Collocations in the L2: Do L2 Proficiency and L1 Congruency Affect Collocation Learning among Chinese High School EFL Learners?

Sihan Fang

MSc in Applied Linguistics and Second Language Acquisition, 2024

DECLARATION BY THE CANDIDATE AS AUTHOR OF THE DISSERTATION

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.
2. I allow the Department to deposit on my behalf a copy of this dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].
3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals (unless permission has been obtained from the individuals) and that such material should be removed before this dissertation is deposited in ORA.
4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [https://ora.ox.ac.uk/deposit_agreements] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.
5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]:	
Date:	15/08/2024

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Anna-Maria. Thank you for always keeping me on track and offering detailed, line-by-line feedback throughout the course of my dissertation. As someone coming from a non-academic background, I have learned immensely from you about the qualities of a good researcher and lecturer. I am deeply grateful for all the time and effort you invested in my academic growth.

I am blessed to have friends, both new and old, who supported me throughout the writing of this dissertation. To my ALSLA course mates and fellow language enthusiasts, I appreciate the discussions we had in room G/H, the shared struggles with assignments at cafes and libraries, the shared fun at squash games, pub quizzes, board games, and the nights out. My Oxford experience was so much more vibrant because of you. A special shoutout to Ivy, my fellow introvert—your proficiency with academic tools and expertise in Excel formulas were not only a lifesaver during my statistical analysis but also a true demonstration of excellence. My research direction was inspired by the doctoral work of the rising star linguist, Zhiyuan Deng from the University of Maryland, who kindly shared his test design for my reference. My college friend and fellow glove hunter Oksana, who always challenges me with fresh perspectives, offered invaluable feedback that helped me refocus during the final stretch of the writing process. Rock, my best friend, thank you for contributing absolutely nothing to this dissertation, but for always being there for me.

I am deeply indebted to Professor Hongchen Xu, whom I am fortunate to also call a friend. Thank you for generously taking the time to listen to my research plan and for giving me advice. I learned so much about critical reading from just one video call with you. Your years of wisdom have continued to guide me well beyond graduation, enriching my teaching career and leading me deeper into the world of academia.

I am grateful to my former colleagues at Reachtop Education in Beijing and Shanghai. Thank you to all the teachers who participated in the pilot study and provided valuable feedback that helped improve the experiment. The directors, Li Pengcheng and Feng Sisi, were exceptionally supportive in assisting with the recruitment of student participants and coordinating with their tutors to supervise the experiment tasks. The participation of the 36 students and the cooperation of 13 teachers—Wen Xin, Cut Limin, Zhou Jiayi, Jie Yuchun, Wang Xinmiao, Fan Wenxiao, Liu Yadan, Kong Ziling, Liu Liping, He Yuheng, Liu Weiran, Lu Miaona, Ma Zixiao—were crucial in gathering valuable data that significantly contributed to the success of this research.

My one-year adventure to Oxford would not have been possible without the support from my wife and family. Thank you for enabling me to pursue my dream.

Last but not least, I want to thank myself for not giving up. Well done.

Abstract

This dissertation investigates the effects of L2 proficiency and L1 congruency (i.e. the presence or absence of a literal translation equivalent in L1) on the learning of collocations among Chinese high school EFL learners. Through a phrasal acceptability judgement task (PAJT) with 34 participants, this study measures accuracy rate (AR) and reaction time (RT) to assess collocational processing. The findings indicate that higher proficiency learners (B2) perform better in accuracy than lower proficiency learners (B1), but no significant difference in processing speed is observed. Additionally, congruent collocations are processed more accurately and quickly than incongruent ones, yet no interaction effect between proficiency and congruency is found. These results suggest that while proficiency enhances accuracy, both congruency and proficiency independently influence collocational learning. The study concludes with pedagogical implications, highlighting the importance of exposure and explicit instruction in improving collocational knowledge.

Keywords: L2 proficiency, L1 congruency, collocations, High School EFL learners, Phrasal Acceptability Judgement Task

Contents

List of Tables

List of Figures

1. Introduction

When learning a second language (L2), building up vocabulary knowledge is fundamental. Its importance in second language acquisition (SLA) has been frequently confirmed in both academic research and classroom practices. However, lexical competence encompasses not only breadth (i.e., the number of words one knows) but also depth (i.e., the extent of one's knowledge about those words), with collocation knowledge being a critical component of vocabulary depth (Schmitt, 2014). According to Nation (2022), knowing a word involves understanding its form, meaning, and use. This includes grammatical functions, constraints on word usage and collocations.

Collocation is usually described as the tendency of certain linguistic items habitually co-occur with certain others. When Firth (1957) introduced the term 'collocation', he said that 'you shall know a word by the company it keeps' (1957:196). Examples of collocations include the adjective *strong* that collocates with the noun *wind* and the adjective *heavy* that collocates with the noun *rain*.

Howarth (1998) suggests that knowing and using collocations positively impacts an L2 speaker's fluency. Pawley and Syder (1983) argue that the proper use of multi-word units enables L2 speakers to sound more 'native-like,' making their speech idiomatic and fluent. However, researchers and teachers have long acknowledged that even advanced L2 speakers struggle to acquire, understand, and use formulaic language, particularly collocations (Nesselhauf, 2005; Wray, 2002). Consequently, collocations have received considerable academic attention in SLA research. Existing work has identified two prominent factors affecting collocation learning: proficiency and congruency, i.e., the availability of direct word-for-word translations of L2 collocations in the L1.

Despite numerous studies (e.g. Ding & Reynolds, 2019; Gyllstad & Wolter, 2015; Wolter & Yamashita, 2015, 2017; Yamashita & Jiang, 2010), the relationship between collocation knowledge and general proficiency or congruency remains inconclusive, often due to methodological limitations in measuring proficiency levels and developing collocation items. Moreover, many collocational studies involve adult learners, while there is a scarcity of SLA research on L2 English with high school participants. Learning collocations is crucial for high school L2 learners who are on the verge of university education, potentially abroad, and whose L2 proficiency can significantly influence their learning experience and integration into a foreign environment. Extending collocational research to high school learners can provide essential insights into classroom instruction and pedagogy, benefiting this demographic.

This paper also highlights an often-overlooked yet important area in education: private tutoring. Private tutoring plays a significant role in SLA and collocation learning due to its complementary

role in education (Bray, 2021) as well as its flexibility and individualised nature (Yung & Bray, 2021), which can be facilitative for collocation learning.

The present study aims to contribute to existing collocation research by investigating whether proficiency and congruency affect collocation learning among Chinese high school learners and if there is an interaction between these factors. To ensure a rigorous methodology, the study triangulates proficiency measurements and meticulously details the item development process. A quantitative approach was employed for data collection and analysis. The target population consisted of high school L1 Chinese learners of L2 English with Common European Framework of Reference for Languages (CEFR) proficiency levels of B1 and B2. General English proficiency was measured using a combination of self-ratings, standardised test scores, and LexTALE (Lemhöfer & Broersma, 2012). Data was collected through a phrasal acceptability judgement task (PAJT), with accuracy rate (AR) and reaction time (RT) recorded to measure collocational processing. The results suggested that while the B2 group outperformed the B1 group in terms of accuracy, there was no difference in their processing speed. Additionally, proficiency was not found to correlate with either AR or RT. Congruent items were processed more accurately and faster than incongruent items, but no interaction was found between proficiency and congruency.

The structure of this dissertation is as follows: Chapter 2 reviews the literature that shaped the research questions and methods. Chapter 3 outlines the research questions and provides a detailed justification for the methods employed to address them. Chapter 4 presents the results of the statistical analyses. Chapter 5 offers a detailed discussion of these results, addresses the study's limitations, and provides recommendations for teaching collocations in the L2 classroom. Finally, Chapter 6 summarises the conclusions derived from this study.

2. Literature Review

This chapter provides an overview of collocational studies in the existing second language acquisition (SLA) literature. Section 2.1 discusses how collocation is defined in current SLA research. Section 2.2 describes the two traditional methods of identifying collocations and how they will be operationalised in the present study. Section 2.3 explores the difficulties L2 learners face in learning collocations. Section 2.4 focuses on two important factors affecting collocation learning: L1 congruency and L2 proficiency. Section 2.5 examines English private tutoring (EPT) as a significant area in language education and its relevance to collocation learning. Finally, Section 2.6 summarises the chapter and presents the gaps the present study aims to address.

2.1. Defining Collocation

When learning a foreign language, the initial focus is naturally placed on vocabulary acquisition, as words are the fundamental building blocks of any language. As vocabulary knowledge expands, the potential combinations of these words also increase, thereby facilitating more complex and meaningful expressions. However, as researchers in the field of SLA have demonstrated, the composition of language is not as unlimited as might initially be assumed. It has been acknowledged in second language vocabulary acquisition research that lexical proficiency extends beyond knowledge of single words, and thus, learners should be given exposure to longer stretches of language.

According to usage-based accounts of L2 acquisition, language is formulaic in nature, and most linguistic information is stored in the form of multi-word units or chunks (Wray, 2002). Through sufficient exposure in meaningful contexts, L2 learners are thought to develop the ability to accurately, rapidly, and subconsciously access these chunks in response to specific contextual and linguistic cues (Ellis, 2012). It is believed that such formulaic chunks can spur morphosyntactic development through abstract, schematic analysis of the chunks' constituent parts (Tomasello, 2003). Corpus-based investigations have shown that a large proportion of oral discourse among L1 speakers (30-40%) appears to be formulaic in nature (Leech, 2000). On the other hand, researchers have observed that L2 users produce spontaneous speech that is comparatively lacking in the use of formulaic language and that often contains recurrent dysfluencies such as filled pauses (Foster, 2001). From this point of view, learners' attainment of collocational knowledge can be seen as instrumental to becoming functional, communicatively successful L2 users (Wray, 2000).

The last two decades have seen a growing body of SLA research on formulaic language, a vast amount of which focuses on collocations. In any line of research, the initial step is to ensure

absolute clarity regarding the definition of terms, many studies suffer from the obfuscation of term usage, and existing work on collocations is no exception. Since Palmer (1933) introduced the concept, a cohesive definition of collocations has been lacking. While Palmer offered a pedagogically oriented definition of collocations as 'successions of words [that] must or should be learnt…as an integral whole or independent entity, rather than by the process of piecing together their component parts' (1933:4), scholars have proposed various interpretations from different angles (Firth, 1957; Sinclair, 1991; Cowie, 1994; Wray, 2002). For instance, Firth (1957) defined collocations as 'words in habitual company' and emphasised the statistical co-occurrence of words. This perspective views collocations primarily through the lens of frequency in language use. Sinclair (1991) proposed another approach, suggesting that collocations involve the occurrence of words together in a predictable manner. He distinguished between the idiom principle and the open-choice principle, emphasising the fixed nature of collocations compared to the more flexible combinations in language. Cowie (1994) provided yet another perspective, defining collocations as semi-fixed expressions that range from fully fixed idioms to more flexible word combinations. His definition highlights the spectrum of fixedness in collocations, acknowledging the variability in how closely words are associated. Benson, Benson, and Ilson (1986) offered a lexicographic approach, categorising collocations into grammatical and lexical types. They emphasised the importance of both grammatical structures and lexical combinations in understanding collocations. More recent research continues to explore and refine these definitions. Nesselhauf (2003) examined learner corpora to better understand how non-native speakers acquire and use collocations, highlighting the complexities involved in collocation learning. McCarthy and O'Dell (2017) focused on teaching collocations to learners, proposing practical strategies for integrating collocation learning into language instruction. Additionally, recent computational approaches, such as those by Evert (2008) and Seretan (2011), have leveraged large corpora and statistical methods to identify and analyse collocations more precisely.

Within the research literature, there is a general consensus concerning two features of collocations. First, a key feature of collocation is co-occurrence, i.e. the tendency for two or more words to appear together. This phenomenon is considered essential by many researchers when discussing collocation (e.g., Biber et al., 2002; Handl, 2008; Nesselhauf, 2005; Schmitt, 2010; Sinclair, 1991). It is important to note that co-occurrence does not necessarily mean the words must be next to each other; the elements of a collocation can be separated (Halliday & Hasan, 1976). For example, while 'wait' is a collocate of 'bus', it often appears with a preposition, an article and possibly other elements in between (*wait for the morning bus*). Researchers generally believe that

such co-occurrence is not random but is influenced by specific relationships between the words, such as topic, register, style, sociolects, semantic prosody, connotation, lexical repulsion, and grammaticality (Bednarek, 2008; Renouf & Banerjee, 2008; Schmitt & Carter, 2004; Whitsitt, 2005).

Second, collocations are typically categorised into two types: grammatical collocation and lexical collocation (e.g., Bahns, 1993; Benson et al., 1986; Cowie, 1981; Durrant, 2009; Gries, 2008; Schmitt, 2000). A grammatical collocation involves a dominant content word (such as a noun, adjective, or verb) and a subordinate grammatical structure (like a preposition, infinitive, gerund, or clause). Examples of grammatical collocations include phrases like 'famous for', 'look forward to', 'in terms of', and 'out of respect'. Conversely, a lexical collocation comprises two content words that contribute almost equally to the overall meaning. Examples of lexical collocations are 'throw+party', 'bee+buzz', and 'heavy+smoker'. Notably, lexical collocations can be systematically classified based on their syntactic patterns. For instance, 'throw a party' is a verb-noun collocation, and 'a heavy smoker' is an adjective-noun collocation. In most cases, the investigations into L2 collocations were narrowed down to collocations of particular structures. For example, Peters (2016) examined verb-noun collocations, Gyllstad and Wolter (2016) focused on delexical verbs, and Vasiljevic (2014) studied adverb-adjective collocations.

2.2. Operationalising Collocations: Phraseological vs. Frequency-based Approaches

When it comes to operationalising collocations in empirical research, there have generally been two approaches. One approach is to define collocations in a phraseological sense (Cowie, 1981; Howarth, 1998; Nesselhauf, 2003). Within this framework, an important distinction is made between combinations restricted solely by semantic properties (known as 'free combinations') and combinations restricted by arbitrary conventions, and this arbitrary restriction on the substitutability of possible word combinations is the defining character of a collocation. Nesselhauf (2003) provides an illustration of this difference:

> For example, in the combination *read a newspaper*, the reason that substitutions resulting in combinations such as **drink a newspaper* or **read water* are not possible or at least highly unusual is that *drink* requires a noun with the semantic property of 'liquid' and *read* requires a noun with the semantic property of 'containing written language'. In the combination *reach a decision*, on the other hand, *decision* can be substituted by a number of nouns denoting 'a particular aim' (*OALD*) such as *conclusion*, *verdict*, *compromise*, or *goal* but not, for example, by *aim*; this restriction does not seem to be a result of the semantic properties of the two elements concerned, but a somewhat arbitrary convention of the language. (Nesselhauf, 2003: 225)

Therefore, by this definition, phrases like *perform an experiment* should be considered restricted collocations because *perform* cannot naturally combine with all possible nouns, such as *survey*, implying arbitrary limitations on substitutability. Conversely, phrases such as *want a car* would be seen as free combinations instead of collocations, since *want* or *car* can collocate with numerous nouns/verbs without arbitrary constraints. The selection criterion has been adopted in many recent studies in existing literature (e.g. Jiang, 2022; Lee, 2021; Wolter & Yamashita, 2018).

However, without clear distinctions between collocations and other linguistic expressions, even the phraseological approach struggles to categorise word combinations as either restricted collocations or free combinations. This challenge has led to the adoption of a frequency-based perspective (Firth, 1957; Sinclair, 1987, 1991; Li & Schmitt, 2010). Halliday et al. (1964) described collocation as the tendency of a lexical item to co-occur with one or more words. Sinclair (1966) emphasised the probability of co-occurrence, acknowledging that while virtually any word combination is possible, some are more likely than others. Later, Sinclair (1991) defined collocation as the occurrence of two or more words within a short distance of each other in a text, differentiating between 'significant' (frequent) and 'casual' (infrequent) collocations. Thus, collocation has been reduced to a statistical concept, where the semantic nature of the words is irrelevant. The development of corpora has greatly supported the frequency-based approach to collocation studies. In this approach, corpus-derived collocational frequency values are crucial.

