. Ram & K. Leonard (Eds.), *Advances in Psychology* (Vol. Volume 94, pp. 85-118): North-Holland.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523-568. doi: 10.1037/0033-295X.96.4.523

Seidenberg, M. S., Petersen, A., MacDonald, M. C., & Plaut, D. C. (1996). Pseudohomophone effects and models of word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 48-62. doi: 10.1037/0278-7393.22.1.48

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 5*(6), 546-554. doi: 10.1037/0278-7393.5.6.546

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

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*(2), 143-174. doi: 10.1348/000712603321661859

Share, D. L. (1995). Phonological recoding and self-teaching: *Sine qua non* of reading acquisition. *Cognition, 55*(2), 151-218. doi: http://dx.doi.org/10.1016/0010-0277(94)00645-2

Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology, 87*(4), 267-298. doi: http://dx.doi.org/10.1016/j.jecp.2004.01.001

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin, 134*(4), 584-615. doi: 10.1037/0033-2909.134.4.584

Share, D. L., Jorm, A. F., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology, 76*(6), 1309-1324. doi: 10.1037/0022-0663.76.6.1309

Share, D. L., & Levin, I. (1999). Learning to read and write in Hebrew. In M. Harris & G. Hatano (Eds.), *Learning to read and write: A cross-linguistic perspective* (pp. 89-111).

Shatil, E., Share, D. L., & Levin, I. (2000). On the contribution of kindergarten writing to grade 1 literacy: A longitudinal study in Hebrew. *Applied Psycholinguistics, 21*(01), 1-21. doi: doi:null

Siegel, L. S., & Ryan, E. B. (1988). Development of grammatical-sensitivity, phonological, and short-term memory skills in normally achieving and learning disabled children. *Developmental Psychology, 24*(1), 28-37. doi: 10.1037/0012-1649.24.1.28

Siegel, L. S., Share, D., & Geva, E. (1995). Evidence for superior orthographic skills in dyslexics. *Psychological Science, 6*(4), 250-254. doi: 10.1111/j.1467-9280.1995.tb00601.x

Siok, W. T., & Fletcher, P. (2001). The role of phonological awareness and visual-orthographic skills in Chinese reading acquisition. *Developmental Psychology, 37*(6), 886-899. doi: 10.1037/0012-1649.37.6.886

Slowiaczek, L. M., Soltano, E. G., Wieting, S. J., & Bishop, K. L. (2003). An investigation of phonology and orthography in spoken-word recognition. *The Quarterly Journal of Experimental Psychology Section A, 56*(2), 233-262. doi: 10.1080/02724980244000323

Snowling, M. J. (1980). The development of grapheme-phoneme correspondence in normal and dyslexic readers. *Journal of Experimental Child Psychology, 29*(2), 294-305. doi: http://dx.doi.org/10.1016/0022-0965(80)90021-1

Snowling, M. J. (2000). *Dyslexia* (Second ed.). Malden: Blackwell Publishing.

So, D., & Siegel, L. S. (1997). Learning to read Chinese: Semantic, syntactic, phonological and working memory skills in normally achieving and poor Chinese readers. *Reading and Writing, 9*(1), 1-21. doi: 10.1023/A:1007963513853

Solman, R. T. (1988). An error analysis of the word-superiority effect. *British Journal of Psychology, 79*(1), 47-62. doi: 10.1111/j.2044-8295.1988.tb02272.x

Stackhouse, J., & Wells, B. (1997). How do speech and language problems affect literacy development. *Dyslexia: Biology, cognition and intervention*, 182-211.

Stage, S. A., & Wagner, R. K. (1992). Development of young children's phonological and orthographic knowledge as revealed by their spellings. *Developmental Psychology, 28*(2), 287-296. doi: 10.1037/0012-1649.28.2.287

Stahl, S. A., Duffy-Hester, A. M., & Stahl, K. A. D. (1998). Theory and research into practice: Everything you wanted to know about phonics (but were afraid to ask). *Reading Research Quarterly, 33*(3), 338-355. doi: 10.2307/748309

Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology, 86*(2), 221-234. doi: 10.1037/0022-0663.86.2.221

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360-407.

Stanovich, K. E. (1992). Speculations on the causes and consequences of individual differences in early reading acquisition. In P. B. Gough, L. C. Ehri & R. Treiman (Eds.), *Reading acquisition* (pp. 307-342). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*(1), 24-53. doi: 10.1037/0022-0663.86.1.24

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24*(4), 402-433. doi: 10.2307/747605

Stanovich, K. E., West, R. F., & Cunningham, A. E. (1991). Beyond phonological processes: Print-exposure and orthographic processing. In S. A. Brady & D. Shankweiler (Eds.), *Phonological processes in literacy* (pp. 219-235). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1987). Most vocabuary is learned from context. In M. McKeown & M. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89-105). Hillsdale, NJ: Erlbaum.

