

ABSTRACT

Title of Dissertation: STATISTICAL LEARNING ACROSS
MODALITIES, DOMAINS AND
LANGUAGES

Jinglei Ren, Doctor of Philosophy, 2024

Dissertation directed by: Professor, Min Wang, Department of Human
Development and Quantitate Methodology

This dissertation establishes a comprehensive line of research that systematically explores statistical learning (SL) across modalities, domains, and languages. The first study delves into SL development, exploring SL changes across different ages, linguistic and non-linguistic domains, visual and auditory modalities, and languages. In the second study, the focus is shifted to the probabilistic regularities embedded in a specific written language. This study investigated the ability to use word endings as a probabilistic cue to lexical stress among those who acquired a second language (L2) in formal learning settings through years of practice. The third study broadens the scope further, delving into whether beginning English learners benefit from a specialized training program emphasizing the association between word endings and lexical stress. The collective body of research in this dissertation makes significant contributions to the fields of cognitive science, linguistics, and education. Ultimately, the insights gleaned have the capacity to positively impact learners of various ages and diverse backgrounds, and to offer valuable implications for educational practices.

STATISTICAL LEARNING ACROSS MODALITIES, DOMAINS AND
LANGUAGES

by

Jinglei Ren

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2024

Advisory Committee:

Professor Min Wang, Chair

Associate Professor Donald J. Bolger

Associate Professor Yi Ting Huang (Dean's Representative)

Professor Jeffery Lidz

© Copyright by
Jinglei Ren
2024

Table of Contents

| | |
|--|-----------|
| TABLE OF CONTENTS | II |
| CHAPTER 1: STATISTICAL LEARNING ACROSS MODALITIES, DOMAINS AND LANGUAGES | 1 |
| CHAPTER 2: DEVELOPMENT OF STATISTICAL LEARNING ABILITY ACROSS MODALITIES, DOMAINS AND LANGUAGES | 13 |
| CHAPTER 3: SENSITIVITY TO WORD ENDINGS AS PROBABILISTIC ORTHOGRAPHIC CUES TO LEXICAL STRESS AMONG ENGLISH AS SECOND LANGUAGE LEARNERS | 49 |
| CHAPTER 4: STATISTICAL LEARNING IN READING: CAN EXPLICIT INSTRUCTION HELP CHILDREN LEARN LEXICAL STRESS IN A SECOND LANGUAGE? | 92 |

Chapter 1: Statistical Learning Across Modalities, Domains and Languages

The environment surrounding us is organized and structured. Instead of being randomly placed, items in our world tend to inhabit particular, predictable positions (e.g., books are typically located on shelves). Similarly, the sounds we hear, such as music or birdsong, often display repetitive patterns (Kalcounis-Rueppel et al., 2006; Lipkind & Tchernichovski, 2011). Moreover, human speech adheres to a structured framework, where specific sounds tend to occur more frequently in certain word positions, such as the more common co-occurrence of “s” and “l” at the beginning of English words compared to the pairing of “t” and “l”. This process of identifying and extracting the patterns from our surroundings is known as statistical learning (SL)."

In this dissertation proposal, SL is defined as the ability to use statistics (e.g., transitional probabilities or frequencies) to discover boundaries of sequentially presented stimuli. SL is present from infancy and is robustly observed across different modalities (e.g., visual, auditory or tactile, Saffran et al., 1997, 1999), domains (e.g., nonlinguistic or linguistic, Raviv & Arnon, 2018; Shufaniya & Arnon, 2018), and languages (e.g., English, Chinese, and Hebrew, Treiman & Kessler, 2022; Tong et al., 2020; Frost et al., 2013). There have been debates on whether SL is age invariant (e.g., Saffran et al., 1997) or can be improved with age (e.g., Arciuli & Simpson, 2011). Evidence thus far suggests that the development of SL is largely dependent on modalities or domains (e.g., Raviv and Arnon, 2018, Shufaniya and Arnon, 2018). However, very limited research took a systematic approach to investigate how the development of SL differs across modalities, domains, and languages simultaneously. Furthermore, SL exists widely across languages and writing systems (e.g., Treiman & Kessler, 2022) and is important for first language (L1) learners acquiring their native language. Nevertheless, it remains a question whether second language (L2) learners are able to extract statistical regularities from languages and writing systems in a newly acquired language. Finally, although SL is generally considered an automatic process that occurs implicitly through exposure to our surroundings, emerging evidence suggests that explicit instruction may play a role in enhancing SL (e.g., Hamrick & Rebuschat; 2012; Kachergis et al.; 2010). Studies are limited and the findings seem to be mixed concerning whether explicit instruction can facilitate the

process of searching for, detecting, or extracting underlying regularities.

The current dissertation sought to address the previously identified research gaps by delineating three core research objectives. The three specific aims are as follows: 1) to investigate the developmental trajectory of SL across modalities (i.e., visual and auditory), domains (i.e., linguistic and nonlinguistic), and languages (Chinese and English) simultaneously; 2) to study whether L2 learners are sensitive to specific statistical regularities embedded within the English writing system; 3) to address whether explicit instruction can enhance the performance of L2 children in learning statistical regularities.

The origin of SL

A pioneering study by Saffran and colleagues in 1996 unveiled that even infants as young as eight months possess the ability to engage in SL. In this research, infants were exposed to a continuous sequence of repetitive three-syllable nonsense words. After this exposure, the infants exhibited the capacity to distinguish between the three-syllable sequences and foil sequences, which were comprised of the same syllables rearranged in a different order. This discovery provides compelling evidence that humans, from a very early age, can detect the statistical properties of the input to identify word boundaries in connected speech. This finding is groundbreaking, reshaping our understanding of how language is acquired (Bates & Elman, 1996, Seidenberg, 1997).

Since this groundbreaking investigation, subsequent studies have yielded corroborative evidence of SL abilities in older children and adults (e.g., Fiser & Aslin, 2002, Saffran et al., 1999, Turk-Browne et al., 2005). In the conventional auditory SL experiments conducted with older children and adults, participants are exposed to a series of repetitive three-syllable nonsense words, resembling the original study conducted with infants by Saffran and the team. Following this exposure, learning is evaluated through a 2-alternative forced-choice (2-AFC) recognition test. During this test, participants are presented with pairs of stimuli consisting of a nonsense word from the initial exposure sequence and a nonword foil created by rearranging syllables in a novel order. Participants are then required to determine which stimulus sounds more familiar based on their earlier exposure in the familiarization phase. The presence of SL is inferred if performance on this recognition measure exceeds the chance level.

SL across modalities and domains

Early on, there was much interest in whether SL is exclusive to language or if it is also applied to nonlinguistic stimuli. Saffran et al. (1999) replicated their previous study to investigate infants' ability to detect statistical regularities in acoustic tone stimuli. Infants were able to identify statistical patterns in sequences of non-linguistic tones, suggesting that statistical structures can be extracted from auditory input regardless of whether it is speech syllables or non-linguistic acoustic information.

Note that Saffran and colleagues' studies (1996, 1999) focused on auditory input, but subsequent research has shown that SL can occur across different modalities. For instance, Kirkham and colleagues (2002) conducted a study using visual stimuli with infants of different ages (2, 5, and 8 months), in which the stimuli were organized in a statistically predictable pattern. Infants were able to detect changes in the pattern and showed greater interest in novel stimuli, indicating the presence of a modality-general SL even in young infants. Researchers have also extended SL to the visual modality in adult research, such as the visual triplet paradigm employed by Fiser and Aslin (2002). In this paradigm, participants are presented with sequences of familiar shapes organized into temporal triplets and are able to distinguish sequences seen frequently during familiarization from novel sequences of familiar shapes and sequences of shapes seen less frequently during familiarization. Overall, the collective findings from these studies suggest that SL is a fundamental cognitive mechanism that operates across various modalities and domains (Conway, 2020; Frost et al., 2015).

Development of SL across modalities and domains

One of the essential questions in this area of research concerns whether SL remains consistent throughout the lifespan (e.g., Saffran et al., 1997) or if it can be enhanced as individuals grow older (e.g., Arciuli & Simpson, 2011). The current literature suggests that the development of SL is predominantly influenced by the specific modalities or domains under consideration (e.g., Raviv & Arnon, 2018, Shufaniya & Arnon, 2018). In the case of auditory SL, it has been considered as a capacity that matures early and is not expected to improve significantly with age (e.g., Saffran et al., 1997). For instance, Saffran et al. (1997) exposed both 6-year-old children and undergraduate students to an unsegmented

artificial language presented auditorily and found no substantial difference in auditory SL between these two age groups. Qi et al. (2019) conducted a similar study and observed that children and adults performed comparably.

When it comes to the linguistic domain, research by Raviv and Arnon (2018) on children aged 5 to 12 showed that linguistic auditory SL did not exhibit significant improvement with age, aligning with the findings of Saffran et al. (1997). However, Shufaniya and Arnon (2018) challenged this perspective and suggested that the lack of age-related improvement in prior studies was attributed to the linguistic nature of the stimuli (syllables) rather than their auditory modality. In their view, the development of SL is contingent on the specific domain (linguistic vs. non-linguistic stimuli) rather than the sensory modality involved. Their results showed that when children aged 5 to 12 were tested with non-linguistic auditory stimuli such as familiar sounds (e.g., door opening, bell ringing), a clear developmental trend emerged, revealing improvement in auditory SL from 5 to 12 years of age.

Regarding visual SL, evidence suggests that it matures at about 10 years of age. Bertels et al. (2015) reported that children (9–12 years) performed similarly to adults on a shape (e.g., arrow) visual SL task. Likewise, Schlichting et al. (2017) showed no significant difference between children (6–11 years) and adolescents (12–17 years) on a similar visual SL task. However, there is also evidence for a clear developmental trend in visual SL. Arciuli and Simpson (2011) examined visual SL in children aged 5 to 12 years in which 12 aliens were divided into four groups of three (four base triplets) referred to as ABC, DEF, GHI, and JKL. The familiarization phase consisted of a continuous stream of aliens displayed on a computer screen one at a time. For each test trial, one base triplet was displayed together with one impossible triplet, and participants were prompted to identify which of the two triplets had appeared previously (during familiarization). A significant effect of age was shown, with older children performing much better than younger children. In addition, Raviv and Arnon (2018), Shufaniya and Arnon (2018) used familiar objects (e.g., house, book, plane) in their visual SL tasks. Both studies showed improvement during childhood from 5 to 12 years of age. Note that all these studies examined visual SL using non-linguistic visual stimuli (e.g., shapes, aliens, objects), with very limited research investigating linguistic

visual SL (e.g., written syllables) using the standardized SL paradigm. Isbilen et al. (2020) employed a statistically induced chunking recall task to tap into SL among adults of visually presented English syllables. In this task, participants were exposed to an artificial language and were later asked to recall strings of syllables that either follow the statistics of the artificial language or comprise the same syllables but presented in a random order. Note that no previous work has examined the developmental trend of linguistic visual SL during childhood. The first study of this dissertation filled this gap by examining the development of linguistic visual SL in both English and Chinese.

Statistical regularities also occur in writing systems (e.g., Arciuli & Cupples, 2006; Arciuli et al., 2010). Recent research in reading has been guided by the idea that proficient reading requires the ability to detect statistical regularities present in the writing system. This statistical view of learning to read influenced different sub-domains of reading research. Multiple studies have documented statistical regularities available to readers in their writing systems as they engage in reading and writing. These statistical regularities include association between letters and speech sounds (Harm & Seidenberg, 2004; Seidenberg & McClelland, 1989), between different orthographic units (e.g., Gingras & Sénéchal, 2019; Fine & Jaeger, 2013), as well as more complex relationships among orthographic units, morphological and semantic representations (Seidenberg & Gonnerman, 2000), and between orthographic units and stress patterns (Arciuli, 2018). This dissertation specifically focused on the association between an orthographic unit (e.g., word ending) and stress patterns.

Word endings as statistical cues

Previous research suggests that there are statistical cues to stress in written English words (e.g., Arciuli & Cupples, 2006; Arciuli et al., 2010; Kelly et al., 1998). The association between word endings and lexical stress is probabilistic (Arciuli & Cupples, 2006). For example, word endings such as *-ette*, *-que*, and *-umb* are strongly associated with second-syllable stress (e.g., *casSETTE*; *phySIQUE*; *sucCUMB*), and *-ior*, *-urn*, and *-em* are strongly associated with first-syllable stress (e.g., *SENior*; *REturn*; *PROBlem*). Readers who are sensitive to the cues to stress positions provided by the word endings may be able to leverage those regularities to guide their stress placement in written words. For

instance, a reader who encounters the unfamiliar word *rennet* may recognize the ending sequence *-et* that would cue them to correctly identify the word's first-syllable stress pattern. Likewise, recognizing the ending sequence *-ee* in *alee* may cue them to determine that the word has a second-syllable stress.

Kelly et al. (1998) was among the first to investigate skilled native language readers' sensitivity to ending spellings as cues to lexical stress in English. In a naming task, the words have either trochaic (first-syllable stress) or iambic (second-syllable stress) patterns and either contain or lack orthographic patterns indicative of iambic stress. For instance, the participants were asked to read iambic words like *cassette* and *cadet*. The former item contains an orthographic ending, *-ette*, associated with iambic stress, whereas the latter does not. Results showed that disyllabic words whose spellings are consistent with their stress positions were named more quickly and were pronounced with higher accuracy. In a separate lexical decision task, two groups of words were selected. In the first group, the word's final syllable was a reliable indicator of lexical stress assignment (e.g., the second syllable of *casSETTE* and the first syllable of *PELlet*). In the other group, the final syllable was an unreliable indicator of stress position (e.g., the second syllable of *caDET* and the first syllable of *PALETTE*). Participants' lexical decision times were significantly shorter, and error rates were significantly lower, for words in which orthographic endings were a reliable indicator of stress assignment. Later, Mundy and Carroll (2013) replicated these findings and, moreover, with the same experimental design, the same pattern of results was shown among adults with dyslexia. Not surprisingly, participants with dyslexia responded significantly slower than their normal controls. Taken together, these previous works were instrumental in demonstrating that skilled and less skilled readers are both sensitive to the links between stress and spelling in English.

Arciuli and Cupples (2006) created nonwords with unique word endings. Native English-speaking adult skilled readers were asked to assign stress when reading these nonwords. The participants responded in predictable ways, showing their sensitivity to word-ending cues. The participants tended to assign first-syllable stress when reading a nonword such as *plosure*, which has an ending that is strongly associated with a first-syllable stress in English words. By contrast, the participants tended to assign a second-syllable stress when reading a nonword, such as *feduct*, which has an ending strongly associated

with a second-syllable stress. However, one major weakness of this study is that the nonwords used were very much like real words of high frequency. Thus, the possibility cannot be ruled out that this performance may be heavily influenced by a process of analogy to high-frequency known words such as *pleasure* and *conduct*.

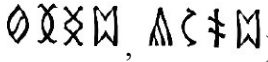
Most of the studies that examined sensitivity to word endings in lexical stress focused on L1 readers, leaving open the question whether L2 readers have such a sensitivity. To our knowledge, the only study related to this question was by Chan and Leung (2014). In the training phase, native Cantonese adult speakers were asked to learn some Spanish words without being informed of the regularities contained in the words, for example, words ending in *-o* have their stress on the penultimate syllable (the second to last syllable), and words that end in *-ar* have their stress on the last syllable. In the test phase, participants completed a two-alternative forced-choice pronunciation judgment task. Participants were instructed to listen to two words and were asked to choose the one that “sounds better” to them. These novel language learners displayed knowledge of stress assignment for novel *-ar* and *-o* ending words and generalized such knowledge to different novel items. Moreover, they were unaware of the underlying regularities based on a post hoc awareness task. It seems that learners acquired ending-phoneme-to-stress regularities in the new language implicitly, consistent with the statistical learning hypothesis. In Chan and Leung (2014), however, only two regularities of mappings between word endings and lexical stress were involved, which made it hard to generalize to a set of broader probabilistic regularities. Furthermore, participants showed learning of the target stress patterns shortly after training in the experimental setting. It is unclear, however, whether these observations can be applied to more complex input materials characteristic of more typical language learners who are in the process of learning and improving their L2 rather than those trained to acquire limited knowledge about the new language in a laboratory setting. It is thus interesting to investigate to what extent the acquisition of untaught regularities can occur in complex input materials among typical L2 learners, and the second study in this dissertation focused on this question.

SL with explicit instruction

An essential characteristic of SL is its ability to occur without explicit instruction or conscious efforts in detecting underlying patterns. In various SL studies, participants demonstrated learning even when stimuli were presented passively without any explicit task requirement (Fiser & Aslin, 2002, Saffran et al., 1999, Toro et al., 2005) or when they were engaged in an unrelated cover task (Saffran et al., 1997, Turk-Browne et al., 2009, Turk-Browne et al., 2005). Furthermore, participants in SL experiments typically exhibit limited explicit knowledge of the statistical structures of the stimuli when questioned during debriefing (Conway & Christiansen, 2005, Turk-Browne et al., 2005). This evidence suggests that SL operates implicitly, with individuals acquiring knowledge about statistical regularities in the environment without explicitly aiming to do so or having conscious awareness of the underlying patterns. This incidental nature of SL underscores its fundamental role in learning and suggests that SL mechanisms are at work even when learners are not actively engaged in explicit learning tasks.

However, some scholars and researchers proposed that although there is evidence that SL occurs incidentally or implicitly, it produces knowledge that is at least partially explicit in nature, at least in adult learners (Franco et al. 2011; Bertels et al.2015). While implicit knowledge, such as abstract grammar knowledge (Reber, 1967), or perceptual-motor skills (Cohen & Squire 1980), requires personal experience over time for acquisition, one characteristic of explicit knowledge is that it can be taught to learners through direct instruction (verbally sharing information from one person to another). Therefore, theoretically, it is possible to provide explicit training to learners to facilitate the acquisition of the concepts and rules. Different SL paradigms have been employed to examine the effect of explicit instruction. For example, some researchers employed the cross-situational word learning paradigm, often considered a type of SL task (Smith & Yu, 2008), to study language learning. In this paradigm, participants are to infer word-picture mapping when presented with one word and multiple possible referents in each trial, a task that can only be successfully completed if relations between referents and words are tracked across multiple trials. Kachergis et al. (2010) investigated the effect of different instruction on cross-situational word learning in adults. Instruction was treated as a within-participant variable so that the same participants received increasingly explicit instruction during the three phases of

the experiment. Results showed that learning was superior when participants were told explicitly to look for word–picture co-occurrence to learn word meanings, suggesting that strategic association improves word learning. Another investigation of cross-situational word learning under incidental versus intentional conditions was carried out by Hamrick and Rebuschat (2012). This study differed from Kachergis et al. (2010) in including a subjective measure of awareness at the end of the experiment and in using instructions as a between-participants variable. The intentional group received explicit instruction to learn word meaning and was told that they would be tested afterward. The incidental group was not informed about the purpose of the experiment nor that they would be tested afterward. Instead, participants in the incidental group were asked to perform a task (how many objects on each slide were animate) that was unrelated to the statistical pattern. Participants in both the intentional and incidental groups displayed significant learning, but the learning effect was larger under intentional than incidental learning condition.

In terms of how explicit instruction functions in SL of written language, some studies employed various mapping paradigms such as spelling-to-sound, spelling-to-meaning, and syllable-to-tone (e.g., Rastle et al., 2021; Taylor et al., 2011; Wiener et al., 2010). Rastle et al. (2021) examined the effects of directly teaching patterns embedded in the writing system rather than relying on learners to independently discover these patterns through reading aloud. Over a period of 10 days, a group of adult readers was taught to read unfamiliar words (e.g., ) written in two artificial writing systems.

The implicit group relied solely on experiencing the novel words to detect the regularities between spelling and sound as well as spelling and meaning, whereas the explicit group received a short session of explicit instruction on these regularities before starting their training. Nearly all participants who received explicit instruction performed exceptionally well on tests assessing their ability to apply these underlying regularities to new situations. In contrast, despite receiving up to 18 hours of training on unfamiliar words, less than 25% of those in the implicit group were able to achieve similar levels of performance to those who had received explicit instruction. Another representative study by Wiener et al. (2010) trained L2 learners of Mandarin syllable-to-tone mapping patterns using an artificial, Mandarin-like tonal

language. Half of the participants were given daily explicit instruction on Mandarin tone contours before training, while the other half were not. Explicit instruction appeared to facilitate adults' production of high-level tone contours.

Contrary to the effect of explicit instructions on adults' SL, the evidence for children's SL seems to be controversial. Witt et al. (2013) conducted a study utilizing an artificial grammar learning (AGL) task among a group of children aged 5-8 years. In this task, participants memorized sequences of images generated by a specific grammar. Half of the participants received explicit instruction about the test phase while the other half did not receive it. The effect of explicit instruction on the AGL task performance varied by age. Specifically, older children exhibited significantly better task performance than younger children when provided with explicit instruction. Moreover, Forest and Finn (under review) tested children aged 5 to 8 years, utilizing a visual triplet SL task. Their findings clearly demonstrated the advantage of explicit instruction, as evidenced in the 2-AFC test, where children in the explicit group performed better than those in the control group. In contrast, Soares et al. (2022) employed a triplet SL task in the auditory modality, involving preschool children of typically developing (TD) and those with developmental language disorder (DLD) (mean age = 5.6 years). Surprisingly, their results did not reveal the benefit of explicit instruction in either group of children, in their 2-AFC test.

Overall, although SL is primarily driven by implicit processes, there is evidence indicating that it can generate knowledge that is partially explicit. However, the findings related to the effectiveness of explicit instruction are inconsistent, particularly in research involving children. Our third proposed study in this dissertation addresses whether explicit instruction facilitates children's SL. This study will shed light on the efficacy of explicit instruction in SL and provide valuable insights for future research and educational practices, particularly in L2 teaching and learning.

Goals of the Present Studies

In this dissertation, a line of research is formed that systematically examines the SL across modalities, domains and languages. In the first study, SL development was examined across ages, domains (linguistic vs. non-linguistic), modalities (visual vs. auditory), and languages (Chinese vs.

English). Comparable experimental tasks across modalities, domains, and languages were designed and administered within the same sample of participants. Findings contribute to our better understanding of whether and how SL develops, like many other cognitive capacities, and the contribution of modality, domain, and language.

In the second study, the focus is shifted to the probabilistic regularities embedded in a specific written language. This study investigated the ability to use word endings as a probabilistic cue to lexical stress among those who acquired L2 in normal learning settings through years of practice. ESL learners with Chinese as L1 were the target population. Chinese is a representative non-alphabetic, tonal language whose prosodic features are different from English (e.g., Sun, 2006; Bellassen, 2012; Gao et al., 1995). The connection between word endings and lexical stress exists commonly among alphabetic, stress assigning languages. There is clearly an absence of explicit discussion of this statistical regularity in English L2 curriculum for these learners (People's Education Press, 2021). Thus, English L2 learners with Chinese L1 provide a unique testing case for examining the implicit acquisition of an unfamiliar linguistic regularity associated with orthography.

In the third study of this dissertation, the focus is extended to investigate whether beginning learners of English would benefit from a training program that emphasizes mapping between word endings and lexical stress. Specifically, we sought to determine whether explicit instruction, compared to self-discovery, would better support L2 learners in acquiring accurate stress placement. Participants in this study were English L2 children with a Chinese language background, a group particularly suitable for this investigation due to the differences between Chinese, a tonal language without phonetic stress, and English, a non-tonal but stress-assigning language. The study aimed to explore not only whether children could learn to use word endings as cues to lexical stress but also whether explicit guidance could facilitate their acquisition more effectively than learning through exposure alone. Through a combination of tasks, including reading aloud, implicit stress judgment, and stress assignment, we evaluated both implicit and explicit learning mechanisms and how these were influenced by explicit instruction. This study, thus, builds on the previous research presented in the dissertation, contributing further insight into the role of

SL and explicit instruction in reading development for L2 learners.

The collection of the three studies in this dissertation can not only deepen our understanding of domain-general and language-specific SL and its developmental trajectory but also have important educational implications. The set of studies included has the potential to influence and advance the fields of cognitive science, linguistics, and education, ultimately benefiting learners of various ages and backgrounds.

Chapter 2: Development of Statistical Learning Ability across Modalities, Domains and Languages

Human beings are sensitive to regularities in environment. One prominent proposed mechanism underlying this sensitivity is statistical learning (SL). In the current study, we defined SL as the ability to

use statistics (e.g., transitional probabilities or frequencies) to discover boundaries of sequentially presented stimuli. SL is present from infancy and is robustly observed across different modalities (e.g., visual, auditory and tactile) during the first year of life (e.g., Emberson et al., 2020; Kirkham et al., 2002; Saffran et al., 1996), and continues to develop throughout childhood at least in the visual modality (e.g., Arciuli & Simpson, 2011; Raviv & Arnon, 2018). Previous research also investigated the differences in SL across linguistic and non-linguistic stimuli. For instance, the classic work by Saffran et al. (1996) showed that both infants and adults were able to detect statistical regularities in non-linguistic sequences of “tone words” and linguistic sequences of English syllables. In terms of development, some studies argued that SL is an early-maturing capacity that is not expected to improve with age (e.g., Saffran et al., 1997; Bertels et al., 2015), while many other studies have shown a clear developmental trend in SL (e.g., Arciuli & Simpson, 2011; Shufaniya & Arnon, 2018). Recent research suggests that the development of SL is modality-dependent and domain-dependent (e.g., Raviv & Arnon, 2018; Shufaniya & Arnon, 2018).

The present study systematically examined development of SL ability across 9-, 11- and 13-year-olds in non-linguistic visual and auditory SL, first language (L1) Chinese visual and auditory SL and second language (L2) English visual and auditory. We focused on the SL in Chinese L1 and English L2 given their contrasting writing systems (non-alphabetic vs. alphabetic) characterized by qualitatively different statistical properties. Limited research has examined Chinese SL in the visual modality (see He & Tong, 2017; Yin & McBride, 2015 for exception) and in the auditory modality (see Wang & Saffran, 2014 for exception). However, no research has systematically studied the developmental trajectory of Chinese SL in comparison to English SL via comparable tasks. Making comparisons in SL across Chinese and English contributes to our understanding of SL beyond the English language and writing system.

The findings from the present study shed new light on our better understanding of the development of SL across modalities, domains and writing systems during late childhood.

SL in Early Infancy

Considering the richness and complexity of a natural language, why and how can infants learn

words so rapidly, especially when speakers do not mark word boundaries with pauses? Saffran et al. (1996) proposed a mechanism for statistical word segmentation: transitional probability (TP) detection. Transitional probability refers to the probability of one syllable occurring given the previous syllable (Enochson, 2015)¹. Syllables within a word have a higher transitional probability than syllables across word boundaries. For example, in the sequence *pretty baby* the likelihood of the syllable *pre* being followed by the syllable *ty* is high, while the probability of the syllable *ty* in *pretty* being followed by the syllable *ba* in *baby* is low. Segmenting the two words using TP would result in a word boundary correctly located between the sequence *pretty* and the sequence *baby*.

Saffran et al. (1996) presented 8-month-olds with a synthesized speech stream consisting of four three-syllable “words” composed of 12 unique syllables (e.g., *tupiro*, *golabu*, *dapiku*, and *tilado*) in a random order (e.g., *dapikutupirotilado-golabutupiro...*) for 2 minutes. The only cues to word boundaries were the TPs between syllable pairs. For example, the TP of *tu-pi* in this corpus was high 1.00, because *pi* always followed *tu* within the word *tupiro* (a within-word syllable pair), whereas the TP of *ro-go* was low 0.33, because *golabu* was one of three words that could follow *tupiro* (a between-word syllable pair). After brief exposure to this language, infants were able to distinguish between high TP syllable sequences and low TP sequences, suggesting that infants are sensitive to the TP information contained in the speech stream, and sensitivity to TPs is a possible mechanism underlying language acquisition (Saffran et al., 1996).