To differentiate statistically significant co-occurrences from random ones, measures of linkage strength are often used (Ellis et al., 2008; Tremblay & Baayen, 2010; Yi, 2018). For instance, mutual information (MI) is a measure used in collocation studies (Durrant & Doherty, 2010) to gauge the strength of the statistical association between words. A higher MI value indicates a stronger statistical association. According to Hunston (2002), an MI score of 3 or above typically indicates a significant collocation, a threshold supported by many scholars (e.g., Durrant & Doherty, 2010; Wolter & Yamashita, 2015). Shin and Chon (2019) suggest that in the Corpus of Contemporary American English (COCA), a minimum frequency of 20 occurrences should be used as the cut-off point due to its large size of over one billion words, which is ten times larger than the British National Corpus (BNC) (2019: 611). Since this research used COCA to identify collocations, the operational definition of collocation is as follows: A collocation is a combination of two or more words that (1) is relatively transparent in meaning, (2) has a restricted range of co-occurrence in at least one of its elements, (3) presents a minimum MI score of 3.0, and (4) has a minimum frequency of 100 tokens in COCA.

2.3. Difficulties in L2 Collocation Learning

The importance of collocational knowledge in L2 competence has gained increasing recognition in recent years. Knowing how words combine to create meaning is essential for all language use. Lewis (2000) highlighted that both native speakers and successful advanced learners possess high collocational competence—a substantial mental lexicon of phrases—that aids in fluent, accurate, and appropriate language use.

Research has identified several key functions of collocations in language acquisition: Firstly, collocations underpin the development of creative language in both first language and childhood second language acquisition (Peters, 2016; Wray, 2000). Secondly, they facilitate fluency in spoken and written communication, as the brain is better equipped for memorising prefabricated chunks than for processing new combinations, thereby reducing cognitive load (Pawley & Syder, 1983). Thirdly, collocations support comprehension, allowing readers to understand text without focusing on every word (Hunston & Francis, 2000). This predictability enhances fluency (Wray, 2002) and overall processing speed (Pawley & Syder, 1983; Conklin & Schmitt, 2008). Lastly, accurate use of collocations helps learners sound more like native speakers (Wray, 2002).

Despite their widely acknowledged importance, collocations pose a major challenge for L2 learners. Many difficulties in L2 collocation learning have been documented in SLA literature. One of the difficulties lies in the idiosyncratic nature of collocational use. For example, while 'strong' has the same meaning as 'powerful' in 'a strong/powerful argument', 'powerful' collocates with 'car' but 'strong' collocates with 'tea'. Also, the arbitrariness of collocational use is distinctly different across languages. Wray observed, 'In English you run a business, but in German you lead it (*ein Geschäft führen*). In English you smoke a cigarette, but in Hindi you drink it (*sigaret piinaa*). In English you lie in the sun, but in Russian you lie on it (*na solntse*) (2002: 73).' Similar to the learning of many other aspects of L2, the biggest challenge for learners is the lack of exposure to the target language. Whereas native speakers acquire the knowledge of collocation subconsciously and gradually as they grow up in their speech community and receive constant natural exposure, most learners do not have this opportunity, especially those who learn the L2 only in the classroom environment (Schmitt & Schmitt, 2020).

Howarth (1998) examined verb-object collocations from native and non-native written corpora, discovering that native speakers used about 50% more restricted collocations (e.g., *reach a conclusion*) than advanced non-native speakers. Notably, deviations from standard collocational forms were present in both native and non-native writing, with a significantly higher proportion of

non-idiomatic language found in non-native essays. Specifically, non-standard collocations constituted approximately 6% of the collocations produced by learners, compared to only 1% in native-speaker texts. Howarth's research also indicates that among the three collocational categories – restricted collocations, free collocations, and idioms – learners struggle the most with producing restricted collocations.

Interestingly, L2 learners do not always recognise the gaps in their knowledge regarding this particular aspect of vocabulary. Martinez and Murphy (2011) presented 101 adult Brazilian learners of English with paired texts; one text contained collocations (of varying degrees of non-transparency) and another text contained the same individual lexical items not in phrases. Not only did participants show lower comprehension of the collocation texts, they also significantly overestimated their own comprehension of passages containing collocations, indicating that learners wrongly assumed they had understood the phrases via the literal interpretation or that they did not notice collocations in the first place. As Boers (2020) explains, L2 readers might not pay attention to collocations during reading if they consist of familiar words (e.g., *do your homework*) and when their meaning is transparent. Learners might attend to the part of the collocation that carries the most meaning, and the collocational patterning might be unnoticed (Boers, 2020). In fact, empirical evidence suggests that L2 learners prioritise individual words over collocational patterning (Hoang & Boers, 2016). Moreover, when non-transparent collocations are presented in a supporting text context, L2 learners sometimes interpret the phrase literally, even when a literal interpretation does not match the context. Bishop (2004) found that L2 English learners used an electronic glossary to look up the meanings of unknown individual words more than unknown phrases while reading. Bishop (2004) interprets this finding as indication that L2 learners do not notice unknown phrases. It is also possible that learners recognised the unknown phrases but failed to use the glossary, incorrectly perceiving that it was a resource exclusively for single word meanings.

Therefore, it could be argued that raising awareness is crucial for recognising and learning collocations. De Bot, Paribakht and Wesche (1997) maintained that recognition of an unfamiliar lexical form is essential, and also the conscious awareness that the form lacks a meaning, before steps can be taken to establish its meaning, as is necessary for subsequent learning. Clearly if a subject does not know that he or she does not know a word, learning is not likely to take place. Thus Schmidt's (1990) noticing hypothesis is an important theoretical part of this argument. Schmidt (1990) argued that if a learner is not consciously aware of a specific language feature, i.e., is unable to articulate that it is problematic, the learner will not be able to learn that language feature whether

grammatical, lexical or pragmatic. It is quite plausible that second language readers do not notice unknown formulaic sequences.

2.4. Factors Affecting L2 Collocation Acquisition: Congruency and Proficiency

To increase learners' awareness and likelihood of noticing collocations in L2 learning, it is crucial for educators to understand the factors that affect collocation acquisition. Research has identified two major factors: L2 proficiency and L1 congruency. This section will explore current studies that have investigated these factors and highlight the limitations in their methodologies.

2.4.1. L2 Proficiency

L2 proficiency refers to the level of language ability in the second language, which can be measured through standardised English proficiency tests or the amount of formal L2 instruction received. The impact of L2 proficiency on collocation learning has been well-documented, although results are mixed. Several studies have found correlations between collocational knowledge and overall proficiency in learners whose L1 is Chinese (Pei, 2008), Swedish (Gyllstad, 2007), and Japanese (Yamashita & Jiang, 2010). Pei (2008) reported increased collocation use from beginner to advanced learners in several Chinese studies. Similarly, Bonk (2001), Gyllstad (2007), and Revier (2009) noted improvements in collocational development across different proficiency levels. However, some studies found opposite patterns. Bahns and Eldaw (1993) did not observe significant differences among learner groups. Ding and Reynolds (2019), who examined collocational use across three proficiency levels, identified some development among advanced learners. Despite this, even advanced learners produced collocations that deviated from native usage and tended to make more errors, although they used more collocations than other groups.

One explanation for these discrepancies could be the use of different proficiency measures and the varying efforts to ensure the validity of these measures. On the more rigorous end, Bonk (2001) used a battery of tests, including a 49-item test adapted from the TOEFL format. In contrast, many other studies seldom reported the reliability values of their proficiency instruments (Gyllstad, 2007). For instance, Ding and Reynolds (2019) simply used scores from two English proficiency tests in China, College English Test (CET) and Test for English Majors (TEM), without providing any justification. Many other bilingual studies within experimental psychology rely on participants' self-ratings of proficiency and language background questionnaires as the only source of proficiency information. Given the central role of proficiency— or vocabulary knowledge, in the case of collocation processing—in L2 research, it is interesting how little consensus there is on how to

measure it. From this, it becomes clear that a practical and valid approach is needed to measure the relevant aspects of proficiency in a quick and easy manner and, if possible, to reach some standardisation across different research groups. Against this background, LexTALE was developed as a quick tool to assess vocabulary knowledge and, by extension, general proficiency. The justification for employing LexTALE will be further discussed in the Methodology chapter.

Another issue that might have led to the inconsistent results might concern the development of collocation items used for research. Different approaches have led to different selection standards for collocations. For example, in Webb, Newton, and Chang (2013), following a frequency-based approach to collocation, an item such as pull strings is treated as a collocation; in the phraseological tradition, the same item would be treated as an idiom, as its meaning (to secretly use your influence with important people in order to get what you want or to help someone else) is not compositionally available. In another example, Wolter and Brent (2013) developed items for their test using an exclusively frequency-based approach. Due to the lack of phraseological consideration, an item such as 'small room' was considered a collocation, whereas it would be regarded as a free combination from a phraseological perspective. The authors acknowledged this limitation, stating that 'any corpus, no matter how large and how divergent its sources, could ever precisely replicate a given language user's experience with the target language' (Wolter & Brent, 2013: 457).

The use of frequency-based approach alone, though reliable in terms of replicability, could also have potential limitations. For example, Ding and Reynolds (2019) found that L1-English native speakers displayed higher accuracy rate for congruent collocations than incongruent collocations, which was an unexpected result since there should be no interference of L1-Chinese for the L1-English native speakers. A post-interview with the L1-English participants revealed that they regarded some incongruent items as 'implausible' 'uncommonly used' and 'incomplete' (2019: 348) The authors attributed this to their attempts to be objective in the selection of the collocations by choosing only the frequency approach in selecting collocations as stimuli used as stimuli for the experiment. In other words, solely relying on the frequency approach is limited, and L1-English speakers should be consulted in the process of item development to make sure the phrases are actually spoken by the native speakers.

2.4.2. L1 Congruency

Collocation congruency between L1 and L2 is typically discussed in terms of translation equivalence (Men, 2018; Nesselhauf, 2005; Wolter & Gyllstad, 2011). Nesselhauf (2005) defines congruency as the presence or absence of a literal translation equivalent. A collocation is considered

congruent if its constituent words in one language have direct translation equivalence.i.e, the kind of equivalence attained by literal word-for-word translation in another language. For example, the phrases *take a drug* and *cut a deal* are considered incongruent collocations in L1 Chinese, because to convey the English meaning requires using alternative word combinations, mainly by switching the constituent verb (e.g., 吃药, liteally *eat a drug**, for *take a drug*; 达成交易, lit. *reach a deal*, for *cut a deal*).

Several studies have investigated the role of L1 in the development and use of collocations, with a general consensus that EFL learners tend to encounter more difficulties with incongruent collocations than congruent ones (Cao & Badger, 2021; Men, 2018; El-Dakhs, Salem, & Al-Haqbani, 2020; Wolter & Gyllstad, 2011, 2013). For instance, using a primed lexical decision task and a test of receptive collocational knowledge (2011), and an acceptability judgment task (2013), Wolter and Gyllstad found that incongruent collocations caused significantly more errors for non-native English speakers (L1 Swedish) than congruent collocations, even at an advanced level of English. Additionally, collocations with L1-L2 equivalents were processed much faster than non-congruent collocations. These findings indicated that EFL learners handled congruent collocations better than incongruent ones, though the results were limited to the L1 Swedish context. This study supported Yamashita and Jiang (2010), who examined the recognition of English verb-noun and adjective-noun collocations among Japanese learners and reported significantly greater accuracy in judging the acceptability of congruent collocations compared to incongruent ones, regardless of proficiency level.

However, some recent studies found no processing advantage for congruent collocations. Wolter and Yamashita (2015; 2017) investigated the impacts of L1-L2 congruency on L2 collocational processing, focusing on the reaction times of university-level EFL students and native English speakers. Using a frequency-based approach, these studies included English-only, Japanese-only (i.e., unacceptable English word combinations translated from Japanese collocations), and baseline collocations (i.e., semantically implausible word combinations). They found no significant differences in learners' reaction times for Japanese-only and baseline items.

Some studies even yielded results that contradicted the consensus. Men (2018), El-Dakhs, Salem, & Al-Haqbani (2020), and Cao & Badger (2021) reported that participants made more mistakes using congruent collocations, which are typically considered less troublesome for L2 learners compared to incongruent ones. Instead of measuring learners' receptive knowledge or form retention of collocations via different tests, Men (2018) used the Chinese Learner English Corpus to examine how Chinese learners of English at different proficiency levels used congruent and

incongruent collocations in their writing. Similarly, Cao and Badger (2021) identified and analysed collocation errors in a database of argumentative essays written by Vietnamese-speaking learners, concluding that unconventional collocations more frequently resulted from congruent collocations. However, it is possible that students tend to avoid using collocations they are uncertain about to minimise mistakes and achieve higher scores, or simply do not need to use certain collocations that are not suitable for the writing topics, using students' writing as the sole resource to evaluate their productive knowledge of collocations may be inadequate.

As mentioned before, the reason for the observed inconsistency of results may arise from methodological limitations, particularly with the operationalisation of congruency. Nesselhauf (2005) notes that the concept of congruency is particularly challenging to grasp adequately, as it is not always clear whether two words are direct translation equivalents in different languages. The challenge inherent in operationalising congruency is that a direct translation of an English phrase can be available in one language but lacking in another. Also, since the translation process involves a certain degree of subjective judgement about the semantic acceptability of L2 output, there might be a discrepancy between the maker and the taker of a test regarding whether a phrase should be considered 'congruent'. This ambiguity leads to study designs that lack rigour in item development. Studies employing purely statistical-based approaches all suffer from the inevitable limitation of neglecting the aforementioned language-specific congruency issues.

For example, in Ding and Reynolds's (2019) study exploring L1 Chinese influence on the processing of L2 English collocations, phrases like 'stand a chance' and 'do justice' were considered collocations. While these phrases lack direct translations and could be seen as incongruent, one involves a singular countable noun and the other includes an uncountable noun, an inconsistency that persists throughout the item list. Additionally, these phrases do not appear in the corpus as standalone expressions; they always occur in specific contexts and combinations. For instance, 'stand a chance' is primarily used in a negative context (e.g., 'The team doesn't stand a chance against their rival') or with an object ('do someone or something justice'). Extracting an item heavily dependent on context increases the difficulty of correct responses, as participants might recognise the phrase but consider it incorrect because it is not used in its most conventional way.

Therefore, it is vital for collocation researchers to take into account the subjectivity of the phraseological approach. Rigorous measures should be taken in the process of item development to minimise the ambiguity of congruency that might invalidate the test results. To the best of my knowledge, most collocation studies involving some form of lexical judgement, even some of the

most cited works, have not taken this factor into consideration, with a few exceptions from recent literature.

Lee's study (2021) provides a good example of methodological rigour. Lee investigated the effect of L1 transfer in the recognition of L2 collocations and unacceptable word combinations across low-intermediate to advanced learners of English, involving learners from two different L1 backgrounds (Korean and Mandarin) and native speakers of English to explore the effect of L1 transfer separate from the effect of intralingual factors. Worth noting in the study's methodology is the rigorous identification of the target verb-noun combinations, using a combination of five measures: phraseological definition of collocations, corpus frequency data, association strength, native speaker judgments, and dictionary checks to select the appropriate items. The present study will follow the item development process outlined in Lee's (2021) design to ensure a rigorous operationalisation of congruency. This process will be discussed in detail in the next chapter.

2.5. English Private Tutoring

Another noteworthy limitation of the existing literature lies in the demographics of participants. While a significant body of SLA research has been conducted in the context of traditional educational settings, involving either learners within the context of formal schooling settings or adult learners in universities, another realm of language learning activity, English private tutoring (PT), has emerged as a dominant player in education but received comparatively little academic attention. PT has become a popular extra-curricular language learning activity worldwide (Hajar & Karakus, 2022; Yung, 2022). PT commonly refers to the 'paid service students used to supplement their learning of academic subjects at school outside school hours' (Yung, 2019: 120). It is widely known as 'shadow education' particularly in the comparative education literature because it operates alongside regular schooling and to some extent mimics its curriculum and instructional practices (Bray, 2021; Yung & Bray, 2021).

Similar to other parts of East Asia, PT has significantly expanded in Mainland China (Lei, 2005; Xue & Ding, 2009). Data from China's 2004 Urban Household Education and Employment Survey shows that 73.8% of primary, 65.6% of junior secondary, and 53.5% of senior secondary students received PT (Xue & Ding, 2009). Known as 'shadow education', PT adapts to the mainstream schooling system (Bray, 2009). In ESL/EFL contexts, English private tutoring (EPT) is especially popular among primary and secondary students due to the significance of English for further studies and employment (Hamid, Sussex, & Khan, 2009; Saito et al., 2018; Yung, 2015). In countries like Bangladesh (Mahmud & Bray, 2017), Japan (Dierkes, 2010), South Korea (Park, Byun, & Kim,

2011), Mainland Chinaand Hong Kong (Zhan et al., 2013), over half of secondary students receive EPT, highlighting its global importance in out-of-school language learning.