Stuart, M., & Coltheart, M. (1988). Does reading develop in a sequence of stages? *Cognition, 30*(2), 139-181. doi: http://dx.doi.org/10.1016/0010-0277(88)90038-8

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*(2), 147-166. doi: http://dx.doi.org/10.1016/S0010-0277(00)00081-0

Szeszulski, P. A., & Manis, F. R. (1987). A comparison of word recognition processes in dyslexic and normal readers at two reading-age levels. *Journal of Experimental Child Psychology, 44*(3), 364-376. doi: http://dx.doi.org/10.1016/0022-0965(87)90040-3

Taft, M., & Chen, H. C. (1992). Judging homophony in Chinese: The influence of tones. In H. C. Chen & O. J. L. Tzeng (Eds.), *Language processing in Chinese*: Elsevier.

Taft, M., & Hambly, G. (1985). The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language, 24*(3), 320-335. doi: http://dx.doi.org/10.1016/0749-596X(85)90031-2

Taft, M., & Russell, B. (1992). Pseudohomophone naming and the word frequency effect. *The Quarterly Journal of Experimental Psychology Section A, 45*(1), 51-71. doi: 10.1080/14640749208401315

Tamaoka, K., & Terao, Y. (2004). Mora or syllable? Which unit do Japanese use in naming visually presented stimuli? *Applied Psycholinguistics, 25*(01), 1-27. doi: doi:10.1017/S0142716404001018

Tan, L.-H., & Perfetti, C. A. (1998). Phonological codes as early sources of constraint in Chinese word identification: A review of current discoveries and theoretical accounts. In C. Leong & K. Tamaoka (Eds.), *Cognitive Processing of the Chinese and the Japanese Languages* (Vol. 14, pp. 11-46): Springer Netherlands.

Tanenhaus, M., Flanigan, H., & Seidenberg, M. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory & Cognition, 8*(6), 513-520. doi: 10.3758/BF03213770

Taylor, I. (1981). Writing systems and reading *Reading research: Advances in theory and practice* (Vol. Vol. 2, pp. 1-51). New York, NY: Academic Press.

Taylor, I., & Taylor, M. M. (1995). *Writing and literacy in Chinese, Korean, and Japanese*. Philadelphia, PA: John Benjamins.

Taylor, M. (2008). *Orthographic and phonological awareness among L1 Arabic ESL learners: A quasi-experimental study.* (Unpublished Doctoral dissertation), University of Phoenix.

Templeton, S., & Bear, D. R. E. (1992). *Development of orthographic knowledge and the foundations of literacy: A memorial festschrift for Edumnd H. Henderson* (S. Templeton & D. R. Bear Eds.). Hillsdale, NJ: Erlbaum.

Tokowicz, N., Michael, E. B., & Kroll, J. F. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition, 7*(03), 255-272. doi: 10.1017/S1366728904001634

Tolchinsky, L., Levin, I., Aram, D., & McBride-Chang, C. (2012). Building literacy in alphabetic, abjad and morphosyllabic systems. *Reading and Writing, 25*(7), 1573-1598. doi: 10.1007/s11145-011-9334-7

Tong, X., McBride-Chang, C., Shu, H., & Wong, A. M. Y. (2009). Morphological awareness, orthographic knowledge, and spelling errors: Keys to understanding early Chinese literacy acquisition. *Scientific Studies of Reading, 13*(5), 426-452. doi: 10.1080/10888430903162910

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40*(1), 7-26. doi: http://dx.doi.org/10.1016/S0022-4405(01)00092-9

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency: TOWRE*: Pro-Ed.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second-to fifth-grade children. *Scientific Studies of Reading, 1*(2), 161-185. doi: 10.1207/s1532799xssr0102_4

Trabasso, T., Sabatini, J., Massaro, D. W., & Calfee, R. C. E. (2005). *From orthogrpahy to pedagogy: Essays in honor of Richard L. Venezky* (T. Trabasso, J. Sabatini, D. W. Massaro & R. C. Calfee Eds.). Mahwah, NJ: Lawrence Erlbaum, Associates.