One may wonder: Is statistical learning only a property of language acquisition, or could it also apply to non-linguistic stimuli? Saffran et al. (1999) modeled directly after Saffran et al. (1996), examined infants’ performance on tone stimuli. The tones were combined into triplets to form six tone words (ADB, DFE, GG#A, FCF#, D#ED, CC#D). These “tone words” were created by translating the

¹ The formula for TP is $TP(A \rightarrow B) = \frac{\Pr(AB)}{\Pr(A)}$. A high transitional probability indicates that B is likely given A, while a low transitional probability indicates that B is unlikely given A.

previous nonsense words (e.g., bu became the musical note D, bupada became DFE). A two-alternative forced-choice task was employed to assess learning, similar to the task in Saffran et al. (1996). In the test phase, each test item consisted of two tone-sequences: a word and a non-word (e.g., ADB and AFG#). Non-words consisted of three-tone sequences made of tones drawn from the previous tone sequences but had never occurred in that order during exposure. Results showed that infants were able to detect statistical regularities in non-linguistic sequences of “tone words”, suggesting that statistical structure can be extracted from auditory input regardless of the domain in which it is presented (syllables or tones).

The aforementioned research has revealed powerful SL ability in infants, as they have the capacity to extract statistical regularities from a variety of linguistic and non-linguistic auditory input. In fact, SL in early infancy is not only limited to auditory information, but also extends to visual modality. A representative study is Kirkham et al. (2002), where 2-, 5-, and 8-month-old infants were habituated to sequences of discrete visual stimuli whose ordering followed a statistically predictable pattern (e.g., pair 1: turquoise square followed by blue cross; pair 2: yellow circle followed by pink diamond; pair 3: green triangle followed by red octagon). The infants subsequently viewed the familiar pattern alternating with a novel sequence of identical stimulus components and exhibited significantly greater interest in the novel sequence at all ages, suggesting that there is a modality general SL that is present among very young infants. Researchers also investigated the developmental trajectory of SL across modalities in infants. For example, Emberson et al. (2020) compared visual and auditory SL in 8-10 months infants and showed that auditory SL increased while visual SL did not for this age range.

In sum, there has been a great deal of evidence that SL is present from early infancy (e.g., Saffran et al., 1996; Teinonen et al., 2009) and is shown across different modalities and domains (e.g., Kirkham et al., 2002; Emberson et al., 2020; Saffran et al., 1999; see Saffran & Kirkham, 2018 for a review). However, relatively fewer studies have examined SL during childhood, and limited research that addressed the developmental trajectory of SL showed a mixed pattern of results.

Development of SL Across Modalities and Domains

One of the key questions in this line of research is whether SL is age-invariant (e.g., Saffran et al.,

1997) or can be improved with age (e.g., Arciuli & Simpson, 2011). Evidence thus far suggests that the development of SL is largely dependent on modalities or domains (e.g., Raviv & Arnon, 2018; Shufaniya & Arnon, 2018).

Auditory SL (ASL) has been considered as an early-maturing capacity that is not expected to improve with age (e.g., Saffran et al., 1997). In the non-linguistic domain, Saffran et al. (1997) exposed both 6-year-olds and undergraduate students to an unsegmented artificial language, presented auditorily, in which the only cues to word boundaries were the TPs between syllables. No significant difference in auditory SL was shown between 6-year-olds and undergraduate students. Qi et al. (2018) employed the same paradigm and found that children and adults performed similarly. In the linguistic domain, Raviv and Arnon (2018) examined linguistic auditory SL in children aged between 5 and 12, and showed that linguistic auditory SL did not improve with age, consistent with Saffran et al. (1997). However, Shufaniya and Arnon (2018) argued that the linguistic nature of the stimuli (syllables), rather than its auditory modality, was responsible for the lack of improvement with age in previous work. The development of SL is dependent upon the domain (linguistic vs. non-linguistic stimuli) rather than the modalities. Children aged between 5 and 12 were tested using a series of non-linguistic auditory stimuli (familiar sounds, e.g., door opening, bell ringing), and a clear developmental trend of improvement in auditory SL from age 5 to 12 was shown.

Regarding *visual SL* (VSL), evidence suggests that it matures at about 10 years of age. Bertels et al. (2015) reported that children (ages 9 to 12) performed similarly to adults on a shape (e.g., arrow) visual SL task. Likewise, Schlichting et al. (2017) showed no significant difference between children (ages 6 to 11) and adolescents (ages 12 to 17) on a similar visual SL task. However, there is also evidence for a clear developmental trend in visual SL. Arciuli and Simpson (2011) examined visual SL in children between the ages of 5 and 12, in which 12 aliens were divided into four groups of three (four base triplets) referred to as ABC, DEF, GHI and JKL. The familiarization phase consisted of a continuous stream of aliens, displayed on a computer screen one at a time. For each test trial, one base triplet was displayed together with one impossible triplet, and participants were prompted to identify which of the two triplets

had appeared previously (during familiarization). A significant effect of age was shown, older children performed much better than younger children. In addition, Raviv and Arnon (2018) and Shufaniya and Arnon (2018) used familiar objects (e.g., house, book, or plane) in their visual SL tasks. Both studies showed improvement during childhood between the ages of 5 and 12. Note that all these studies examined visual SL using non-linguistic visual stimuli (e.g., shapes, aliens, or objects), very limited research investigated linguistic visual SL (e.g., syllables or characters) using the standardized SL paradigm. Isbilen et al. (2020) employed a statistically induced chunking recall task to tap into SL among adults of visually presented English syllables. However, none of previous work has examined the developmental trend of linguistic visual SL in childhood. The current study filled this gap by examining the linguistic visual SL in both English and Chinese.

The current study aimed to systematically examine SL development across ages, domains (linguistic vs. non-linguistic), modalities (visual vs. auditory) and languages (Chinese vs. English). Most previous research focused on only a single modality (e.g., Saffran et al., 1997 on auditory; Arciuli & Simpson, 2011 on visual). Broadbent et al. (2017, 2018) examined the role of task modality on incidental category learning in 6- to 10-year-olds. Their results revealed that engaging in an auditory concurrent task led to poorer performance on incidental category learning compared to an audiovisual or visual concurrent task. Raviv and Arnon (2018) and Shufaniya and Arnon (2018) compared the developmental trajectory of non-linguistic visual, non-linguistic auditory and linguistic auditory SL, but the development of linguistic visual SL (e.g., English syllables) has not been examined in their two studies. Moreover, compared to the aforementioned two studies that recruited different participant samples in comparing different SL types, our study had same participants complete a set of different types of SL tasks for a more direct comparison across modality. Finally, while most previous studies used linguistic stimuli in one language with a heavy focus on English, we compared the developmental pattern of SL in two different languages (Chinese L1 and English L2).

In summary, the current study investigated developmental changes in SL across visual and auditory modalities, across linguistic and non-linguistic domains and across Chinese and English writing

systems simultaneously with comparable experimental tasks within the same sample of participants. Findings contribute to our better understanding of whether and how SL develops, like many other cognitive capacities, and the contribution of modality, domain, and language.

SL in Chinese Writing System

The Chinese writing system provides a fascinating window into the visual statistical learning mechanisms due to its unique orthography and statistical consistencies. Unlike English and other alphabetic orthographies, the basic orthographic unit in Chinese is a character that maps onto a syllable and a morpheme. A Chinese spoken syllable consists of an optional initial consonant, a vowel (accompanied by tone) and an optional final nasal consonant (n or ng). There are four tones in Chinese syllables: level (tone 1, ā), rising (tone 2, á), falling-rising (tone 3, ǎ), falling (tone 4, à). For example, the syllable segment *ma* accompanied by four different tones carries four different meanings: *mā* ‘mother’, *má* ‘hemp’, *mǎ* ‘horse’, *mà* ‘scold’ (Chao, 1968). Most Chinese characters (more than 80%) are semantic-phonetic compounds containing a semantic radical that provides a cue to the meaning of the whole character and a phonetic radical that provides a cue to the pronunciation of the whole character (Shu et al., 2003). For example, the character 枝 (*branch*) /tʂi1/ consists of a semantic radical 木 (*wood*) indicating a *wood*-related concept and a phonetic radical 支 /tʂi1/ providing a cue to the pronunciation of the whole character. Moreover, radicals exhibit certain positional consistencies: semantic radicals are usually on the left or top position and phonetic radicals are on the right or bottom position, in the left–right or top–bottom structured character, respectively (e.g., Shu et al., 2003).

Beyond the character level, most Chinese vocabularies are compound words which consist of two or three characters (Shu & Anderson, 1997). For example, the Chinese word for ‘train’ is 火车 (*huǒ chē*), which literally means “*fire car*”. Likewise, a helicopter – 直升机 (*zhí shēng jī*) – is a “*straight-rising machine*”. There are three types of typical Chinese compounds: subordinate, attributive and coordinated (Ceccagno & Basciano, 2007). In subordinate compounds, one word clearly subordinates (or supports) another, such as 房型 (*fáng xíng*) means “*layout of house*” (house + model). In attributive compounds, the

descriptive character precedes the noun/verb, such as 互动(hù dòng) means “*interaction*” (mutual + to move). In coordinated compounds, all the characters are equally important in determining the meaning, such as 高矮(gāo’ǎi) means “*height*” (high + low). Therefore, the sequential patterns between-characters are useful information in determining the meanings of Chinese words.

Previous research investigated children’s SL of positional consistencies within individual Chinese characters. For example, the corpus analysis of Chinese characters revealed that the consistency with which 木 appears on the left of left–right characters is 64.5%. Native Chinese-speaking children can extract statistical regularities from Chinese visual stimuli starting at a young age. Yin and McBride (2015) showed that native Chinese-speaking children are sensitive to visual-orthographic regularities before they receive formal literacy instruction in Chinese. In a character learning task, 4-year-old children performed better when phonetic cues were available compared to when the cues were absent, and the 5-year-olds performed better when radicals in the stimuli were positioned legally than when they were positioned illegally. Such sensitivities demonstrate that native Chinese kindergarteners can detect statistical patterns in Chinese characters without being taught explicitly. Furthermore, He and Tong (2017) employed a modified classical statistical learning paradigm (i.e., artificial orthography learning), in which a set of Chinese-like logographic characters was created using an ideographic script (Dongba) and a syllabic script (Geba). After a short exposure to a subset of novel logographic characters, children in grades 3-5 (about 8-10 years of age) were able to distinguish characters containing radicals in legal positions from those in illegal positions. Moreover, children with developmental dyslexia were less sensitive to the statistical regularities (i.e., positional consistencies of radicals) compared to normal achieving learners (Lee & Tong, 2020). Clearly, the previous research related to SL in Chinese has focused on within-character statistical patterns. The current study focused on sequential patterns in combining characters at the Chinese word level. We employed character triplets similar to English triplets in terms of the combination of written syllables to tap into children’s sensitivity to between-character sequential regularities.

The Present Study

We examined the development of statistical learning using comparable tasks across domains and modalities aged 9, 11 and 13 in Chinese L1 children who learn English L2. There are two primary reasons for choosing this age range. First, we were interested in comparing our results with previous studies, therefore we decided on this age range which is within the most studied in the previous work (5-12 years) (Arciuli & Simpson, 2011; Saffran et al., 1997; Raviv & Arnon, 2018; Shufaniya & Arnon, 2018). Furthermore, since we included L2 English linguistic SL tasks, we wanted to make sure that our participants are able to complete the L2 tasks. Children by the age of 9 in China received one year of formal English instruction, are a good starting point to be included in the current study.

There are three primary research questions guiding this study: 1) Does SL ability improve across ages? 2) Does the development of SL differ across domains (non-linguistic vs. linguistic) and across modalities (visual vs. auditory)? 3) Is the developmental trajectory different in SL between Chinese L1 and English L2?

For the first two research questions, based on previous evidence that both non-linguistic visual and non-linguistic auditory SL improved with age, but the linguistic nature of auditory SL stimuli resulted in age-invariance since linguistic knowledge of auditory stimuli is expected to be well-established in early childhood (e.g., Raviv & Arnon, 2018; Shufaniya & Arnon, 2018), we thus hypothesized that the developmental trajectory for non-linguistic and linguistic SL would be different in that both non-linguistic visual and non-linguistic auditory SL would improve with age. For linguistic SL, on the other hand, only visual would improve with age, auditory would not in both Chinese and English.

For the third research question, given that Chinese characters contain salient visual features, Chinese children would rely more on visual processing in reading Chinese compared to auditory processing (e.g., Yang & Meng, 2020). Furthermore, Chinese children may adopt their

visual processing strategies in Chinese to read and process English words (e.g., Wang & Geva, 2003). Hence, we hypothesized that there may be a modality-based advantage in both English and Chinese, that is, children would perform better in English visual than English auditory and in Chinese visual than Chinese auditory.

In addition, within Chinese visual SL, we were interested in whether the specific Chinese orthographic structure (top-bottom vs. left-right) plays a role. Tong and McBride (2014) suggested that Chinese top-bottom structure is more predictable than left-right structure, thus we hypothesized that our participants are likely to perform better on the visual SL of top-bottom structure. We were also interested in examining the correlation between L2 proficiency and L2 SL. We hypothesized that there should be a significant correlation between L2 proficiency and SL in L2. The higher the L2 vocabulary is, the better the sensitivity is to the statistical patterns of the L2 constituent syllables, which then could be linked to better SL with both auditory and visual stimuli.

Following the standardized SL triplet paradigm (Arciuli & Simpson, 2011), we designed six SL tasks in the current study. Specifically, there were two SL tasks in non-linguistic domain, which included a non-linguistic visual and a non-linguistic auditory. The well-established triplet paradigm includes a familiarization phase and a forced-choice test phase. The familiarization phase consists of four triplets, displayed as a continuous stream of aliens, shown one at a time. There are four new triplets created in the test phase and participants are asked to identify the triplet they saw previously. The non-linguistic visual stimuli were adapted from Arciuli and Simpson (2011), and 12 alien images were created to form four base triplets. The non-linguistic auditory stimuli were modeled after Qi et al. (2018), including 12 pure tones divided into four groups. In addition, four tasks in the linguistic domain, English visual and auditory, Chinese

visual and auditory comparable to the non-linguistic tasks were administered. The English visual task was modeled after Kidd et al. (2020) in which four triplets kibudu, modipa, takapo, lomari were employed. The English auditory task was adapted based on Schneider et al. (2020), consisting of four English syllable triplets. Children in the current study were familiar with syllables in English VSL and ASL, which were included in their English textbooks and taught in the classroom. The Chinese visual task stimuli were half borrowed from He and Tong (2017) and half located via online source (<https://www.pinterest.ca/pin/491455378072678746/>). Twelve pseudo-Chinese characters were divided into four base triplets. These characters represent Chinese basic orthographic structures, half top-bottom and half left-right. We also created Chinese auditory stimuli, to be comparable to the English auditory stimuli. Chinese tones were incorporated in the stimuli to reflect the Chinese tonal feature. Each syllable in a triplet was assigned a tone out of the four basic Mandarin tones, and three syllables within the triplet were ensured to carry different tones. In addition, children received an English receptive vocabulary test as an index of their English L2 proficiency.

Method

Participants

We recruited 9-year-old (mean age = 9.2, male = 17), 11-year-old (mean age = 11.1, male = 15), and 13-year-old (mean age = 13.5, male = 18) children, 40 in each age group, who were from a mid-sized city in northern China. All the participants were native Chinese speakers and had at least one year of formal English instruction. **We recruited students whose scores on their English exams in the past academic year were among top 50% in the class.** Parental consents were obtained. Parents were asked to fill out a questionnaire for the information about their children's home language exposure and experiences. Most parents (76%) had high school degree, while the remaining had college degree. Twenty-two percent of the parents had CET-4 (College English Test – Level 4, a common standard test

assessing English proficiency for college students in China), while 78% had only high school English proficiency or lower. Eighty-seven percent of the children started to learn English in grade 3, while 7% started from grade 2 onwards and 6% started from grade 1 onwards. Eighty-two percent only spoke English in their English classes at school and 18% received private lessons taught by non-native English speakers ranging from 2 to 5 (average: 3.75) hours per week outside school. All of these children never spoke English at home.

We also asked three teachers who taught English to our participants to fill out a questionnaire regarding their teaching approach employed in English class instruction. All of them applied whole-language based instruction. The *whole-language based* instruction means that the curriculum does not include any phonics instruction. Two of them had English proficiency at CET-4 level while the other one had CET-6 level. Each English class lasted for 45 minutes, and children received 5 lessons each week. Reading, writing and vocabulary were ranked as the most important tasks by the three teachers for their instruction. Speaking and listening were ranked less important. One teacher reported that she sometimes emphasized the importance of specific spelling patterns such as double consonants and plural suffixes, and the other two did not emphasize any spelling patterns.

All the following SL tasks were composed of a familiarization and testing phase. Children either saw (in visual SL tasks) or heard (in auditory SL tasks) a sequence of stimuli during the familiarization phase. There were 32 two-alternative forced choice trials in each SL test phase. The experimenter provided verbal instruction to the child throughout the two phases and monitored to make sure the children stayed on-task. There was no time constraint for responses and no feedback on accuracy of answers in the test phase. All the SL tasks were written in Jspsych (de Leeuw, 2015) and run in by a R package named “Jspsych” in R Studio (RStudio Team, 2020).





Non-linguistic VSL

Modeled after Arciuli and Simpson (2011), 12 unique alien images were divided into 4 groups of 3 images to create 4 base triplets (ABC, DEF, GHI, and JKL). For the test phase, four foil triplets were created. The relative position of each image in a foil triplet was the same as the base triplets, but the

images were grouped into four foil triplets referred to AEI, DHL, GKC, and JBF.

Children saw and heard at the beginning “Hi there, today you are going to see some aliens line up to enter a cool spaceship. We need you to help us keep track of a very special alien as the aliens line up to enter their spaceship. We will show you the alien now.” In the familiarization phase, we repeated each of the four base triplets 24 times for a total of 96 triplets. Aliens were presented one at a time on a computer screen for 800 ms each with 200 ms of interstimulus interval. During the familiarization phase, children were instructed to press the spacebar as quickly as possible whenever the specific target alien appears on the screen. The target alien image was always the third alien of one of the four base triplets (see Figure 1 upper part for the example of image presentation order). The 2AFC test phase was introduced after the familiarization phase. At the beginning of test phase, children heard and saw “Now you are going to see two groups of aliens, three as a group, and one of the two groups was friends! The friend group always followed each other to enter the spaceship. Please select the group that you think they were friends before.” Participants were asked to identify which of the two triplets (one was the base triplets and the other the foil) seemed more like what they saw during the familiarization phase.

Chinese VSL

Modeled after He & Tong (2017), 12 pseudo-characters were created by combining two Geba characters (e.g., ) . Because some Geba characters originate from Chinese loan characters, they are visually similar to Chinese characters (Zhou, 2014). Gerba characters have the same structure (e.g., top-bottom or left-right) as real Chinese characters and share many similar strokes with the modern Chinese characters. In addition, they are meaningless, thus offering a unique tool for testing children’s implicit learning of sequential pattern of written Chinese. There were 6 left-right pseudo-characters (e.g., ) and 6 top-bottom pseudo-characters (e.g., ) , reflecting the two commonly used character structures in Chinese (Shu et al., 2003). These 12 pseudo-characters were divided into 4 base triplets, for example,  . See Table 1 for the full set of stimuli. These characters were visually presented in a

sequence on the screen.

Table 1

Examples of pseudo-characters in Chinese VSL task

| | Pseudo Chinese characters | | | | | |
|---------------|---------------------------|---|---|---|---|---|
| Top-bottom | 𠄎 | 𠄎 | 𠄎 | 𠄎 | 𠄎 | 𠄎 |
| Left-right | 𠄎 | 𠄎 | 𠄎 | 𠄎 | 𠄎 | 𠄎 |
| Triplets | 𠄎 | 𠄎 | 𠄎 | 𠄎 | 𠄎 | 𠄎 |
| Foil Triplets | 𠄎 | 𠄎 | 𠄎 | 𠄎 | 𠄎 | 𠄎 |

Chinese VSL's procedure was the same as the non-linguistic VSL, children heard and saw at the beginning “Hi there, we are going to learn a new language today! We need you to help us keep track of a special character as you see different characters appearing on the screen. We will show you the characters now.” At the beginning of the test phase, children heard and saw “Now you are going to see two groups of characters, three as a group, please select the group of characters that always followed each other appearing on the screen before.” See Figure 1 upper part for the example of presentation order of Chinese VSL and English VSL described below.

English VSL

Modeled after Kidd et al. (2020), 12 written English syllables (ki, bu, du, mo, di, pa, ta, ka, po, lo, ma, ri) were used to create four triplets (kibudu, modipa, takapo, lomari). These syllables were visually presented in a sequence on the screen. Foils were constructed from the same set of syllables (e.g., the foil taburi consisted of syllables taken from the three target triplets takapo, kibudu, and lomari) but crucially differed from the targets in their statistical structure. The four alternative forced choice foils were: taburi, kimapo, mokari, lodidu. English VSL's procedure was the same as the Chinese VSL.

Non-linguistic ASL

Modeled after Qi et al. (2018), 12 pure tones within the same octave (a full chromatic scale

starting from middle C were divided into four base triplets (FGD, G#C#B, CF#D, EAA#). For the two alternative choice task, four additional foils were created including F#BF, AA#G#, C#D#E, GDC.

Children saw and heard "Hi! We're going to listen to some alien folk music today. We need you to help us keep track of a special sound as you hear different sounds." During the familiarization phase, we repeated each of the four triplets 48 times for a total of 192 triplets (twice as much as the visual conditions, following the standard paradigm). The duration of each tone is 460 ms with a 20 ms inter-stimulus interval. Children were instructed to press the spacebar as quickly as possible whenever they hear the target sound in the familiarization phase. The target sound was always the third one of the four base triplets (see Figure 1 lower part for the example of sound presentation order for non-linguistic ASL, English ASL and Chinese ASL tasks described below). At the beginning of the test phase, children heard and saw "Now you are going to hear two groups of sound, three as a group, please select the group of sounds that always followed each other you heard before." Participants were asked to identify which of the two triplets seemed more like what they heard during the familiarization phase.

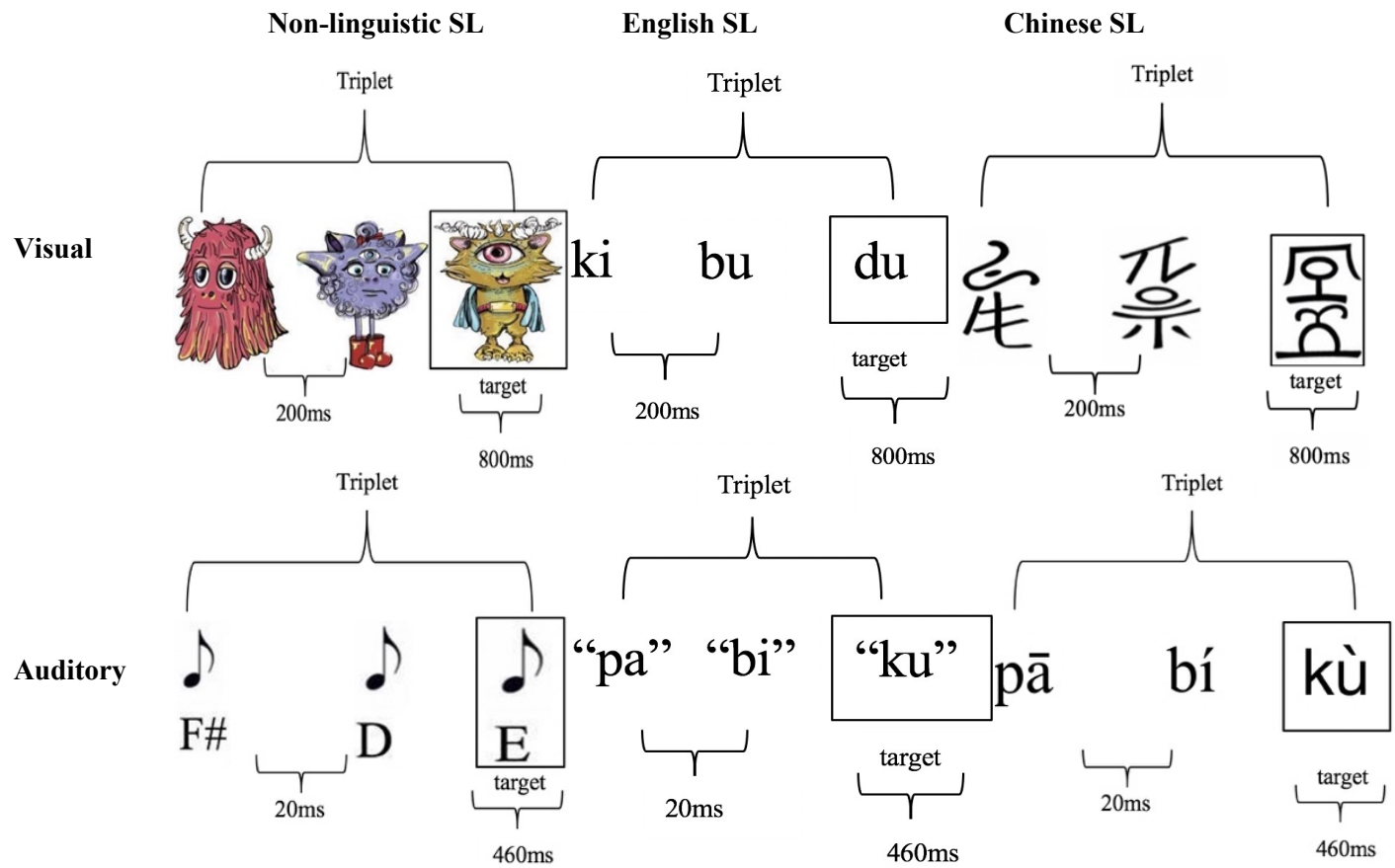
English ASL

The artificial language used in this experiment was adapted from Schneider et al. (2020) and consisted of 12 English spoken syllables: pa, bi, ku, go, la, tu, da, ro, pi, ti, bu, do. These were used to construct 4 tri-syllabic based words: pabiku, golatu, daropi, and tibudo. For 2AFC, four additional nonword foils were created, using the same syllables as the input language. The four foil words were gobutu, parodo, tilapi, dabiku. English ASL's procedure was the same as the non-linguistic ASL.

Chinese ASL

Chinese auditory syllables were made comparable to English with four tones evenly distributed across syllables. They are pā, bí, kù, gōu, lǎ, tù, dě, ròu, pī, tī, bú, dòu. The four tri-syllabic words are pābíkù, gōulǎtù, děròupī, tībúdòu. For 2AFC, the four nonword foils are gōubútù, pāròudòu, tīlǎpì, děbíkù. Chinese ASL's procedure was the same as English ASL.

Figure 1



Familiarization phase of all six statistical learning tasks.

Note. Example triplets across each task are depicted in this figure. Each visual stimulus appeared for 800 ms with a 200 ms interval, and each auditory stimulus was heard for 460 ms with a 20 ms interval.

English receptive vocabulary

To evaluate participants' current English language proficiency, a shortened version of PPVT (the Peabody Picture Vocabulary Test, Fourth Edition; Dunn & Dunn, 2007) was administered to all the participants in the classroom as a group. To maintain the same progression of item difficulty as the original test, 30 items from the original test were selected (items 1-6, 13-18, 37, 39-43, 49-53, 61-64, 73, and 75-76). The experimenter played recording of each item once, and children circled the picture in their response booklets that best described the word presented.

Procedure

Two sessions were involved, each lasted around 35 minutes. Children completed three SL tasks in each session, separated by a one-week interval. In the first session, children completed non-linguistic visual, Chinese visual, Chinese auditory SL tasks and English proficiency test. In the second session, children completed non-linguistic auditory, English visual and English auditory SL tasks. For each age group, the order of completing SL tasks were counterbalanced within each session. There were six different possible orders in each session, so children were roughly divided into six groups and received one of the orders. Also, half of the children received session 1 first and the other half received session 2 first for each age group.