PT is characterised by three main dimensions (Yung & Bray, 2021). The first is privateness. This dimension limits tutoring to that provided by individuals or organisations in exchange for a fee and excludes unpaid tutoring offered by families, friends or volunteers, or extra lessons provided by teachers free of charge. The second dimension is supplementation. Shadow education supplements the provision of mainstream schools and is provided outside school hours. Thirdly, the shadow education focuses on academic subjects, which in many education systems mean national languages, English, mathematics and other subjects that feature in public exams. Domains that are learnt mainly for leisure and/or personal development such as music, art and sports are excluded from the focus.

For the third dimension, although English is typically treated as a major academic subject in the school curriculum in many education systems in both English as the first language or EAL contexts, some EPT may not solely focus on teaching English that shadows the mainstream curriculum. An example is tutorial courses for the International English Language Testing System (IELTS), Test of English as a Foreign Language (TOEFL) and Scholastic Aptitude Test (SAT) (Allen, 2016; Buchmann, Condron & Roscigno, 2010). Many tutorial companies offer courses for these international standardised tests in addition to those that prepare tutees for the national English exams in the school curriculum. These test-prep courses are usually popular among senior secondary students because the test scores may help them apply for a university place. On the other hand, some EPT courses do not feature the exam component but aim to raise learners' English proficiency for authentic communication.

The context of L2 language learning in the domain of private tutoring represents a distinct variety of education, one not driven by curiosity, exploration and creativity, but rather dictated by single standards, clear objectives and measurable outcomes. Given the focus on academic subjects, students enrolled in EPT programmes are primarily exposed to a formal English lexicon and low-frequency vocabulary (Yung & Bray, 2021). Moreover, due to the exam-oriented nature of EPT courses, instructors are incentivised to provide form-focused, rather than meaning-focused guidance on English grammar, and to emphasise intentional vocabulary building through word lists and drilling (Zhang & Bray, 2020).

The phenomenon of shadow education has elicited considerable criticism from various stakeholders within the field of education. A prominent concern is the belief that private tutoring centres negatively impact the overall educational landscape by exacerbating inequalities in the

distribution of educational resources, allowing affluent families to gain unfair advantage through the provision of supplementary instruction outside the mainstream school system. Nonetheless, it is undeniable that students derive substantial benefits from EPT, not merely in terms of improved test performance, but also in the form of tangible gains in English language proficiency, whether intentionally or incidentally acquired (Yung, 2015). The lucrative compensation offered by EPT centres serves to attract language instructors with higher-than-average qualifications, thereby enhancing the overall quality of the curriculum. Moreover, the private and individualised nature of shadow education enables students to receive personalised attention tailored to their specific needs and challenges during the language learning process, thus ensuring more substantial gains. Unlike traditional schooling, which is often delivered in large classes of 50 or more students, private tutoring is generally provided in small groups or on a one-on-one basis (Zhang & Bray, 2020). This format allows for greater flexibility in instruction, as it can be adapted to the individual student's pace, needs, and objectives, and this flexibility can be particularly beneficial for learners requiring a more personalised approach to their language studies.

Empirical research has found positive findings regarding L2 learners' perceived effectiveness and enjoyment of extracurricular private tutoring. A study conducted by Yung and Chiu (2020) explored the factors influencing secondary school students' enjoyment of EPT in Hong Kong through the lens of the L2 Motivational Self System (L2MSS) framework. The results indicate that the majority of students (80%) reported enjoying their EPT experiences. Students were more likely to enjoy EPT if they perceived greater financial resources within their families, attended schools with Chinese as the medium of instruction, had internalised instrumental goals, exhibited positive dispositions towards the English language, were not influenced by advertisements to enrol in EPT, participated in face-to-face tutoring rather than video-based instruction, had a designated tutor, or expressed a stronger affinity for their EPT tutor compared to their school-based English teacher. Therefore, more academic attention should be paid to this overlooked extracurricular English language teaching (ELT) setting, as the pedagogical considerations in this subfield can inform the field of SLA.

In particular, private tutoring may significantly enhance collocational learning in second language acquisition due to its personalised approach, flexibility in curriculum, and focused interaction. Personalised learning in private tutoring allows instruction to be tailored to the specific needs and pace of the learner, thereby focusing on the individual challenges a student might have with particular collocations. Studies such as those by Bloom (1984) have demonstrated that individualised instruction can significantly improve learning outcomes. Additionally, the flexibility

in curriculum that private tutoring offers enables tutors to adapt lessons dynamically, ensuring that collocational teaching remains relevant and engaging as students' needs evolve, a concept supported by Baker et al. (2002) in their cross-national analysis of educational effects. This adaptability is crucial for collocational learning, which often requires context-specific teaching. Furthermore, the focused interaction in one-on-one or small group settings allows for immediate feedback and correction, which is essential for mastering collocations. Research by Long (1996) highlights the importance of interaction in language learning, emphasising how timely feedback can solidify understanding and usage of collocations. Together, these factors make private tutoring a powerful tool for effective collocational learning, providing an environment where students can practice, receive personalised guidance, and adaptively learn collocations in meaningful contexts.

2.6. The Present Study

This study aims to investigate the effect of proficiency and congruency on L2 collocation processing. Findings from existing literature remain inconclusive. Research on collocation proficiency shows mixed results, largely due to inconsistent methods for measuring proficiency and selecting collocation items. While some studies indicate a positive correlation between L2 proficiency and collocational knowledge, others find no significant differences across proficiency levels. This discrepancy can be attributed to the varied measures of proficiency employed, ranging from extensive test batteries to self-ratings, often without reliability reports. Furthermore, different standards for selecting collocation items contribute to inconsistent findings, as items considered collocations in frequency-based approaches might be viewed as idioms in phraseological traditions, complicating the comparison of results across studies.

In terms of congruency, the challenge lies in the subjective nature of defining and operationalising the concept. While many studies have found that learners struggle more with incongruent collocations, some recent research shows no processing advantage for congruent collocations or even more errors with them. This inconsistency underscores the methodological limitations in defining and selecting congruent and incongruent collocations rigorously, which can lead to ambiguous and non-generalisable findings about the impact of congruency on collocation learning. Additionally, much research focused on only few English-language learning contexts (e.g. Japanese, Swedish), limiting the generalisability to other L2 settings.

This study aimed to enhance the understanding of L2 collocational competence among EFL learners in the Chinese EPT context by examining their performance on a phrase judgment task. The test items were developed using a combination of phraseology-based and frequency-based

approaches. Unlike traditional studies that focus on adult learners, this research uniquely examined high school learners studying in EPT context, addressing a gap in the SLA literature regarding data on this demographic.

3. Methods

This chapter details and provides justification for the methods employed in this study. First, section 3.1 outlines the research design. Section 3.2 then describes the sample population and their requirements. The materials utilised in the study are discussed in section 3.3. Section 3.4 outlines the administration of the study. Finally, the chapter concludes with an ethics statement in section 3.5.

3.1. Research Design

The research utilised a quantitative, cross-sectional design to investigate two factors that influence how L1 Mandarin high school students in China who learn English as a Foreign Language (EFL) process English collocations, i.e. L1 congruency of the collocations and L2 proficiency of the learners, measured by a proficiency test LexTALE. A phrase-semantic judgment test was administered on participants to extract Accuracy Rates (AR) and Reaction Times (RT) as dependent variables. A multivariate analysis of variance (MANOVA) was conducted on the AR and RT data, with four levels of congruency (acceptable in both Mandarin and English, English only, Mandarin only, baseline) as within-subject variables, and LexTALE score (measured on a continuous scale) as a between-subject variable. Based on the literature review presented in Chapter 2, the study proposed the following research questions (RQs) along with their corresponding hypotheses:

RQ1 Does L2 proficiency affect collocation processing?

RQ1.1 Is LexTALE valid in testing proficiency?

RQ1.2 Is there a statistically significant difference in collocation processing accuracy between B1 and B2 learners?

RQ1.3 Is there a statistically significant difference in collocation processing speed between B1 and B2 learners?

RQ1.4 Is there a statistically significant correlation between proficiency (as measured by LexTALE) and collocation processing?

RQ2 Does L1 congruency affect collocation processing?

RQ2.1 Is there a difference in AR between responses to congruent and incongruent items?

RQ2.2 Is there a difference in RT between responses to congruent and incongruent items?

RQ3 Is there an interaction effect between L2 proficiency and L1 congruency?

3.2. Participants

The study was conducted with 60 high school students in China, aged 16-18 years old. Their first language was Mandarin Chinese, and English was considered a foreign language. They were recruited through a private tuition centre specialising in enhancing academic English skills and preparing students for standardised English proficiency tests such as Test of English as a Foreign Language (TOEFL) and International English Language Testing System (IELTS).

The study focused on intermediate-level learners (B1-B2) because at this stage, learners have already acquired sufficient vocabulary to enable basic-level communication. However, they may encounter challenges in developing the collocational knowledge necessary to enhance their proficiency to an advanced level.

As each student is enrolled in a course at the centre, they are required to undergo a mock test to verify their reported proficiency level upon programme registration or submit a certificate displaying results from their latest test. Consequently, the scores from the mock test or the most recent test can be conveniently utilised in the current study as a pre-selection criterion for participants' general proficiency. According to the mapping of TOEFL and IELTS scoring scales onto the Common European Framework of Reference for Languages (CEFR), an IELTS score ranging from 5.5 to 6.5 and a TOEFL score of 72–94 are considered equivalent to CEFR upper-intermediate level B2. On the other hand, an IELTS score falling between 4.0 and 5.0 and a TOEFL score of 42–71 are aligned with CEFR lower-intermediate level B1 (Papageorgiou, Tannenbaum, Bridgeman, & Cho, 2015). Hence, we enlisted participants with TOEFL scores ranging from 42 to 94 and IELTS scores falling between 4.0 and 6.5.

In contrast to previous studies (e.g. Wolter & Yamashita, 2018; Ding & Reynolds, 2019), this research did not include L1-English Native Speakers as a comparison group. The decision to exclude an L1-English comparison group was based on the consideration that the focus of the current study was on exploring nuances in the impact of L1 congruency and L2 proficiency specifically among L1-Chinese speakers, rendering a native-speaking comparison group unnecessary. In other words, the primary objective of the study is to compare learners' performances within in the L1 Chinese group rather than across distinct language demographics.

The directors of two tuition centres within a private tutoring company in China received an email outlining the research background and objectives. This email included an attached information sheet. Each director subsequently forwarded the email to the teachers at their respective centres to inform them about the research. The teachers then received another email (Appendix 2)

providing further details about the research, along with the information sheet, a link to the test, and participant IDs.

These teachers, who tutor students individually, conveyed the research details to their students and provided them with the PDF information sheet. 39 interested students signed up and received a participant ID from their supervising teacher. They accessed the test at their convenience via an internet-connected computer, and completed the experiment under the teacher's supervision.

3.3. Materials

3.3.1. LexTALE as a Proficiency Test

Operationalising L2 proficiency of participants in an experimental setting has long been a challenging issue (Grosjean, 1998). The elicitation of self-reports through questionnaires is a common practice not only in the realm of collocation research but also across studies in the broader field of second language acquisition (SLA). However, a wide variety of measurements are used as indications of L2 proficiency in these questionnaires, ranging from self-report proficiency scores of English abilities (Ding & Reynolds, 2019; Gyllstad & Wolter, 2016; Wolter & Yamashita, 2018), the length of immersion (Jiang et al., 2019), to a 100-point scale of average daily use of L2 (Chen, 2023). Self-ratings were not used for this study for three reasons. First, without a prior introduction of the questionnaire, the participants did not have a unifying understanding of the scoring standards, so the same score could receive distinct perceptions. Also, high-school participants are more likely than adult learners to overestimate their current proficiency level due to limited language experience. Finally, since the setting is anonymous they might not take the questions seriously and randomly fill in numbers, affecting the quality of the data.

While each choice has its justifications, it is important to note that relying on a single measurement, particularly subjective self-ratings, of English proficiency puts a limit on the data's validity, given the vastly different individual differences. The current study aims to address the lack of data triangulation observed in previous studies by introducing a proxy test to accompany the self-ratings, thus giving more accurate insight into the participants proficiency levels.

While employing self-reports to control for language proficiency is a common practice in SLA research, the accuracy of such assessments remains uncertain (Tomoschuk, Ferreira & Gollan 2018). As a proxy for participants' L2 English proficiency in the study, Lexical Test for Advanced Learners of English (LexTALE). was administered before the main collocation task. LexTALE is a validated and standardised vocabulary test tailored for individuals with medium to high proficiency in English as a foreign language (Lemhöfer & Broersma, 2012). The test consists of an unspeeded

lexical decision task in which participants have to make word/nonword decisions to 60 items (40 words and 20 nonwords). The list includes words ranging from moderately well-known to very well-known to native speakers. In this way, various proficiency levels can be discerned in L2 learners. Nonwords are included to correct the test for false positives (i.e., the tendency that some participants have to indicate that they "know" words they have never encountered before). The number of nonwords is smaller than the number of words to make the subjective proportions of words and nonwords more equal, given that most participants do not know all the words. The final score is computed by taking into account both the number of correct words identified and the "yes" responses to nonwords.

Within the last 10 years, the LexTALE has been employed in a substantial 551 studies (Puig-Mayenco, Chaouch-Orozco, Liu & Martín-Villena, 2023). Furthermore, it is regarded as a reliable gauge of general English proficiency, evidenced by a significant correlation between LexTALE scores and both the Quick Placement Test (QPT) and Test of English for International Communication (TOEIC) scores (Lemhöfer & Broersma, 2012). Despite the extensive use of LexTALE in linguistics research as a proxy for EFL general proficiency, it is essential to acknowledge its limitations. Puig-Mayenco et al. (2023) enlisted 266 L1 Chinese young adults to partially replicate Lemhöfer and Broersma's (2012) work. Their conclusion raised doubts about LexTALE's reliability as a global proficiency indicator due to the observed low and moderate correlations with a standardised measure of global proficiency. However, many studies have also found support for the validity of LexTALE. For example, a recent study involving Japanese university students (Tatsuya, Yu and Scott, 2020) established significant correlations between LexTALE and TOEFL scores, the latter of which happens to coincide with the screening criterion used in the current study. Additionally, due to geographical constraints preventing in-person experiments, the ease and efficiency of online LexTALE make it an optimal instrument for this research.

As per Lemhöfer and Broersma (2012), LexTALE scores below 60% align with the B1 proficiency on the CEFR proficiency scale. Scores ranging from 60% to 80% correspond to the B2 proficiency level, while scores above 80% correspond to the advanced (C1 and C2) proficiency levels. These correlations were established based on equivalences between the LexTALE and QPT score bands, compared with the score bands of the QPT and the CEFR proficiency levels. Therefore, data from individuals scoring above 80% were excluded from the further analysis.

3.3.2. Collocation Test

Although the study of learners' production of collocations is indeed interesting, assessing this type of knowledge in a test involves several practical considerations. A productive where prompt words are provided and participants are expected to produce common collocates, would require more time per item for the participants. This would likely limit the number of participants that could be tested. Additionally, the scoring procedure would be significantly more complex, as a system would need to be developed to quantify the participants' responses. In contrast, measuring receptive collocation knowledge allows for a larger number of items to be tested in each session, and an objective scoring key can be easily created for the test.

Following the design in many previous studies (e.g. Ding & Reynolds, 2019; Wolter and Yamashita, 2015; Yamashita and Jiang, 2010), the main task was a speeded phrase-acceptability judgment task wherein participants were required to make binary YES/NO choices within a limited timeframe to determine the acceptability of presented items in English. All items were presented on a computer screen in a randomised order. Test takers were instructed to press one key (the '1' key) if they felt 'the word combinations [were] common in English' and another key (the '0' key) if they felt 'the word combinations [were] NOT common in English.'

3.3.3. Item Development

Following Lee (2021), this study utilised a comprehensive item development process that combined five measures: phraseological definition of collocations, corpus frequency data, association strength, native speaker judgments, and dictionary checks. This approach ensured that both traditional methods—phraseological and frequency-based—were not only followed but also integrated to select the most appropriate items.

To avoid a potential confounding factor associated with noun phrase grammar, specifically concerning articles and the countability of nouns, the study exclusively utilised a single type of verb-noun pairing with indefinite articles, namely V-a-N (e.g. *meet a friend*). Because this study used L1 Mandarin participants, all items were developed with both English and Mandarin phraseology taken into consideration. Based on their congruency, the items were categorised into four types: (1) congruent collocations, i.e. those that can be directly translated from English to Mandarin (EM, e.g., *play a game*), (2) English-only (incongruent) collocations (E-only, e.g., *catch a cold*), (3) Mandarin-only (translated) collocations (M-only, e.g., *eat a drug*), and (4) baseline items (implausible, e.g. *write a cold*).