Tranel, B. (1987). *The sounds of French: An introduction*. New York: Cambridge University Press.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition, 15*(1–3), 49-74. doi: http://dx.doi.org/10.1016/0010-0277(83)90033-1

Treiman, R. (1984). Individual differences among children in spelling and reading styles. *Journal of Experimental Child Psychology, 37*(3), 463-477. doi: http://dx.doi.org/10.1016/0022-0965(84)90071-7

Treiman, R. (1985). Onsets and rimes as units of spoken syllables: Evidence from children. *Journal of Experimental Child Psychology, 39*(1), 161-181. doi: http://dx.doi.org/10.1016/0022-0965(85)90034-7

Treiman, R., & Baron, J. (1981). Segmental analysis ability: Development and relation to reading ability. *Reading research: Advances in theory and practice, 3*, 159-198.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General, 124*(2), 107-136. doi: 10.1037/0096-3445.124.2.107

Treiman, R., & Zukowski, A. (1990). Toward an understanding of English syllabification. *Journal of Memory and Language, 29*(1), 66-85. doi: http://dx.doi.org/10.1016/0749-596X(90)90010-W

Treiman, R., & Zukowski, A. (1991). Levels of phonological awareness. In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy* (pp. 67-84): Routledge.

Tunmer, W. E., Herriman, M. L., & Nesdale, A. R. (1988). Metalinguistic abilities and beginning reading. *Reading Research Quarterly, 23*(2), 134-158.

Tunmer, W. E., & Hoover, W. A. (1993). Phonological recoding skill and beginning reading. *Reading and Writing, 5*(2), 161-179. doi: 10.1007/bf01027482

Tunmer, W. E., & Nesdale, A. R. (1985). Phonemic segmentation skill and beginning reading. *Journal of Educational Psychology, 77*(4), 417-427. doi: 10.1037/0022-0663.77.4.417

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127-154.

Tzeng, O. J., Hung, D. L., & Wang, W. S. Y. (1977). Speech recoding in reading Chinese characters. *Journal of Experimental Psychology: Human Learning and Memory, 3*(6), 621-630. doi: 10.1037/0278-7393.3.6.621

Ussishkin, A. (2006). Semitic morphology: Root-based or word-based? *Morphology, 16*(1), 37-40. doi: 10.1007/s11525-006-0002-6

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition, 15*(3), 181-198. doi: 10.3758/bf03197716

van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97*(4), 488-522. doi: 10.1037/0033-295X.97.4.488

Vance, T. (1987). *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.

Velan, H., & Frost, R. (2007). Cambridge University versus Hebrew University: The impact of letter transposition on reading English and Hebrew. *Psychonomic Bulletin & Review, 14*(5), 913-918. doi: 10.3758/BF03194121

Velan, H., & Frost, R. (2009). Letter-transposition effects are not universal: The impact of transposing letters in Hebrew. *Journal of Memory and Language, 61*(3), 285-302. doi: http://dx.doi.org/10.1016/j.jml.2009.05.003

Vellutino, F., Scanlon, D., & Tanzman, M. (1994). Components of reading ability: Issues and problems in operationalizing word identification, phonological coding, and orthographic coding. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 279-332). Baltimore, MD: Paul Bookes Publishing.

Venezky, R. L. (1979). Orthographic Regularities in English Words. In P. Kolers, M. Wrolstad & H. Bouma (Eds.), *Processing of Visible Language* (Vol. 13, pp. 283-293): Springer US.

Venezky, R. L., & Massaro, D. W. (1979). The role of orthographic regularity in word recognition. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and practice of early reading* (Vol. Vol. 1, pp. 85-108). Hillsdale, NJ: Lawrence Erlbaum.

Ventura, P., Kolinsky, R., Pattamadilok, C., & Morais, J. (2008). The developmental turnpoint of orthographic consistency effects in speech recognition. *Journal of Experimental Child Psychology, 100*(2), 135-145. doi: http://dx.doi.org/10.1016/j.jecp.2008.01.003

Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes, 19*(1), 57-95. doi: 10.1080/01690960344000134

Verhoeven, L. (1999). Second language reading. In D. A. Wagner, R. L. Venezky & B. V. Street (Eds.), *Literacy: An international handbook* (pp. 143-147). Boulder, Colorado: Westview Press.

Wade-Woolley, L. (1999). First language influences on second language word reading: All roads lead to Rome. *Language Learning, 49*(3), 447-471. doi: 10.1111/0023-8333.00096

Wagner, R. K., & Barker, T. A. (1994). The development of orthographic processing ability. In V. Berninger (Ed.), *The Varieties of Orthographic Knowledge* (Vol. 8, pp. 243-276): Springer Netherlands.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*(2), 192-212. doi: 10.1037/0033-2909.101.2.192

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology, 30*(1), 73-87. doi: 10.1037/0012-1649.30.1.73

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing: CTOPP*: Pro-Ed.

Wallach, M. A., & Wallach, L. (1976). Teaching all children to read.