Results

Arciuli and Simpson (2011) suggested that participants who failed to identify half of the targets in the familiarization phase should be excluded from data analysis. All of participants in the current study were retained for data analysis as their score ranged from 67% to 100%, and their performance in familiarization phase did not differ across SL tasks ($p > .05$). Table 2 lists the accuracy rates for all SL tasks across ages. All participants' SL accuracy were within 2.5 standard deviations. Children across all ages performed significantly better than chance level (50%) across all SL tasks including non-linguistic VSL, non-linguistic ASL, Chinese VSL, Chinese ASL, English VSL and English ASL tasks (all $ps < .05$). Children also appeared to improve their SL with age in all the tasks (see Figure 2). The mean accuracy of SL tasks in session 1 was 0.646 ($sd = .14$) and that of session 2 was 0.649 ($sd = .15$). The mean accuracy

did not differ across sessions ($p = .76$). Children's performance was highly stable across the two sessions ($r = 0.62$, $t = 8.66$, $p < .001$). The internal consistency reliabilities (i.e., Cronbach's alpha coefficients) of SL tasks were 0.74 for non-linguistic visual, 0.76 for non-linguistic auditory, 0.90 for English visual, 0.89 for English auditory, 0.81 for Chinese visual, and 0.83 for Chinese auditory.

Table 2

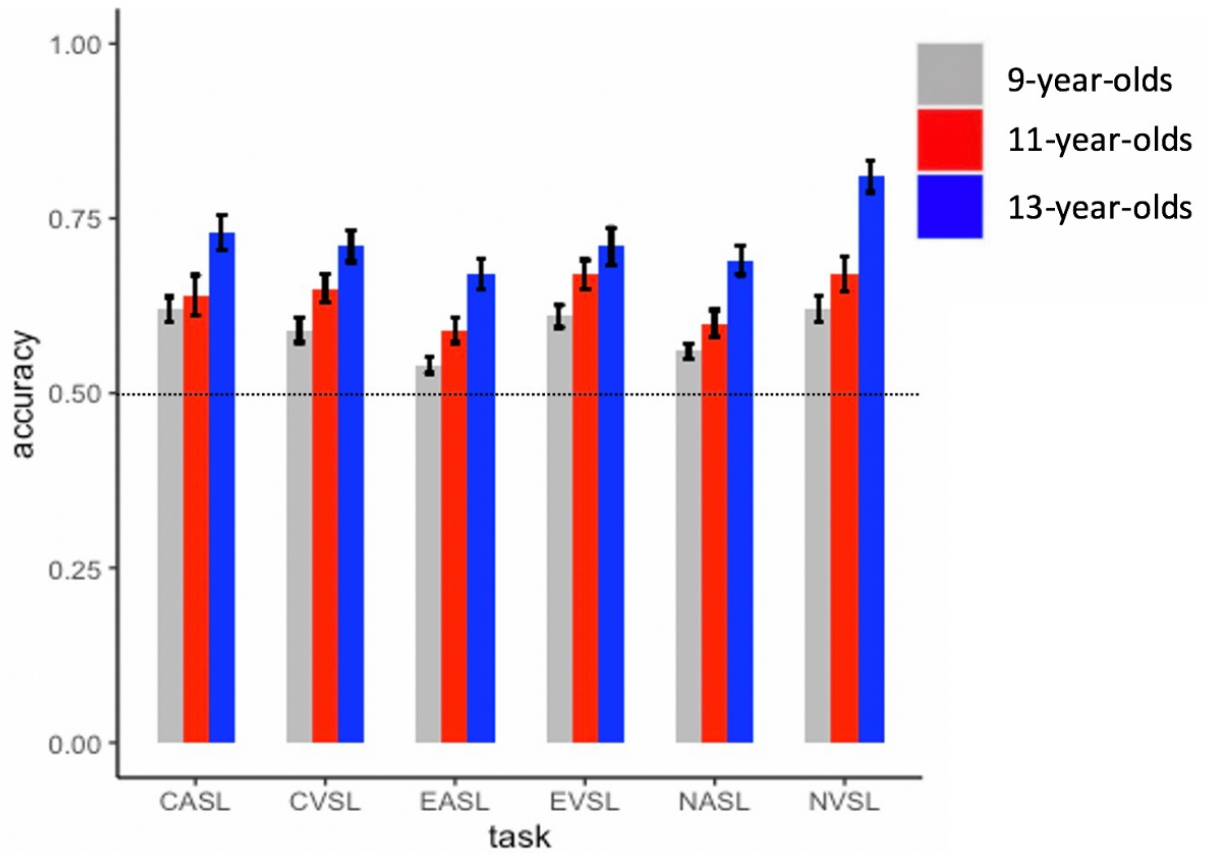
SL accuracy rates across tasks by grades

| | 9-year-old | | | 11-year-old | | | 13-year-old | | |
|-------------------------|------------|------|---------------------|-------------|------|---------------------|-------------|------|---------------------|
| | Mean | SD | <i>t</i> | Mean | SD | <i>t</i> | Mean | SD | <i>t</i> |
| Non-linguistic Visual | 0.62 | 0.12 | 6.35 ^{***} | 0.67 | 0.16 | 6.74 ^{***} | 0.81 | 0.14 | 13.3 ^{***} |
| Non-linguistic Auditory | 0.56 | 0.07 | 5.28 ^{***} | 0.60 | 0.12 | 5.16 ^{***} | 0.69 | 0.13 | 9.07 ^{***} |
| English Visual | 0.61 | 0.10 | 5.82 ^{***} | 0.67 | 0.13 | 8.24 ^{***} | 0.71 | 0.17 | 7.74 ^{***} |
| English Auditory | 0.54 | 0.07 | 3.24 ^{**} | 0.59 | 0.11 | 5.06 ^{***} | 0.67 | 0.14 | 7.32 ^{***} |
| Chinese Visual | 0.59 | 0.11 | 4.67 ^{***} | 0.65 | 0.13 | 6.97 ^{***} | 0.71 | 0.14 | 9.03 ^{***} |
| Chinese Auditory | 0.62 | 0.11 | 7.02 ^{***} | 0.64 | 0.18 | 5.76 ^{***} | 0.73 | 0.16 | 8.71 ^{***} |

p* < .01; *p* < .001

Note. Forty children were tested in each grade for each task.

Figure 2



Mean SL accuracy rates across all the tasks by age.

Note. Accuracy rates in the 2AFC task. Dotted line represents the chance level. CASL denotes Chinese ASL; CVSL denotes Chinese VSL; EASL denotes English ASL; EVSL denotes English VSL; NASL denotes Non-linguistic ASL; NVSL denotes Non-linguistic VSL.

We used mixed-effect logistic regression model (Barr et al., 2013) to examine the effects of age, modality, and linguistic type in the statistical programming language R (R Core Team, 2017). Our dependent binominal variable was success in a single test trial. We started with a full model with a maximum random effects structure justified by the experimental design and removed random factors until the model converged (Barr et al., 2013). The final model included fixed effects for age (centered), modality (visual vs. auditory), linguistic type (non-linguistic vs. Chinese vs. English), the interactions among the three independent variables and a random intercept for participants (see Table 3). Results showed that the two-way interaction of age and modality was significant ($b = 0.14$, $SE = 0.06$, $z = 2.46$, $p = 0.014$). Post-hoc comparisons revealed that visual SL was associated with a significant increase from 9 to 11 years ($b = 0.25$, $SE = 0.10$, $z = 2.55$, $p = 0.01$) and a significant increase from 11 to 13 years ($b = 0.47$, $SE = 0.10$, $z = 4.67$, $p = 3.02e-06$). However, the increase in the auditory SL accuracy between 9 and 11 years was not significant ($b = 0.14$, $SE = 0.10$, $z = 1.46$, $p = 0.145$), whereas there was a significant increase from 11 to 13 years ($b = 0.44$, $SE = 0.10$, $z = 4.45$, $p = 8.66e-06$).

The two-way interaction of modality and linguistic type was also significant ($b = -0.15$, $SE = 0.05$, $z = -2.93$, $p = 0.003$). Post-hoc comparisons revealed that while children performed significantly better in non-linguistic visual than non-linguistic auditory ($b = 1.25$, $SE = 0.15$, $z = 8.28$, $p = 2.22e-16$) and in English visual than English auditory ($b = 0.88$, $SE = 0.15$, $z = 6.06$, $p = 1.38e-9$), there was no difference in children's performance across Chinese visual and auditory ($b = -0.16$, $SE = 0.15$, $z = -1.08$, $p = 0.278$). The two-way interaction between age and linguistic type and the three-way interaction of age, modality and linguistic type were both not significant ($b = -0.02$, $SE = 0.03$, $z = -0.71$, $p = 0.475$; $b = -0.07$, $SE = 0.04$, $z = -1.69$, $p = 0.092$, respectively).

Table 3*Regression model for SL across age groups, modalities, and linguistic types.*

| | Estimate | Std. Error | z value | p-value |
|--------------------------------|----------|------------|---------|------------|
| Intercept | 0.14 | 0.07 | 1.91 | 0.05677 |
| age group | 0.31 | 0.06 | 5.38 | 7.43e-8*** |
| modality | 0.31 | 0.07 | 4.51 | 6.42e-6*** |
| language | 0.11 | 0.04 | 3.06 | 0.00221** |
| age group: modality | 0.14 | 0.06 | 2.46 | 0.01382* |
| age group: linguistic type | -0.02 | 0.03 | -0.71 | 0.47527 |
| modality: linguistic type | -0.15 | 0.05 | -2.93 | 0.00337** |
| age: modality: linguistic type | -0.07 | 0.04 | -1.69 | 0.09169 |

* $p < .05$; ** $p < .01$; *** $p < .001$

The main effects of age, modality, and linguistic type were all significant ($b = 0.31$, $SE = 0.06$, $z = 5.38$, $p = 7.39e-8$; $b = 0.31$, $SE = 0.07$, $z = 4.51$, $p = 6.42e-6$; $b = 0.11$, $SE = 0.04$, $z = 3.06$, $p = 0.002$, respectively). Overall, 13-year-olds performed significantly better than 11-year-olds ($b = 2.71$, $SE = 0.55$, $z = 4.89$, $p = 1.02e-06$), 11-year-olds performed significantly better than 9-year-olds ($b = -1.17$, $SE = 0.55$, $z = -2.13$, $p = 0.033$), and 13-year-olds performed significantly better than 9-year-olds ($b = 3.88$, $SE = 0.55$, $z = 7.01$, $p = 2.33e-12$). Participants overall performed significantly better in visual SL than auditory SL, better in non-linguistic SL than English SL ($b = 0.97$, $SE = 0.21$, $z = 4.62$, $p = 3.84e-6$), and better in Chinese SL than English SL ($b = 0.68$, $SE = 0.21$, $z = 3.29$, $p = 0.001$). Nevertheless, the difference in participants' performance between non-linguistic SL and Chinese SL was not significant ($b = 0.29$, $SE = 0.21$, $z = 1.36$, $p = 0.174$).

To examine the effect of Chinese orthographic structure on Chinese visual SL accuracy, we ran another mixed-effect logistic regression model including Chinese orthographic structure (top-bottom vs. left-right), age (centered), and the interaction between them as fixed factors. The model had a random intercept for participants, and Chinese visual SL accuracy in each trial served as a dependent binominal variable. Results showed that there was a significant interaction between Chinese orthographic structure and age ($b = 0.11$, $SE = 0.04$, $z = 2.54$, $p = 0.011$). Post-hoc comparisons revealed that while 9- and 11-year-olds performed significantly better in Chinese top-bottom structure than left-right structure ($b = 0.40$, $SE = 0.12$, $z = 3.42$, $p = 0.001$; $b = 0.50$, $SE = 0.12$, $z = 4.13$, $p = 3.66e-5$, respectively), 13-year-olds did not perform differently between the two Chinese orthographic structures ($b = -0.06$, $SE = 0.13$, $z = -0.45$, $p = 0.657$). The main effects of Chinese orthographic structure and age were also significant ($b = -0.28$, $SE = 0.07$, $z = -4.03$, $p = 5.61e-5$; $b = 0.09$, $SE = 0.04$, $z = 2.26$, $p = 0.024$, respectively). Children overall performed better in top-bottom structure than left-right structure. 13-year-olds performed significantly better than 11-year-olds ($b = -0.63$, $SE = 0.28$, $z = -2.22$, $p = 0.026$), 11-year-olds performed significantly better than 9-year-olds ($b = -0.56$, $SE = 0.28$, $z = -2.03$, $p = 0.043$), and 13-year-olds performed significantly better than 9-year-olds ($b = -1.19$, $SE = 0.28$, $z = -4.24$, $p = 2.23e-5$).

Finally, English vocabulary and English visual SL were significantly correlated with each other (r

= 0.42, $p < .001$), and so were English vocabulary and English auditory SL ($r = 0.50, p < .001$).

Furthermore, English vocabulary and Chinese visual SL were significantly correlated with each other ($r = 0.42, p < .001$), and so were English vocabulary and Chinese auditory SL ($r = 0.41, p < .001$). English vocabulary and non-linguistic visual SL were significantly correlated with each other ($r = 0.49, p < .001$), and so were English vocabulary and non-linguistic auditory SL ($r = 0.51, p < .001$).

Discussion

The goal of the present study was to investigate the effects of modality (visual vs. auditory) and linguistic type (non-linguistic vs. Chinese vs. English) in the developmental trajectory of SL across children aged 9, 11, and 13. There are several important findings. First, we showed three main effects of age, modality, and linguistic type. Overall, children developed their SL abilities across the three age groups, and they performed better in visual than auditory modality. Moreover, children performed better in non-linguistic and Chinese SL compared to English SL. Most importantly, we showed a significant interaction between age and modality, that the developmental trajectory was different across visual and auditory modalities. While visual SL constantly improved from 9 years of age to 11, and to 13 years, auditory SL improves only from the age of 11 to 13, but not from 9 to 11. The same pattern of the age group and modality interaction occurred across non-linguistic, Chinese, and English SL. A significant interaction between modality and linguistic type was shown, and that better learning was achieved in visual as compared to auditory SL in both non-linguistic and English stimuli. However, children performed similarly across the two modalities in Chinese. The same pattern of modality and linguistic type interaction occurred across 9-, 11- and 13-year-olds. In addition, while children aged 9 and 11 demonstrated better learning in Chinese top-bottom structure, 13-year-old children did not show this learning bias. Finally, the correlations between English vocabulary and SL measures were all significant.

The Interaction between Age and Modality

Our findings were consistent with our hypotheses and those in previous studies in terms of the significant improvement of non-linguistic visual SL during childhood (e.g., Arciuli & Simpson, 2011; Raviv and Arnon, 2018; Shufaniya & Arnon, 2018). For example, Raviv and Arnon (2018) tested

children aged 5-12 on matching visual and auditory SL tasks, and they also showed a development trajectory of visual SL in the non-linguistic domain. This finding suggests that the developmental nature of non-linguistic visual SL is robust during childhood. More importantly, our study is one of the first to demonstrate that the developmental nature of visual SL is not specific to non-linguistic input, but it also applies to linguistic stimuli including both Chinese L1 and English L2. This finding is contradictory to the speculation by Shufaniya and Arnon (2018) that linguistic visual SL, like linguistic auditory SL, may not grow with age due to its linguistic nature. Our finding suggests that children across early, middle, and late childhood can extract the statistical patterns from visual input regardless of the domain in which it is presented (images, English syllables, or Chinese characters), and this skill increases with age.

The auditory modality, however, was different from our hypotheses and previous findings that mentioned only non-linguistic auditory SL increases as a function of age and linguistic auditory SL appears to be age-invariant (Shufaniya & Arnon, 2018). Our study demonstrated that regardless of auditory stimuli type (non-linguistic, Chinese L1 or English L2), auditory SL increased from 11 to 13 years of age but not from 9 to 11 years of age. Shufaniya and Arnon (2018) argued that the auditory knowledge of syllable co-occurrence has already matured in early childhood (see also Storkel, 2001), so the ability of extracting linguistic auditory SL information will not change after early childhood. Then, why did our data show that there is still a developmental improvement from ages 11 to 13? One possible explanation for this inconsistent finding may lie in the difference of presentation speed between auditory (460 ms per stimulus) and visual (800 ms per stimulus) stimuli in the current study. Emberson et al. (2011) revealed that auditory and visual SL have opposite timing constraints where visual SL performs better in slow presentation and auditory SL performs better in fast presentation. Both slow visual and fast auditory presentation rate may demand less attention and working memory while fast visual and slow auditory presentation rate may demand greater attention and working memory. Multiple works showed that SL should be thought of as a componential, rather than a unified, theoretical construct, and attention and working memory can be underlying components of SL (e.g., Arciuli, 2017; Arciuli & Conway, 2018). Note that the SOA (Stimulus Onset Asynchrony: each stimulus+ interstimulus rate) in our study was very

slow-1000ms compared to other studies (e.g., 600 ms in Arciuli & Simpson, 2012; 500 ms in Bertels et al., 2015; 600 ms in Raviv & Arnon, 2018), so our visual SL may require relatively less attention and working memory; however, our auditory presentation rate was mid-way, 480ms, between how quickly they can be presented (e.g., 222ms in Saffran et al., 1996, 1997) and how slowly they could be presented (e.g., 750ms in Emberson et al., 2011). Therefore, we speculate that compared to the slow presentation of our visual stimuli, the relatively faster auditory stimuli in the auditory SL task may demand more attention and working memory to help encode and process the information. Furthermore, as suggested by Broadbent et al. (2017, 2018), the use of visual information may help learners focus their attention on the visual features of a learning task, leading to better incidental learning of category information. In our study, the presence of visual symbols may also serve as an aid to focus attention and reduce working memory load in the visual SL tasks. Taken together, these may help explain why we did not observe significant change between 9 and 11 years of age but a significant increase in children's auditory SL was evident between 11 and 13 years of age. Future research would benefit from having a working memory task to directly measure working memory in children across these ages alongside SL tasks and see if it is associated with children's SL across modalities.

The Interaction between Modality and Language

Our results showed that children performed better in visual compared to auditory SL in non-linguistic and English SL. The visual advantage shown in our study is similar to the findings from Siegelman and Frost (2015), Raviv and Arnon (2018), and Shufaniya and Arnon (2018). Raviv and Arnon (2018) argued that the visual advantage in their study could be due to the fact that the alien visual stimuli is more attractive than auditory sounds and easier to encode for children. Nevertheless, **our study showed that the visual advantage was not limited to non-linguistic alien stimuli but also to English syllable stimuli. One possible explanation is that as L2 learners of English, native Chinese speaking children rely more on visual information compared to auditory in learning and processing English words (e.g., Wang & Geva, 2003), and Chinese language teachers of our participants put emphasis on recognizing visual patterns of English syllables than listening and speaking skills based on our survey.** The consequence of

this teaching style is that children performed much better in reading and writing than listening and speaking in English (Zhou, 2005).

For Chinese SL, on the other hand, there was no significant difference between Chinese visual and auditory SL, which is contrary to our hypothesis. The average accuracy of Chinese visual SL and English visual SL were similar, but children performed much better in Chinese auditory SL than English auditory SL. **One explanation is that our children had a relatively weaker auditory SL in L2 compared to that in L1, which is not surprising given that English is their L2.** Another potential explanation is that the presence of tonal information in the auditory Chinese stimuli improves Chinese auditory SL, thus minimizing the advantage of visual SL since the syllable structures and phoneme types are very similar across the Chinese and English linguistic stimuli, and the only difference is the presence of tones in the Chinese auditory stimuli. Our hypothesis is that not only do children rely on syllable segmental information to detect the correct sequence of the triplets in Chinese, but they also utilize the additional tonal information superimposed onto the segmental sounds for the statistical pattern identification. Future research is needed to test this hypothesis directly.

The Effect of Character Structure and the Role of English Proficiency

The finding that children aged 9 and 11 learn Chinese top-bottom structure better than left-right structure is in accordance with that in Tong and McBride (2014) in which children's invented character production accuracy rate was higher for top-bottom structure than those of left-right structure, and children made more positional errors on the left-right structure than the top-bottom structure. The first explanation offered by Tong and McBride (2014) on this top-bottom structure preference was the greater prevalence of the left-right structure (> 70%) compared to the top-bottom structure in the Chinese writing system, and children tend to pay more attention to less common structure. The second explanation was that the position of stroke patterns embedded in top-bottom structured characters is highly predictable and easier to learn. There are a relatively small portion of top-bottom structured characters and the stroke patterns making up the top-bottom characters are less extensive. But why does this learning bias disappear by age 13 in our study? We speculate that after 11 years of age, children's Chinese character reading is

much enhanced in both top-bottom and left-right structured characters and the preference for a particular type of structure fades gradually.

In addition, we showed a significant positive correlation between all the six types of SL and English L2 vocabulary. This result supports and extends previous findings on the important role of different forms of SL in vocabulary acquisition. SL has been shown to be one important, early contributor to vocabulary development (e.g., Saffran et al., 2009; Evans et al., 2009; Kemény & Lukács, 2021; Reuter et al., 2018; Hedenius et al., 2011). For example, Evans et al. (2009) revealed a link between auditory SL and language proficiency including vocabulary in children aged 6–14. Kemény and Lukács (2021) demonstrated that SL skills had an independent contribution to vocabulary size over and above age, receptive grammatical ability, and short-term memory in childhood. Our finding further suggests that SL in L2 across modalities and domains are strongly associated with children's L2 vocabulary knowledge. This association is consistent with the hypothesis that SL plays a critical role in word segmentation. Word segmentation is crucial in lexical development (Erickson & Thiessen, 2015). After all, lexical acquisition not only in L1 but also in L2 relies on distributional information and is supported by SL.

Limitations

There are several limitations in the current study. First, we only investigated children aged 9, 11, and 13. It would be valuable to track SL performance across modalities and domains across a greater range of age, such as from infancy to late childhood. To achieve this goal, researchers need to be able to design a SL paradigm suitable for both infants and children. Second, as a cross-sectional study, our research was not able to detect individual changes in SL ability. Longitudinal studies are extremely helpful in this area that sets out to provide a more comprehensive understanding of individual differences in the developmental trajectory of SL. **Also, the different patterns of SL between Chinese and English could be a combined effect of the writing system (alphabetic vs. non-alphabetic) and language status (L1 vs. L2). Our current participants sample would not allow us to tease these two effects apart cleanly. Future research needs to recruit monolingual Chinese- and monolingual English-speaking children for a more direct comparison in terms of their linguistic SL to address the language effect.** Moreover, we did

not include a measure of reading fluency as a predictor to allow us to address the role of individual differences in reading skills in developing SL. Future research may include this measure to address this question. Finally, the three syllables in each English triplet are equally stressed without a specific stress assignment in the English auditory SL task. This may result in unnatural English spoken multisyllabic words. Future research needs to consider a specific stress assignment in each triplet in the auditory English SL task, to make it comparable with the Chinese triplets in the auditory Chinese SL task in terms of the presence of the suprasegmental information.

Conclusions

Our study makes a unique contribution to the growing body of literature on the development of SL across visual and auditory modalities in non-linguistic and linguistic domains. Our findings suggest that the developmental trajectory of SL differs across visual and auditory modalities. While visual SL constantly improves from age 9 to 11 to 13, auditory SL only improves significantly from age 11 to 13, probably due to the faster temporal processing speed in auditory stimuli as it requires a higher demand on attention and working memory. Furthermore, visual modality advantage was shown in non-linguistic and English SL but not in Chinese SL, possibly due to the contribution from tonal information. Together, our findings point to the joint function of age, modality, and linguistic type in SL development.

References

- Arciuli, J. (2017). The multi-component nature of statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 20160058.
<https://doi.org/10.1098/rstb.2016.0058>
- Arciuli, J., & Conway, C. M. (2018). The promise—and challenge—of statistical learning for elucidating atypical language development. *Current Directions in Psychological Science*, 27(6), 492-500.
<https://doi.org/10.1177/0963721418779977>
- Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental Science*, 14(3), 464-473.
<https://doi.org/10.1111/j.1467-7687.2009.00937.x>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.
<https://doi.org/10.1016/j.jml.2012.11.001>
- Bertels, J., Boursain, E., Destrebecqz, A., & Gaillard, V. (2015). Visual statistical learning in children and young adults: how implicit?. *Frontiers in Psychology*, 5, 1541.
<https://doi.org/10.3389/fpsyg.2014.01541>
- Broadbent, H. J., Osborne, T., Rea, M., Peng, A., Mareschal, D., & Kirkham, N. Z. (2018). Incidental category learning and cognitive load in a multisensory environment across childhood. *Developmental Psychology*, 54(6), 1020. <https://doi.org/10.1037/dev0000472>
- Broadbent, H. J., White, H., Mareschal, D., & Kirkham, N. Z. (2018). Incidental learning in a multisensory environment across childhood. *Developmental science*, 21(2), e12554.
<https://doi.org/10.1111/desc.12554>
- Ceccagno, A., & Basciano, B. (2007, January). Classification of Chinese compounds. In *Mediterranean Morphology Meetings* (Vol. 6, pp. 71-83).
- Chao, Y. R. (1968). *Grammar of Spoken Chinese*. Berkeley, CA: University of California Press.
- Cowan, N. (1997). The development of working memory. In N. Cowan (Ed.), *The development of*

- memory in childhood* (pp. 163–199). Psychology Press/Erlbaum (UK) Taylor & Francis.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1), 1-12.
<https://doi.org/10.3758/s13428-014-0458-y>
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.
- Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology*, 64(5), 1021-1040.
<https://doi.org/10.1080/17470218.2010.538972>
- Emberson, L. L., Misyak, J. B., Schwade, J. A., Christiansen, M. H., & Goldstein, M. H. (2020). Comparing statistical learning across perceptual modalities in infancy: An investigation of underlying learning mechanism (s). *Developmental Science*, 22(6), e12847.
<https://doi.org/10.1111/desc.12847>
- Enochson, K. (2015). *Adaptation as statistical learning: An individual differences study* (Doctoral dissertation, George Mason University).
- Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, 37, 66-108. <https://doi.org/10.1016/j.dr.2015.05.002>
- Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52, 321–335.
[https://doi:10.1044/1092-4388\(2009/07-0189\)](https://doi:10.1044/1092-4388(2009/07-0189))
- Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language?. *Psychological Science*, 24(7), 1243-1252.
<https://doi.org/10.1177/0956797612472207>
- He, X., & Tong, X. (2017). Statistical learning as a key to cracking Chinese orthographic codes. *Scientific Studies of Reading*, 21(1), 60-75.

<https://doi.org/10.1080/10888438.2016.1243541>

Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Veríssimo, J., Dye, C. D., Alm, P., Jennische, M., Bruce Tomblin, J., & Ullman, M. T. (2011). Grammar predicts procedural learning and consolidation deficits in children with Specific Language Impairment. *Research in Developmental Disabilities, 32*(6), 2362–2375.

<https://doi-org.ezproxy.its.uu.se/10.1016/j.ridd.2011.07.026>

Isbell, E., Fukuda, K., Neville, H. J., & Vogel, E. K. (2015). Visual working memory continues to develop through adolescence. *Frontiers in Psychology, 6*, 696.

<https://doi.org/10.3389/fpsyg.2015.00696>

Kemény, F., & Lukács, Á. (2021). The role of statistical learning and verbal short-term memory in impaired and typical lexical development. *Frontiers in Communication, 6*, 148.

<https://doi.org/10.3389/fcomm.2021.700452>

Kidd, E., Arciuli, J., Christiansen, M. H., Isbilen, E. S., Revius, K., & Smithson, M. (2020). Measuring children's auditory statistical learning via serial recall. *Journal of Experimental Child Psychology, 200*, 104964.

<https://doi.org/10.1016/j.jecp.2020.104964>

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*(2), B35-B42.

[https://doi.org/10.1016/s0010-0277\(02\)00004-5](https://doi.org/10.1016/s0010-0277(02)00004-5)

Lee, S. M. K., & Tong, X. (2020). Spelling in developmental dyslexia in Chinese: Evidence of deficits in statistical learning and over-reliance on phonology. *Cognitive Neuropsychology, 37*(7-8), 494-510. <https://doi.org/10.1080/02643294.2020.1765754>

Tong, X., & McBride, C. (2014). Chinese children's statistical learning of orthographic regularities: Positional constraints and character structure. *Scientific Studies of Reading, 18*(4), 291-308.