Congruency was operationalised through a word-for-word translation approach (Yamashita & Jiang, 2010). Congruent collocations allow a straightforward translation of their essential meaning from the corresponding constituents into the L1, exemplified by phrases like *buy a flower*. On the other hand, incongruent collocations necessitate the use of different lexical items in Chinese or require paraphrasing for translation, as seen in examples such as *take a drug* or *cut a deal*. In other words, for incongruent collocations, to convey the English meaning requires using alternative word combinations, mainly by switching the constituent verb (e.g., 吃药 *lit. eat a drug** for *take a drug;* 达成交易 *lit. reach a deal* for *cut a deal*). The core meaning of the individual word in each collocation was obtained from *Oxford Collocations Dictionary* (online version: http:// oxforddictionary.so8848.com).

To achieve inter-rater reliability on the appropriateness of collocation aforementioned classifications, the nouns from the selected collocations were shown in to six native speakers of English, who are adult Masters students in applied linguistics at the University of Oxford. The nouns were preceded with a blank (e.g. '____ a secret') to prompt the elicitation of verbs that naturally pair with the given nouns. Participants were requested to enumerate several verbs commonly associated with each provided noun. If at least five out of the six native speakers of English suggested the same verb for a given noun, that specific verb-noun combination was considered appropriate for the study. The selection threshold was established following Lee's (2021) study, wherein a singular expression eliciting five out of six confirmations from native speakers was deemed a conventional expression. In addition, a compilation of English verb-noun collocations was presented in written form to six L1 Mandarin speakers with advanced English proficiency and a Master's degree in translation. They were queried about whether the presented English collocations can be directly translated into their L1. If at least five out of the six indicated the absence of a direct translation, the collocation was designated as an E-only collocation. Conversely, when at least five Mandarin speakers affirmed the existence of a translational equivalent in their L1, the collocation was classified as a congruent EM collocation.

The participants were asked to indicate the variety of English that they spoke. Since most of the participants received training in preparation for TOEFL test, a substantial number of participants reported speaking American English. This clearly justifies the preference of an American English corpus over a British one. Thus, all items were checked against the Corpus of Contemporary American English (COCA; Davies, 2008) to ensure the appropriateness of their classifications. COCA was chosen also because it is the largest existing corpus of American English that that is

available without cost. The COCA was chosen over competing corpora, namely the British National Corpus (BNC), for two reasons. First, American English is the preferred variety taught in most academic contexts in China. Secondly, Durrant (2014) discovered that frequency values from the COCA had a stronger correlation with learner knowledge compared to frequency values from the BNC. It also needs to be pointed out that our use of frequency values from a corpus is not without limitations. Although corpora have often been used to approximate frequency of input for NSs and NNSs alike, it is always highly unlikely that these values accurately replicate any language user's collective experience with the target language. There will likely be vast difference in input both across groups and among the individuals who comprise these groups. Nonetheless, we felt that, in the absence of better information regarding individual participants' language histories, corpus frequencies provided the best possible estimation.

In theory, E-only and EM collocations are expected to be present in the corpus, whereas M-only and baseline items should not be present. In practice, this was challenging given the large corpus size of COCA. Consequently, a limited number of occurrences in the corpus was permitted, provided that these instances were sufficiently idiosyncratic to be considered highly improbable constructions.

It should be noted that in order to assemble lists that were parallel and fulfilled our L1–L2 or L2-only criteria, sometimes it was necessary to use the same word form two times. In total, there were five words that appeared two times. However, the same form was never used more than once in the same block of items on the test.

To ensure only high-frequency collocations were included, collocation frequencies and MI scores were obtained for the items from the COCA. MI scores were included as an extra level of analysis because these scores consider the frequency of the words that comprise the collocations rather than the frequency for the entire collocation. As a result, MI scores served as more of a measure of strength of association between the two words comprising the collocations. Only collocations with a frequency of over 100 and an MI score of at least 3.0 in COCA were included. Items that failed either criterion were eliminated, since two words could be close collocates but they did not necessarily co-occur frequently as a 'VERB-a-NOUN' construction. For example, the collocation 'bend a rule' was eliminated because while 'bend' and 'rule' have a high MI score of 3.89, the phrase 'bend a rule' only has a frequency of 0.004, because the more common collocation is 'bend rules / bend the rules' as they are conventionally used in authentic contexts.

The baseline items were included not only for comparison purposes, but also to ensure that participants did not develop a familiarity effect for the task. The anomalous baseline items consisted

of random combinations of the words used in the M-only and E-only collocations. This was done to ensure that all resulting items are semantically implausible and that the lexical frequency of the individual words would be the same between the baseline items and the collocational items. All recombined items were checked against COCA in the same manner used for the M-only items. If any occurrence did occur in the corpus, the MI scores were checked to make sure they were inverse.

In total, there were 80 items: 20 in each of EM, E-only, and M-only conditions, and 20 baseline items. Scores are measured by calculating the total amount of correct responses. Responses were categorised as incorrect if the response was different from the expected one (i. e. a 'no' response for a collocation or a 'yes' response for an unacceptable word combination) or if an item was timed-out.

3.4. Administration

The experiment was designed and deployed using Gorilla Experiment Builder. On initiating the experiment, participants were presented a participant information sheet, wherein a brief explanation about the study is provided and it was made clear that participation is voluntary and anonymous. Then participants were asked to fill out the consent form. Once consent is given, a questionnaire concerning age and L2 language background was administered. Upon completion of this questionnaire, LexTALE was administered to assess the L2 participants' proficiency.

After LexTALE was the main task, where participants were asked whether or not the word combinations that appeared on the screen are acceptable in English. Similar to many previous studies (e.g. Yamashita & Jiang, 2010) in my investigation the participants were required to complete a phrase-acceptability judgment task by determining whether the collocations are commonly used in English. Participants used their own laptop computers to access and complete the lexical decision task. First, the instructions were displayed on the screen in Chinese and English for the L1-Chinese speakers. There was no time limit, which allowed the participants to fully understand the instruction. The participants were instructed that each time a word combination was displayed on the screen, they should press the YES key if the combination was common in English and the NO key if the combination was not common in English. Each collocation was displayed individually. To familiarise the participants with the procedure, they were given six practice items before the formal experiment began.

After the practice session, participants began a timed phrase-acceptability judgment task. For the presentation of the 120 items, first there was a blank screen with a fixation target (a plus sign)

lasting for 500 ms for eye fixation before the item appeared on the screen. Once a participant pressed a button for a phrase, the plus sign was presented briefly to signal a new item, and the next phrase was presented (See Appendix 7). The item remained on the screen until a response was made or the item timed out at 4000 ms, following Wolter and Gyllstad (2013). RTs (in milliseconds) and response patterns (i.e., a 'yes', 'no' or 'time-out' response to each item) were logged. All items were presented in a randomised order. Two self-timed breaks were provided at equal intervals (after 40 items).

3.5. Ethical Considerations

The present study was approved by the University of Oxford's Central University Research and Ethics Committee (CUREC) (See Appendix 1). Data collection commenced upon receipt of approval. Information outlining the nature of the study, its objectives, and procedures was provided to all potential participants. It was made clear that participation was voluntary, anonymous, and that participants had the right to withdraw at any time during the task by closing their browser tab/ window.

4. Findings

This chapter presents the findings of the current study, contextualised by the research questions (RQs) outlined in section 3.1. Section 4.1 discusses the reliability of LexTALE and compares the performance of the two proficiency groups on the phrasal acceptability judgement task (PAJT) in terms of the dependent variables: mean reaction time (RT) and accuracy rate (AR). Section 4.2 compares the same dependent variables across the four types of items—congruent (EM), incongruent (EO), Mandarin-only (MO), and baseline (BL). Next, the combined effect of collocation type and proficiency will be explored to investigate potential interaction between the two factors in section 4.3. Finally, the findings are summarised in section 4.4.

Two individuals were excluded from the initial pool of 36 participants, resulting in a sample comprising 34 participants, with. One participant received 49 points in the LexTALE test, resulting in an accuracy rate of 81.67%, which corresponds to advanced level proficiency (C1-C2), thus failing to meet the selection criterion for data analysis. One participant with mean RT shorter than 1000ms were eliminated since it suggested the participant was not attentively engaged with the task.

Table 1. Participants' background information

	Group	
Variable	**B1 (*n* = 17) M (SD)**	**B2 (*n* = 17) M (SD)**
Age	16.88 (0.78)	16.65 (0.61)
Gender (M/F)	8/9	7/10
Handedness (L/R)	2/15	0/17
Age starting English study	5.76 (2.08)	4.59 (1.06)
English self-ratings		
Reading	2.58 (1.00)	3.59 (0.71)
Listening	2.76 (1.03)	3.29 (0.77)
Speaking	2.59 (1.00)	3.24 (0.66)
Writing	2.29 (0.77)	3.12 (0.60)
TOFEL scores	75.73 (15.82)	89.94 (19.94)

4.1. RQ1: L2 Proficiency's Effect on Collocation Processing

RQ1 inquired whether there would be a statistically significant difference in processing between two proficiency groups. Prior to this, RQ1.1 addressed the validity of LexTALE as a proficiency measurement tool. Given that LexTALE was chosen as the basis for participant recruitment, it was crucial to examine its reliability. Specifically, it was necessary to determine if LexTALE scores correlate with self-rated proficiency and standardised test scores. This validation is essential as it

will guide the decision-making process regarding which measure to use when investigating the relationship between proficiency and collocations.

First, assumptions for the correlation test are checked. The skewness and kurtosis values of LexTALE scores are within the range of -1 to 1, while participants' self-ratings showed slight skewness (-1.220, *SE* = 0.403) as well as significant kurtosis (1.028, *SE* = 0.788), suggesting that the self-ratings of proficiency are not normally distributed (Field, 2013).

Table 2. Measures of participants' L2 proficiency

	N	Mean (SD)	Skewness (SD)	Kurtosis (SD)
Self-rated general proficiency	34	2.93 (0.75)	-1.22 (0.40)	1.03 (0.79)
LexTALE score	34	58.97 (8.66)	0.31 (0.40)	-0.13 (0.79)
TOEFL score	30	82.97 (16.88)	-0.29 (0.43)	-0.99 (0.83)

Therefore, a Spearman's rank-order correlation was conducted to assess the relationship between self-rated general proficiency and LexTALE scores (N = 34). There was a moderate positive correlation between self-rated general proficiency and LexTALE scores, which was statistically significant (ρ = .422, p = .013). This indicates that higher self-rated general proficiency is associated with higher LexTALE scores among the participants. The correlation was significant at the 0.05 level (2-tailed).

Table 3. Correlation between LexTALE and self-rated proficiency

		LexTALE score	Self-rated general proficiency
Self-rated general proficiency	Correlation Coefficient	0.422	1.000
	Sig. (2-tailed)	0.13	.
	N	34	34
LexTALE score	Correlation Coefficient	1.000	0.42
	Sig. (2-tailed)	.	0.13
	N	34	34

To triangulate the measurement of proficiency, participants' standardised test scores (TOEFL) were also collected for each student. This was to further ensure the validity of LexTALE. With both variables normally distributed, a Pearson correlation analysis was conducted to examine the relationship between LexTALE scores and TOEFL scores. The sample sizes for LexTALE scores and TOEFL scores were n = 34 and n = 30, respectively. The results indicated a moderate positive correlation, which was statistically significant, (r(28) = .500, p = .005). This suggests that higher

LexTALE scores are associated with higher standardised proficiency test scores. The correlation was significant at the 0.01 level (2-tailed).

Table 4. Correlation between LexTALE and TOEFL scores

		LexTALE score	**TOEFL score**
TOEFL score	Pearson correlation	0.50	1.00
	Sig. (2-tailed)	0.05	.
	N	30	30
LexTALE score	Pearson correlation	1.00	0.50
	Sig. (2-tailed)	.	0.05
	N	34	34

To sum up, these findings supported the validity of LexTALE scores as a measure of English language proficiency. They also indicate that self-rated reports were generally accurate, and that they are also in line with standardised tests that the learners were taking.

Having established the validity of LexTALE, RQ1.2 examined whether B2 learners have a different accuracy of processing collocations than B1 learners. To answer this question, the first variable to investigate was AR. This was measured by calculating the proportion of correct responses out of 80 items. An independent-samples t-test was conducted to compare the AR of the collocation test for B1 and B2 learners. The results showed a significant difference in the AR of collocation judgment for B1 learners (M = 59.85, SD = 8.50) and B2 learners (M = 67.06, SD = 6.11; $t(32)$ = -2.839, p = 0.004, two-tailed), with B1 learners accurately identifying correct collocations at a higher rate than B2 learners. The magnitude of the differences in the means (mean difference = -7.21, Cl [-12.98, -2.66]) was large (η^2 = 0.224), suggesting a substantial impact of proficiency on collocation judgment accuracy rate.

Table 5. Comparison of AR between B1 and B2 learners

	B1 (n = 17)		**B2 (n = 17)**		***MD***	***t*(32)**
	M	***SD***	***M***	***SD***		
Accuracy Rate	59.85	8.50	67.06	6.11	-7.21	-2.839

Next, RQ1.3 aims to assessed the difference in the average speed of collocation processing between two groups. An independent-samples t-test was conducted to compare the RT of all responses between B1 and B2 proficiency level learners based on LexTALE scores. There was no significant difference in the average reaction times for B1 learners (M = 2087.58, SD =587.63) and

B2 learners (M = 1790.13, SD = 384.06); $t(32) = 1.747$, $p = .092$. Levene's test indicated unequal variances ($F = 4.742$, $p = .037$), so the equal variances not assumed t-test was used. The mean difference in reaction times was 297.44 ms, with a 95% confidence interval ranging from -51.57 to 646.46 ms. These results suggest that there is no significant effect of proficiency level on average RT.

Table 6. Comparison of RT between B1 and B2 learners

	B1 (n = 17)		B2 (n = 17)		*MD*	*t*(32)
	M	***SD***	***M***	***SD***		
Mean Reaction Time	2087.58	587.63	1790.13	384.06	297.44	1.747

In conclusion, B2 learners performed better than B1 learners in terms of AR, but there was no significant difference in their RT, indicating that proficiency knowledge might not play a role in this sample in promoting accuracy but not the speed with which decisions were made.

Having examined the differences in AR and RT between two categorical proficiency groups, the next question RQ1.4 explores whether there exists a relationship between proficiency as measured by LexTALE and collocation processing, i.e., whether LexTALE score positively correlate with collocation. LexTALE was chosen as the sole measure for proficiency in this case because it has been established to highly correlate with other scores. Having established the normality of LexTALE scores and slight deviation from normality for AR values, a non-parametric Spearman rank-order correlation was conducted to assess the relationship between LexTALE scores and collocation test accuracy rates among the 34 participants. There was a weak positive correlation between LexTALE scores and collocation test accuracy rates, ($\rho(32) = .304$) but with no statistical significance ($p = .080$). These results do not support the notion that LexTALE scores are associated with higher collocation test accuracy rates.

Table 7. Correlation between LexTALE score and AR

		LexTALE score	**Accuracy Rate**
LexTALE score	Correlation Coefficient	1.000	0.304
	Sig. (2-tailed)	.	0.080
	N	34	34
Accuracy Rate	Correlation Coefficient	0.304	1.00
	Sig. (2-tailed)	0.080	
	N	34	34

The next variable to look at is mean reaction time. Having established the normality of both LexTALE scores and RT values, a Pearson correlation was conducted to assess the relationship between LexTALE scores and mean response time for all responses among 34 participants. Table 8 shows there was a very weak negative correlation between LexTALE scores and mean reaction time, which was again not statistically significant, ($r(32) = -0.083$, $p = 0.640$). These results suggest that there is no significant linear relationship between LexTALE scores and the RT of all responses in this sample.

Table 8. Correlation between LexTALE score and RT

		LexTALE score	**Mean reaction time**
LexTALE score	Pearson correlation	1.00	-0.083
	Sig. (2-tailed)	.	0.640
	N	34	34
Mean reaction time	Pearson correlation	-0.083	1.00
	Sig. (2-tailed)	0.640	.
	N	34	34

In summary, the data did not yield sufficient evidence to support any correlation between LexTALE and either AR or RT of collocation processing.