Walton, P. D., Bowden, M., Kurtz, S., & Angus, M. (2001). Evaluation of a rime-based reading program with Shuswap and Heiltsuk First Nations prereaders. *Reading and Writing, 14*(3-4), 229-264. doi: 10.1023/A:1011125315377

Walton, P. D., & Walton, L. M. (2002). Beginning reading by teaching in rime analogy: Effects on phonological skills, letter-sound knowledge, working memory, and word-reading strategies. *Scientific Studies of Reading, 6*(1), 79-115. doi: 10.1207/S1532799XSSR0601_04

Walton, P. D., Walton, L. M., & Felton, K. (2001). Teaching rime analogy or letter recoding reading strategies to prereaders: Effects on prereading skills and word reading. *Journal of Educational Psychology, 93*(1), 160-180. doi: 10.1037/0022-0663.93.1.160

Wang, M., & Geva, E. (2003). Spelling performance of Chinese children using English as a second language: Lexical and visual–orthographic processes. *Applied Psycholinguistics, 24*(01), 1-25. doi: 10.1017/S0142716403000018

Wang, M., & Koda, K. (2005). Commonalities and differences in word identification skills among learners of English as a second language. *Language Learning, 55*(1), 71-98. doi: 10.1111/j.0023-8333.2005.00290.x

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*(2), 129-149. doi: 10.1016/s0010-0277(02)00232-9

Wang, M., & Yang, C.-L. (2008). Learning to read Chinese: Cognitive consequences of cross-language and writing system differences. In K. Koda & A. M. Zehler (Eds.), *Learning to read across languages: Cross-linguistic relationships in first- and second-language literacy development*. New York, NY: Routledge.

Waters, G. S., Bruck, M., & Seidenberg, M. (1985). Do children use similar processes to read and spell words? *Journal of Experimental Child Psychology, 39*(3), 511-530. doi: http://dx.doi.org/10.1016/0022-0965(85)90054-2

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology Section A, 50*(2), 439-456. doi: 10.1080/713755710

West, M. (1953). *A general service list of English words*. London, UK: Longman.

Williams, J. P. (1980). Teaching decoding with an emphasis on phoneme analysis and phoneme blending. *Journal of Educational Psychology, 72*(1), 1-15. doi: 10.1037/0022-0663.72.1.1

Wimmer, H., Landerl, K., & Schneider, W. (1994). The role of rhyme awareness in learning to read a regular orthography. *British Journal of Developmental Psychology, 12*(4), 469-484. doi: 10.1111/j.2044-835X.1994.tb00648.x

Yates, M., Locker, L., & Simpson, G. (2003). Semantic and phonological influences on the processing of words and pseudohomophones. *Memory & Cognition, 31*(6), 856-866. doi: 10.3758/BF03196440

Yoon, H.-K., Bolger, D. J., Kwon, O.-S., & Perfetti, C. A. (2002). Subsyllabic units in reading: A difference between Korean and English. In L. Verhoeven, C. Elbro & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 139-164). Philadelphia, PA: John Benjamins.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23*(2), 159-177. doi: 10.2307/747800

Zhang, J., & McBride-Chang, C. (2010). Auditory sensitivity, speech perception, and reading development and impairment. *Educational Psychology Review, 22*(3), 323-338. doi: 10.1007/s10648-010-9137-4

Zhang, Y., Tardif, T., Shu, H., Li, H., Liu, H., McBride-Chang, C., . . . Zhang, Z. (2013). Phonological skills and vocabulary knowledge mediate socioeconomic status effects in predicting reading outcomes for Chinese children. *Developmental Psychology, 49*(4), 665-671. doi: 10.1037/a0028612

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., . . . Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science, 21*(4), 551-559. doi: 10.1177/0956797610363406

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review, 5*(4), 683-689. doi: 10.3758/BF03208845

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

Ziegler, J. C., & Goswami, U. (2006). Becoming literate in different languages: similar problems, different solutions. *Developmental Science, 9*(5), 429-436. doi: 10.1111/j.1467-7687.2006.00509.x

Ziegler, J. C., & Muneaux, M. (2007). Orthographic facilitation and phonological inhibition in spoken word recognition: A developmental study. *Psychonomic Bulletin & Review, 14*(1), 75-80. doi: 10.3758/BF03194031

Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language, 48*(4), 779-793. doi: http://dx.doi.org/10.1016/S0749-596X(03)00006-8

Ziegler, J. C., Stone, G., & Jacobs, A. (1997). What is the pronunciation for -ough and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, & Computers, 29*(4), 600-618. doi: 10.3758/bf03210615

Ziegler, J. C., Tan, L. H., Perry, C., & Montant, M. (2000). Phonology matters: The phonological frequency effect in written Chinese. *Psychological Science, 11*(3), 234-238. doi: 10.1111/1467-9280.00247

Ziegler, J. C., van Orden, G. C., & Jacobs, A. M. (1997). Phonology can help or hurt the perception of print. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 845-860. doi: 10.1037/0096-1523.23.3.845