<https://doi.org/10.1080/10888438.2014.884098>

Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D., & Arciuli, J. (2019). Hearing matters more

- than seeing: A cross-modality study of statistical learning and reading ability. *Scientific Studies of Reading*, 23(1), 101-115.
<https://doi.org/10.1080/10888438.2018.1485680>
- R Core Team (2017) R: A Language and Environment for Statistical Computing.
<https://www.R-project.org/>
- Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developmental Science*, 21(4), e12593. <https://doi.org/10.1111/desc.12593>
- Reuter, T., Emberson, L., Romberg, A., & Lew-Williams, C. (2018). Individual differences in nonverbal prediction and vocabulary size in infancy. *Cognition*, 176, 215-219.
<https://doi.org/10.1016/j.cognition.2018.03.006>
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928. <https://doi.org/10.1126/science.274.5294.1926>
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1), 27-52.
[https://doi.org/10.1016/s0010-0277\(98\)00075-4](https://doi.org/10.1016/s0010-0277(98)00075-4)
- Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual review of psychology*, 69, 181-203. <https://doi.org/10.1146/annurev-psych-122216-011805>
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, 8(2), 101-105. <https://doi.org/10.1111/j.1467-9280.1997.tb00690.x>
- Saffran, J. R., Pelucchi, B., & Hay, J. F. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, 80(3), 674-685.
<https://doi.org/10.1111/j.1467-8624.2009.01290.x>
- Schlichting, M. L., Guarino, K. F., Schapiro, A. C., Turk-Browne, N. B., & Preston, A. R. (2017). Hippocampal structure predicts statistical learning and associative inference abilities during

- development. *Journal of Cognitive Neuroscience*, 29(1), 37-51.
https://doi.org/10.1162/jocn_a_01028
- Schneider, J. M., Hu, A., Legault, J., & Qi, Z. (2020). Measuring statistical learning across modalities and domains in school-aged children via an online platform and neuroimaging techniques. *JoVE (Journal of Visualized Experiments)*, (160), e61474.
<https://doi.org/10.3791/61474>
- Shu, H., & Anderson, R. C. (1997). Role of radical awareness in the character and word acquisition of Chinese children. *Reading Research Quarterly*, 32(1), 78-89. <https://doi.org/10.1598/RRQ.32.1.5>
- Shu, H., Chen, X., Anderson, R. C., Wu, N., & Xuan, Y. (2003). Properties of school Chinese: Implications for learning to read. *Child Development*, 74(1), 27-47.
<https://doi.org/10.1111/1467-8624.00519>
- Shufaniya, A., & Arnon, I. (2018). Statistical learning is not age-invariant during childhood: Performance improves with age across modality. *Cognitive Science*, 42(8), 3100-3115.
<https://doi.org/10.1111/cogs.12692>
- Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, 81, 105-120.
<https://doi.org/10.1016/j.jml.2015.02.001>
- Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, 44(6), 1321-1337. [https://doi.org/10.1044/1092-4388\(2001/103\)](https://doi.org/10.1044/1092-4388(2001/103))
- Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, 10(1), 1-8.
<https://doi.org/10.1186/1471-2202-10-21>
- Tong, X., & McBride, C. (2014). Chinese children's statistical learning of orthographic regularities: Positional constraints and character structure. *Scientific Studies of Reading*, 18(4), 291-308.
<https://doi.org/10.1080/10888438.2014.884098>

- Wang, M., & Geva, E. (2003). Spelling performance of Chinese children using English as a second language: Lexical and visual-orthographic processes. *Applied Psycholinguistics*, 24(1), 1–25. <https://doi.org/10.1017/S0142716403000018>
- Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, 5, 953. <https://doi.org/10.3389/fpsyg.2014.00953>
- Yang, X., & Meng, X. (2020). Visual processing matters in Chinese reading acquisition and early mathematics. *Frontiers in Psychology*, 11, 462. <https://doi.org/10.3389/fpsyg.2020.00462>
- Yin, L., & McBride, C. (2015). Chinese kindergartners learn to read characters analytically. *Psychological Science*, 26(4), 424-432. <https://doi.org/10.1177/0956797614567203>
- Yip, M. (2002). *Tone*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139164559>
- Zhou, N. (2005). How English as a Second Language Affects Chinese Students Giving Presentations During Class in US (Doctoral dissertation, Marietta College).
- Zhou, Y. (2014). Writing evolution initiated by Naxi-Sino Writing Contact. *Studies in Literature and Language*, 8(3), 156–159. <https://doi.org/10.3968/n>

Chapter 3: Sensitivity to Word Endings as Probabilistic Orthographic Cues to Lexical Stress Among English as Second Language Learners

Abstract

Assigning stress to polysyllabic words is a crucial aspect of reading aloud in English. Previous research demonstrated that native English speakers are sensitive to word endings as probabilistic orthographic cues to stress assignment (e.g., Arciuli & Cupples, 2006; Kelly et al., 1998). However, little is known about if second language (L2) learners of English are sensitive to word endings as cues to lexical stress. The current study investigated whether native Chinese-speakers learning English as a second language (ESL) are sensitive to word endings as probabilistic orthographic cues to lexical stress. Our ESL learners demonstrated sensitivity to word endings as cues in a stress assignment task and a naming task. With the increase in language proficiency, ESL learners responded more accurately in the stress assignment task. Moreover, stress position and language proficiency moderated the strength of the sensitivity, with a trochaic bias and better proficiency leading to better sensitivity in the stress assignment task. However, as language proficiency increased, participants' naming speed became faster in iambic but slower in trochaic reflecting the learners' fledgling knowledge about the specific stress patterns associated with varying orthographic cues, especially in a demanding naming task. Taken together, the evidence from our ESL learners fits in the proposed statistical learning mechanism, that is, L2 learners are able to implicitly extract statistical regularities from linguistic materials, the orthographic cues to lexical stress in our study. Stress position and language proficiency both play a role in developing this sensitivity.

Keywords: word endings; orthographic cues; lexical stress; second language; statistical learning

Sensitivity to Word Endings as Probabilistic Orthographic Cues to Lexical Stress Among English as Second Language Learners

Lexical stress refers to the opposition of strong and weak syllables within words and is exhibited only in polysyllabic words. Assigning stress to polysyllabic words is a crucial aspect of reading aloud in English since words cannot be articulated until stress is assigned. Previous research demonstrated that native English speakers are sensitive to word endings as probabilistic orthographic cues to stress assignment (e.g., Arciuli & Cupples, 2006; Kelly et al., 1998). In the line of research concerning second language (L2) learners, converging evidence suggests that L2 learners of English are able to utilize orthographic cues in English to aid word recognition, phonological and morphological processing (e.g., Carrasco-Ortiz & Frenck-Mestre, 2014; Hayes-Harb et al., 2018; Vaid & Frenck-Mestre, 2002). For example, native French-speaking adults who learn English as a second language (ESL) were able to identify English words that contain orthographic cues (e.g., KICK, which contains English-specific digram CK) faster than unmarked words (e.g., DROP, which contains digram DR that are equally frequent in both English and French) (Vaid & Frenck-Mestre, 2002). However, little is known about whether L2 learners of English are sensitive to word endings as cues to lexical stress. Chinese, a representative tonal language, has distinct prosodic features compared to English. Studying whether Chinese English L2 learners are sensitive to word endings as cues to English lexical stress will contribute to our better understanding of implicit learning of statistical distribution of stress patterns in English L2 when the prosodic system of first language (L1) does not support this aspect of L2 learning.

Word endings as probabilistic orthographic cues to lexical stress

Previous research suggests that there are cues to stress in written English words (e.g., Arciuli & Cupples, 2006; Arciuli et al., 2010; Kelly et al., 1998). The association between word endings and lexical stress is probabilistic, due to the statistical regularities in the written language (Arciuli & Cupples, 2006). Specifically, the spellings of many word endings are statistically associated with particular stress patterns. For example, word endings such as *-ette*, *-que*, and *-umb* are strongly associated with second syllable stress (e.g., casSETTE; phySIQUE; sucCUMB), and *-ior*, *-urn*, and *-em* are strongly associated with first

syllable stress (e.g., SENior; REturn; PROBLEM). Readers who are sensitive to the cues to stress positions provided by the word endings may be able to leverage those regularities to guide their stress placement in written words. For instance, a reader who encounters the unfamiliar word *rennet* may recognize the ending sequence *-et* that would cue them to correctly identify the word's first-syllable stress pattern. Likewise, recognizing the ending sequence *-ee* in *alee* may cue them to determine that the word has second-syllable stress.

An influential dictionary analysis by Arciuli and Cupples (2006) provided an operational definition of word endings as those that are comprised of “the letter string beginning at the second phonemic vowel of a disyllabic word (and including any following consonants)”. For example, the ending of *planet* is *-et*. A key advantage of this definition is that it can be applied straightforwardly to nearly all disyllabic words in English, allowing for a comprehensive analysis of the extent to which written word endings are associated with stress. Arciuli and Cupples (2006)'s analysis confirmed that a wide range of word endings are associated with stress in English. Particularly, the strength of the association varies on a continuous scale. For example, 92% of words ending with *-et* have first-syllable stress, whereas this was only true of 62% of words ending with *-ew*; both endings are primarily associated with first syllable stress, but *-et* serves as a more reliable stress cue.

Kelly et al. (1998) was among the first to investigate skilled native language readers' sensitivity to ending spellings as cues to lexical stress in English. In a naming task, the words have either trochaic (first syllable stress) or iambic (second syllable stress) patterns and either contain or lack orthographic patterns indicative of iambic stress. For instance, the participants were asked to read iambic words like *cassette* and *cadet*. The former item contains an orthographic ending, *-ette*, associated with iambic stress, whereas the latter does not. Results showed that disyllabic words whose spellings are consistent with their stress positions were named more quickly and were pronounced with higher accuracy. In a separate lexical decision task, two groups of words were selected. In the first group, the word's final syllable was a reliable indicator of lexical stress assignment (e.g., the second syllable of *casSETTE* and the first syllable of *PELlet*). In the other group, the final syllable was an unreliable indicator of stress position (e.g., the

second syllable of caDET and the first syllable of PALette). Participants' lexical decision times were significantly shorter, and error rates significantly lower, for words in which orthographic endings were a reliable indicator of stress assignment. Later, Mundy and Carroll (2013) replicated these findings and moreover, with the same experimental design, the same pattern of results was shown among adults with dyslexia. Not surprisingly, participants with dyslexia responded significantly slower than their normal controls. Taken together, these previous works were instrumental in demonstrating that skilled and less skilled readers are both sensitive to the links between stress and spelling in English.

Arciuli and Cupples (2006) created nonwords with unique word endings. Native English-speaking adult skilled readers were asked to assign stress when reading these nonwords. The participants responded in predictable ways, showing their sensitivity to word ending cues. The participants tended to assign first-syllable stress when reading a nonword such as *plosure*, which has an ending that is strongly associated with first-syllable stress in English words. By contrast, the participants tended to assign second-syllable stress when reading a nonword, such as *feduct*, which has an ending strongly associated with second-syllable stress. However, one major weakness of this study is that the nonwords used were very much like real words of high frequency. Thus, a possibility cannot be ruled out that this performance may be heavily influenced by a process of analogy to high frequency known words such as *pleasure* and *conduct*.

Subsequent cross-linguistic research has revealed pervasive probabilistic orthographic cues in the final parts of polysyllabic words to stress in large-scale corpus analyses of Italian, Greek, Dutch, Spanish, and German, as well as English (Monaghan et al., 2016). The analysis of English, much larger than that conducted by Arciuli and Cupples (2006)—and with the addition of trisyllabic words—confirmed the rich orthographic cues to stress embedded in the “word endings” as defined by Arciuli and Cupples (2006). Subsequent studies have gathered consistent evidence for skilled readers' sensitivity to probabilistic orthographic cues in languages other than English (e.g., Italian: Burani & Arduino, 2004; Greek: Grimani & Protopapas, 2017; Russian: Jouravlev & Lupker, 2014)

Types of stress position

There are three representative lexical stress patterns in disyllabic English words. The first one is the strong–weak (SW) pattern, such as in the word WATer, is described as “trochaic”. The second is the weak–strong (WS) pattern, such as in the word deDUCT, is described as “iambic”. The third is the strong–strong pattern, such as in the word BASEBALL, is described as “spondaic” and is relatively rare in English. A commonly known hypothesis regarding stress placement is the *universal trochaic bias* hypothesis, first proposed by Allen and Hawkins (1979, 1980), which posits that the SW pattern is favored over other types of stress patterns in all languages. Researchers observed that L1 English-speaking children tend to first preserve stressed and final syllables in the truncated forms they produce. For example, when producing WSW words such as baNAna or toMAto, children showed a strong tendency to delete the first unstressed syllable to produce a SW, trochaic form. In addition, Fikkert (1994) investigated the development of prosodic sensitivity in L1 Dutch learners. Similar to the English pattern described above, Dutch children tend to reduce long word forms to SW forms during the early stage of phonological development. McGregor and Johnson (1997) further extended *trochaic bias* from words to nonwords and claimed that *trochaic bias* is a default strategy to assign lexical stress to disyllabic pseudowords in English (also see Ballard et al., 2012).

There are two major accounts for the *trochaic bias*. First, this bias reflects the statistical make-up of the language. The *trochaic bias* may simply reflect a greater exposure to this stress pattern (Jusczyk, et al., 1993). According to Cutler and Carter (1987), 78% of a corpus of 20,000 American English words begin with a strong syllable. Alternatively, it may reflect sensitivity to the physiological constraint involved in the production of WS words. Snow (2007) suggested that the physiological demand of producing a rising tone (as in WS words) is greater than producing a falling tone (as in SW words). Such physiological constraint is universal and responsible for the *trochaic bias*.

Mechanisms underlying sensitivity to cues from word endings to stress

Given the evidence that skilled readers are sensitive to the links between word endings and stress in English (e.g., Arciuli et al., 2010; Kelly et al., 1998), it is important to study how this sensitivity is acquired and developed. One well-known hypothesis suggests that statistical learning is

the underlying mechanism that can help explain the development of this sensitivity (e.g., Colombo et al., 2014; Sulpizio et al., 2013). Statistical learning refers to human beings' ability to extract regularities from environment implicitly (Saffran et al., 1996; Aslin & Newport, 2012). The central premise of statistical learning theory is that readers gradually and implicitly learn regularities through exposure to input (Perruchet & Pacton, 2006). In the case of acquiring the links between word ending and stress assignment, exposure to co-occurring spelling and stress pattern gradually shapes the sensitivity. More specifically, as readers encounter ample words with the same ending cues (e.g., COMet, JACKet, CARpet; caDET), they come to recognize broader probabilistic regularities embedded in these cues (e.g., the ending *-et* usually occurs in words with first syllable stress).

The statistical learning account is in line with the connectionist model of disyllabic word reading (e.g., Perry et al., 2010; Ševa et al., 2009), which operates on the statistical regularities present in the input to which they are exposed. The central focus of this model is to determine whether stress can be assigned to written words based solely on orthography. The model consists of slot-based letter units for orthographic input, hidden units that learn the association between words' orthography and stress patterns, and a single output unit identifying stress placement. The model does not contain stores of affixes or lexical entries; instead, cues relevant to stress placement are learned through a training process designed to mimic readers' exposure to written words (Ševa et al., 2009). As a result, the model draws on statistical regularities in the segmental orthography in English to assign stress, without a need to directly specify the relevant stress cues in advance.

Word endings as cues to lexical stress in L2

Most of the studies that examined sensitivity to word endings in lexical stress focused on L1 readers, leaving open the question whether L2 readers have such a sensitivity. Previous research has shown that the ability to detect statistical regularities also plays a role in L2 learning. For example, Kaufman et al. (2010) investigated the relation between statistical learning and L2 learning among native English-speaking adolescents of 16-17 years of age. Their results revealed a significant positive relation between statistical learning using serial reaction time and artificial grammar learning tasks and

performance in L2 French and German. Participants' French or German levels were based on their scores on the exams for the General Certificate of Secondary Education (GCSE) in French or German which include written, listening, speaking, and reading sessions. Furthermore, Siyanova-Chanturia and Janssen (2018) showed that late L2 learners of English from a wide range of L1 backgrounds were sensitive to the frequency with which phrases occur in English, for example, they processed multiword expressions (e.g., idioms and collocations) faster and more easily than novel phrases in a naming task. In addition, a few studies showed that statistical learning contributes to L2 reading (e.g., Frost et al., 2013; Onnis et al., 2016). Frost et al. (2013) tested native English-speaking adults learning Hebrew L2 and showed that non-linguistic statistical learning was significantly associated with Hebrew reading.

If statistical learning is the mechanism behind learners' sensitivity to word endings as stress cues, L2 learners may also be able to develop such sensitivity. To our knowledge, the only study related to this question was by Chan and Leung (2014). In the training phase, native Cantonese adult speakers were asked to learn some Spanish words without being informed of the regularities contained in the words, for example, words ending in 'o' have their stress on the penultimate syllable (the second to last syllable), and words that end in 'ar' have their stress on the last syllable. In the test phase, participants completed a two-alternative forced choice pronunciation judgement task. Participants were instructed to listen to two words and were asked to choose the one that 'sounds better' to them. These novel language learners displayed knowledge of stress assignment for novel *-ar* and *-o* ending words and transferred such knowledge to novel items. Moreover, they were unaware of the underlying regularities based on a post-hoc awareness task. It seems that learners acquired ending-phoneme-to-stress regularities in the new language implicitly, consistent with the statistical learning hypothesis. In Chan and Leung (2014), however, only two regularities of mappings between word endings and lexical stress were involved, which made it hard to generalize to a set of broader probabilistic regularities. Furthermore, participants showed learning of the target stress patterns shortly after training in the experimental setting. It is unclear, however, whether these observations can be applied to more complex input materials characteristic of

more typical language learners who are in the process of learning and improving their L2 rather than those trained to acquire limited knowledge about the new language in a laboratory setting. From this perspective, it is thus interesting to investigate to what extent the acquisition of untaught regularities can occur in complex input materials among typical L2 learners.

Given the scarcity of research on L2 learners in this field, the current study investigated the ability to use word endings as probabilistic cues to lexical stress based on broader probabilistic regularities and among those who acquired L2 in normal learning settings through years of practice. We specifically focused on ESL learners with Chinese as L1. Chinese is a representative non-alphabetic, tonal language whose prosodic features are different from English. The connection between word endings and lexical stress exists commonly among alphabetic, stressed languages, and is not explicitly taught to these L2 learners. There is clearly an absence of explicit discussion of the regularities in English L2 curriculum for these learners (People's Education Press, 2021). Thus, English L2 learners with Chinese L1 provide a unique testing case for examining the implicit acquisition of an unfamiliar linguistic regularity associated with orthography.

Role of L2 proficiency

The proficiency of ESL learners plays an important role in various aspects of English learning including phonological and orthographic processing. According to Strange (2011), L2 phonological processing is less automatic and requires more attentional resources compared to L1 phonological processing. When L2 learners' proficiency increases, the processing becomes more automatic and demands fewer attentional resources. White et al. (2017) showed that highly proficient L2 learners of English with French L1 are able to discriminate the English-specific /h/-/Ø/ contrast, similarly to native English speakers, even if /h/ does not exist in their French L1. However, this is not the case for L2 learners with lower proficiency. English proficiency has also been shown to be related to L2 orthographic skills, for instance, only when L2 English proficiency achieves a certain level can the transfer from L1 French orthographic knowledge (i.e., distinguish the correct spelling of a word from its pseudohomophone) to L2 English orthographic skills occurs (e.g., Commissaire et al., 2011). Zhang et al.

(2017) tested Chinese ESL learners' sensitivity to stress typicality (i.e., nouns tend to be stressed more often on the first syllable, whereas verbs are more likely to be stressed on the second syllable). Their results showed that while highly proficient ESL learners had this sensitivity, just like native English speakers, intermediate ESL learners did not. The researchers explained their findings from the perspective of statistical learning of English. Statistical learning can occur as reading experiences increase. Stress assignment is learned by mapping orthographic cues onto stress positions. Given the importance of L2 proficiency in orthographic processing, in the current study, we investigated whether language proficiency affects ESL learners' sensitivity to word endings as probabilistic orthographic cues to English lexical stress.

The present study

The present study sought to examine sensitivity to word endings as cues to L2 lexical access by speakers whose L1 and L2 are typologically different. We focused on ESL learners who have Chinese as their L1, a tonal language without phonetic stress (Duanmu, 2007) and English as L2, a non-tonal, but stress-timed language. Chinese tones work very differently from stress in English. For single-syllable words in Chinese, a tone is assigned to every syllable word, for example, 球 /qiú/, which means "ball". Compound words are prevalent in Chinese in which two or more syllables are combined to form a word. In a compound word, each syllable carries a tone without the need to decide on the specific location of the tone on a specific syllable, for example, 房型 /fáng xíng/, which means "room type". In learning English L2, our native Chinese speakers may face the challenge of an unfamiliar linguistic feature compared to those learners whose L1s are involved with the similar demand of stress assignment among disyllabic or multisyllabic words (e.g., Italian: Sulpizio & McQueen, 2012; Spanish: Gutiérrez-Palma & Palma-Reyes, 2008; Russian: Molczanow et al., 2013). The distinct phonetic features between these two languages allow us to investigate the acquisition of an unfamiliar L2 regularity. Two tasks were employed in this study, one is a stress assignment task in which participants were asked to choose either the first or second syllable to assign the stress, and the other one was a naming task in which participants were asked to read

words aloud.

In our stress assignment task, we employed nonwords to minimize the confound from participants' existing vocabulary knowledge. Concerning the cognitive processes involved in stressing assignment with unknown words without reference to lexically stored information, previous research focusing on English and Italian proposed two different types of distributional knowledge that the readers rely on to aid their stress assignment, based on a connectionist view of reading (e.g., Sulpizio et al., 2013; Paizi et al., 2011; Colombo, 1992; Rastle & Coltheart, 2000). The first type is the distribution of the stress patterns in the lexicon (e.g., English has more trochaic than iambic patterns). This knowledge may induce readers, following a default stress bias, to assign the most common pattern to unknown words (Colombo, 1992; Rastle & Coltheart, 2000). The second type of distributional knowledge rests on the cues provided by certain orthographic and phonological units to stress assignment, with word ending being a strong predictor of the stress pattern (e.g., Arciuli et al., 2010; Kelly et al., 1998).

In our naming task, we employed real, familiar words due to our participants' L2 status. Several cognitive processes are likely to be involved in stress assignment in this task. First, readers can resort to lexically stored information for the stress position. Second, readers may utilize the same two different types of distributional knowledge to decide the stress assignment used on nonwords. The different processes between our stress assignment task and the naming task could be: 1) lexically stored information may be used to assign the stress in the naming task, but not in the stress assignment task; 2) the naming task requires an overt phonological production including stress production, whereas the stress assignment task does not demand such overt production.

In sum, the stress assignment task tapped into ESL learners' implicit knowledge about lexical stress patterns in English, without the involvement of these patterns in overt pronunciation. If word endings as probabilistic cues to lexical stress are learnt implicitly, we expect that word endings that reliably predict stress patterns should show advantages than endings with unreliable cues in the stress assignment task. The naming task, on the other hand, entails overt pronunciation of the stress patterns and readers may tap into lexically stored information as well as the distributional knowledge about stress

patterns. Similar results may be shown across the two experimental tasks, that is, ESL learners will demonstrate their sensitivity to word ending cues in their stress assignment as well as in their reading aloud. Alternatively, they may show a relatively weaker sensitivity in the reading aloud experiment given the potential reliance on the lexical information and higher demand for an overt construction of phonological output. We were also interested in whether ESL learners' sensitivity to word ending cues differs between trochaic and iambic patterns. Since there is a bias toward first syllable stress in English (CELEX Database: Baayen et al., 1993), that is, there are far more English words with first syllable stress (Cutler & Carter, 1987), we expected that ESL learners encountered more trochaic patterns in learning English and would be more likely to assign stress to first syllables, so a trochaic advantage may occur. Furthermore, we assessed learners' English proficiency to see if L2 proficiency plays a role in their sensitivity to word ending cues. Given that language proficiency facilitates L2 orthographic processing (e.g., Commissaire et al., 2011; Zhang et al., 2017), we expected that in both tasks, learners with higher L2 proficiency would show higher sensitivity to word endings as stress cues than those with lower L2 proficiency. In sum, the following three main research questions guided the current study:

- (1) Are Chinese ESL learners sensitive to word endings as cues of English disyllabic word stress?
- (3) Will these ESL learners' sensitivity to word ending cues differ across stress positions (trochaic vs. iambic)?
- (2) Does language proficiency play a role in L2 learners' sensitivity to word ending cues?

Experiment 1: Stress assignment task

The goal of this experiment was to see if ESL learners are sensitive to word endings as probabilistic orthographic cues in stress assignment. More specifically, we investigated whether ESL learners are able to utilize word endings' reliability to decide lexical stress in nonwords.

Participants

Based on G*Power 3 analysis (Faul et al., 2007), a minimum sample size of 36 was recommended for multiple linear regression analysis with 80% power to detect a medium effect size (0.3) and an alpha level of 0.05. We recruited 40 college students (24 females, mean age = 25.85 years, $SD =$

3.28) enrolled in a Mid-Atlantic university, who are ESL learners with Chinese as their native language. All the participants had not lived in an English-speaking country for more than five years at the time of experimental testing. According to their responses in the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian & Kaushanskaya, 2007), these participants had started learning English at the mean age of 8.2 years ($SD = 2.4$) and became fluent in English at the mean age of 17.8 years ($SD = 4.2$). The majority (87.5%) were graduate students and others were undergrads. The average self-rating of English proficiency was 7.6 out of 10 with a range from 5.8 to 9.3. In terms of their experiences with any other languages beyond English, two of them reported learning basic Japanese previously, three of them were able to speak basic Korean, and one of them had experience learning Spanish. These participants showed a similar pattern in their performance as other participants who did not report any other language experiences. Eighteen participants spoke Northern Mandarin and 22 spoke Southern Mandarin. None of them had vision, hearing, language or learning disability. We measured participants' English proficiency using an English lexical decision task (LexTALE; Lemhöfer & Broersma, 2012) and a written cloze test (Brown, 1980).

LexTALE

The LexTale is an English lexical decision task, commonly used to measure ESL learners' English proficiency. The participants were asked to judge whether a sequence of letters was a real English word or a pseudoword. They were instructed to press the right button to indicate if it was a real word, and to press the left button if they believed that it was not a real word. They were reminded that they did not necessarily need to know the meaning of the stimuli to press the right button, as long as they believed that it was a real word. The task consisted of 60 trials and the ratio of real words to pseudo-words was 2:1. The participants were required to answer all questions without time limit.

Cloze Test

The cloze test measured ESL learners' vocabulary, spelling and an understanding of the English language, also commonly used as an index of English L2 proficiency. Participants were asked to fill in 50 blanks distributed throughout a reading passage while reading within a limited time (20 minutes). This

test has been demonstrated to have high reliability and validity across a broad range of learner groups.

Stress Assignment Task

Materials and design

Arciuli and Cupples (2006) and subsequent studies treated word ending's reliability as a cue to the stress assignment as a categorical variable (reliable vs. unreliable, using 50% as the borderline). Reliability here is defined as to what extent a word ending can successfully predict the stress position (either trochaic or iambic). However, this approach failed to attend to the more fine-grained variation along the continuum of the reliability. Our design provided a more nuanced investigation of readers' sensitivity to word ending cues using a continuous scale of the reliability of word endings. Eighty two-syllable pseudowords were constructed, and each of them had a unique ending. Forty of them had trochaic pattern, and the remaining 40 had iambic pattern. The word ending stress information was obtained from the corpus analysis of Arciuli and Cupples (2006). Trochaic endings are those endings who are more than 50% of chance to be associated with the trochaic pattern, and iambic endings are those that are more than 50% of chance to be associated with the iambic pattern. For trochaic endings, their reliability ranges from 54.55% to 100% with an average of 76.50%. For iambic endings, their reliability ranges from 51.85% to 100% with an average of 76.67%. Although our word ending's reliability was treated as a continuous variable, we roughly divided pseudowords into 5 intervals with equal items (i.e., 50-60%, 60-70%, 70-80%, 80-90%, 90-100%) for both trochaic and iambic endings to make sure that our pseudowords were roughly evenly distributed across the reliability continuum. As a result, there were 8 pseudowords in each interval. Trochaic and iambic endings were matched on word length ($p > .1$). Within trochaic and iambic endings, separately, the five intervals were also matched on word length ($p > .1$). When constructing the pseudowords, we carefully chose those word beginnings which did not provide cues to word stress based on Arciuli and Cupples (2007) to minimize the possible influence of word beginnings. Word beginning is defined as the letter string up to and including the first vowel or vowel cluster (Arciuli & Cupples, 2007). There were 40 unique word beginnings in our design, and they followed different syllable structures (i.e., "CV", "CCV" or "V"), see supplementary Table A for the description of the different structures of word

beginnings. The nonwords with the same beginning were separated by at least three other nonwords to avoid practice effect.