4.2. RQ2: L1 Congruency's Effect on Collocation Processing

RQ2 concerned whether the congruency of collocation would affect collocation processing. First, the assumption of homogeneity of variances was tested for AR and RT and satisfied using Levene's Test ($p > .05$). To answer RQ2.1, which aimed to understand whether congruent items would receive a different AR (i.e. easier/harder to process) than incongruent items, a one-way ANOVA was conducted to compare the proportion of participants who made an accurate judgment across four different collocation types: congruent (EM) collocation, incongruent (EO) collocation, Mandarin-only (MO) collocation, and baseline (BL) collocation. The results of the ANOVA showed a significant effect of collocation type on the proportion of accurate judgments ($F(3,76) = 16.090$, $p < .001$).

Post hoc comparisons using the Tukey HSD test indicated that the proportion of accurate judgments for EM collocations ($M = .8420$, $SD = .13779$) was significantly higher than for EO collocations ($M = .5550$, $SD = .20902$, $p < .001$), MO collocations ($M = .4565$, $SD = .21967$, $p < .001$), and BL collocations ($M = .6835$, $SD = .16551$, $p = .042$). Additionally, BL collocations had a significantly higher proportion of accurate judgments compared to MO collocations ($p = .001$).

However, there was no significant difference between EO and BL collocations (p = .137), nor between EO and MO collocations (p = .344).

The mean proportion of participants who made an accurate judgment varied significantly across collocation types, as shown in the descriptive statistics and illustrated in the means plot. The findings suggest that congruent collocations are identified more accurately than other collocation types. Worth noting is the fact that Mandarin-only items received a significantly fewer accurate responses, suggesting the impact of L1 influence on EFL learners' judgment of phraseological acceptability.

Table 9. Accuracy rates by proficiency level and item type

Collocation type	**Respondent Proficiency level**	**Mean**	**Std. Deviation**
Congruent (EM)	B1	0.80	0.18
	B2	0.89	0.16
	Total	0.84	0.17
Incongruent (EO)	B1	0.55	0.21
	B2	0.56	0.23
	Total	0.55	0.21
Mandarin-only (MO)	B1	0.42	0.21
	B2	0.49	0.25
	Total	0.46	0.23
Baseline (BL)	B1	0.63	0.19
	B2	0.74	0.16
	Total	0.68	0.18

Figure 1. Accuracy rates by item type

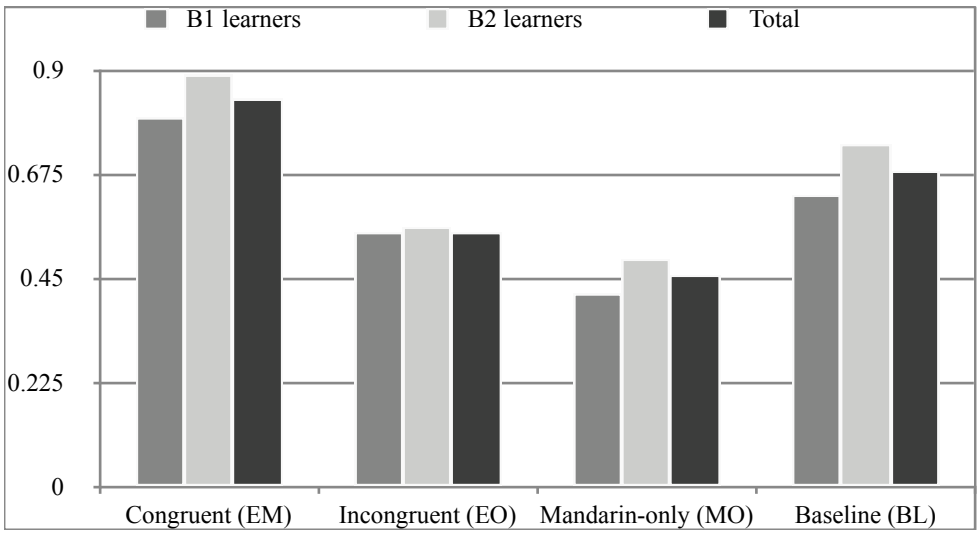


Table 10. Multiple comparisons of accuracy rates

(I) Collocation type	(J) Collocation type	Mean difference (I-J)	Std. Error	Sig.
Congruent (EM)	Incongruent (EO)	.29*	0.06	<.001
	Mandarin-only (MO)	.39*	0.06	<.001
	Baseline (BL)	.16*	0.06	0.042
Incongruent (EO)	Congruent (EM)	-.29*	0.06	<.001
	Mandarin-only (MO)	0.10	0.06	0.344
	Baseline (BL)	-0.13	0.06	0.137
Mandarin-only (MO)	Congruent (EM)	-.39*	0.06	<.001
	Incongruent (EO)	-0.10	0.06	0.344
	Baseline (BL)	-.23*	0.06	0.001
Baseline (BL)	Congruent (EM)	-.16*	0.06	0.042
	Incongruent (EO)	0.13	0.06	0.137
	Mandarin-only (MO)	.23*	0.06	0.001

* The mean difference is significant at the 0.05 level.

Next, to answer RQ2.2, the question of whether congruent items would receive a different RT (i.e. be processed in longer/shorter time) than incongruent items, the one-way ANOVA results showed a significant effect of collocation type on reaction time, $F(3, 76) = 9.510$, $p < .001$. Post hoc comparisons using the Tukey HSD test indicated that the RT for EM collocations was significantly lower than those for EO collocations (mean difference = -398.29, $p < .001$), MO collocations (mean difference = -437.09, $p < .001$), and BL collocations (mean difference = -389.14, $p < .001$), while there were no significant differences in reaction times between the latter three types of collocations. These results suggested that participants responded significantly faster to congruent collocations compared to other types of collocations.

Table 11. Mean reaction time by by proficiency level and item type

Collocation type	Respondent Proficiency level	Mean	Std. Deviation
Congruent (EM)	B1	1837.69	451.89
	B2	1423.83	344.59
	Total	1630.76	448.61
Incongruent (EO)	B1	2143.68	358.79
	B2	1914.50	377.11
	Total	2029.09	381.40
Mandarin-only (MO)	B1	2159.91	266.46
	B2	1977.95	319.50
	Total	2068.93	304.65
Baseline (BL)	B1	2196.47	284.87
	B2	1844.26	364.61
	Total	2020.37	368.93

Figure 2. Mean reaction time (in milliseconds)

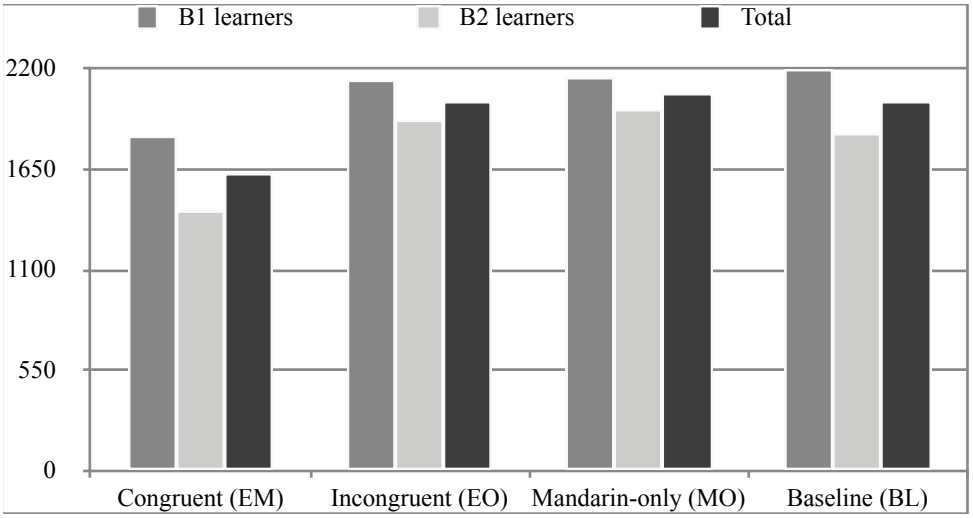


Table 12. Multiple comparisons of mean reaction time

(I) Collocation type	(J) Collocation type	Mean difference (I-J)	Std. Error	Sig.
Congruent (EM)	Incongruent (EO)	-398.29*	94.07	<.001
	Mandarin-only (MO)	-437.09*	94.07	<.001
	Baseline (BL)	-389.14*	94.07	<.001
Incongruent (EO)	Congruent (EM)	398.29*	94.07	<.001
	Mandarin-only (MO)	-38.81	94.07	0.976
	Baseline (BL)	9.14	94.07	1.000
Mandarin-only (MO)	Congruent (EM)	437.09*	94.07	<.001
	Incongruent (EO)	38.81	94.07	0.976
	Baseline (BL)	47.95	94.07	0.957
Baseline (BL)	Congruent (EM)	389.14*	94.07	<.001
	Incongruent (EO)	-9.14	94.07	1.000
	Mandarin-only (MO)	-47.95	94.07	0.957

* The mean difference is significant at the 0.05 level.

4.3. RQ3: Interaction between L2 proficiency and L1 congruency

RQ3 investigated whether there would an interaction effect between proficiency and congruency. To answer RQ3, A two-way MANOVA was run with two independent variables – proficiency level (B1, B2) and collocation type (EM, EO, MO, BL) – and two dependent variables – mean accuracy rate (AR) and mean reaction time (RT). Again, AR was calculated as the proportion of correct responses on a particular item.

First, assumptions for MANOVA were checked. There was homogeneity of covariance matrices, as assessed by Box's M test (p = .103), and homogeneity of variances, as assessed by Levene's Test of Homogeneity of Variance (p > .05). There was a correlation between the dependent variables, as assessed by scatterplot, and no evidence of multicollinearity, as assessed by Pearson correlation (|r|

< 0.9). There were no univariate outliers in the data, as assessed by inspection of a boxplot, and no multivariate outliers in the data, as assessed by Mahalanobis distance ($p > .001$) (see Appendix 9 for the full statistical output). Among the variety of measures that can be used to test for multivariate outliers, Mahalanobis distance is often used when assessing multivariate outliers in MANOVA (e.g., Tabachnick & Fidell, 2014).

While there were two types of congruency (EM and EO), two other types were included for comparison and baseline (MO and BL). Therefore, there were a total of eight groups of variable combinations (four types of collocations combined with two levels of proficiency). To test for normality, the Shapiro-Wilk test is recommended when sample sizes are smaller than 50 participants (Mishra, Pandey, Singh, Gupta, Sahu, & Keshri, 2019). The test produced 16 rows of results, and 13 of 16 Shapiro-Wilk results were not statistically significant. In other words, in these cases, there is no evidence to suggest that the data deviates from a normal distribution. However, for 3 of the 16 group combinations, the Shapiro-Wilk test indicated a significant deviation from normality, suggesting that for these specific groups, the data may not be normally distributed.

Table 13. Tests of normality for eight variable combinations

		Kolmogorov-Smirnov			**Shapiro-Wilk**		
Combination	**Dependent variable**	**Statistic**	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
EM-B1	Accuracy Rate	0.203	20.00	0.030	0.880	20	0.018
	Reaction Time	0.134	20.00	.200*	0.887	20	0.024
EM-B2	Accuracy Rate	0.254	20.00	0.002	0.721	20	<.001
	Reaction Time	0.127	20.00	.200*	0.943	20	0.279
EO-B1	Accuracy Rate	0.127	20.00	.200*	0.964	20	0.635
	Reaction Time	0.143	20.00	.200*	0.957	20	0.486
EO-B2	Accuracy Rate	0.148	20.00	.200*	0.959	20	0.520
	Reaction Time	0.150	20.00	.200*	0.941	20	0.245
MO-B1	Accuracy Rate	0.17	20.00	0.118	0.958	20	0.501
	Reaction Time	0.130	20.00	.200*	0.975	20	0.848
MO-B2	Accuracy Rate	0.130	20.00	.200*	0.964	20	0.625
	Reaction Time	0.13	20.00	.200*	0.937	20	0.208
BL-B1	Accuracy Rate	0.17	20.00	0.148	0.939	20	0.226
	Reaction Time	0.20	20.00	0.038	0.945	20	0.303
BL-B2	Accuracy Rate	0.17	20.00	0.148	0.939	20	0.226
	Reaction Time	0.20	20.00	0.038	0.945	20	0.303

Due to many number of tests being run, Pituch & Stevens (2016) suggested applying a Bonferroni correction to the level at which statistical significance is accepted. That is, the new level at which statistical significance is accepted is .05 divided by the number of tests run. In this study,

there were eight Shapiro-Wilk tests being run, so the new level would be .05 ÷ 8 = .00625 (i.e., accept statistical significance if $p < .00625$). With this in mind, two results were noteworthy. First, for EM items processed by B1 learners, AR and RT are considered normally distributed even with a p value smaller than .05. Second, while RT in EM-B2 combination was not significant, AR showed deviation from normality even considering the Bonferroni correction ($p = 0.002$). However, Weinfurt (1995) notes that in practice MANOVAs tend to be performed even if the data is not normal due to a general consensus that MANOVA is robust to non-normality. Therefore, this result should not affect the MANOVA.

Having checked all the assumptions for MANOVA, the test statistic used for interpretation of results is Wilks' Lambda (Λ) since it is the most widely used multivariate test statistic (Bray & Maxwell, 1985). According to the results, the interaction effect between collocation type and proficiency level was not statistically significant, Wilks' Lambda = .974, $F(6, 302) = .675$, $p = .670$, partial $\eta^2 = .013$. This means that there is no significant combined effect of collocation type and proficiency level on the dependent variables. In other words, the way that collocation type and proficiency level work together does not have a notable impact on the outcomes being measured. The interaction effect between these two factors is very small and not statistically meaningful. The partial eta squared value (partial $\eta^2 = .013$) also indicates that only 1.3% of the variance in the dependent variables can be explained by the interaction of collocation type and proficiency level, which is a very small effect size.

Table 14. Multivariate Tests

		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Type * Level	Pillai's Trace	0.026	0.68	6.000	304.000	0.667	0.013
	Wilks' Lambda	0.974	.675b	6.000	302.000	0.670	0.013
	Hotelling's Trace	0.027	0.67	6.000	300.000	0.673	0.013
	Roy's Largest Root	0.02	.938c	3.000	152.000	0.424	0.018

4.5. Summary

This chapter presents findings addressing the three RQs on how L2 proficiency and L1 congruency might affect collocation processing. First, the study validated LexTALE as a reliable proficiency measure through significant correlations with self-rated proficiency. B2 learners significantly outperformed B1 learners in terms of accuracy rate. However, no significant difference was found in mean reaction time between B1 and B2 learners.

Second, the effect of L1 congruency on collocation processing was examined. A one-way ANOVA showed that congruent collocations were processed significantly more accurately and quickly than incongruent, Mandarin-only, and baseline collocations. No difference was found between the latter three groups.

Finally, the interaction between L2 proficiency and L1 congruency was explored using MANOVA. The results indicated no significant interaction effect, meaning that the combined effect of congruency type and proficiency level does not significantly impact AR and RT.

5. Discussion

This chapter addresses the findings for each research questions in light of existing research. The implications of the findings will also be discussed within the broader context of current research on collocations. Subsequently, pedagogical recommendations will be suggested to Chinese EFL learners on how to enhance collocation learning during vocabulary building. Finally, the limitations of the study, and areas for future research will be considered.

5.1. LexTALE Validity in Testing Proficiency

An important contribution of this study is the validation of the LexTALE test, a widely used tool for quickly assessing participants' L2 English proficiency. Results from the study demonstrated that, in the context of Chinese L2 English learners the LexTALE test is a robust tool for gauging language proficiency. The significant correlations between LexTALE scores and both participants' self-ratings and standardised test scores (TOEFL) underscore its reliability, a finding which aligns with Lemhöfer and Broersma (2012). The positive correlation with self-ratings suggests that learners' perceptions of their language ability correspond well with their LexTALE performance. Furthermore, the correlation with TOEFL scores, a widely recognised standardised test, is corroborated by a study on Japanese students by Tatsuya, Yu, and Scott (2020).

However, it must be acknowledged that the limited sample size of the study can reduce the generalisability of the findings. Therefore, caution is warranted when interpreting the results. Additionally, some recent studies employing LexTALE have raised doubts about its reliability as an L2 proficiency measure for intermediate level learners (e.g., Puig-Mayenco et al., 2023). These studies suggest that LexTALE may not consistently reflect the proficiency levels of learners who fall within the intermediate range, potentially due to a narrower scope of vocabulary assessment or varying degrees of lexical knowledge among learners.