To ensure that none of the pseudowords had an obvious real-word analogy that might encourage participants to think of this particular word, we asked 7 native English speakers who were not involved in this experiment to rate each of the 80 experimental pseudowords from “1” to “5”, where “1” indicated that this pseudoword was very similar to a real word, and “5” was very dissimilar to a real word. With a mean of 3.81 and a standard deviation of 0.47, these pseudowords were not similar to real words. Also, native speakers’ ratings did not differ across stress positions and across reliability intervals ($p > .1$).

Procedure

Participants completed the tasks in a single session, held in a quiet room. The stress assignment task was run on PsychoPy v2022.1.3 (Peirce et al., 2019). Participants were given nonsense words one at a time and asked to choose whether more emphasis should be put on the beginning or the end of each nonsense word in reading it aloud, press “*F*” if they thought the beginning should be emphasized and press “*J*” if they thought the ending should be emphasized. There was no time limit for response, but participants were instructed to make response as soon as they had an answer, and interstimulus delay was response contingent. The whole experiment took around 8 minutes. After the stress assignment task, participants were instructed to complete LexTALE (5 minutes) and Cloze Test (20 minutes). At the end of the experiment, participants were asked to fill out the LEAP-Q. We also surveyed participants’ explicit knowledge about word endings as lexical stress cues (i.e., if they were explicitly taught and aware of how word endings cued English stress).

Coding

For accuracy of the stress assignment task, if participants’ response was consistent with the correct position for the specific ending type, we coded as “1”, otherwise, as “0”.

Results and Discussion

The overall accuracy across all the conditions was 0.60, significantly above chance level ($p < .001$). For trochaic, the average accuracy across reliabilities was 0.62, and for iambic, the average

accuracy across reliabilities was 0.58. As Figure 1 depicted, as endings' reliability increased, participants' accuracy increased, and trochaic showed a faster increasing rate compared to iambic. We conducted further analysis to confirm our observations.

<Insert Figure 1 here>

We used mixed-effect logistic regression model (Barr et al., 2013) to examine the effects of ending's reliability (ranging from 50% to 100%, continuous), stress position (trochaic vs. iambic, categorical) in the statistical programming language R (R Core Team, 2017). Our dependent binomial variable was success in a single test trial. Our model included fixed effects for ending's reliability, stress position (contrast coded, -0.5 vs 0.5), and the interaction among these two independent variables, a random intercept for participants, and a random intercept for test items (see Table 1). The two-way interaction of stress position and endings' reliability was significant ($b = -4.24$, $SE = 1.23$, $z = -3.44$, $p = 0.0006$). As endings' reliability increased, participants' accuracy rate increased significantly faster in trochaic compared to iambic. The main effect of stress position was significant ($b = 2.76$, $SE = 0.99$, $z = 2.79$, $p = 0.005$). Overall, participants performed significantly better in trochaic compared to iambic. The main effect of endings' reliability was also significant ($b = 5.57$, $SE = 0.81$, $z = 2.89$, $p = 0.004$). As endings' reliability increased, participants' accuracy increased significantly.

<Insert Table 1 here>

Role of L2 proficiency

The average score of LexTALE is 80.75 with a standard deviation of 8.58, and the average score of the cloze test is 41.88 (out of 50) with a standard deviation of 7.47. To examine the effect of English language proficiency on ESL learners' sensitivity to the stress cues provided by the orthographic endings, scores from the two written-based proficiency measures (LexTALE, cloze) were both transformed to z-score, and then we took the average of these two z-scores as the indicator of their English proficiency. This model included fixed effects for ending's reliability, stress position (contrasted coded, -0.5 vs 0.5), English proficiency, and the interactions among the three independent variables, a random intercept for participants, and a random intercept for test items (see Table 2). The interaction between English

proficiency and ending's reliability was significant ($b = 4.95$, $SE = 0.47$, $z = 8.12$, $p < .001$). As word ending's reliability increased, more proficient ESLs' accuracy increased faster than less proficient peers (see Figure 2). In addition, the interaction between position and ending's reliability was significant ($b = -5.23$, $SE = 1.31$, $z = -3.99$, $p = 6.73e-05$). Participants' accuracy increased faster as word endings' reliability increased in trochaic than iambic. The main effects of position, reliability, and English proficiency were all significant ($b = 3.46$, $SE = 1.04$, $z = 3.32$, $p = 0.0009$; $b = 2.43$, $SE = 0.79$, $z = 2.85$, $p = 0.004$; $b = -1.93$, $SE = 0.32$, $z = -2.60$, $p = 0.009$, respectively), with accuracy higher in trochaic than iambic, and higher in more reliable than less reliable conditions. More proficient ESL learners were more sensitive to word endings as stress cues compared to less proficient ESL learners. The two-way interaction between position and English proficiency, and the three-way interaction between position, reliability, and English proficiency were both not significant ($p > .1$)

In the post-experiment survey, none of the participants indicated that they received any explicit instruction or information about the regularities underlying word ending and lexical stress association, nor were they aware of these regularities.

<Insert Table 2 here>

<Insert Figure 2 here>

Our results showed that the ESL learners in this experiment, whose L1 was different from English in terms of suprasegmental phonology, had sensitivity to word endings as cues to lexical stress, consistent with the previous results from Chan and Leung (2014). This finding suggests that L2 learners are able to utilize statistical information from word endings to guide stress assignment even when their L1 has very distinct phonetic features compared to L2. Our results also showed that participants were more sensitive to word ending cues in trochaic than iambic, likely due to the higher prevalence of trochaic patterns in English. This finding lends support to the view that trochaic bias is a default strategy to assign lexical stress to disyllabic pseudowords (McGregor & Johnson, 1997; Ballard et al., 2012). Furthermore, L2 proficiency played an important role that more proficient ESL learners grasping the word ending cues better than less proficient ESL peers. We suggest that the better sensitivity by more proficient ESL

learners may be due to their larger English lexicon size. Vocabulary size has been shown closely associated with statistical learning, readers with larger lexicon size tend to have higher statistical learning ability (e.g., Evans et al., 2009; Kemény & Lukács, 2021; Reuter et al., 2018; Hedenius et al., 2011). Our finding supports and extends previous findings on the importing role of L2 proficiency in learning wording ending cues to stress.

Note that this stress assignment task did not require a participant's phonological representation to be accessed or constructed explicitly, it is unknown whether this sensitivity will be shown in an overt phonological output as well. Our second experiment utilized a naming task to further investigate ESL learners' sensitivity to word endings cues in an overt, reading aloud context.

Experiment 2: Naming

The aim of this experiment was to determine whether the endings of disyllabic words served as stress cues in ESL learners, in other words, whether the reliable stress cues were processed differently from those less reliable ones during reading aloud when explicit phonological access and construction are mandated. Naming latency and stress placement accuracy in reading words aloud served as dependent measures.

Participants

We recruited another group of 40 college students (females $N = 28$, mean age = 22.51 years, $SD = 2.78$) enrolled in the same university as those in Experiment 1, who are ESL learners with Chinese as their native language. All the participants have not lived in an English-speaking country for more than five years at the time of experimental testing. According to their responses in the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian & Kaushanskaya, 2007), these participants had started learning English at the mean age of 8.9 years ($SD = 1.7$) and became fluent in English at the mean age of 16.2 years ($SD = 4.9$). The majority (72.5%) were graduate students and others were undergrads. The average self-rating of English proficiency was 7.2 out of 10 with a range from 5.4 to 8.9. Four of them reported learning basic Japanese previously, and one of them was able to speak basic Italian. The performance of these five participants did not differ from others. Sixteen participants spoke Northern

Mandarin and 24 spoke Southern Mandarin. None of them had vision, hearing, language, or learning disability. English proficiency was measured through an English lexical decision task (LexTALE; Lemhöfer & Broersma, 2012) and a written cloze test (Brown, 1980), as in Experiment 1.

Naming Task

Materials and design

We have two variables; one is word ending's reliability (reliable vs. less reliable) and the other one is stress position (trochaic vs. iambic). The entire set of stimuli comprises 75 disyllabic words, and each word has a unique ending. Of the 75 disyllabic words, 38 of them are iambic, and 37 are trochaic. Among the 38 iambic words, 18 of them have reliable and 20 less reliable word endings. Among the 37 trochaic words, 19 of them have reliable and 18 less reliable word endings. The word ending stress information was obtained from the corpus analysis of Arciuli and Cupples (2006). We consider reliable word endings as those that are 100% associated with first or second-syllable stress. Less reliable word endings are those that are 50% to 75% associated with first or second syllable stress.

There are four experimental conditions: reliable trochaic, less reliable trochaic, reliable iambic, and less reliable iambic. According to the English Lexicon Project (Balota et al., 2007), words across these four conditions were matched on word length, word frequency, phonological neighborhood size, orthographic neighborhood size, phonological neighborhood frequency, orthographic neighborhood frequency, and bigram frequency ($ps > .1$, see supplementary Table B for lexical characteristics of words across four conditions). The *Log_Freq_HAL* of all the words we selected was higher than 8.5. Based on the English Lexicon Project (Balota et al., 2007), the average of *Log_Freq_HAL* is 4.49, and *Log_Freq_HAL* higher than 8.5 is considered high-frequency English words. The initial consonants across four conditions were controlled for the number of stops, fricatives, affricates, nasals, and liquids, see supplementary Table C for articulation manners of initial consonants across four conditions.

Prior to the formal experiment, we asked five ESL learners who did not participate in this experiment to rate familiarity of each word. The familiarity rating was overall high ($M = 3.92$ out of 5) and did not differ significantly across the four conditions ($p > .05$).

Procedure

At the beginning of the experiment, participants saw instructions “In this experiment, you will see some English words. Please read words aloud as quickly, clearly, and accurately as possible. Please try to avoid coughing or saying uh or um. Please press the space bar to start practice!” There were five practice items before the formal experiment. A Toner TC-777 microphone was used as the voice key of the DMDX program for recording participants’ reaction time while participants’ responses were recorded via a voice recorder. After each successive trial, the following trial began 200ms after a response was given, or after 1500ms with no response. The naming task took around 8 minutes. After the naming task, participants were instructed to complete LexTALE (5 minutes) and the cloze test (20 minutes). At the end of the experiment, participants were asked to fill out the LEAP-Q and complete a survey on whether they were explicitly taught and aware of how word endings cued English stress.

Coding

In addition to naming latency, we also coded participants’ accuracy of stress placement and word segment during reading words aloud. For stress placement, if the participants stressed the correct syllable, we coded it as “1”; if participants stressed the wrong syllable, we coded it as “0”. For word segment, if the pronunciation of the segment was correct, we coded it as “1”, and if it was incorrect, we coded it as “0”. In both stress placement and word segment pronunciation, we coded as “99” if participants completely did not know how to read the word, and these data (1.53%) were removed before analysis.

Results and Discussion

For naming latency, incorrect stress placement trials were removed. Furthermore, those naming latency above or below 2.5 SDs were removed. Table 3 shows the mean response time, mean stress placement accuracy, and SD for each condition.

<Insert Table 3 here>

Naming latency

We used mixed-effect linear regression model to examine the effects of word ending cue’s reliability (reliable vs. less reliable, categorical) and stress position (trochaic vs. iambic, categorical) in

the statistical programming language R (R Core Team, 2017). Our dependent variable was naming latency for each word. Our model included fixed effects for ending cue's reliability (contrast coded, -0.5 vs 0.5), stress position (contrast coded, -0.5 vs 0.5), and the interaction among these two independent variables, a random intercept for participants and a random intercept for test item (see Table 4). The main effect of reliability was significant ($b = -22.32$, $SE = 8.58$, $t = -2.60$, $p = 0.011$). Participants responded significantly faster in reliable than less reliable condition. The interaction between stress position and ending cue's reliability and the main effect of position were both not significant ($ps > .1$).

<Insert Table 4 here>

We further treated reliability within 50%-75% as a continuous variable and reran the analysis of the less reliable condition. **Our results showed that the main effect of reliability was significant ($b = 188.83$, $SE = 89.63$, $t = 2.11$, $p = 0.035$), participants read the words faster when the word ending was a more reliable indicator of stress position.** The interaction between reliability and position, and the main effect of position were not significant ($ps > .1$).

Stress placement accuracy

We used the same statistical approach as in the response times analysis, with the same independent variables and interactions except that accuracy data were used instead of the response data. The main effects of reliability, position and the interaction between position and reliability were all not significant ($ps > .1$). After we combined stress placement accuracy and segment accuracy scores, the patterns did not change.

To address whether segmental accuracy affects stress accuracy, we re-coded the naming accuracy in a more fine-grained manner. We added scoring 0.50 for those pronunciations that were correct for more than half of the phonemes/sounds. We checked the correlation between segment accuracy and stress assignment accuracy. Our results showed a positive, significant, and moderate correlation ($r = 0.28$, $p < .001$). This result suggests that if the learners struggle with pronouncing graphemes correctly, their ability to accurately assign stress is likely to be impacted.

Effect of dialect

Previous studies suggest that the dialect used by Chinese participants may affect English stress learning (e.g., Guo et al., 2017, Guo, 2022). We further ran analyses after adding dialect (Northern vs Southern) as a variable in both Experiments 1 and 2. We did not find dialect interacting with other variables (position, reliability, and English proficiency), nor did we find the main effect of dialect (all $ps > .1$). This result seems to suggest that the Chinese dialect background may have little impact on our participants' stress sensitivity in English L2, however, caution is needed in interpreting this result given the small sample size in this analysis.

Effect of grapheme-to-phoneme mapping

To address whether regularity and consistency of grapheme-to-phoneme mapping would affect naming latency and stress accuracy, we further checked our stimuli in the naming experiment based on the grapheme-to-phoneme correspondence databases (Gontijo et al., 2003; Ziegler et al., 1997). We found that in the reliable-trochaic condition, 9 of the words are grapheme-to-phoneme inconsistent, in the reliable-iambic condition, 11 are inconsistent, in the less reliable-trochaic condition, 8 are inconsistent, and in the less reliable-iambic condition, 6 are inconsistent. For grapheme-to-phoneme regularity, in the reliable-trochaic condition, 4 of the words are grapheme-to-phoneme irregular, in the reliable-iambic condition, 5 are irregular, in the less reliable-trochaic condition, 5 are irregular, and in the less reliable-iambic condition, 4 are irregular. We included consistency, regularity, and the interaction between them as predictors, where naming latency and stress accuracy served as dependent variables, separately. Our results showed that for each model in every condition, the interaction between consistency and regularity, the main effect of consistency, and the main effect of regularity were all not significant ($ps > .1$).

Role of L2 proficiency

The average score of LexTALE is 78.78 with a standard deviation of 8.51, and the average score of the cloze test is 42.35 (out of 50) with a standard deviation of 8.23. To examine the effect of English language proficiency on ESL learners' sensitivity to the stress cues provided by the orthographic endings, scores from the two independent proficiency measures (LexTALE, Cloze test) were first transformed to z-scores (as continuous variables), and then the average of the two z-scores as the indicator of English

proficiency was added to the model.

Naming latency. This model included fixed effects for ending cue's reliability, stress position (dummy coded), English proficiency, and the interactions among the three independent variables, a random intercept for participants, and a random intercept for test items. Naming latency was the outcome variable (see Table 5). The interaction between position and English proficiency was significant ($b = 26.29$, $SE = 10.57$, $t = 2.49$, $p = 0.013$). As ESL learners' proficiency increased, participants responded faster in iambic while slower in trochaic (see Figure 3). The main effect of reliability was significant ($b = -22.23$, $SE = 8.57$, $t = -2.59$, $p = 0.012$). As reliability increased, ESL learners responded faster. The interactions between position, reliability, and English proficiency, and between reliability and position, between reliability and English proficiency, and the main effect of position were all not significant ($p > .1$).

<Insert Table 5 here>

Stress placement accuracy. We used the same statistical approach as in the response times analysis, with the same independent variables and interactions except that accuracy data were used instead of the response data. None of the main effects or interactions was significant ($ps > .05$). In the post-experiment survey, none of the participants indicated that they received any explicit instruction or information about the regularities underlying word ending and lexical stress association, nor were they aware of these regularities.

Our naming task showed that as word endings' reliability increased, participants responded faster, suggesting that ESL learners are sensitive to word endings as stress cues at the phonological output level. Participants responded significantly faster as English proficiency was increased, consistent with previous findings that the naming latency for highly proficient L2 learners is shorter (e.g., Issa et al., 2002) likely due to their more rapid lexical activation. Participants' naming latency, however, was getting slower in trochaic while it was faster in iambic as English proficiency was increased. See the general discussion for this surprising finding. Together with the stress assignment task, the two experiments made a strong case that ESL learners are sensitive to word endings as cues to lexical stress.

General Discussion

The goal of the present study was to investigate if Chinese ESL learners are sensitive to word endings as cues to stress assignment in English disyllabic words. We also examined if this sensitivity differs across stress positions (trochaic vs. iambic), and whether L2 proficiency plays a role. Several important findings were obtained. First, Chinese ESL learners are sensitive to word endings as probabilistic cues to English lexical stress. As word endings' reliability increased, learners' accuracy increased in the stress assignment task, and they responded faster in the naming task. Moreover, this sensitivity was moderated both by stress position and English proficiency. Specifically, when word ending's reliability increased in the stress assignment task, participants' accuracy increased faster in trochaic than iambic, and more proficient ESL learners' accuracy increased faster than less proficient peers. As ESL learners' proficiency increased, participants' naming latency became slower in trochaic while it became faster in iambic. Overall, more proficient ESL learners responded faster in the naming task and more accurately in the stress assignment task compared to less proficient peers.

Effect of reliability of word ending cues

Consistent with previous studies from English L1 speakers (e.g., Kelly et al., 1998; Smith et al., 1982; Zevin & Joanisse, 2000; Arciuli & Cupples, 2006; Arciuli et al., 2010) and L2 learners (Chan & Leung, 2014), our study showed that Chinese speakers whose L1 is different from English at the suprasegmental phonology are sensitive to word endings as cues to lexical stress of disyllabic words in English. Although these ESL learners were never explicitly taught these regularities, they demonstrated implicit statistical learning. Previous research suggests that the average LexTale rate of proficient learners of English is about 71% (Lemhöfer & Broersma, 2012), our participants received about 80% accuracy rate in LexTale, thus could be considered proficient ESL learners. Our findings provided evidence that proficient ESL learners are sensitive to the mapping between orthography to suprasegmental aspect of phonology, such as lexical stress. Such a sensitivity could be explained from the perspective of an implicit statistical learning. As learners gain experiences with the written words in L2, they become sensitive to

word endings as stress cues. Stress assignment can be learned by mapping orthographic cues onto stress positions. After encountering a number of words with the same ending cue, language learners implicitly learn the association between the word ending and stress position. This account is in line with the L2 acquisition model proposed by Schmidt (2010), specifically, implicit learning of generalized regularities may take place through a basic human learning mechanism that automatically detects patterns across stimuli, leading to unaware knowledge of rules, often hard to verbalize. For example, readers learn that in the word *conscribe*, ending *-ibe* is associated with second syllable stress, and for the word *subscribe*, ending *-ibe* is also associated with second syllable stress. After some cases, readers will learn that ending *-ibe* is always associated with second syllable stress beyond particular instances.

Role of stress position

Ninety percent of the disyllabic nouns follow a trochaic stress pattern in English (Arciuli, 2017). Given the dominance of the trochaic pattern, one would expect that words with this stress pattern are processed more easily. McGregor and Johnson (1997) proposed that trochaic bias is a default strategy to assign lexical stress to disyllabic pseudowords in English (also see Ballard et al., 2012). Our findings support a trochaic advantage in the stress assignment using nonword materials. We did further analysis of our data and found that consistent with McGregor and Johnson (1997) and Ballard et al. (2012), participants assigned more first syllable than second syllable stress across the range of reliabilities in the current study (first syllable: 55.38%, second syllable: 44.62%). The Chinese tone pattern that may influence English L2 learners' trochaic bias could be the neutral tone in the second syllable of a two-syllable word in Chinese such as 孩子/ hái zi/, which means “child”. In 孩子, 子 carries the neutral tone, and 孩 is stressed, resulting in a stressed-unstressed pattern. A neutral tone has considerable vowel reduction (Chao, 1968) and weaker articulatory strength (Chen & Xu, 2006). The Chinese neutral tone pattern resembles the English trochaic pattern, which may have an influence on these learners' trochaic bias.

The trochaic bias, however, was not shown in the naming task. A possible explanation is inspired

by the Dual-Route Computational Model (DRC) proposed by Rastle and Coltheart (2000). According to the DRC model, the access of lexical stress for highly familiar words is through a lexical route while for unfamiliar/nonwords is through a sub-lexical route. In the lexical route, the lexical representation of familiar words is first activated in the orthographic lexicon. Then the corresponding phonological lexicon including lexical stress information is retrieved when translating the word from print to speech. In the sub-lexical route, on the other hand, readers follow a phonological assembly procedure and are likely to adopt the dominant stress pattern of English disyllabic words to assign stress to first syllable. Our stress assignment task involved nonword materials while the naming task utilized real, highly familiar words, we speculate that the participants may utilize a lexical route in the naming task whereas a sub-lexical route may be adopted in the assignment task. That is, in reading the real and familiar words, readers are likely to rely on the lexical information in addition to the word ending cues to decide on the stress position. In the stress assignment task, readers may rely heavily on the sub-lexical information including the word ending cues to assign stress based on the dominant pattern, leading to a trochaic bias.

Importance of L2 proficiency

In the current study, English proficiency played a significant role in ESL learners' sensitivity to word ending as stress cues. More proficient ESL learners demonstrated higher sensitivity in the stress assignment task but not in the naming task. Given the high familiarity of words used in the naming task, it is not surprising to see that English proficiency did not affect ESL learners' performance. So, why did more proficient ESL learners demonstrate better sensitivity to word ending as cues than less proficient peers in the stress assignment task? Perhaps more proficient ESL learners have better statistical learning ability, in other words, higher statistical learning ability is associated with higher language proficiency. Indeed, previous research has revealed that statistical learning is linked to L1 language proficiency in childhood and adults (e.g., Newman et al., 2006; Kemény & Lukács, 2021; Reuter et al., 2018; Evans et al., 2009), and in L2 learning (e.g., Godfroid & Kim, 2021; Potter et al., 2017). The important role of language proficiency shown in our study extends previous findings to the L2 learners' sensitivity to the word ending cues to stress, a written language-based statistical learning. Another possible explanation

which is also relevant to the statistical learning account is that since more proficient ESL learners have a larger English lexicon, as indicated by their performance in the LexTale test, they have been exposed to more words and have more accumulative experiences on the relationship between word endings and stress positions, and this accumulative process leads to implicit statistical learning. This explanation is in line with previous researchers' view that the assignment of lexical stress develops through statistical learning of lexical information (Arciuli et al., 2010; Ševa et al., 2009). As learners' lexicon grows, their ability to generalize the orthographic cues (e.g., word endings) to stress assignment becomes better via statistic learning.

The finding is surprising that as L2 proficiency increases, the speed in naming iambic words becomes faster, whereas it gets slower in naming trochaic words. We speculate that with the increase in English L2 proficiency, participants become more sensitized to the less dominant stress pattern, leading to faster responses. There is more room for improvement within iambic words. At the same time, as L2 proficiency and the sensitivity to iambic pattern increases, L2 learners may become cautious in the process of deciding whether to assign a trochaic stress position, leading to a slowing-down speed in naming words with a trochaic pattern. This surprising result may reflect L2 learners' fledgling knowledge about the specific stress patterns associated with varying orthographic cues, especially in a naming task when they need to overtly produce the words. Consequently, learners may need more time for making accurate decisions. Further improvement in their L2 proficiency will be most likely to support their full grasp of varying stress positions associated with different word ending cues, regardless of the different task demands.

Note that our English proficiency measures (LexTALE and Close test) primarily focus on written language and reading. While we did not include a spoken language measure in testing, we collected participants' spoken English proficiency through self-reported TOEFL (Test of English as a Foreign Language) scores. The average of participants' self-reported TOEFL spoken language score was 22.02 (out of 30) with a standard deviation of 3.51 in Experiment 1. The average of participants' self-reported TOEFL spoken score was 22.13 (out of 30) with a standard deviation of 2.68 in Experiment 2. We further

transformed spoken English score to z-score, and ran statistical analysis on stress position, reliability, and spoken English proficiency as key predictors in Experiments. Our results showed that for stress assignment accuracy, the interaction between spoken English proficiency and endings' reliability was significant ($p < .001$). The main effects of position, reliability and spoken English proficiency were all significant ($ps < .01$). For naming latency, the interaction between position and spoken English proficiency was significant ($p < .05$) and the main effect of reliability was significant ($p < .05$). The results treating spoken language as an indicator of English proficiency were consistent with treating LexTALE and Cloze test as the English proficiency indicator, suggesting that both spoken and written language proficiency contribute to English L2 stress sensitivity.

Limitations and future directions

The current study has several limitations. First, our experiments focused solely on disyllabic words and nonwords, the experimental materials in future research could be extended to longer words and nonwords, since orthographic cues to stress are present in those words as well. For example, like disyllabic words, most trisyllabic English words have stress on the first syllable, but there are many exceptions. Orthographic cues would be useful in marking trisyllabic words that have atypical stress position. For example, a double letter can mark second syllable stress (e.g., saVANna). Second, our naming task treated reliability as a categorical (reliable vs. less reliable) instead of a continuous variable, which limited the possibility to investigate the nuanced picture of English L2 learners' sensitivity to stress cue reliability in word endings in the reading aloud context. Future research may consider adopting a similar design with continuous reliability data in the naming task as that in the stress assignment task. This will allow for a direct comparison across different tasks. Finally, given the evidence from our study that English L2 learners are sensitive to word endings as stress cues, future research can study how to explicitly teach orthographic endings as cues to stress in the classroom, to enhance L2 reading competence.

Conclusions

Our study made a significant contribution to the limited literature on whether ESL learners are

sensitive to word endings as orthographic cues to lexical stress. Our L2 learners demonstrated such sensitivity in both a stress assignment task that did not require explicit phonological representation and a naming task that required direct access and construction of phonological representation. Our findings also suggest that both stress position and English proficiency moderate the strength of the sensitivity, with a trochaic bias and better proficiency leading to better sensitivity in the nonword stress assignment task. Taken together, the evidence from our ESL learners fits in the proposed statistical learning mechanism, that is, L2 learners are able to implicitly extract statistical regularities from linguistic materials, the orthographic cues to lexical stress in our study.

Open Practices Statement

The data that support the findings of this study are available upon request to the first author. None of the experiments was preregistered.