Nonetheless, these concerns can be partially mitigated by efforts to triangulate the measurement of proficiency. In this study, the inclusion of self-reports serves as an additional data point. The observed correlation between these self-reports and LexTALE scores provides supplementary evidence supporting the validity of the tool. By demonstrating the alignment between self-assessments and their LexTALE performance, this reinforces the argument that LexTALE can be a reliable measure of L2 proficiency.

5.2. Collocation Processing Differences Between B1 and B2 Learners

Contrary to previous studies that have compared a single group of L2 language learners (i.e., ESL learners) to L1-English native speakers, the current study adopted a nuanced approach by comparing between two groups of learners that have often been amalgamated in previous research, B1 and B2. The analysis reveals a difference in collocation processing between the two groups of learners: B2 learners significantly outperformed B1 learners in terms of accuracy. This suggests that higher proficiency learners have a better grasp of collocational knowledge, which enhances their processing accuracy. This, to some degree, aligns with the notion that as learners' language proficiency improves, so does their ability to accurately process collocations (Laufer & Waldman, 2011). Even though this study focuses solely on a sliver of proficiency, i.e. intermediate-level learners, this particular finding resonates with the well-documented notion that advanced learners are more adept at recognising collocations due to their extensive exposure to and practice with the language (Nesselhauf, 2003). The existing body of empirical research on collocation has primarily been concerned with how advanced EFL learners deal with multi-word units or making comparisons between native speakers and non-native speakers (Ding, Reynolds, Csaba and Boone, 2023), with an implication that EFL learners with high proficiency are more equipped to acquire collocation knowledge. This study, however, suggests that it is reasonable to begin learning collocations at the intermediate stage, rather than assuming that such knowledge is too challenging for B1/B2 learners and should be deferred until the advanced phase, and it is indeed possible for these learners to be able to identify collocations after having learned them and process them as chunks.

However, it is important to note that the higher accuracy rate (AR) among B2 learners could simply indicate better task performance, reflecting an advantage in test strategy or judgment, rather than signifying a deeper collocational knowledge in their mental lexicon. Research by Cohen (2014) shows that higher proficiency learners often employ more effective test-taking strategies. Additionally, B2 learners may be more familiar than B1 counterparts with the types of tasks and formats used in language tests, giving them an advantage in test performance. Bachman and Palmer (2010) highlighted that test familiarity and experience with test formats can significantly influence test performance, suggesting that B2 learners' higher AR could stem from their greater familiarity with the testing process. Consequently, the observed performance might reflect strategic competence rather than a substantial difference in collocational knowledge between B1 and B2 learners..

Whereas AR differed significantly between B1 and B2 groups, no significant difference was found in terms of RT. In other words, even though B2 learners made more accurate judgments about the acceptability of collocations, they needed about the same time as B1 learners to make the decision. This could be attributed to several possible explanations. First, there could be a speed-accuracy trade-off, where B2 learners prioritise making accurate judgments over making quick decisions. This trade-off might result in them taking similar amounts of time as B1 learners to ensure their responses are correct (Skehan, 2009). However, this may not apply to the acceptability judgment task employed in this study. The speed-accuracy trade-off is generally more common and pronounced in complex tasks than in simple tasks, since complex tasks typically require more cognitive resources and involve higher levels of processing.

Another explanation is although B2 learners are more accurate, their level of automaticity in processing collocations might not be fully developed yet. Both groups may still rely on conscious, deliberate processing rather than automatic processing, resulting in similar reaction times. Automaticity, the ability to recognise and use language components quickly and effortlessly, develops gradually through stages (Segalowitz & Hulstijn, 2005). It is possible that B2 learners, likely at an intermediate stage of automaticity, process collocations more accurately but not yet fully automatically, resulting in similar RTs as B1 learners. This intermediate stage involves partial automaticity, where some language components are processed quickly while others still require conscious effort. Factors such as frequency of exposure, practice, and complexity of collocations influence the development of automaticity. The pedagogical implications of this will be further explored in the next section of the chapter.

5.3. Lack of Correlation between Proficiency and Collocation Knowledge

While higher proficiency might be expected to associate with better collocational knowledge, this relationship was not found in the participants' performance in the judgement task, either in accuracy rate or processing speed, a result corroborated by previous studies (e.g. Bartfield, 2003; Howarth, 1996). Bartfield's (2003) study aimed at testing decontextualised verb + noun collocations for recognition. Barfield observed a relation between the recognition levels of the verbs and the nouns individually, but no correlation was established between general English proficiency and collocation recognition. The same result was found in an earlier study by Howarth (1996), who analysed the English essays of university students (with different L1s) and examined their production of verb + noun collocations.

An interesting interpretation of this finding is that it highlights an important yet often neglected aspect of vocabulary knowledge in L2 education: depth (Nation, 2022). In this case, LexTALE provides a broad measure of language proficiency, primarily focusing on the amount, or breadth, of L2 vocabulary. While vocabulary is an important component of language proficiency, it is not the only factor that affects collocation accuracy. Collocation accuracy involves not just vocabulary knowledge but also syntactic and semantic knowledge, as well as familiarity with fixed expressions and phrases (Bahns & Eldaw, 1993; Howarth, 1998; Schmitt, 2014). LexTALE did not adequately capture these specific skills, which might have led to weaker correlations.

It could be argued that gaining vocabulary knowledge involves different cognitive processes than acquiring collocation knowledge. Knowing individual words is a foundational skill, but understanding and using word combinations requires additional syntactic and semantic processing. Gyllstad (2007) argues that collocations involve deeper levels of processing, including the understanding of word relationships and contexts, which are not necessarily covered by general vocabulary tests. Existing SLA literature offers no conclusive answer to the question whether collocation knowledge is closely related to general proficiency. In some studies, a clear relationship has been observed (Gitsaki, 1999; Bonk, 2001), whereas in other studies, no relationship was established (Howarth, 1996; Barfield, 2003). It seems that results from this study adds to the latter.

This could suggest that significant improvements in collocation accuracy may only become apparent once learners cross a certain proficiency threshold. The categorical groups (B1 and B2) capture these threshold effects, showing clear differences between these levels. However, when proficiency is measured continuously, the small incremental gains within each proficiency band might not show a strong correlation with accuracy rates. Barfield (2009) posited that collocational competence might plateau at a certain proficiency level, beyond which additional improvements are minimal. Durrant and Schmitt (2009) found that even advanced learners sometimes struggle with less frequent or more idiomatic collocations, suggesting that proficiency alone does not guarantee perfect collocational competence. Additionally, Peters (2016) highlighted that the development of collocational knowledge can be influenced by factors such as the learner's first language, the type of collocations, and the context in which they are learned. Peters' study showed that while advanced learners performed better overall, there were still significant individual differences in their collocational knowledge and usage. Research by Qian (2002) on vocabulary size and reading comprehension supports the idea that certain linguistic competencies improve significantly once a proficiency threshold is crossed.

In conclusion, the findings of this study, along with those from previous research, suggest that while proficiency is an important factor, it does not solely determine collocational competence. This highlights the need for effective instruction that not only enhance collocation usage but also increases awareness of collocations, which will be explored in the implications section.

5.4. Differences in Processing Congruent and Incongruent Items

As anticipated, congruent collocations were processed much faster than incongruent collocations by both groups of intermediate learners. Previous studies comparing native and non-native speakers have yielded similar results: non-native speakers responded to congruent items faster, while native speakers were found to process both types at nearly the same speed. This suggests a processing advantage for congruent collocations among L2 learners. Congruent collocations benefit from positive transfer, where existing L1 knowledge aids in the recognition and processing of similar structures in L2. In contrast, incongruent collocations require additional cognitive resources to process, leading to lower accuracy and longer reaction times (See Figure 1 & Figure 2 in Chapter 4). These findings are consistent with the growing body of evidence in previous studies discussed in chapter 2 (Yamashita & Jiang, 2010; Wolter & Gyllstad, 2011, 2013; Wolter & Yamashita, 2015, 2017).

This study builds on previous research by providing evidence for the influence of L1 congruency on collocation acquisition for L2 learners with such logographic L1s as Chinese. The difficulty in acquisition of incongruent collocations by non-native speakers can be explained by the L2 vocabulary development hypothesis proposed by Jiang (2000) which postulates that there exists a long-lasting reliance on the L1 to acquire L2 vocabulary. At the initial stage of L2 vocabulary acquisition, there is a strong reliance on the L1 when the L2 form first enters the L2 lexicon. With the increase in L2 exposure, the reliance on the L1 to acquire the L2 will decrease. However, rarely can L2 learners fully reach a stage where the L1 is not needed to assist in L2 vocabulary acquisition. It is thus understandable that lower proficiency L2 learners do not respond at an equal speed and accept incongruent collocations easily as valid if no direct semantic connection is present in the L1.

Another explanation for the effect of congruency on L2 collocational processing could be related to the Age of Acquisition (AoA) or Order of Acquisition (OoA)—the sequence in which words are learned influences how quickly their representations are activated, independent of frequency (Wolter & Gyllstad, 2013). The notion is that the age or order of learning significantly affects how deeply words become entrenched in the language system, particularly as the system

stabilises. Wolter and Yamashita (2015, 2018) also support this view. Numerous studies have shown that words learned earlier are more easily accessed than those learned later (e.g., Izura et al., 2011; Stewart & Ellis, 2008). This learning framework seems to explain the discrepancies in L2 processing of congruent and incongruent collocations, as well as the lack of accelerated processing observed for Mandarin-only collocations in this study and Japanese-only collocations in Wolter and Yamashita (2015). At the beginning of learning, L2 learners have access to all collocational knowledge from their L1, including both congruent and nontransferable L1 collocations (like the Mandarin-only collocations in the present study). As learners gain L2 experience, congruent collocations become reinforced through repeated exposure in the L2, while nontransferable collocations gradually diminish in prominence due to lack of reinforcement, ideally becoming inactive in the L2 collocational network. Although these may still be deemed acceptable in the L2, they are not readily activated as part of the collocational network. Meanwhile, learners will inevitably encounter incongruent L2 collocations—those acceptable in the L2 but not in the L1—but these will not surpass the dominance of congruent collocations unless there is significantly disproportionate exposure to high-frequency incongruent collocations over low-frequency congruent ones.

This explanation not only has solid empirical and theoretical support but also accounts for a wide range of empirical results in L2 collocational processing. It clarifies why the congruency effect has been consistently observed in studies involving L2 learners with different L1s and explains the comparative lag in processing incongruent collocations, even among advanced learners (e.g., Wolter & Gyllstad, 2011, 2013). Furthermore, it elucidates why nontransferable collocations, such as the Mandarin-only collocations in this study, do not show a processing advantage over baseline items, regardless of whether the task is a lexical decision task (LDT) (Wolter & Yamashita, 2015) or an acceptability judgement task, as in this study.

Since this study only recruited participants aged 16-18, it would be interesting for future research to explore the specific differences in collocation acquisition between this age group and adults. For example, future studies could replicate this research with adult subjects to compare the findings.

5.5. Interaction Between L2 Proficiency and L1 Congruency

The study did not find any significant interaction effect between L2 proficiency and L1 congruency on collocation processing. Whether learners are at B1 or B2 levels, congruent collocations are processed more accurately and quickly than incongruent ones. This result is

consistent with the findings of Wolter and Gyllstad (2013) and Wolter and Yamashita (2018), who found no interaction between proficiency and congruency during real-time recognition. This indicates that the advantages of L1 congruency in processing collocations are consistent across different these proficiency levels. These findings align with the notion that L1 influence is a pervasive factor in SLA, affecting learners regardless of their proficiency level (Jarvis & Pavlenko, 2008). These results may also lend credibility to the lasting-effect of the L1 as even when L2 proficiency increases the effect still appears subordinate to that of L1 influence. The present study adds to previous findings through (a) including EFL participants only, thus controlling for L2 context, and (b) treating proficiency as a continuum to allow for a more fine-grained depiction of its effect. While the non-natives with low estimated levels of proficiency showed dependence on the L1, the L1 effect remains strong even as estimated proficiency increases, at least within the band of intermediate EFL learners.

However, there are also robust studies that did find an interaction effect (Yamashita & Jiang, 2010; Sonbul & El-Dakhs, 2020). Sonbul and El-Dakhs (2020), for example, examined the impact of congruency and proficiency on L2 collocation recognition. Their study involved two experiments with 228 female Saudi English learners: a traditional multiple-choice test for untimed recognition and a timed acceptability judgment task, the same tools as in the present study. They found that both congruency and proficiency significantly influence collocation recognition, and there was an interaction effect: as proficiency increases, the influence of the L1 on collocation processing decreases, approaching native-like processing. Sonbul and El-Dakhs pointed out that a significant limitation of previous studies which did not yield an interaction effect is that they tested participants at a single proficiency level, a limitation shared to some degree by the present study. Future research will contribute to a more comprehensive picture by incorporating more diverse proficiency levels.

Also important to note here is that it is not possible to directly compare the proficiency level of our participants to those in previous research employing the same judgment task format (e.g., Wolter & Gyllstad, 2013; Wolter & Yamashita, 2018; Yamashita & Jiang, 2010) due to variation in proficiency measures. However, operationalising proficiency as both a continuous and a categorical variable seems to have allowed for a clearer depiction of its effect on collocation recognition.

5.6. Pedagogical Implications

Given the potential challenges L2 learners may face with both congruent and incongruent collocations, as highlighted in this study, it is important for EFL teachers and students to focus on

both types. While incongruent collocations are noticeable due to their distinct differences, the findings of this study suggest that more effort should be devoted to learning congruent collocations. These can cause subtler yet more frequent problems that are harder to resolve, even after students have been taught well-formed congruent collocations. Therefore, teachers may benefit from raising students' awareness of the difficulties posed by congruent collocations and provide ample practice, particularly with the more confusing ones.

One major problem learners and teachers encounter is how to address the large number of collocations present in English. Vasiljevic (2014) stressed that the influence of the L1 must be considered when selecting which collocations should be studied, practiced, or incorporated into language learning tasks. Research has shown that if a contrastive approach to language learning is taken to select only incongruent collocations for further study by learners, the number of collocations that require direct instruction and practice can be greatly reduced (Bahns, 1993). However, a recent study by Pham (2023) found a weak or non-existent relationship between learners' knowledge of congruent and incongruent collocations, refuting the idea that focusing on one group would automatically improve knowledge of the other. It would seem essential, therefore, for both congruent and incongruent collocations to receive adequate attention from teachers and students.

However, a more critical concern is whether collocations can be effectively taught at all. Mackin (1978) expresses doubt about the feasibility of teaching collocations, arguing that their sheer number makes systematic teaching or learning impractical. He suggests that foreign language learners can only develop collocational competence through extensive study, reading, and observation over many years (Mackin, 1978: 150-151). In contrast, Bahns (1993) proposes a different approach, focusing on incongruent collocations. He believes that these should be the primary focus of instruction, whereas those with direct L1 equivalents do not need to be explicitly taught. Bahns also suggests that the selection of collocations for TOEFL preparation materials should vary based on the learners' L1 backgrounds.

Literature from recent research has also been more optimistic, proposing various methods for effectively teaching collocations while considering learners' L1. Webb and Kagimoto's research (2009) investigates the effects of receptive and productive vocabulary tasks on learning collocation and meaning with Japanese EFL students. The results showed that both tasks led to significant gains in knowledge with little difference between the size of the gains. When participants were grouped according to level, the productive task was more effective for higher level learners, and the receptive task was more effective for lower level learners. Mean scores on the productive tests were

slightly higher for both tasks on the test of meaning than on the test of collocation. However, the findings indicate knowledge of collocation may be acquired at a rate similar to that of meaning, and that tasks which focus on collocation, as well as meaning, may be effective. Two specific areas of focus may lead to optimal gains in collocational competence: exposure and awareness.

5.6.1. Exposure

In terms of the lack of accelerated processing in B2 learners, studies by Segalowitz and Hulstijn (2005) and DeKeyser (2001) highlight the importance of extensive practice and varied exposure for achieving automaticity. To foster automatic processing, language instruction should emphasise frequent practice in diverse contexts, fluency activities, and incremental complexity in collocation use. These strategies help learners move from deliberate to automatic processing, ultimately enhancing their language proficiency and processing speed. As argued by Ellis (2012), both the recognition and production of words is a function of their frequency of occurrence in the language. The same thing goes for sequences of words. It stands to reason that through repeated exposure to authentic sources, learners gradually figure out what sequences of words are normally used in certain situations.