References

- Allen, G. D., & Hawkins, S. (1979). Trochaic rhythm in children's speech. *Current Issues in the Phonetic Sciences*, 927-933. <https://doi.org/10.1075/cilt.9.103all>
- Allen, G. D., & Hawkins, S. (1980). Phonological rhythm: Definition and development. In *Child phonology* (pp. 227-256). Academic Press.
- Arciuli, J. (2017). The relationship between children's sensitivity to dominant and non-dominant patterns of lexical stress and reading accuracy. *Journal of Experimental Child Psychology*, 157, 1-13. <https://doi.org/10.1016/j.jecp.2016.11.016>
- Arciuli, J., & Cupples, L. (2006). The processing of lexical stress during visual word recognition: Typicality effects and orthographic correlates. *Quarterly Journal of Experimental Psychology*, 59(5), 920-948. <https://doi.org/10.1080/02724980443000782>
- Arciuli, J., & Cupples, L. (2007). Would you rather 'embert a cudsert' or 'cudsert an embert'?. *Mental States: Language and Cognitive Structure V. 2: Language and Cognitive Structure*, 213. <https://doi.org/10.1075/slcs.93.12arc>
- Arciuli, J., Monaghan, P., & Seva, N. (2010). Learning to assign lexical stress during reading aloud: Corpus, behavioral, and computational investigations. *Journal of Memory and Language*, 63(2), 180-196. <https://doi.org/10.1016/j.jml.2010.03.005>
- Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, 21(3), 170-176. <https://doi.org/10.1177/0963721412436806>
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database* [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
- Ballard, K. J., Djaja, D., Arciuli, J., James, D. G., & van Doorn, J. (2012). Developmental trajectory for production of prosody: Lexical stress contrastivity in children ages 3 to 7 years and in adults. *Journal of Speech, Language, and Hearing Research*, 55(6), 1822-35. [https://doi.org/10.1044/1092-4388\(2012/11-0257\)](https://doi.org/10.1044/1092-4388(2012/11-0257))

- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, 39(3), 445-459.
<https://doi.org/10.3758/bf03193014>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *The Modern Language Journal*, 64(3), 311-317. <https://doi.org/10.1111/j.1540-4781.1980.tb05198.x>
- Burani, C., & Arduino, L. S. (2004). Stress regularity or consistency? Reading aloud Italian polysyllables with different stress patterns. *Brain and Language*, 90(1-3), 318–325. [https://doi.org/10.1016/S0093-934X\(03\)00444-9](https://doi.org/10.1016/S0093-934X(03)00444-9)
- Carrasco-Ortiz, H., & Frenck-Mestre, C. (2014). Phonological and orthographic cues enhance the processing of inflectional morphology. ERP evidence from L1 and L2 French. *Frontiers in Psychology*, 5, 888. <https://doi.org/10.3389/fpsyg.2014.00888>
- Chan, R. K., & Leung, J. H. (2014). Implicit learning of L2 word stress regularities. *Second Language Research*, 30(4), 463-484. <https://doi.org/10.1177/0267658313510169>
- Chao, Y. R. (1968). A grammar of spoken Chinese. Berkeley and Los Angeles: University of California Press.
- Chen, Y., & Xu, Y. (2006). Production of weak elements in speech—evidence from f₀ patterns of neutral tone in Standard Chinese. *Phonetica*, 63(1), 47-75. <https://doi.org/10.1159/000091406>
- Colombo, L. (1992). Lexical stress effect and its interaction with frequency in word pronunciation. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4), 987.
- Colombo, L., Deguchi, C., & Boureux, M. (2014). Stress priming and statistical learning in Italian nonword reading: Evidence from children. *Reading and Writing: An Interdisciplinary Journal*,

- 27(5), 923–943. <https://doi.org/10.1007/s11145-013-9476-x>
- Commissaire, E., Duncan, L. G., & Casalis, S. (2011). Cross-language transfer of orthographic processing skills: A study of French children who learn English at school. *Journal of Research in Reading*, 34(1), 59-76. <https://doi.org/10.1111/j.1467-9817.2010.01473.x>
- Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, 2(3-4), 133-142. [https://doi.org/10.1016/0885-2308\(87\)90004-0](https://doi.org/10.1016/0885-2308(87)90004-0)
- Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, 2(3-4), 133-142. [https://doi.org/10.1016/0885-2308\(87\)90004-0](https://doi.org/10.1016/0885-2308(87)90004-0)
- Duanmu, S. (2007). *The phonology of standard Chinese*. OUP Oxford.
- Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52(2), 321-35. [https://doi.org/10.1044/1092-4388\(2009/07-0189\)](https://doi.org/10.1044/1092-4388(2009/07-0189))
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175-191. <https://doi.org/10.3758/bf03193146>
- Fikkert, P. (1994). *On the acquisition of prosodic structure*. [Sl: sn].
- Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, 24(7), 1243–1252. <https://doi.org/10.1177/0956797612472207>
- Godfroid, A., & Kim, K. M. (2021). The contributions of implicit-statistical learning aptitude to implicit second-language knowledge. *Studies in Second Language Acquisition*, 43(3), 606-634. <https://doi.org/10.1017/S0272263121000085>
- Gontijo, P. F., Gontijo, I., & Shillcock, R. (2003). Grapheme—phoneme probabilities in British English. *Behavior Research Methods, Instruments, & Computers*, 35(1), 136-157.

<https://doi.org/10.3758/BF03195506>

Grimani, A., & Protopapas, A. (2017). Derivational suffixes as cues to stress position in reading

Greek. *Journal of Research in Reading*, 40(Suppl 1), S23–S41. <https://doi.org/10.1111/1467-9817.12092>

Guo, X. (2022). Acoustic correlates of English lexical stress produced by Chinese dialect speakers compared to native English speakers. *Frontiers in Psychology*, 13.

<https://doi.org/10.3389/fpsyg.2022.796252>

Guo, X. R., Chen, X. X., & Guo, Y. M. (2017, February). Perception of English Lexical Stress by English, Beijing Dialect, and Guangzhou Cantonese Speakers. In *2017 2nd International Conference on Humanities and Social Science (HSS 2017)* (pp. 252-257). Atlantis Press.

Gutiérrez-Palma, N., & Palma-Reyes, A. (2008). On the use of lexical stress in reading Spanish. *Reading and Writing*, 21, 645-660. <https://doi.org/10.1007/s11145-007-9082-x>

Hayes-Harb, R., Brown, K., & Smith, B. L. (2018). Orthographic input and the acquisition of German final devoicing by native speakers of English. *Language and Speech*, 61(4), 547–564. <https://doi.org/10.1177/0023830917710048>

Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Veríssimo, J., Dye, C. D., ... & Ullman, M. T. (2011). Grammar predicts procedural learning and consolidation deficits in children with specific language impairment. *Research in Developmental Disabilities*, 32(6), 2362-2375. <https://doi.org/10.1016/j.ridd.2011.07.026>

Issa, S. H., Awadh, F. H., & Ahmed, H. R. (2022). The role of proficiency level in the speed of lexical activation. *Cogent Arts & Humanities*, 9(1), 1999613. <https://doi.org/10.1080/23311983.2021.1999613>

Jouravlev, O., & Lupker, S. J. (2014). Stress consistency and stress regularity effects in Russian. *Language, Cognition and Neuroscience*, 29(5), 605–619. <https://doi.org/10.1080/01690965.2013.813562>

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns

- of English words. *Child Development*, 64(3), 675–687. <https://doi.org/10.2307/1131210>
- Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, 116(3), 321-340.
<https://doi.org/10.1016/j.cognition.2010.05.011>
- Kelly, M. H., Morris, J., & Verrechia, L. (1998). Orthographic cues to lexical stress: Effects on naming and lexical decision. *Memory & Cognition*, 26(4), 822–832. <https://doi.org/10.3758/BF03211401>
- Kemény, F., & Lukács, Á. (2021). The Role of Statistical Learning and Verbal Short-Term Memory in Impaired and Typical Lexical Development. *Frontiers in Communication*.
<https://doi.org/10.3389/fcomm.2021.700452>
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44(2), 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Marian, V., & Kaushanskaya, M. (2007). Language context guides memory content. *Psychonomic Bulletin & Review*, 14(5), 925-933. <https://doi.org/10.3758/bf03194123>
- McGregor, K. K., & Johnson, A. C. (1997). Trochaic template use in early words and phrases. *Journal of Speech, Language, and Hearing Research*, 40(6), 1220-1231.
<https://doi.org/10.1044/jslhr.4006.1220>
- Molczanow, J., Domahs, U., Knaus, J., & Wiese, R. (2013). The lexical representation of word stress in Russian: Evidence from event-related potentials. *The mental lexicon*, 8(2), 164-194.
<https://doi.org/10.1075/ml.8.2.03mol>
- Monaghan, P., Arciuli, J., & Seva, N. (2016). Cross-linguistic evidence for probabilistic orthographic cues to lexical stress. *Linguistic Rhythm and Literacy*, 215-236.
<https://doi.org/10.1075/tilar.17.10mon>
- Mundy, I. R., & Carroll, J. M. (2013). Spelling–stress regularity effects are intact in developmental dyslexia. *Quarterly Journal of Experimental Psychology*, 66(4), 816-828.
<https://doi.org/10.1080/17470218.2012.719530>

- Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology, 42*(4), 643–655. <https://doi.org/10.1037/0012-1649.42.4.643>
- Onnis, L., Frank, S. L., Yun, H., & Lou-Magnuson, M. (2016). *Statistical learning bias predicts second-language reading efficiency*. Poster presented at the 38th Annual Meeting of the Cognitive Science Society.
- Paizi, D., Zoccolotti, P., & Burani, C. (2011). Lexical stress assignment in Italian developmental dyslexia. *Reading and Writing, 24*, 443-461. <https://doi.org/10.1007/s11145-010-9236-0>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195-203. <https://doi.org/10.3758/s13428-018-01193-y>
- People's Education Press. (2021). Chinese Junior high school English textbook.
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*(5), 233-238. <https://doi.org/10.1016/j.tics.2006.03.006>
- Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology, 61*(2), 106-151. <https://doi.org/10.1016/j.cogpsych.2010.04.001>
- Potter, C. E., Wang, T., & Saffran, J. R. (2017). Second language experience facilitates statistical learning of novel linguistic materials. *Cognitive Science, 41*(Suppl 4), 913–927. <https://doi.org/10.1111/cogs.12473>
- R Core Team (2017) R: A Language and Environment for Statistical Computing. <https://www.R-project.org/>
- Rastle, K., & Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory and Language, 42*(3), 342–

364. <https://doi.org/10.1006/jmla.1999.2687>
- Reuter, T., Emberson, L., Romberg, A., & Lew-Williams, C. (2018). Individual differences in nonverbal prediction and vocabulary size in infancy. *Cognition*, *176*, 215–219. <https://doi.org/10.1016/j.cognition.2018.03.006>
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928. <https://doi.org/10.1126/science.274.5294.1926>
- Schmid, M. S. (2010). Languages at play: The relevance of L1 attrition to the study of bilingualism. *Bilingualism: Language and Cognition*, *13*(1), 1-7. <https://doi.org/10.1017/S1366728909990368>
- Siyanova-Chanturia, A., & Janssen, N. (2018). Production of familiar phrases: Frequency effects in native speakers and second language learners. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(12), 2009–2018. <https://doi.org/10.1037/xlm0000562>
- Smith, P. T., Baker, R. G., & Groat, A. (1982). Spelling as a source of information about children's linguistic knowledge. *British Journal of Psychology*, *73*(3), 339–350. <https://doi.org/10.1111/j.2044-8295.1982.tb01816.x>
- Snow, D. (2007). Polysyllabic units in the vocalizations of children from 0; 6 to 1; 11: Intonation-groups, tones and rhythms. *Journal of Child Language*, *34*(4), 765-797. <https://doi.org/10.1017/S030500090700815X>
- Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456–466. <https://doi.org/10.1016/j.wocn.2010.09.001>
- Sulpizio, S., Arduino, L. S., Paizi, D., & Burani, C. (2013). Stress assignment in reading Italian polysyllabic pseudowords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 51–68. <https://doi.org/10.1037/a0028472>
- Sulpizio, S., & McQueen, J. M. (2012). Italians use abstract knowledge about lexical stress during spoken-word recognition. *Journal of Memory and Language*, *66*(1), 177-193.

<https://doi.org/10.1016/J.JML.2011.08.001>

Vaid, J., & Frenck-Mestre, C. (2002). Do orthographic cues aid language recognition? A laterality study with French–English bilinguals. *Brain and Language*, 82(1), 47-53.

[https://doi.org/10.1016/S0093-934X\(02\)00008-1](https://doi.org/10.1016/S0093-934X(02)00008-1)

White, E. J., Titone, D., Genesee, F., & Steinhauer, K. (2017). Phonological processing in late second language learners: The effects of proficiency and task. *Bilingualism: Language and Cognition*, 20(1), 162-183. <https://doi.org/10.1017/S1366728915000620>

Zevin, J., & Joanisse, M. (2000). Stress assignment in nonword reading. *Journal of Cognitive Neuroscience B*, 41, S5.

Zhang, J., Meng, Y., Fan, X., Ortega-Llebaria, M., & Jeong, S. L. (2018). Stress typicality effect in Chinese advanced and intermediate ESL learners. *Educational Psychology*, 38(5), 617-632.

<https://doi.org/10.1080/01443410.2017.1380169>

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

<https://doi.org/10.3758/BF03210615>

Table 1

Regression model for accuracy across stress position and endings' reliability.

| | Estimate | Std. Error | z value | p-value |
|---------------------------------------|----------|------------|---------|------------|
| Intercept | -1.28 | 0.49 | -0.63 | 0.529 |
| Stress position | 2.76 | 0.99 | 2.79 | 0.005 ** |
| Endings' reliability | 5.57 | 0.81 | 2.89 | 0.004** |
| Stress position: Endings' reliability | -4.24 | 1.23 | -3.44 | 0.0006 *** |

* $p < .05$; ** $p < .01$; *** $p < .001$

Table 2

Regression model for accuracy across stress position, endings' reliability and English proficiency.

| | Estimate | Std. Error | z value | p-value |
|------------------------------------|----------|------------|---------|--------------|
| Intercept | -1.13 | 0.46 | -1.23 | 0.219 |
| Position | 3.46 | 1.04 | 3.32 | 0.0009 *** |
| Reliability | 2.43 | 0.79 | 2.85 | 0.004 ** |
| Proficiency | -1.93 | 0.32 | -2.60 | 0.009 *** |
| Position: Reliability | -5.23 | 1.31 | -3.99 | 6.73e-05 *** |
| Position: Proficiency | 0.63 | 0.58 | 1.07 | 0.284 |
| Reliability: Proficiency | 4.59 | 0.47 | 8.12 | 3.89e-12 *** |
| Position: Reliability: Proficiency | -0.63 | 0.77 | -0.82 | 0.410 |

Table 3

Naming latency and stress placement accuracy across four conditions.

| Stress position/Reliability | Reliable M(SD) ms | Less Reliable M(SD) ms |
|-----------------------------|-----------------------|------------------------|
| Naming latency | | |
| Trochaic M(SD) ms | 758.87 ms (224.99 ms) | 773.35 ms (235.75 ms) |

| | | |
|---------------------------|-----------------------|-----------------------|
| Iambic M(<i>SD</i>) ms | 760.71 ms (232.40 ms) | 788.67 ms (238.00 ms) |
| Stress placement accuracy | | |
| Trochaic M(<i>SD</i>) | 0.85 (0.36) | 0.87 (0.33) |
| Iambic M(<i>SD</i>) | 0.89 (0.31) | 0.88 (0.32) |
| Segment accuracy | | |
| Trochaic M(<i>SD</i>) | 0.99 (0.15) | 0.97 (0.13) |
| Iambic M(<i>SD</i>) | 0.98 (0.14) | 0.98 (0.12) |

Table 4

Regression model for naming latency across stress position, ending cue's reliability and the interaction.

| | Estimate | Std. Error | t-value | p-value |
|-----------------------|----------|------------|---------|---------|
| Intercept | 46.88 | 4.32 | 2.16 | 0.031* |
| Position | -15.31 | 11.98 | -1.28 | 0.205 |
| Reliability | -22.32 | 8.58 | -2.60 | 0.011* |
| Position: Reliability | 13.48 | 17.05 | 0.79 | 0.432 |

Table 5

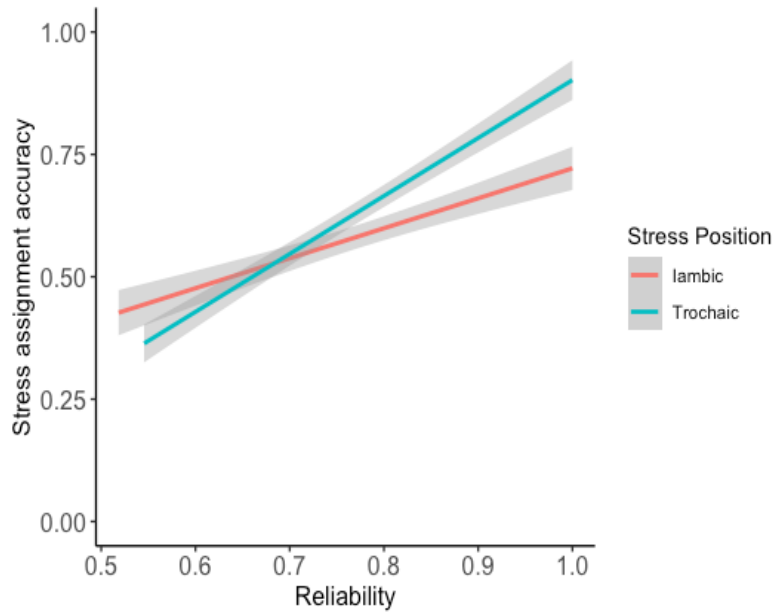
Regression model for naming latency across stress position, endings' reliability and English proficiency.

| | Estimate | Std. Error | t-value | p-value |
|--------------------------|----------|------------|---------|-------------|
| Intercept | 771.08 | 9.97 | 77.38 | < 2e-16 *** |
| Position | - 7.39 | 8.32 | -0.89 | 0.375 |
| Reliability | -22.23 | 8.57 | -2.59 | 0.012* |
| Proficiency | 2.67 | 12.29 | 0.22 | 0.829 |
| Position: Reliability | 13.19 | 17.14 | 0.77 | 0.444 |
| Position: Proficiency | 26.29 | 10.57 | 2.49 | 0.013** |
| Reliability: Proficiency | 8.57 | 10.26 | 0.84 | 0.404 |

| | | | | | | |
|-----------|--------------|-------------|--------|-------|-------|-------|
| Position: | Reliability: | Proficiency | -32.74 | 21.14 | -1.55 | 0.122 |
|-----------|--------------|-------------|--------|-------|-------|-------|

Figure 1

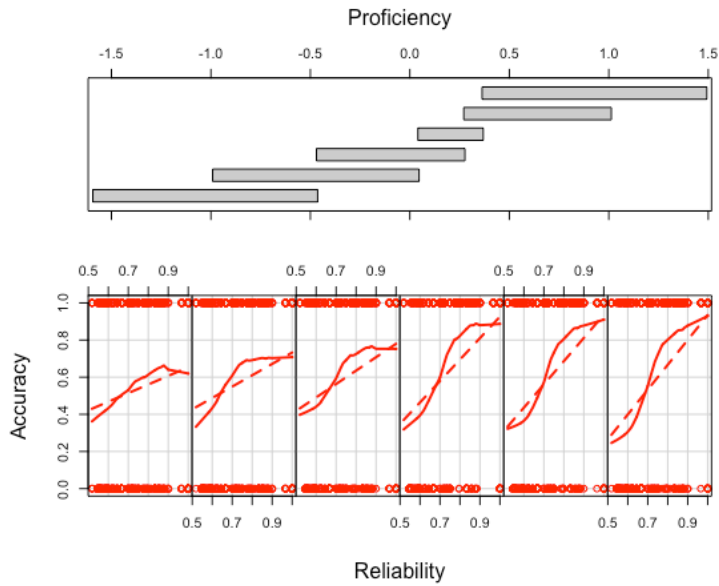
Stress assignment accuracy as a function of word endings' reliability and stress position.



Note. The solid red and blue lines represent the fitted regression lines. The grey areas represent confidence intervals for each regression line.

Figure 2

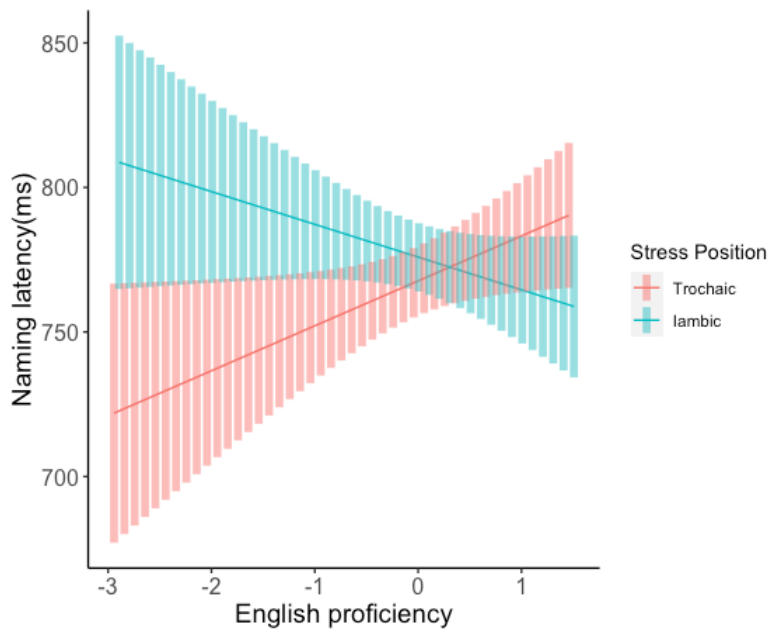
Stress assignment accuracy as a function of word endings' reliability and English proficiency.



Note. The solid lines represent the average of stress assignment accuracy, the dotted lines represent the fitted linear regression lines, and proficiency is the standardized z-value of the average of LexTale and cloze test.

Figure 3

Naming latency as a function of position and English proficiency.



Note. The solid red and blue lines represent the fitted regression lines. The dotted areas represent confidence intervals for each regression line. English proficiency is indexed by the standardized average score of LexTale and cloze test for each participant.

Supplementary Table A

Different structures of word beginnings.

| Structure of beginnings | 50-60% | 60-70% | 70-80% | 80-90% | 90-100% |
|-------------------------|---------|----------|---------|--------|----------|
| CCV | prukave | brezell | cruzile | grafex | brozerce |
| CV | burvoad | corpenge | calvix | tarfip | pesvike |
| V | imgeed | ilmase | ipzine | inpord | ulbem |

Supplementary Table B

Lexical characteristics of words across four conditions.

| Group/Lexical characteristics M(<i>SD</i>) | Less Re-iambic | Re-iambic | Less Re-trochaic | Re-trochaic |
|--|----------------|-------------|------------------|--------------|
| Word length | 6.60 (0.88) | 7.12 (0.99) | 7.28 (1.32) | 6.84 (1.26) |
| Word frequency (Log (Freq_HAL)) | 9.53 (1.23) | 9.14 (1.67) | 9.18 (1.88) | 10.20 (0.85) |
| Phonological neighborhood size | 0.75 (1.25) | 0.65 (0.93) | 0.5 (1.04) | 0.79 (0.98) |
| Orthographic neighborhood size | 0.40 (0.75) | 0.41 (0.62) | 0.33 (0.59) | 0.47 (0.70) |
| Phonological neighborhood frequency | 2.50 (3.21) | 1.87 (2.96) | 1.69 (2.90) | 3.77 (3.86) |
| Orthographic neighborhood frequency | 1.89 (3.44) | 2.23 (3.48) | 1.64 (2.81) | 2.72 (3.86) |
| Bigram frequency | 2056 (892) | 2003 (957) | 1664 (857) | 1914 (1014) |

Note. Less Re-iambic = Less reliable iambic condition; Re-iambic = Reliable iambic condition; Less Re-trochaic = Less reliable trochaic condition; Re-trochaic= Reliable trochaic condition.

Supplementary Table C

Articulation manners of initial consonants across four conditions.

| Group/ Articulation manners (N) | Less Re-iambic | Re-iambic | Less Re-trochaic | Re-trochaic |
|---------------------------------|----------------|-----------|------------------|-------------|
|---------------------------------|----------------|-----------|------------------|-------------|

| | | | | |
|-----------|---|---|---|---|
| Stop | 8 | 8 | 8 | 7 |
| Fricative | 5 | 3 | 5 | 6 |
| Liquid | 2 | 2 | 2 | 1 |
| Nasal | 1 | 1 | 2 | 3 |
| Vowel | 4 | 4 | 1 | 2 |

Note. Less Re-iambic = Less reliable iambic condition; Re-iambic = Reliable iambic condition; Less Re-trochaic = Less reliable trochaic condition; Re-trochaic= Reliable trochaic condition.

Supplementary Table D

Sample words in the naming experiment across four conditions

| Less Re-iambic | Re-iambic | Less Re-trochaic | Re-trochaic |
|----------------|-----------|------------------|-------------|
| support | subscribe | software | setup |
| reveal | receive | native | module |
| divorce | dismiss | matrix | context |

Note. Less Re-iambic = Less reliable iambic condition; Re-iambic = Reliable iambic condition; Less Re-trochaic = Less reliable trochaic condition; Re-trochaic= Reliable trochaic condition.

Chapter 4: Statistical Learning in Reading: Can Explicit Instruction Help Children Learn Lexical Stress in a Second Language?

One of the core challenges in learning to read in a second language (L2) is understanding the relationship between orthographic patterns and prosodic features of the language. Lexical stress is a key prosodic feature that significantly affects word recognition, pronunciation, and overall reading fluency in English (e.g., Cutler, 2015; Kelly, 2004; Field, 2025). Correct stress placement is not only important for accurate pronunciation but also for effective word identification, which subsequently impacts comprehension. This challenge is particularly pronounced for learners of English as L2, who must navigate the intricate rules of stress assignment while simultaneously acquiring new vocabulary and grappling with a different orthographic system (e.g., Lord, 2007; Sagarra et al., 2024).

Word endings can serve as a statistical cue for stress placement in English, with certain endings being probabilistically associated with specific stress patterns (e.g., *-ette* often signaling second-syllable stress). Through statistical learning (SL), native speakers can internalize these regularities (Arciuli & Cupples, 2006; Arciuli et al., 2010; Kelly et al., 1998). The process of recognizing and extracting patterns from the environment is referred to as SL. SL is often considered an implicit process (e.g., Christiansen, 2019; Batterink et al., 2019) where individuals extract patterns from repeated exposure, aiding in the acquisition of prosodic features. The majority of research on word endings as a cue for lexical stress has focused on native English speakers, leaving a question about whether L2 learners, especially children, can similarly utilize these cues. More importantly, while explicit instruction has been shown to enhance learning outcomes (e.g., phonological awareness or grammatical rules) in language learning (e.g., Goo et al., 2015; Rahimpour & Salimi, 2010), it remains unclear if it plays a significant role in SL for L2 learners of English. The current study explored whether explicit instruction could enhance L2 children's sensitivity to word endings as a cue for lexical stress. We focused on children learning English as an L2 with a Chinese language background, aiming to shed light on effective instructional methods for supporting prosodic awareness and learning in L2 learners.

Word endings as a statistical cue to lexical stress

Previous research suggests that word endings serve as an important SL cue that guides lexical stress in written English words (e.g., Arciuli & Cupples, 2006; Arciuli et al., 2010; Kelly et al., 1998). The association between word endings and lexical stress is probabilistic, due to the statistical regularities in the written language (Arciuli & Cupples, 2006). Specifically, the spellings of many word endings are statistically associated with particular stress patterns. For example, word endings such as *-ette*, *-que*, and *-umb* are strongly associated with second-syllable stress in two-syllable words (e.g., *casSETTE*; *phySIQUE*; *sucCUMB*), and *-ior*, *-urn*, and *-em* are strongly associated with first syllable stress (e.g., *SENIor*; *RETurn*; *PROBlem*). Readers who are sensitive to the cue to stress position provided by the word endings may be able to leverage those regularities to guide their stress placement in written words. For instance, a reader who encounters the unfamiliar word *rennet* may recognize the ending sequence *-et* that would cue them to correctly identify the word's first-syllable stress pattern. Likewise, recognizing the ending sequence *-ee* in *alee* may cue them to determine that the word has second-syllable stress.

Given their limited exposure and experience with the new language, could L2 learners of English be also able to utilize word endings as an orthographic cue to guide stress placement? A representative study by Chan and Leung (2014) tested native Cantonese-speaking adult speakers, who were asked to learn some Spanish words without being informed of the regularities contained in the words, for example, words ending in *-o* have their stress on the penultimate syllable (the second to last syllable), and words that end in *-ar* have their stress on the last syllable. After 15 minutes of training, these novel language learners displayed knowledge of lexical stress placement on novel words with *-ar* and *-o* ending. More recently, Ren and Wang (2023) investigated whether adult native Chinese speakers who have learned English for about 10 years as L2 are sensitive to word endings as a cue to lexical stress. Their results showed that these L2 speakers demonstrated sensitivity to such cues in both a stress assignment and a reading aloud task. Furthermore, more proficient English learners demonstrated stronger sensitivity. In sum, these findings highlight that word endings can serve as a reliable statistical cue to lexical stress, not only for native speakers but also for L2 learners. However, the scarcity of research in this area calls for

further investigation. In addition, previous studies' focus on adults limits insights into how children with minimal L2 experience might perceive the association between word endings and lexical stress.

Role of Explicit Instruction in SL

Some previous research demonstrated that SL seems to be unaffected by previous instructions given to participants (e.g., Arciuli et al. 2014; Batterink et al. 2015). However, emerging evidence suggests that explicit instruction may play a role in enhancing SL. Specifically, some studies demonstrated that when learners are provided with explicit instruction about the statistical regularities present in the environment, their ability to detect and learn these regularities improves (e.g., Hamrick & Rebuschat; 2012; Kachergis et al.; 2014). Indeed, according to the two-systems view of SL (Conway, 2020), learning can occur through both implicit and explicit routes. The implicit system operates automatically and is well-suited for learning simple and frequently occurring patterns through repeated exposure, while the explicit system involves conscious, deliberate efforts and is more effective for acquiring complex or less intuitive patterns. In the context of SL, explicit instruction can activate the explicit system, helping learners acquire challenging language rules that may not be easily discovered through implicit processes alone.

Given that word endings can serve as an orthographic cue to English stress, learning the mapping between word endings and English lexical stress should help beginning learners read English words accurately via appropriate assignment of lexical stress. One important question is how this learning could be supported. In reading instruction, there have been two major instructional approaches. One approach advocates explicit instruction of regularities. For example, in systematic phonics instruction, a child may be taught that letter *d* makes the sound /d/ and then asked to read aloud examples such as dog, dad, and dig. An alternative approach suggests that these regularities may be discovered without explicit instruction, simply through text experiences (SL). Which approach works better is dependent on whether explicit instruction can facilitate learning. One recent meta-analysis showed a significant, positive, moderate effect of explicit instruction in boosting SL performance (Ren et al., 2024), and this effect is more significant for linguistic SL.

Several artificial language learning studies have shown the benefits of explicit instruction in acquiring knowledge of underlying letter-sound mapping patterns. Rastle et al. (2021) examined the effects of directly teaching patterns embedded in the writing system rather than relying on learners to independently discover these patterns through reading alone. A group of 48 adults was taught to read unfamiliar words (e.g., $\phi \chi \xi \eta$, $\Lambda \zeta \Phi \eta$) written in two artificial writing systems. The implicit group relied solely on experiencing the novel words to detect the regularities between spelling and sound (e.g., η corresponds to /g/) as well as spelling and meaning (e.g., η means *old*), whereas the explicit group received a short session of explicit instruction on these regularities before starting their training. Nearly all participants who received explicit instruction performed exceptionally well on tests assessing their ability to apply these underlying regularities to new situations. In contrast, despite receiving up to 18 hours of training on unfamiliar words, less than 25% of those in the implicit group were able to achieve similar levels of performance to those who had received explicit instruction. Another representative study by Wiener et al. (2010) trained L2 learners of Mandarin syllable-to-tone mapping patterns using an artificial, Mandarin-like tonal language. Half of the participants were given daily explicit instruction on Mandarin tone contours before training, while the other half were not. Their results showed that explicit instruction appears to facilitate adults' production of high-level tone contours.

The previous studies concerning the effect of explicit instruction in linguistic materials mostly focused on basic letter-sound mapping process (e.g., Rastle et al., 2021; Wiener et al., 2010). Very limited if any evidence is available that addressed whether and how explicit instruction influences the learning of more subtle and complex underlying regularities such as word endings as a cue for lexical stress placement. It is also crucial to note that the previous studies employed artificial orthographies, which limited the generalizability of their findings to real languages and writing systems. This limitation inherently restricts the scope of the conclusions that can be drawn. To address these gaps, the current study sought to provide more direct evidence on the role of explicit instruction in learning complex linguistic regularities in a real-world language context.

Current study

The current study aimed to investigate whether beginning learners of English L2 would benefit from a training program that emphasizes mapping between word endings and lexical stress. We focused on English L2 children with Chinese as a native language background. Chinese is a tonal language without phonetic stress (Duanmu, 2007) and English is a non-tonal, but stress-assigning language. The distinct phonetic features between these two languages allow us to investigate the acquisition of an unfamiliar L2 regularity. This training study was composed of a pre-training session, two training sessions, a post-training testing session (including generalization), and a retention test session. One group received explicit instruction on the underlying regularities related to the orthographic cues (explicit-instruction group), whereas the other group discovered these regularities through experience with whole words (discovery-learning group).

We employed three primary tasks—reading aloud, implicit stress judgment, and explicit stress assignment—to evaluate different facets of children’s acquisition of lexical stress in L2. The reading aloud task was used during the pre-training, post-training, and retention phases to assess participants’ ability to accurately pronounce words with the correct stress placement. This task allowed us to observe how training methods (explicit instruction or self-discovery) impact the application of stress patterns in a naturalistic reading context. The implicit stress judgment task involved presenting pairs of nonwords, where participants had to judge which nonword sounded better. This task helped determine whether children had internalized the statistical regularities between word endings and stress positions, and whether they could generalize these regularities to new items. By including a generalization phase, we could assess the robustness of learning and its transfer to novel stimuli. Lastly, the explicit stress assignment task was included in the post-training and retention phases, where participants were asked to assign stress to new nonwords. This task was crucial in examining whether participants could explicitly use their learned knowledge of word endings to make accurate stress assignment. Together, these tasks provided a comprehensive assessment of both the implicit and explicit aspects of stress pattern acquisition, helping to shed light on the differential benefits of explicit instruction versus self-discovery in

learning complex linguistic features.

Specifically, we aimed to answer: 1) Are L2 children sensitive to word endings as an orthographic cue to English stress after training independent of their learning condition (i.e., with explicit instruction or self-discovery)? Previous research showed that adult L2 learners have the capacity to extract statistical regularities from writing systems (e.g., Chan & Leung, 2014; Ren & Wang, 2023). If SL is a powerful mechanism that requires minimal exposure, for instance, children were shown to be able to detect regularities in grammatical structures with less than two minutes of exposure to an artificial grammar (Gómez et al., 2000), we anticipated that L2 children who are beginning learners of English may also be able to effectively employ word endings as an orthographic cue to English lexical stress; 2) Would explicitly teaching regularities rather than relying on learners to discover help L2 children better learn English lexical stress? Drawing upon previous research on adults that explicit instruction plays a role in improving learning syllable-to-sound and syllable-to-tone mappings (e.g., Rastle et al., 2021; Wiener et al., 2010), we hypothesized that explicitly teaching regularities may also notably facilitate English lexical stress acquisition among L2 children. Furthermore, from a theoretical perspective, if SL occurs through both implicit and explicit mechanisms, then explicit instruction can serve as additional support to direct children's attention and further improve SL.

Methods

Participants

We recruited 180 normally developing children aged 12 years (Grade 6) in a medium size city in China. An a priori power analysis using G*Power (Faul et al., 2007) to test for differences between two independent group means using a one-tailed test ($d = .50$, $\alpha = .05$) suggests that 176 students with two equal-sized groups of $n = 88$ are required to achieve a power of .95. By the age of 12 years, native Chinese-speaking children have received around 2.5 years of formal English L2 instruction. Their English L2 receptive vocabulary size in this age is around 600 (Li et al., 2011). Children were randomly assigned to either the explicit learning or the self-discovery group. The groups did not differ significantly in terms of their English vocabulary size, as measured by a shortened version of Peabody Picture Vocabulary Test,

Fourth Edition (PPVT-4; Dunn & Dunn, 2007). The explicit instruction group had a mean vocabulary score of $M = 11.40$, $SD = 4.79$, and the self-discovery group had a mean score of $M = 13.02$, $SD = 3.61$ ($p = .792$). Similarly, there were no significant differences between the groups in their English word decoding skill, assessed using the Test of Word Reading Efficiency-Second Edition (TORWE-II; Torgesen et al., 2012). For real word decoding, the explicit instruction group had a mean score of $M = 11.34$, $SD = 6.09$, and the self-discovery group had a mean score of $M = 10.12$, $SD = 5.43$ ($p = .654$). For nonword decoding, the explicit instruction group had a mean score of $M = 5.12$, $SD = 2.66$, and the self-discovery group had a mean score of $M = 4.87$, $SD = 3.10$ ($p = .720$).

Parents completed a language background and experience questionnaire. This questionnaire asked children's language experience at home, *e.g.*, if children receive English instruction at home, and if yes, the average duration of English instruction each day. Of the participants, 21.7% of parents reported that their children received English instruction at home. Among those who reported home instruction, the average duration of English instruction per day was 21 minutes ($SD = 10.27$). Teachers completed a questionnaire targeting the approaches employed in teaching English in class. In addition to questions regarding whether a whole language or phonics approach is adopted in teaching English, teachers were also asked whether they explicitly teach word endings as lexical stress cues in class, and to what degree the endings used in the current study are familiar to their students. Of the teachers surveyed ($n = 6$), five reported using a whole language approach, while only one teacher indicated using a phonics-based approach. Additionally, no teacher reported that they explicitly teach word endings as lexical stress cues in their classrooms. Regarding the familiarity of the word endings used in the current study, teachers reported that their students were, on average, 3.27 (on a scale of 1 to 5, where 1 is "not at all familiar" and 5 is "very familiar") familiar with these endings ($SD = 1.93$). The questionnaires from parents and teachers together provide the context in which our participants learnt English as L2.

Research Design

Pre-training phase

Reading aloud. On Day 1, children in the two groups were asked to read aloud 12 two-syllable

real English words (Table 1). These words are embedded with four endings *-ior*, *-ial*, *-ude*, and *-ibe*. According to previous corpus analysis (Arciuli & Cupples, 2006, based on CELEX database, Baayen et al., 1996), all disyllabic English words ending in *-ior and -ial* have first syllable stress, and all disyllabic English words ending in *-ude and -ibe* have second syllable stress. Thus, endings *-ior*, *-ial*, *-ude and -ibe* are reliable predictors of lexical stress. These word endings are treated as relatively clean orthographic word endings without morphemic information based on Arciuli & Cupples (2006)’s corpus analysis. Table 2 listed length, frequency, orthographic neighborhoods, phonological neighborhoods and phonemes of these words across two conditions (first syllable stress, second syllable stress). Due to the limited number of items in each condition, we did not conduct a t-test. However, we observed a pattern suggesting that the frequency in the first-syllable stress condition is higher than in the second-syllable stress condition.

Implicit stress judgment task. On day 1, children were asked to listen to a pair of nonwords (8 pairs in total, e.g., PEAnior, firNIOR, see Table 1), and judge which nonword sounds better. The nonwords used in this study were created using Praat, a tool for phonetic analysis and manipulation (Boersma & Weenink, 2021). Each nonword was composed of two syllables, with lexical stress placed on either the first or second syllable. Recordings were made using a female voice, with a typical rate of speech set to approximately 4.5 syllables per second to maintain natural presentation. The fundamental frequency (F0) was manipulated in Praat to emphasize the stressed syllable, with an average increase of around 20-30 Hz compared to the unstressed syllable. Additionally, intensity was adjusted, with stressed syllables being about 4-6 dB louder than their unstressed counterparts. Duration was also controlled, with stressed syllables extended by approximately 50-100 ms relative to the unstressed syllables to reflect natural patterns of stress duration in English.

Table 1
Real English words and nonwords in the pre-training and post-training phase

| | First-Reliable | | Second-Reliable | |
|------------|-----------------------|-------------|------------------------|-------------|
| | <i>-ior</i> | <i>-ial</i> | <i>-ude</i> | <i>-ibe</i> |
| Real words | | | | |
| (pretest) | junior | social | exclude | describe |
| (posttest) | senior | racial | include | prescribe |

| | | | | |
|----------------------|----------|-----------|----------|-----------|
| | warrior | facial | conclude | subscribe |
| Which sounds better | PEAnior | peaCIAL | PEAclude | peaSCRIBE |
| (pretest) (posttest) | firNIOR | FIRcial | firCLUDE | FIRscribe |
| | beaCIOR | beaNIAL | BEAclude | BEAscribe |
| Which sounds better | GOAcior | GOAnial | goaCLUDE | goaSCRIBE |
| (generalization) | foaHIOR | FOAhial | foaHUDE | FOAhibe |
| | GEAkior | geaKIAL | GEAkude | geaKIBE |
| | PEEzior | PEEzial | peeZUDE | peeZIBE |
| | teeRIOR | teeRIAL | TEErude | TEEkibe |
| Explicit stress | dooscior | dooclial | dooclude | doonibe |
| assignment | paiclior | paiscrial | painude | pairribe |
| | geakior | geakial | geakude | geakibe |

Table 2
Descriptive statistics of characteristics of words

| Measure | First-Syllable M (SD) | Second-Syllable M (SD) |
|--------------------------------|-----------------------|------------------------|
| Word Length | 6.17 (0.41) | 8.00 (0.89) |
| Frequency | 56.67 (10.80) | 43.17 (5.64) |
| Orthographic Neighborhood Size | 5.00 (0.89) | 4.67 (0.52) |
| Phonological Neighborhood Size | 4.00 (0.89) | 3.50 (0.55) |
| Number of Phonemes | 5.17 (0.41) | 7.50 (0.55) |

In addition to word endings, prefixation, vowel length (the perceived length of a vowel sound), and orthographic weight (the number of letters in a syllable) are three important predictors of word stress assignment (e.g., Hayes, 1982; Kelly et al., 1998; Mousikou et al., 2017; Smith & Baker, 1976; Rastle & Coltheart, 2000). To avoid confounding effects from these three predictors, these nonwords do not contain prefixations. These items were matched on word length, orthographic weight, orthographic neighborhood size, and bigram frequency across conditions (Table 3).

Table 3
Descriptive statistics of characteristics of nonwords

| Measure | First-Syllable M (SD) | Second-Syllable M (SD) | t |
|--------------------------------|-----------------------|------------------------|------|
| Word Length | 7.00 (0.00) | 8.31 (0.87) | 1.07 |
| Orthographic Weight | 5.44 (0.51) | 5.50 (0.52) | .51 |
| Orthographic Neighborhood Size | 6.44 (1.09) | 6.50 (1.15) | .13 |
| Bigram Frequency | 12.38 (1.09) | 12.50 (1.15) | .32 |

Training Phase

Prior to formal training, participants in the explicit instruction group received a PowerPoint presentation explicitly showing the association between word ending and stress position (see Figure 1 for example of words ended with *-ence* and *-ieve*). The discovery group, on the other hand, heard the pronunciation of the words used in the explicit instruction slides. They were asked to practice copying the spelling of each word. The word endings used in the pre-training, training, and post-training phases did not appear in the PowerPoint presentation. All instructions were given in Chinese L1 to ensure comprehension of the task demands. At the end of this phase, the explicit instruction group was checked if they understood the mapping between word endings and stress position (e.g., listen to the pronunciation of *absence*, what is the word ending, and which syllable is being emphasized?).

Over the *four training sessions*, the explicit instruction and discovery groups were trained on the four endings used in the pretraining phase, *-ior*, *-ial*, *-ude*, and *-ibe*. Each word ending was associated with 8 two-syllabic English nonwords, thus thirty-two nonwords in total. Since these were reliable word endings, all the nonwords had default stress position, either on the first or the second syllable. These nonwords did not differ significantly in word length, orthographic weight, orthographic neighborhood size, and bigram frequency between conditions ($ps > .4$). Children were asked to sit in front of a computer. The experiment was run on PsychoPy v2022.1.3 (Peirce et al., 2019). There were four training sessions conducted over four consecutive days, with one session (about 35 minutes) per day. During each session, children listened to the pronunciation of nonwords and were asked to repeat them. The pronunciation was recorded via PsychoPy v2022.1.3. Each nonword was presented four times within a session, and the same set of materials was used across all four sessions. The nonwords' pronunciations were played in random order during the training.

Post-training phase

Children completed the same *reading aloud task* and *implicit stress judgment task* as in the pre-training phase.

Survey of pattern awareness. The discovery group only then completed a short survey to assess whether the children recognized any patterns related to the association between word endings and stress

assignment during the training. The survey included questions asking children to indicate whether they noticed any patterns, with response options of "Yes," "No," or "I'm not sure." For those who responded affirmatively, an open-ended question prompted them to describe the patterns they perceived.

Generalization task. Additionally, both groups of children were provided with a list of entirely new nonwords (8 pairs in total, Table 1) which had the same word endings as the training items (e.g., foaHUDE, GEAkude). The design and procedure were the same as the implicit stress judgment task. The purpose of this generalization phase was to assess whether children could apply their knowledge of lexical stress patterns to novel words.

Lexical stress tutorial. After the generalization phase, children in both groups were introduced to the concept of lexical stress. Since we were not certain whether children had prior knowledge of lexical stress, this brief tutorial was provided to ensure their basic knowledge of the concept. In the tutorial, lexical stress was explained as follows: "Lexical stress is when one part of a word is said louder or with more emphasis than the other parts. For example, in the word apple, we say AP louder than ple (APple). Let's clap together to feel the stress: clap harder on AP and softer on ple. Now try it with the word banana—we say NA louder than the other parts (ba-NA-na). Can you hear how some parts of the word sound stronger? That's what we call lexical stress!".

Explicit stress assignment task. After the lexical stress tutorial, children were given 12 new nonwords (Table 1) but with the same endings as those from the pre-training phase. These nonwords were pre-divided into syllables. Children were instructed to choose a syllable on which they wanted to assign stress. The purpose of this task was to tap into children's explicit knowledge of the mapping between word endings and lexical stress placement.

Retention

One week later, children were asked to complete all the tasks from the post-training phase again including reading aloud, implicit stress judgement (not including generalization), and explicit stress assignment tasks to assess how well they retained the knowledge.

Data Scoring

Participants' responses during the reading aloud, implicit stress judgment, and explicit stress assignment tasks were scored for accuracy. In the reading aloud task, participants received one point only if they read the segment sounds correctly and assigned stress correctly. Children were scored as incorrect if they produced incorrect segment sounds but correct stress positions, correct segment sounds but incorrect stress positions, or incorrect segment sounds and incorrect stress positions. For the implicit stress judgment task, participants' responses were scored based on whether their choice matched the expected stress pattern. In the explicit stress assignment task, accuracy was determined by whether participants correctly assigned stress to the appropriate syllable based on the word ending.

Figure 1

Training and testing protocol

Pre-training (30 minutes)

| Task | Stimulus | Response | Task | Stimulus | Response |
|---------------|----------|---------------|--------------------------|--------------------|---------------------|
| Reading aloud | social | Read it aloud | Implicit stress judgment | peaNiOR FiRnior | Which sounds better |

Explicit-instruction (10 minutes)

--ence words
 Listen to the pronunciation of “silence”, “essence”, and “presence”.
 Do you find something in common for these three?

--ence words
 They were all emphasized on the first syllable. Words that end in **-ence** are usually emphasized on the first syllable. Let’s read these words together.

--ieve words
 Listen to the pronunciation of “believe”, “achieve”, and “relieve”.
 Do you find something in common for these

--ieve words
 They were all emphasized on the second syllable. Words that end in **-ieve** are usually emphasized on the second syllable. Let’s read these words together.

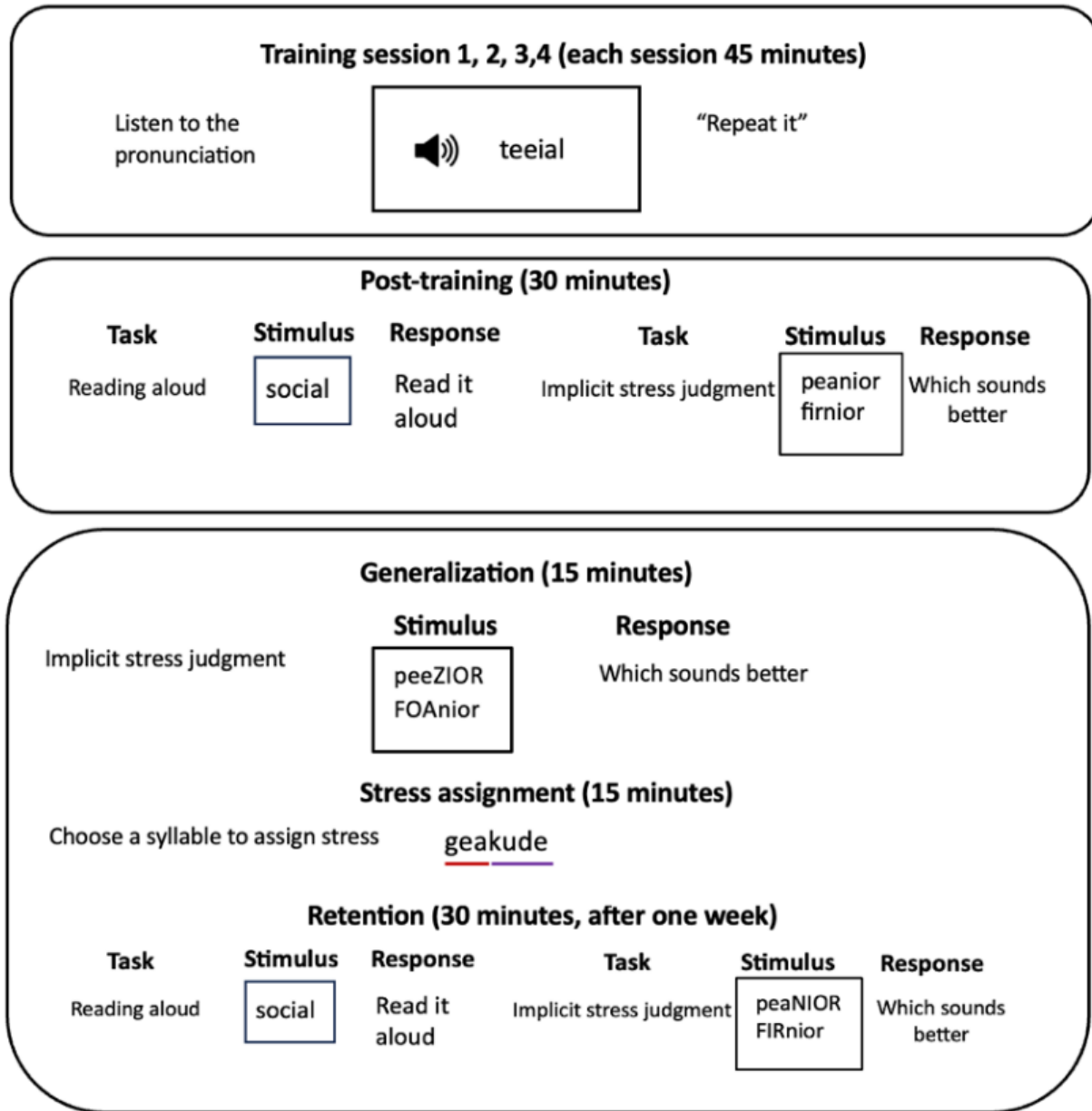
Discovery-learning (10 minutes)

Listen to the pronunciation of “silence”, “essence”, and “presence”.
 Can you read aloud these three words?

Can you write down “silence”, “essence”, and “presence”?

Listen to the pronunciation of “believe”, “achieve”, and “relieve”. Can you read aloud these three words?

Can you write down “believe”, “achieve”, and “relieve”?



Note. The training tasks were constructed to build knowledge of individual novel words. The test tasks probed both knowledge of the trained novel words (reading aloud, implicit stress judgement) and knowledge of underlying regularities used to generalize (explicit stress assignment).

Results

We summarized children's performance in the reading aloud task (pretest, posttest, and retention), implicit stress judgment task (pretest, posttest, generalization, and retention), and explicit stress assignment task (posttest and retention). The results are presented in Table 4.

Table 4

Children's Mean Accuracy (SD) Across Reading Aloud, Implicit Stress Judgment, and Stress Assignment Tasks (Explicit Instruction vs. Discovery Learning Groups)

| Test/ Group | Explicit instruction M (SD) | Discovery-learning M (SD) |
|---|-----------------------------|---------------------------|
| Reading aloud (pretest, first syllable) | 0.59 (0.49) | 0.58 (0.49) |
| (pretest, second syllable) | 0.69 (0.46) | 0.67 (0.47) |
| (posttest, first syllable) | 0.83 (0.38) | 0.80 (0.40) |
| (posttest, second syllable) | 0.83 (0.37) | 0.79 (0.41) |
| (retention, first syllable) | 0.78 (0.42) | 0.71 (0.45) |
| (retention, second syllable) | 0.77 (0.42) | 0.7 (0.46) |
| <hr/> | | |
| Implicit stress judgment | | |
| (pretest, first syllable) | 0.58 (0.50) | 0.61 (0.48) |
| (pretest, second syllable) | 0.52 (0.44) | 0.56 (0.49) |
| (posttest, first syllable) | 0.91 (0.29) | 0.86 (0.35) |
| (posttest, second syllable) | 0.91 (0.28) | 0.87 (0.33) |
| (generalization, first syllable) | 0.73 (0.38) | 0.72 (0.44) |
| (generalization, second syllable) | 0.72 (0.38) | 0.71 (0.45) |
| (retention, first syllable) | 0.85 (0.36) | 0.83 (0.38) |
| (retention, second syllable) | 0.83 (0.37) | 0.75 (0.43) |
| <hr/> | | |
| Explicit stress assignment | | |
| (posttest, first syllable) | 0.79 (0.32) | 0.77 (0.38) |
| (posttest, second syllable) | 0.74 (0.37) | 0.63 (0.32) |
| (retention, first syllable) | 0.71 (0.41) | 0.68 (0.39) |
| (retention, second syllable) | 0.71 (0.30) | 0.64 (0.40) |

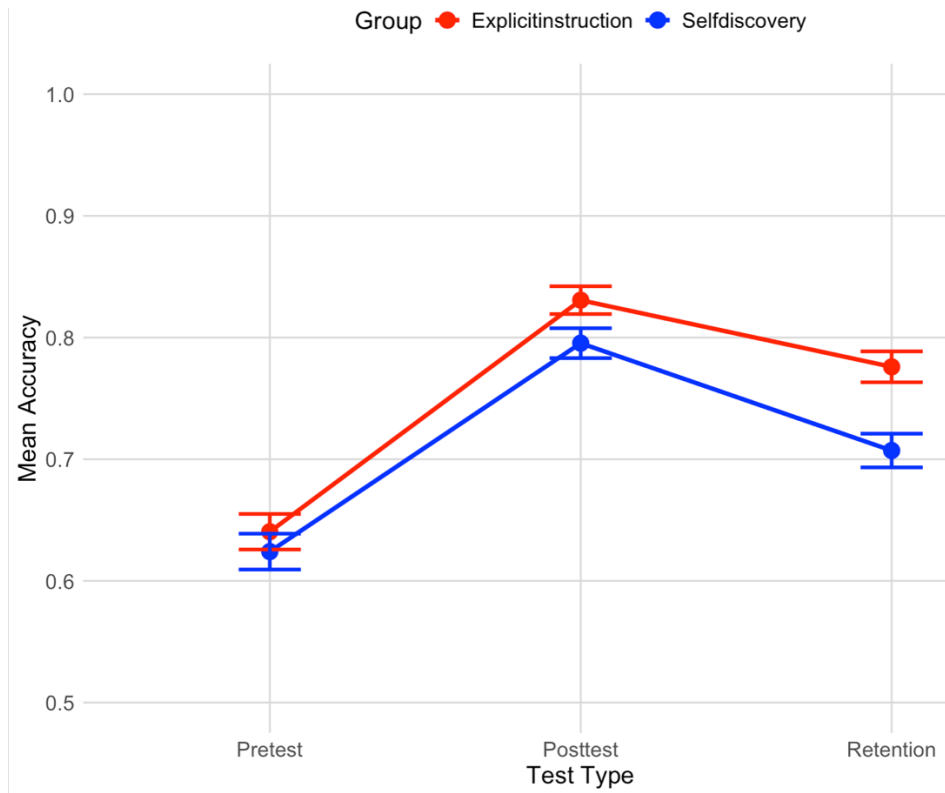
Reading aloud

A mixed-effects logistic model, with random intercepts by participant and by item, was used to analyze the effects of test (pretest, posttest, retention), group (explicit instruction, self-discovery), and stress position (first, second) on naming accuracy, while controlling for word frequency and number of phonemes. The interaction between group and test was significant for the pretest to posttest comparison ($b = 0.48$, $SE = 0.20$, $z = 2.36$, $p = 0.02$), the explicit instruction group showed a greater improvement from pretest to posttest compared to the self-discovery group. However, the interaction between group and test

for the retention phase was not significant ($p = 0.65$). Main effects revealed that both groups performed significantly better in the posttest than in the pretest ($b = -2.03, SE = 0.15, z = -13.31, p < 2e-16$), and there was a significant decline in accuracy from the posttest to the retention test ($b = -0.68, SE = 0.17, z = -3.99, p < 0.001$). Additionally, the explicit group outperformed self-discovery group ($b = -0.46, SE = 0.18, z = -2.56, p = 0.01$). There was no significant effect of stress position on performance ($p = 0.38$). Figure 2 presents the comparison of pretest, posttest, and retention performance.