Much empirical evidence has been found to support this type of learning. Durrant and Schmitt (2010) provided evidence for the incidental learning of collocations encountered once or twice in sentence contexts. Items encountered twice were learned better than those encountered once in a form recall test. Webb, Newton, and Chang (2013) suggested positive effects of reading-while-listening on the incidental learning of collocations. Learners acquired the form and meaning of the target items when encountering them 5, 10, and 15 times through reading-while-listening to a graded reader. Pellicer-Sánchez (2017) also found significant incidental learning of adjective-pseudoword collocations appearing 4 and 8 times in a short story. Additionally, students should be taught to use corpora and collocation dictionaries to communicate more effectively and efficiently, thereby enhancing learner autonomy both inside and outside the classroom.

5.6.2. Awareness

While incidental learning facilitates exposure, explicit instructions are equally essential for another important aspect of collocation learning: awareness raising. In an L1 acquisition setting, language—collocations in particular—is learned from mere exposure. However, in an L2 setting where exposure is much more limited, it may be beneficial to include some sort of explicit learning of collocations. An interesting finding revealed by item-by-item analysis in the present study serves

as a good example: while most incongruent collocations received inaccurate judgements, one item in particular, ‘make a difference’ was correctly judged by 94% of all participants, the highest among all incongruent collocations (see No. 29 in Appendix 11). The familiarity of this phrase to the participants can be attributed to its high frequency of appearance in their curriculum. The participants are students preparing for the TOEFL exam, and as a result, they are primarily exposed to TOEFL-related course materials and practice tests. While most items in the PAJT are irrelevant to or absent from their course materials, the phrase ‘make a difference’ was prominently featured in practice materials for all four sub-skills (e.g., appearing in a vocabulary question in one practice test). This led to more explicit instruction from the tutors, resulting in a familiarity effect. This particular case provides evidence for the important role of explicit exposure in L2 collocation acquisition.

Because of their sheer number, it is probably unrealistic that collocations should be taught en masse in a structured way, just like it is unrealistic for teachers to focus on vocabulary material beyond the high-frequency words of the language, due to the lack of time (Nation, 2001), which means that the learners have to be responsible for continuously exposing themselves to collocations after classes. However, educators can draw students’ attention to collocations, formulaicity and idiomaticity through classroom activities and teaching and learning materials. What is needed is first and foremost an awareness of collocations as linguistic items and the problems they may cause on the part of the learners (Howarth 1996; Hill, 2000). This raising of awareness could be accompanied by some teaching of collocations that are typically problematic, frequent and occur in a wide range (Nation 2001; Hill 2000). This will be particularly critical for EFL learners with L1 Chinese, as Sun, Chen and Zhu (2023) discovered that many Chinese students paid more attention to new words (words that they did not know the definitional meaning of) when reading, while frequently ignoring more common and familiar words and how such words are combined to make meaning.

Previous research has also demonstrated the potential for explicit learning of collocations. Alali and Schmitt (2012) taught 30 collocations over six weeks and found that participants could recognise the form and meaning of nearly all the collocations taught, although they could recall only 8% of them. Zyzik (2011) reported significant gains in both recognition and production tasks after teaching 38 Spanish collocations over ten sessions. In a study by Al Hassan and Wood (2015), 65 collocations were taught over ten weekly sessions, with 80% of class time dedicated to presentation and practice and 20% to uncontrolled production. They observed that participants could use a significantly higher number of collocations in their post-test writing compared to the

pre-test. Furthermore, research suggests that the frequency of exposure influences the explicit learning of collocations. For instance, Peters (2014) and Webb and Kagimoto (2009) found significantly better learning outcomes when the number of explicit encounters increased. Even though the effects of various instructional methods were not examined in the present study, the case of 'make a difference' as discussed earlier did provide an insight into the potentially facilitative role of explicit instruction.

A significant challenge for L1-Chinese learners of English is effectively learning collocations through awareness and practice exercises. Research has shown that both beginner (Wang, 2011) and advanced EFL learners (Peters, 2009) who become familiar with collocations tend to actively seek them out in texts, even without explicit instruction. Therefore, the first step in improving L2 learners' collocation knowledge is to introduce the concept and raise their awareness (Wang, 2011). Additionally, exposure to collocations in EFL contexts may be insufficient, as they are not typically highlighted in learning materials (Reynolds, 2016). Highlighting collocations in these materials can create induced salience (Bruton, López & Mesa, 2011), and this strategy can be particularly effective in private tuition settings.

While pushing for revisions and top-down reforms in EFL curricula can be challenging, the PT context, the focus of this particular study, can play an especially helpful and complementary role in awareness raising, due to its flexibility of curriculum design and personalised support, as well as other features. In a private tutoring setting, educators have the freedom to tailor their teaching methods and materials to meet the specific needs and learning pace of individual students. This flexibility allows tutors to focus on collocations that are particularly challenging or relevant to the learner, which might not be possible in a standard classroom setting with a fixed curriculum. Therefore, PT teachers are well suited to respond to the calls for explicit teaching of collocations in the L2 classroom (Tsai, 2020; Vedder & Benigno, 2016). It is already common for private tutoring centres to produce reference word lists used in preparation for standardised proficiency exams; however, such lists found in the Chinese EPT market do not usually include references to collocations. Efforts should be made to incorporate more collocations and collocation practice exercises in EFL textbooks and to raise the status of collocations in vocabulary lists.

5.7. Limitations and Suggestions for Future Research

This study has several limitations. First, considering this study employs a quantitative method, the sample size is obviously limited. This limitation has also impacted the exclusion criteria for dealing with outliers. According to Jiang (2013), the traditional practice for handling outliers in

reaction time-based research involves setting the low cutoff at 300ms and the high cutoff at 2500ms. Jiang also recommends that this treatment should affect no more than 10% of the data (2013: 70-71). Given that the study only includes 36 participants, removing any participants is not ideal, even though two participants had to be removed from analysis. Consequently, the results are limited in their interpretability.

Another limitation was that the data were collected online, which meant that the environment in which participants performed the task could not be controlled. Consequently, some participants completed the task in a classroom setting with a supervising teacher present, while others might have accessed the study and performed the task at home or in a café. This variability is problematic because environmental distractions could impact processing. In fact, 8 participants reported experiencing distractions, 6 participants reported internet connectivity issues, and 13 reported response delay or failures. Future research should address these distractions in the analysis, e.g. by adding them as a covariate and test for their significant effect.

Additionally, the study did not adequately control for the frequency effect. Since the collocation frequencies and MI scores of collocation items are prioritised over the lexical frequencies of component words, some component words in the collocation set exceeded the 2,000-word lexical frequency range (see Appendix 12). Although the proportion of low-frequency words is limited (12 out of 113 content words) and some describe everyday objects that are arguably familiar to participants (e.g. computer, file, photo, drug, video), nearly all (33 of the total 34) participants reported encountering unfamiliar words in the debrief questionnaire (see Appendix 8), indicating a potential confounding effect due to infrequent words. It would have been valuable to explore through follow-up interviews or surveys what participants considered 'unfamiliar words' to better understand this issue. It has long been established that people react faster to high-frequency words than to low-frequency words (Monsell, Doyle, & Haggard, 1989; Scarborough, Cortese, & Scarborough, 1977). Future research should address this issue and ensure better control of word frequency during the item development stage.

Moreover, the cross-sectional nature of the study meant that future research should also consider conducting longitudinal studies to map the development of collocational knowledge over time. This would provide insights into how collocational knowledge evolves with increased exposure and practice, particularly across different proficiency levels. Including a range of proficiency bands and measures of vocabulary depth can offer a more nuanced understanding of learners' collocational competence. Additionally, comparing the effectiveness of different learning environments, such as PT versus mainstream schooling, could reveal context-based differences in collocation acquisition.

Such comparisons might highlight the benefits of personalised support and flexible curriculum design in PT settings, as discussed earlier. By addressing these areas, future studies can provide a more comprehensive view of collocational learning and its influencing factors.

6. Conclusion

Despite its significance in second language acquisition, collocation has often been overlooked in L2 classrooms. The purpose of this dissertation was to examine the role of L2 proficiency and L1 congruency in the processing of L2 collocations, thereby providing insights into how this aspect of language competence can be more effectively taught. Data were analysed from 34 high school learners whose first language is Chinese. Their proficiency levels were assessed using a combination of measures, and they were subsequently divided into two proficiency groups. To test their processing of different types of collocations, a phrasal acceptability judgement task was administered. This task included congruent collocations, incongruent collocations, Mandarin-only items, and baseline items. Independent-samples t-tests were conducted to identify differences in processing between the two proficiency groups and the congruency group. Furthermore, a MANOVA was carried out to explore the interaction between proficiency and congruency.

Firstly, regarding the effect of L2 proficiency, learners with B2 proficiency demonstrated higher accuracy in the collocation judgement tasks compared to B1 learners, although there were no significant differences in reaction times between the groups. Existing research has shown mixed results, with some studies suggesting that proficiency positively affects processing, while others do not. This study provides a more nuanced perspective by concentrating on a narrower proficiency range. While proficiency improved learners' accuracy in processing collocations, the lack of increased speed indicates a lack of automaticity, highlighting the need for extensive practice and varied exposure to achieve automatic mastery of collocations.

Regarding the effect of L1 congruency, the results indicated that L1 proficiency aids collocational processing in L2, with congruent collocations being processed with higher accuracy and speed than incongruent ones. This finding aligns with numerous other studies in the field (e.g., Wolter & Gyllstad, 2011, 2013) and highlights the challenges L2 speakers face when processing incongruent collocations. Based on these findings, explicit teaching of incongruent collocations in L2 classrooms is recommended to raise students' awareness, particularly in small-sized, private tuition settings where individualised attention can enhance learning effectiveness.

References

Al Hassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: A quantitative research study. *Journal of English for Academic Purposes*, 17, 51–62. https://doi.org/10.1016/j.jeap2015.02.001

Alali, F.A., & Schmitt. N. (2012). Teaching formulaic sequences: The same as or different from teaching single words? *TESOL* 3(2): 153–180. https://doi.org/10.1002/tesj.13

Allen, D. (2016). Japanese cram schools and entrance exam washback. *The Asian Journal of Applied Linguistics*, 3(1), 54–67.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford: Oxford University Press.

Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal*, 47(1), 56–63. https://doi.org/10.1093/elt/47.1.56

Bahns, J., Eldaw, M. (1993). Should we teach EFL students collocations? *System,* 21 (1), 101–114. https://doi.org/10.1016/0346-251X(93)90010-E

Baker, D. P., Goesling, B., & Letendre, G. K. (2002). Socioeconomic status, school quality, and national economic development: A cross-national analysis of the "Heyneman-Loxley effect" on mathematics and science achievement. *Comparative Education Review*, 46(3), 291-312.

Barfield, A. (2003). *Collocation Recognition and Production: Research Insights*. Tokyo: Chuo University.

Barfield, A. (2003). *Collocation Recognition and Production: Research Insights*. Tokyo: Chuo University.

Barfield, A. (2009). Exploring productive L2 collocation knowledge. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in language learners: Papers and perspectives in honour of Paul Meara* (pp. 95-110). Clevedon, UK: Multilingual Matters.

Barfield, A. (2009). Following individual L2 collocation development over time. In A. Barfield and H. Gyllstad (eds.) *Researching collocations in another language*. New York: Palgrave Macmillan. pp. 208–223. https://doi.org/10.1057/9780230245327_16

Bednarek, M. (2008). Semantic preference and semantic prosody re-examined. *Corpus Linguistics and Linguistic Theory*, 4(2), 119-139.

Benson, M., Benson, E., & Ilson, R. (1986). *Lexicographic description of English*. Philadelphia: John Benjamins.

Benson, P. (2011). Language learning and teaching beyond the classroom: An introduction to the field. In P. Benson & H. Reinders (Eds.), *Beyond the language classroom* (pp. 7–16). Palgrave Macmillan.

Biber, D., Conrad, S., & Leech, G. (2002). *Longman student grammar of spoken and written English*. Harlow, Essex: Longman.

Bishop, H. (2004). The effect of typographic salience on the look up and comprehension of unknown formulaic sequences. In Schmitt, N. (Ed.), *Formulaic sequences* (pp. 227–247). Amsterdam, the Netherlands: John Benjamins.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4-16. https://doi.org/10.3102/0013189X013006004

Boers, F. (2020). Factors affecting the learning of multiword items. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 143–157). New York: Routledge.

Bonk, W. J. (2001). Testing ESL learners' knowledge of collocations. In T. Hudson & D. Brown (Eds.), *A focus on language test development: Expanding the language proficiency construct across a variety of tests* (pp. 113-142). Honolulu: University of Hawaii.

Bray, M. (2021). Shadow Education in Europe: Growing Prevalence, Underlying Forces, and Policy Implications. *ECNU Review of Education*, 4(3), 442-475. https://doi.org/10.1177/2096531119890142

Bruton, A., López, M. G., & Mesa, R. E. (2011). Incidental L2 Vocabulary Learning: An Impracticable Term? *TESOL Quarterly*, 45(4), 759–768. https://doi.org/10.5054/tq.2011.268061

Buchmann, C., Condron, D. J., & Roscigno, V. J. (2010). Shadow education, American style: Test preparation, the SAT and college enrollment. *Social Forces*, 89(2), 435–461.

Cao, D. & Badger, R. (2021). Cross-linguistic influence on the use of L2 collocations: the case of Vietnamese learners. *Applied Linguistics Review 2021*, 1-26. https://doi.org/10.1515/applirev-2020- 0035

Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition*, 17(4), 784–797. doi: 10.1017/S1366728913000795

Chi, M.L., Wong, P.Y., Wong, C.P. (1994). Collocational problems amongst ESL learners: a corpus-based study. *Proceedings Joint Seminar on Corpus Linguistics and Lexicology,*

Guangzhou and Hong Kong (Language Centre, HKUST, Hong Kong, 1994). https://repository.hkust.edu.hk/ir/Record/1783.1-1088

Cieślicka, A. B., & Heredia, R. R. (2017). How to Save Your Skin when Processing L2 Idioms: An Eye Move- ment Analysis of Idiom Transparency and Cross-language Similarity among Bilinguals. *Iranian Journal of Language Teaching Research*, 5(June), 81–107.

Cohen, A. D. (2014). *Strategies in Learning and Using a Second Language*. Routledge. https://doi.org/10.4324/9781315833200

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72–89. https://doi.org/10.1093/applin/amm022

Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 2(3), 223-235.

Cowie, A. P. (1994). Phraseology. In R. E. Asher (Ed.), *The encyclopedia of language and linguistics* (pp. 3168-3171). Pergamon Press.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. http://corpus.byu.edu/coca/

De Bot, K., Paribakht, T. S., & Wesche, M. B. (1997). Toward a Lexical Processing Model for the Study of Second Language Vocabulary Acquisition. *Studies in Second Language Acquisition*, 19(03). https://doi.org/10.1017/s0272263197003021

De Rycker, A. Encouraging Collocational and Colligational Fluency: Pedagogical Chunking, Word and Verb Mapping, Pause Reading and Other Strategies. J Interdiscip Res Educ 4, 2 (2014). https://doi.org/10.7603/s40933-014-0002-4

DeKeyser, R. (2001). Automaticity and Automatization. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 125-151). Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.007

Dierkes, J. (2010). Teaching in the shadow: Operators of small shadow education institutions in Japan. *Asia Pacific Education Review*, 11(1), 25-35. https://doi.org/10.1007/s12564-009-9059-3.

Ding, C., & Reynolds, B. L. (2019). The effects of L1 congruency, L2 proficiency, and the collocate-node relationship on the processing of L2 English collocations by L1-Chinese EFL learners. *Review of Cognitive Linguistics*, 17(2), 331–357.

Ding, C., Barry Lee Reynolds, Szabo, C. Z., & Boone, G. (2023). Assessing English language learners' collocation knowledge: a systematic review of receptive and productive

measurements. *International Review of Applied Linguistics in Language Teaching/ IRAL.*0(0). https://doi.org/10.1515/iral-2022-0163

Durrant P., & Schmitt N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching*, 47(2), 157–177. https://doi.org/10.1515/iral-2009-0156

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, 28, 157-169.

Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations: A meta-analysis. *International Journal of Corpus Linguistics*, 19, 443–477. https://doi.org/ 10.1075/ijcl.19.4.01dur

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the collocational competence of proficient L2 users. *Applied Linguistics*, 31(4), 455-475.

El-Dakhs, D., Salem, M. & Al-Haqbani, J. (2020). The Influence of L1 Congruency, L2 Exposure and Word Class on Collocation Learning: The Case of Arab Learners of English. *Asian EFL Journal,* 27, 62-88.

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17–44.

Ellis, N. C., Rita, S.-V., and Carson, M. (2008). Formulaic language in native and second-language speakers: psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*. 42, 375–396.