Figure 2

Mean Accuracy in Reading Aloud Task Across Pretest, Posttest, and Retention for Explicit Instruction and Self-Discovery Groups



Implicit stress judgement: which one sounds better

Another mixed-effects logistic model, with random intercepts by participant and by item, was employed to investigate the effects of test (pretest, posttest, retention), group (explicit instruction, self-discovery), and stress position (first, second) on implicit stress judgement accuracy. The interaction

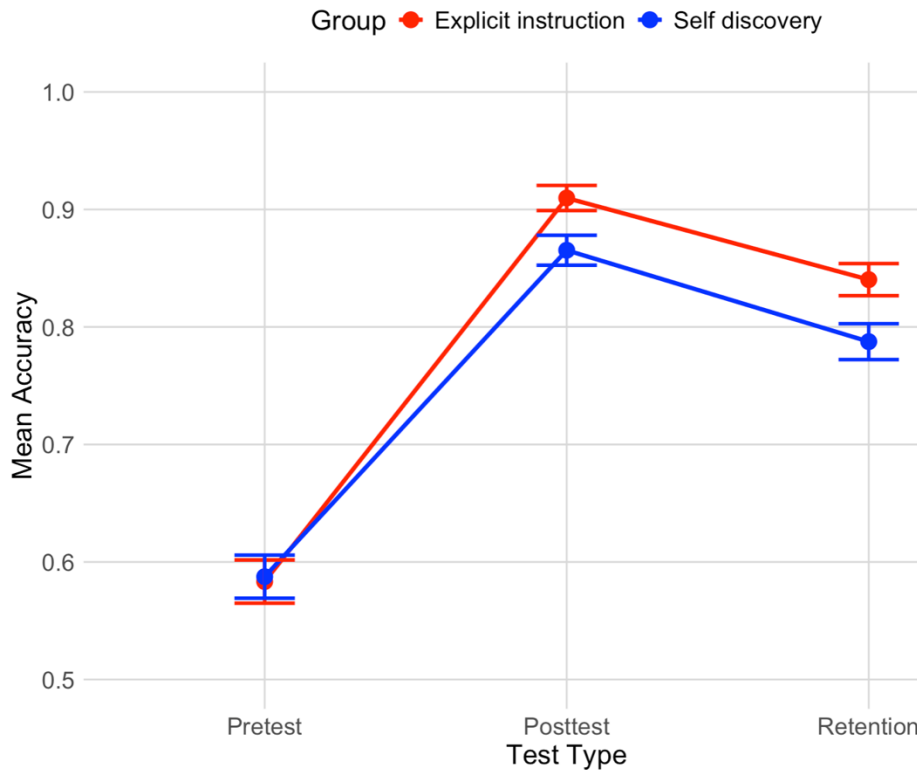
between group and test for the pretest-posttest comparison was significant ($b = 0.48$, $SE = 0.20$, $z = 2.36$, $p = 0.0184$). The explicit instruction group showed a significantly larger improvement in accuracy from pretest to posttest compared to the self-discovery group. The interaction between group and test type for the posttest-retention comparison was not significant ($p = 0.6493$). Both groups demonstrated significantly higher accuracy in the posttest compared to the pretest ($\beta = -2.02$, $SE = 0.15$, $z = -13.28$, $p < 2e-16$), indicating a substantial improvement following training regardless of whether explicit instruction was offered. There was also a significant decline in accuracy from the posttest to the retention test ($b = -0.66$, $SE = 0.17$, $z = -3.98$, $p < 0.001$), showing some loss of learning over time. Additionally, the self-discovery group performed significantly worse overall compared to the explicit instruction group ($b = -0.46$, $SE = 0.18$, $z = -2.56$, $p = 0.01$). However, there was no significant effect of stress position on performance ($p = 0.3812$), the accuracy did not differ significantly between first and second stress.

Generalization test

The interaction between group and stress position and the main effects of group and stress position were all not significant ($p = 0.972$, $p = 0.629$, and $p = 0.586$, respectively). Both groups showed above chance level performance.

Figure 3

Mean Accuracy in Implicit Stress Judgement Across Pretest, Posttest, and Retention for Explicit Instruction and Self-Discovery Groups



Explicit stress assignment

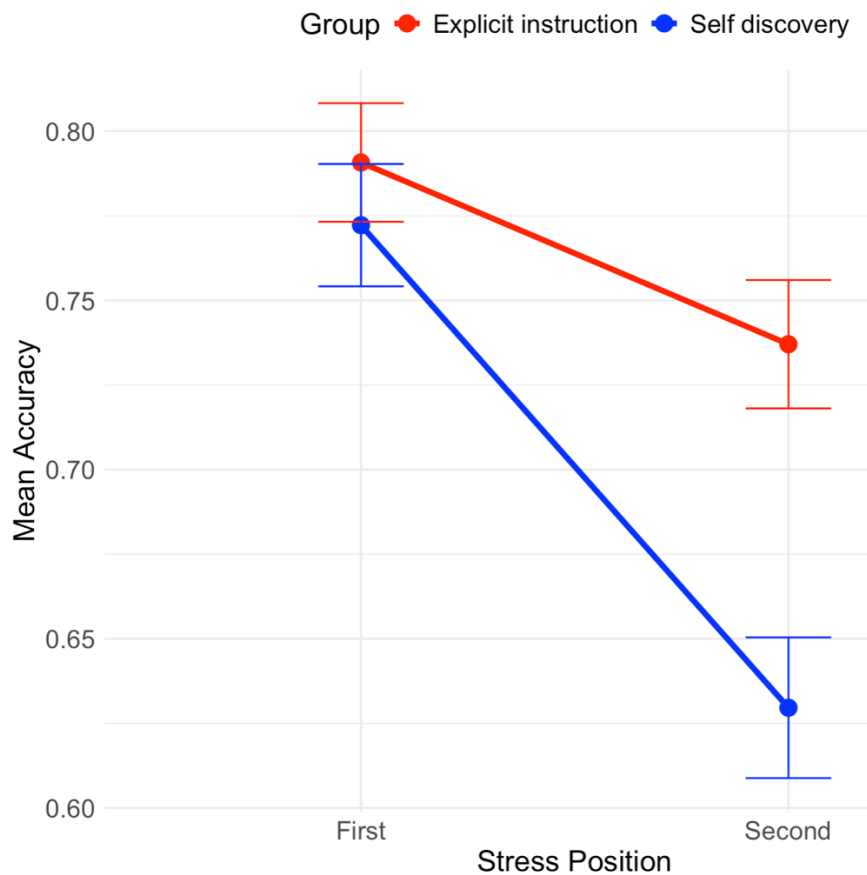
A mixed-effects model was conducted to examine the effects of test (posttest, retention), stress position (first, second), and group (explicit instruction, self-discovery) on stress assignment accuracy. The final model showed that the interaction between stress position and group was significant ($b = -0.39$, $SE = 0.20$, $z = -2.00$, $p = 0.045$). Post hoc analyses suggested that the explicit instruction group performed significantly better on the second stress compared to the self-discovery group. There was no significant difference between groups for the first stress. Participants' accuracy was higher in posttest than retention ($b = -0.61$, $SE = 0.31$, $z = -2.98$, $p = 0.008$). The main effects of stress position and group were not significant ($p = 0.36$, $p = 0.45$, respectively).

To verify syllable perception in our nonword stimuli, we recruited 10 native English-speaking college students to rate the number of syllables. For the implicit stress judgment task, native English speakers listened to the pronunciation of the nonwords, whereas, for the explicit stress assignment task, they visually assessed nonwords in the written form. Results indicated that most nonwords were consistently judged as two syllables. However, there were a few notable exceptions in the explicit stress

assignment task: *dooscior* was rated as having three syllables by 6 participants, *paiclior* by 7 participants, and *geakior* by 4 participants. We further excluded *dooscior*, *paiclior*, and *geakior* from the analysis for the explicit stress assignment task to exclude the potential influence of three-syllable perception. After removing these items, the pattern of results remained the same.

Figure 4

Mean Accuracy in Explicit Stress Assignment for Explicit Instruction and Self-Discovery Groups



Training over days

A mixed-effects model was conducted to examine the effects of stress position (first, second), group (explicit instruction, self-discovery), and training day (Day 1 to Day 4), and the interaction between group and training day on repeating accuracy. The interaction between group and training day was significant ($b = -1.57$, $SE = 0.10$, $z = -16.29$, $p < 0.001$). Post hoc analyses showed that the explicit instruction group's performance increased from Day 1 to Day 2 ($b = 0.85$, $SE = 0.12$, $z = 7.08$, $p <$

0.001), but not the self-discovery group ($p = 0.17$). From Day 2 to Day 3, the self-discovery group increased ($b = 0.72$, $SE = 0.13$, $z = 5.54$, $p < 0.001$) but not the explicit instruction ($p = 0.40$). From Day 3 to Day 4, both groups' performance did not increase ($ps = 0.65$, 0.53 , respectively, see Figure 5). These results demonstrate that although the self-discovery group showed improvement over time, their progress was slower compared to the explicit instruction group, which exhibited a rapid increase in performance during the early training.

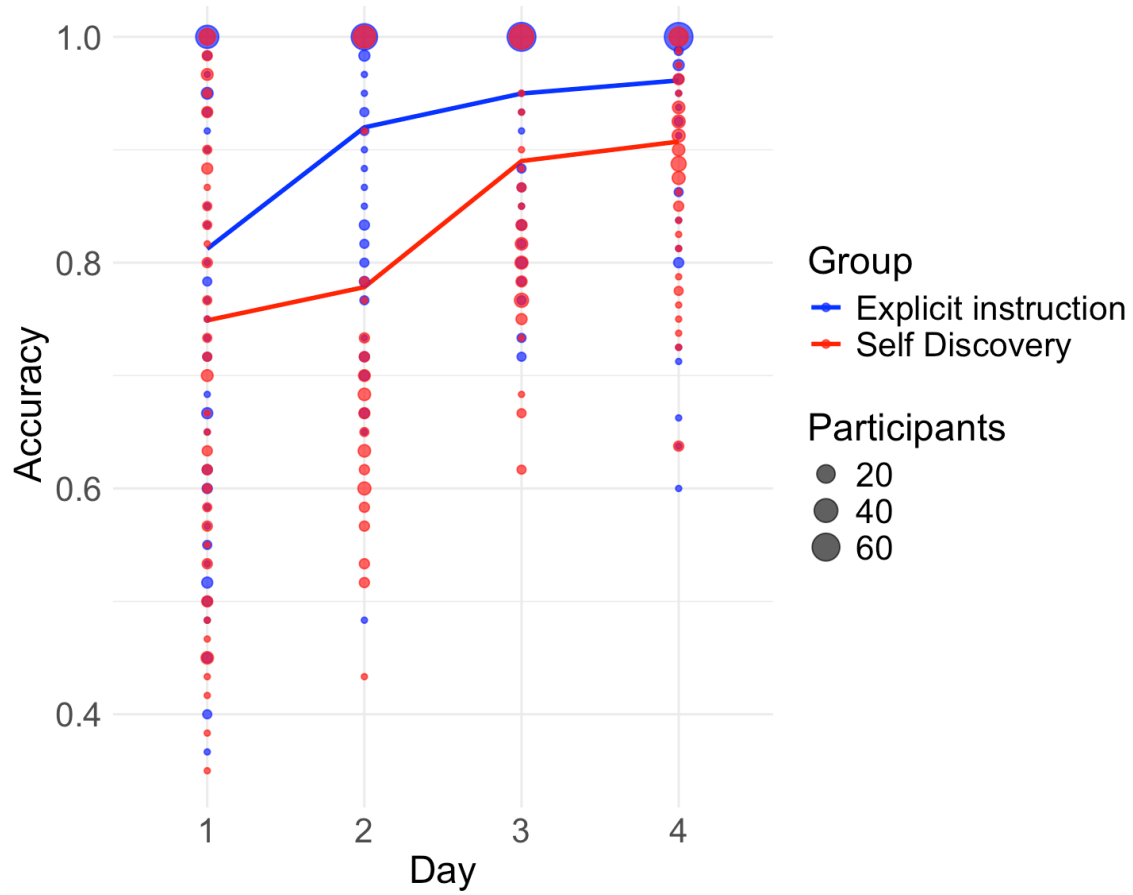
The main effect of group was significant, indicating that participants in the self-discovery group had lower accuracy compared to those in the explicit instruction group ($b = -0.93$, $SE = 0.22$, $z = -4.25$, $p < 0.001$). The training days had significant effects on accuracy, with participants showing improved performance from Day 1 to subsequent days. Specifically, accuracy improved significantly on Day 2 ($b = 1.65$, $SE = 0.08$, $z = 19.95$, $p < 0.001$), Day 3 ($b = 1.42$, $SE = 0.08$, $z = 18.30$, $p < 0.001$), and Day 4 ($b = 1.83$, $SE = 0.08$, $z = 23.31$, $p < 0.001$), compared to Day 1. There was no significant main effect of stress position ($p = 0.53$).

Post-training survey on pattern awareness by discovery group

The survey results for the discovery learning group indicated that most children, 63 (70%), did not notice any patterns, while 16 children (17.8%) were uncertain about whether they had perceived any. Only 11 children (12.2%) reported that they noticed patterns. Among these children, 3 children (3.3%) noticed "stress changes with word endings", 2 children (2.2%) mentioned "some endings have different sounds". Additionally, 3 children (3.3%) reported that they noticed patterns but could not provide further details.

Figure 5

Accuracy Over Four Training Days for Explicit Instruction and Self-Discovery Groups



Discussion

The present study investigated whether L2 children are sensitive to word endings as an orthographic cue to English stress after training, regardless of their learning condition (explicit instruction or self-discovery). Additionally, it sought to determine if explicitly teaching regularities would help L2 children learn English lexical stress more effectively than relying on learners to discover these regularities independently. Our results indicated that explicit instruction led to greater immediate gains in learning compared to self-discovery in both reading aloud and implicit stress judgment on both stress positions. Both groups showed improvement from pretest to posttest, although participants’ performance in both learning conditions decreased in the retention test after one week. Moreover, the explicit instruction group performed better than the self-discovery group in the explicit stress assignment task for the second syllable but not the first syllable.

Our results revealed that both groups improved significantly from pretest to posttest in their

ability to accurately produce stress patterns, make implicit and explicit stress judgments, suggesting that learning—whether through explicit instruction or self-guided discovery—plays a crucial role in gaining sensitivity to lexical stress. The explicit instruction group benefited from direct guidance in associating word endings with specific stress patterns, leading to more immediate improvements. Meanwhile, the self-discovery group, although not receiving direct instruction and were mostly not aware of the regularities embedded in the stimuli (based on the results from post survey), was still able to extract patterns through exposure and practice, which resulted in their learning over time. The self-discovery group employed SL to identify probabilistic cues present in the language input, such as the association between specific word endings and stress positions. Their gradual improvement over time suggests that, through exposure, participants were able to detect and internalize these statistical relationships, similar to how adults acquire phonological structures naturally without explicit instruction (Kachergis et al., 2014). Although this learning process was slower compared to explicit instruction, it led to the development of an implicit and generalized understanding of the stress patterns (Ren & Wang, 2023; Chan & Leung, 2014).

The decline in retention for both groups may be attributed to the absence of continuous reinforcement or practice after the initial training. Lexical stress patterns as complex linguistic features, often require repeated exposure and consistent practice to be retained effectively. This finding is consistent with the notion that retention of complex linguistic information, such as lexical stress, often requires extended practice and exposure (Baddeley, 1997). Without continued reinforcement, learners may struggle to maintain the level of accuracy achieved immediately after training.

The role of explicit instruction

The group that received explicit instruction demonstrated a greater improvement compared to the self-discovery group in reading aloud, implicit and explicit stress judgement. This difference underscores the potential advantage of explicit teaching approaches in facilitating the rapid acquisition of complex language rules, such as those governing lexical stress. According to the two-systems view of SL (Conway, 2020), learning can occur through two distinct routes: an implicit system that is automatic and

operates without conscious awareness, and an explicit system that involves deliberate, conscious efforts to learn. Our findings suggest that explicit instruction effectively leverages the explicit learning system, helping learners acquire complex language rules more efficiently than they could through implicit, self-discovery processes alone.

The two-systems view of SL provides a useful framework for understanding the differences in outcomes between explicit instruction and self-discovery groups. The implicit system is well-suited for learning simple, frequently occurring patterns through passive exposure, while the explicit system is more effective for complex, novel, or less intuitive patterns that require focused attention and conscious effort. In our study, explicit instruction provided targeted guidance that activated the explicit system, enabling learners to quickly grasp the association between word endings and stress positions. In contrast, the self-discovery group relied more heavily on the implicit system, which may have been less effective for the complexity of orthographic cues involved.

The explicit instruction directly pinpointed the associations between word endings and stress positions, allowing children to establish quick connections without relying solely on inference or trial and error. This finding aligns with prior research that has demonstrated the effectiveness of explicit instruction in enhancing learning outcomes such as grammatical rules when the material is novel and complex (e.g., Ellis, 2005). Furthermore, our findings align with previous studies that have demonstrated the benefits of explicitly teaching regularities embedded in writing systems, particularly in the alphabetic systems where learners benefit from systematic guidance in recognizing orthographic cues, such as grapheme-phoneme correspondences, syllable boundaries, and morphological patterns (e.g., Castles et al., 2018; Galuschka et al., 2014). These studies have shown that systematic, explicit teaching of orthographic rules helps learners quickly identify and apply consistent patterns. This approach has the potential to accelerate literacy acquisition and provides structured support, especially for younger learners or those with limited exposure to the language.

The finding that explicit instruction group performed better than the self-discovery group in the explicit stress assignment task for the second syllable suggests that explicit instruction may be

particularly effective for learning more challenging or less intuitive aspects of lexical stress, such as assigning stress to the second syllable. Prior research has indicated that explicit instruction can help learners focus their attention on specific linguistic features, enhancing their ability to learn complex patterns (Ellis, 2005; Rastle et al., 2021). The second-syllable stress pattern is less common in English (Hayes, 1985) and thus may require more focused attention or explicit learning strategies, which can be accomplished by direct instruction. In contrast, first-syllable stress may be more intuitive or easier to learn through exposure alone, making the benefit of explicit instruction less pronounced in this context. This pattern indicates that explicit instruction is particularly advantageous when dealing with complex or less predictable aspects of language learning, whereas self-discovery might suffice for simpler, more intuitive regularities.

The lack of significant difference in overall group performance during the generalization phase suggests that while explicit instruction may provide an initial boost in accuracy when combined with SL, it does not significantly enhance the ability to generalize beyond what SL alone can achieve. This finding may suggest that explicit instruction has a limited role in promoting generalization in SL, with its focus on implicit pattern extraction, is essential for achieving transferable learning outcomes (Rastle et al., 2021).

Educational implications

The findings of this study offer several important educational implications for L2 learning. Firstly, our results emphasize that SL can serve as a powerful mechanism for enabling L2 learners to internalize the relationships between orthographic cues and lexical stress patterns. This suggests that educators should consider providing ample opportunities for learners to implicitly extract these statistical regularities from reading experiences. By exposing students to varied examples that highlight consistent orthographic patterns, teachers can help learners build a more intuitive understanding of the language, even without direct instruction. Secondly, our findings suggest that explicit instruction can complement implicit SL, particularly in the initial phase of learning. However, the benefit of explicit instruction appears to diminish over time, as it does not significantly contribute to learners' ability to generalize

learned rules to novel stimuli. Therefore, educators should use explicit instruction strategically, focusing on providing initial clarity and guidance that facilitates an early understanding of complex orthographic features, such as lexical stress patterns. Explicit instruction should be followed by opportunities for students to reinforce and generalize these rules through practice and self-discovery. Finally, the decline in performance during the retention phase indicates the importance of consistent practice and probably with feedback to maintain the gains. Teachers should consider incorporating spaced repetition and review sessions to help learners retain their knowledge of lexical stress patterns over time. Varied activities, such as reading aloud, peer practice, and game-based exercises, could help solidify students' understanding of stress placement, ultimately supporting better reading fluency and comprehension in L2 learners.

Limitations and future directions

This study has several limitations that should be addressed in future research. First, our sample was limited to L2 learners with a Chinese language background, which restricts the generalizability of our findings to learners from other linguistic backgrounds. Future studies should include a more diverse set of participants to understand how different L1 phonological and orthographic systems impact the acquisition of English lexical stress. Second, the retention period was relatively short, and the observed decline in performance suggests that longer-term follow-up is necessary to understand the sustainability of the learning effects. Future research should include extended retention intervals and additional booster sessions to explore how reinforcement over time might help maintain learning gains. Third, the study employed controlled experimental tasks using nonwords in the implicit stress judgment and explicit stress assignment task, which may not fully reflect the complexities of real-world reading contexts. Future research should investigate whether the observed effects of explicit instruction and self-discovery can be applied to more naturalistic reading tasks. Finally, the word endings used in this study may inherently carry morphemic information when including the consonants prior to them such as the ending *-ude*, when paired with *cl*, forms *-clude*, which may limit the assessment of clean orthographic influence on lexical stress acquisition.

Conclusion

This study provides evidence that both explicit instruction and self-discovery play significant roles in L2 children's acquisition of English lexical stress patterns. Explicit instruction led to more immediate improvements, by providing structured guidance that facilitated a clear understanding of the association between word endings and stress position. On the other hand, the self-discovery group showed gradual progress, reflecting the power of SL through repeated exposure. Moreover, the decline observed in retention across both groups highlights the necessity of ongoing reinforcement for maintaining learning gains. Overall, a blended approach that incorporates both explicit instruction and opportunities for self-discovery may be most effective in supporting L2 learners in mastering complex orthographic and prosodic rules, leveraging the strengths of both methods to optimize learning and retention.

Bibliography

- Arciuli, J., & Cupples, L. (2006). The processing of lexical stress during visual word recognition: Typicality effects and orthographic correlates. *Quarterly Journal of Experimental Psychology*, 59(5), 920-948. <https://doi.org/10.1080/02724980443000782>
- Arciuli, J., Monaghan, P., & Seva, N. (2010). Learning to assign lexical stress during reading aloud: Corpus, behavioral, and computational investigations. *Journal of Memory and Language*, 63(2), 180–196. <https://doi.org/10.1016/j.jml.2010.03.005>
- Arciuli, J., Torkildsen, J. V. K., Stevens, D. J., & Simpson, I. C. (2014). Statistical learning under incidental versus intentional conditions. *Frontiers in Psychology*, 5, 747. <https://doi.org/10.3389/fpsyg.2014.00747>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.
- Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, 83, 62-78. <https://doi.org/10.1016/j.jml.2015.04.004>
- Chan, R. K., & Leung, J. H. (2014). Implicit learning of L2 word stress regularities. *Second Language Research*, 30(4), 463-484. <https://doi.org/10.1177/0267658313510169>
- Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, 11(3), 468-481.
- Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience & Biobehavioral Reviews*, 112, 279-299.
- Cutler, A. (2015). Representation of second language phonology. *Applied Psycholinguistics*, 36(1), 115–128. <https://doi.org/10.1017/S0142716414000459>
- Duanmu, S. (2007). *The phonology of standard Chinese*. OUP Oxford.

- Dunn, L. M., & Dunn, D. M. (2007). Peabody Picture Vocabulary Test--Fourth Edition (PPVT- 4) [Database record]. APA PsycTests. <https://doi.org/10.1037/t15144-000>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191. <https://doi.org/10.3758/bf03193146>
- Field, J. (2025). Cognitive Insights Into First And Second Language Listening. In *The Routledge Handbook of Second Language Acquisition and Listening* (pp. 13-28). Routledge.
- Goo, J., Granena, G., Yilmaz, Y., Novella, M., & Rebuschat, P. (2015). Implicit and explicit instruction in L2 learning. *Implicit and Explicit Learning of Languages*, 48, 443-482
- Hamrick, P., & Rebuschat, P. (2012). How implicit is statistical learning. *Statistical Learning and Language Acquisition*, 365-382. <https://doi.org/10.1515/9781934078242.365>
- Kachergis, G., Yu, C., & Shiffrin, R. M. (2014). Cross-situational word learning is both implicit and strategic. *Frontiers in Psychology*, 5, 588. <https://doi.org/10.3389/fpsyg.2014.00588>
- Kalcounis-Rueppell, M. C., Metheny, J. D., & Vonhof, M. J. (2006). Production of ultrasonic vocalizations by *Peromyscus* mice in the wild. *Frontiers in Zoology*, 3(1), 1-12. <https://doi.org/10.1186/1742-9994-3-3>
- Kelly, M. H. (2004). Word onset patterns and lexical stress in English. *Journal of Memory and Language*, 50(3), 231–244. <https://doi.org/10.1016/j.jml.2003.12.002>
- Kelly, M. H., Morris, J., & Verrekia, L. (1998). Orthographic cues to lexical stress: Effects on naming and lexical decision. *Memory & Cognition*, 26(4), 822–832. <https://doi.org/10.3758/BF03211401>
- Lipkind, D., & Tchernichovski, O. (2011). Quantification of developmental birdsong learning from the subsyllabic scale to cultural evolution. *Proceedings of the National Academy of Sciences*, 108(supplement_3), 15572-15579. <https://doi.org/10.1073/pnas.101294110>
- Lord, G. (2007). The role of the lexicon in learning second language stress patterns. *Applied Language Learning*, 17(1/2), 1.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... & Lindeløv, J.

K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195-203. <https://doi.org/10.3758/s13428-018-01193-y>

Rahimpour, M., Salimi, A., & Farrokhi, F. (2012). The Effect of Intensive and Extensive Focus on Form on EFL Learners' Written Accuracy. *Theory & Practice in Language Studies (TPLS)*, 2(11).

Rastle, K., Lally, C., Davis, M. H., & Taylor, J. S. H. (2021). The dramatic impact of explicit instruction on learning to read in a new writing system. *Psychological Science*, 32(4), 471-484. <https://doi.org/10.1177/0956797620968790>

Ren, J., & Wang, M. (2023). Sensitivity to word endings as probabilistic orthographic cues to lexical stress among English as second language learners. *Memory & Cognition*, 1-17. <https://doi.org/10.3758/s13421-023-01432-4>

Ren, J., Wang, M., & Conway, C. M. (2024). Can explicit instruction boost statistical learning? A meta-analytical review. *Journal of Educational Psychology*, 116(7), 1215–1237. <https://doi.org/10.1037/edu0000897>

Sagarra, N., Fernández-Arroyo, L., Lozano-Argüelles, C., & Casillas, J. V. (2024). Unraveling the complexities of second language lexical stress processing: The impact of first language transfer, second language proficiency, and exposure. *Language Learning*.

Torgesen, J. K., Wagner, R., & Rashotte, C. (2012). *Test of Word Reading Efficiency* (2nd ed.). Austin, TX: Pro-d. <https://doi.org/10.1177/0829573515594334>

Treiman, R., & Kessler, B. (2022). Statistical learning in word reading and spelling across languages and writing systems. *Scientific Studies of Reading*, 26(2), 139-149. <https://doi.org/10.1080/10888438.2021.1920951>

Wiener, S., Chan, M. K., & Ito, K. (2020). Do explicit instruction and high variability phonetic training improve nonnative speakers' Mandarin tone productions?. *The Modern Language Journal*, 104(1), 152-168. <https://doi.org/10.1111/modl.12619>