Evert, S. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), Corpus linguistics: An international handbook (pp. 1212-1248). Walter de Gruyter.

Fan, M.Y. (1991). A Study of the Company Kept by a Selection of English Delexical Verbs and the Implications for Teaching of English in Hong Kong. Ph.D. University of Durham, UK.

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications - Books.

Firth, J. R. (1957). A synopsis of linguistic theory 1930-1955. In F. R. Palmer (Ed.), *Selected Papers of J.R. Firth 1952-1959* (pp. 168-205). Longman.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 75–94). Harlow, UK: Pearson Education.

Gitsaki, C. (1999). *Second Language Lexical Acquisition: A Study of the development of collocational knowledge*. San Francisco: International Scholars Publications.

Granger, S.. (1998). Prefabricated patterns in advanced EFL writing: collocations and lexical phrases. In A. P. Cowie (Ed), *Phraseology: Theory, Analysis and Applications* (pp. 145–160). Oxford: Clarendon Press.

Gries, S. T. (2008). Phraseology and linguistic theory: A brief survey. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 3-25). Amsterdam: John Benjamins.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, 1(2), 131-149. https://doi.org/10.1017/S136672899800025X.

Gyllstad, H. (2007). *Testing English collocations: Developing receptive tests for use with advanced Swedish learners*. Lund University.

Gyllstad, H., & Wolter, B. (2015). Collocational Processing in Light of the Phraseological Continuum Model: Does Semantic Transparency Matter? *Language Learning*, 66(2), 296–323. https://doi.org/10.1111/lang.12143

Gyllstad, H., and Wolter, B. (2016). Collocational processing in light of the phraseological continuum model: does semantic transparency matter? *Lang. Learn*. 66, 296–323 doi: 10.1111/lang.12143

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) (3 ed.). Thousand Oaks, CA: Sage.

Hajar, A., & Karakus, M. (2023). An inevitable phenomenon: A bibliometric analysis of shadow education research (1982–2022). In *Annual Review of Comparative and International Education 2022* (Vol. 46a, pp. 193-213). (International Perspectives on Education and Society; Vol. 46A). Emerald Group Publishing Ltd.. https://doi.org/10.1108/S1479-36792023000046A014.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Halliday, M. A. K., McIntosh, A., & Strevens, P. (1964). *The Linguistic Sciences and Language Teaching*. Longman.

Hamid, M. O., Sussex, R., & Khan, A. (2009). Private Tutoring in English for Secondary School Students in Bangladesh. *TESOL Quarterly*, 43(2), 281–308. https://doi.org/10.1002/j.1545-7249.2009.tb00168.x

Handl, S. (2008). Essential collocations for learner of English: The role of collocational direction and weight. In F. Meunier & S. Granger (Eds.), Phraseology in foreign language learning and teaching (pp. 43-66). Amsterdam: John Benjamins.

Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (eds.). *Teaching collocation: Further developments in the lexical approach* (pp. 47-67). Hove: LTP.

Hoang, H., & Boers, F. (2016). Re-telling a story in a second language: How well do adult learners mine an input text for multiword expressions? *Studies in Second Language Learning and Teaching*, 6(3), 513–535. https://doi.org/10.14746/ssllt.2016.6.3.7.

Hohenberger, A. & Peltzer-Karpf, A. (2009). Language learning from the perspective of nonlinear dynamic systems. Linguistics. 47. 10.1515/LING.2009.017.

Howarth, P. A. (1996). *Phraseology in English Academic Writing*. DE GRUYTER. https://doi.org/10.1515/9783110937923.

Howarth, P. (1998). Phraseology and Second Language Proficiency. *Applied Linguistics*, 19(1), 24–44.

Hunston, S., & Francis, G. (2000). *Pattern Grammar*. Amsterdam, Philadelphia, PA: John Benjamins. https://doi.org/10.1075/scl.4

Izura, C., Pérez, M. A., Agallou, E., Wright, V. C., Marín, J., Stadthagen-González, H., & Ellis, A. W. (2011). Age/order of acquisition effects and the cumulative learning of foreign words: A word training study. *Journal of Memory and Language*, 64(1), 32–58. https://doi.org/10.1016/j.jml.2010.09.002

Jarvis, S., & Pavlenko, A. (2007). *Crosslinguistic Influence in Language and Cognition* (1st ed.). Routledge. https://doi.org/10.4324/9780203935927

Jeon, J., & Choe, Y. (2018). Cram schools and English language education in East Asian contexts. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1–14). Wiley.

Jiang, J. (2009). Pedagogic materials for L2 collocation use. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language* (pp. 99–113). London: Palgrave Macmillan. https://doi.org/10.1057/9780230245327_8

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21(1), 47–77. https://doi.org/10.1093/applin/21.1.47

Jiang, N. (2013). *Conducting Reaction Time Research in Second Language Studies*. https://doi.org/10.4324/9780203146255

Jiang, N., & Zhang, Y. (2019). Measuring L2 proficiency in research on bilingualism. *Bilingualism: Language and Cognition*, 22(3), 661-675.

Jiang, N., Li, M., & Guo, T. (2020). A tale of two frequency effects: Toward a verification model of L2 word recognition. *Applied Psycholinguistics*, 41(1), 215–236. https://doi.org/10.1017/S0142716419000481

Jiang, Y. (2022). Effects of L1-L2 congruency, collocation type, and restriction on processing L2 collocations. *Frontiers in Psychology*, 13, 947725. https://doi.org/10.3389/fpsyg.2022.947725

Laufer, B., & Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning*, 61(2), 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x

Lee, S. (2021). L1 transfer, proficiency, and the recognition of L2 verb-noun collocations: A perspective from three languages. *International Review of Applied Linguistics in Language Teaching*, 59(2), 181–208. https://doi.org/10.1515/iral-2018-0220

Leech, G. (2000). Grammars of spoken English: New outcomes of corpus-oriented research. *Language Learning,* 50, 675–724.

Lei, W. (2005). Expenditure on Private Tutoring for Senior Secondary Students: Determinants and Policy Implications. *Education and Economy*. 1, 39–42. (in Chinese)

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

Li, J., and Schmitt, N. (2010). *The Development of Collocation Use in Academic Texts by Advanced L2 Learners: A Multiple Case Study Approach.* https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.707.4078&rep=rep1&type=pdf

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413-468). Academic Press.

Lorenz, T. R. (1999). *Adjective Intensification – Learners Versus Native Spaekers. A Corpus Study of Argumentative Writing*. Amsterdam and Atlanta, Rodopi.

Mackin, R. (1978). On collocations: Words shall be known by the company they keep. In P. Strevens (Ed.), *In honour of A. S. Hornby* (pp. 149-165). Oxford, UK: Oxford University Press.

Mahmud, R., & Bray, M. (2017). School factors underlying demand for private supplementary tutoring in English: Urban and rural variations in Bangladesh. *Asia Pacific Journal of Education*, 37(3), 299-309.

Martinez, R., & Murphy, V. A. (2011). Effect of Frequency and Idiomaticity on Second Language Reading Comprehension. *TESOL Quarterly*, 45(2), 267–290. https://doi.org/10.5054/tq.2011.247708

McCarthy, M., & O'Dell, F. (2017). English collocations in use. Cambridge University Press.

Men, H. (2018). *Vocabulary increase and collocation learning : a corpus-based cross-sectional study of Chinese learners of English*. Springer; Shanghai.

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. Annals of cardiac anaesthesia, 22(1), 67–72. https://doi.org/10.4103/aca.ACA_157_18

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: where are they? *Journal of Experimental Psychology. General*, 118(1), 43–71. https://doi.org/10.1037//0096-3445.118.1.43

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambrigde: Cambridge University Press.

Nation, I. S. P. (2022). *Learning vocabulary in another language*. Cambrigde: Cambrigde University Press.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.

Palmer, H. E. (1933). *Second interim report on English collocations*. Kaitakusha.

Papageorgiou, S., Tannenbaum, R., Bridgeman, B., & Cho, Y. (2015). *The Association Between TOEFL iBT ® Test Scores and the Common European Framework of Reference (CEFR) Levels*. https://www.ets.org/Media/Research/pdf/RM-15-06.pdf

Park, H., Byun, S.-Y., & Kim, K.-K. (2011). Parental involvement and students' cognitive outcomes in Korea: Focusing on private tutoring. *Sociology of Education*, 84(1), 3-22.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication*, 191–226.

Pellicer-Sánchez, A. (2017). Incidental learning of L2 collocations: A classroom study. *Language Teaching Research*, 21(3), 381–402. https://doi.org/10.1177/1362168815618428.

Peters, E. (2009). Learning Collocations through Attention-Drawing Techniques: A Qualitative and Quantitative Analysis. In: Barfield, A., Gyllstad, H. (eds) *Researching Collocations in Another Language*. Palgrave Macmillan, London. https://doi.org/10.1057/9780230245327_15

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research*, 18(1), 75–94. https:// doi.org/10.1177/1362168813505384

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, 20(1), 113-138.

Pham, N. (2023). The Role of Congruency in Collocation Acquisition: A Case Study of Vietnamese Students Learning English Adjective + Noun Collocations. *Rupkatha Journal on Interdisciplinary Studies in Humanities*, 15(1). https://doi.org/10.21659/rupkatha.v15n1.05

Pituch K. A., Stevens J. P. (2016). *Applied statistics for the social sciences* (6th ed.). Routledge.

Puig-Mayenco, E., Chaouch-Orozco, A., Liu, H., & Martín-Villena, F. (2023). The LexTALE as a measure of L2 global proficiency. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.22048.pui.

Qian, D. D. (2002). Investigating the Relationship Between Vocabulary Knowledge and Academic Reading Performance: An Assessment Perspective. *Language Learning*, 52(3), 513–536. https://doi.org/10.1111/1467-9922.00193

Read, J., & Nation, P. (2004). Measurement of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 23-35). Amsterdam: John Benjamins.

Renouf, A., & Banerjee, J. (2007). Lexical repulsion between sense-related pairs. *International Journal of Corpus Linguistics*, 12(3), 415-443.

Renouf, A., & Banerjee, J. (2008). The phenomenon of lexical repulsion in text. *Lingvisticæ Investigationes*, 31(2), 213-225.

Revier, R. L. (2009). Evaluating a new test of whole English collocations. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 125-138). Basingstoke: Palgrave Macmillan

Saito, K., Dewaele, J.-M., Abe, M., & In'nami, Y. (2018). Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning*, 68(3), 709-743.

Salsbury, T., Crossley, S. A., & McNamara, D. S. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research*, 27, 343–360.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, 3(1), 1–17. https://doi.org/10.1037/0096-1523.3.1.1

Schmidt, R. W. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, 11(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York, NY: Palgrave Macmillan.

Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge: What the Research Shows. *Language Learning*, 64(4), 913–951. https://doi.org/10.1111/lang.12077

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 1-22). Amsterdam: John Benjamins.

Schmitt, N., & Schmitt, D. (2020). *Vocabulary in language teaching*. Cambridge University Press.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in Bilingualism and Second Language Learning. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford University Press.

Seretan, V. (2011). *Syntax-based collocation extraction*. Springer.

Shin, D., & Chon, Y. V. (2019). Revisiting collocational patterns in COCA and implications for EFL teaching. *Journal of Linguistics and Language Teaching*, 10(2), 593-622.

Sinclair, J. (1966). Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In Memory of J. R. Firth* (pp. 410-430). Longman.

Sinclair, J. (1991). Corpus, concordance, collocation. Oxford: Oxford University Press.

Sinclair, J.. (1987). *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London: Collins ELT.

Skehan, P. (2009). Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. *Applied Linguistics*. https://doi.org/30. 10.1093/applin/ amp047.

Sonbul, S., & El-Dakhs, D. (2020). Timed versus untimed recognition of L2 collocations: Does estimated proficiency modulate congruency effects? *Applied Psycholinguistics*, 41(5), 1197–1222. https://doi.org/10.1017/S014271642000051X

Stewart, N., & Ellis, A. W. (2008). Order of acquisition in learning perceptual categories: A laboratory analogue of the age-of-acquisition effect? *Psychonomic Bulletin & Review*, 15(1), 70–74. https://doi.org/10.3758/PBR.15.1.70

Sun, D., Chen, Z., & Zhu, S. (2023). What affects second language vocabulary learning? Evidence from multivariate analysis. *Frontiers in Education*, 8. https://doi.org/10.3389/feduc.2023.1210640

Szudarski, P. (2015). Formal Instruction in Collocations in English: Mixed Methods Approach. In: Piasecka, L., Adams-Tukiendorf, M., Wilk, P. (eds) *New Media and Perennial Problems in Foreign Language Learning and Teaching. Second Language Learning and Teaching*. Springer, Cham. https://doi.org/10.1007/978-3-319-07686-7_9

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Pearson

Tatsuya, N., Yu, T., & Scott, A. (2020). Examining the Validity of the LexTALE Test for Japanese College Students. *The Journal of Asia TEFL*, 17(2), 335–348. https://doi.org/10.18823/asiatefl.2020.17.2.2.335

Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting Bilingual and Monolingual Idiom Processing. In A. B. Cieślicka & R. R. Heredia (Eds.), *Bilingual figurative language processing* (pp. 171–207). Cambridge: Cambridge University Press. https://doi.org/doi: 10.1017/CBO9781139342100.011.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomoschuk, B., Ferreira, V. S., & Gollan, T. H. (2018). When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition*, 22(3), 516–536. https://doi.org/10.1017/s1366728918000421.

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 151-173). Continuum.

Tsai, M. (2020). Teaching L2 collocations through concept-based instruction: The effect of L2 proficiency and congruency. *International Journal of Applied Linguistics*, 30(3), 553–575. https://doi.org/10.1111/ijal.12311

Vasiljevic, Z. (2014). Teaching collocations in a second language: Why, what and how?. *ELTA Journal*, 2(2), 48-73.

Vedder, I., & Benigno, V. (2016). Lexical richness and collocational competence in second-language writing. *International Review of Applied Linguistics in Language Teaching*, 54(1). https://doi.org/10.1515/iral-2016-0015

Wang, D. (2011). Language Transfer and the Acquisition of English Light Verb + Noun Collocations by Chinese Learners, *Chinese Journal of Applied Linguistics*, 34 (2), 107 – 125.

Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*, 43(1), 55–77. https://doi.org/10.1002/j.1545-7249.2009.tb00227.x.

Webb, S., Newton, J., & Chang, A. (2012). Incidental Learning of Collocation. *Language Learning*, 63(1), 91–120. https://doi.org/10.1111/j.1467-9922.2012.00729.x.

Whitsitt, S. (2005). A critique of the concept of semantic prosody. *International Journal of Corpus Linguistics*, 10(3), 283-305.

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, 32(4), 430–449.

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing. *Studies in Second Language Acquisition,* 35(3), 451–482.

Wolter, B., & Yamashita, J. (2015). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, 36(5), 1193-1221.

Wolter, B., & Yamashita, J. (2017). Word Frequency, Collocational Frequency, L1 Congruency, And Proficiency In L2 Collocational Processing. *Studies in Second Language Acquisition*, 40(2), 395–416.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, 21, 463–489.

Wray, A. (2002). Formulaic language and the lexicon. Cambridge: Cambridge University Press.

Xue, H., & Ding, X. (2009). A Study on Additional Instruction for Students in Cities and Towns in China. *Educational Research*. 2009;348(1):39–46. (in Chinese)

Yamashita, J., & Jiang, N. (2010). L1 Influence on the Acquisition of L2 Collocations: Japanese ESL Users and EFL Learners Acquiring English Collocations. *TESOL Quarterly*, 44(4), 647–668.

Yi, H. (2018). The effect of collocational frequency and transitional probability on the processing of English adjective-noun collocations by L2 learners. *Language Learning & Technology*, 22(3), 187-204.

Yi, W.. (2018). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Stud. Second Lang. Acquis.* 40, 1–16.

Ying, Y., O'Neill, M. (2009). Collocation Learning through an 'AWARE' Approach: Learner Perspectives and Learning Process. In: Barfield, A., Gyllstad, H. (eds) *Researching Collocations in Another Language*. Palgrave Macmillan, London. https://doi.org/10.1057/9780230245327_14

Yung, K. W. H. (2019). Exploring the L2 selves of senior secondary students in English private tutoring in Hong Kong. *System*, 80, 120–133. https://doi.org/10.1016/j.system.2018.11.003

Yung, K. W. H., & Bray, M. (2021). Globalisation and the Expansion of Shadow Education: Changing Shapes and Forces of Private Supplementary Tutoring. *Third International Handbook of Globalisation, Education and Policy Research*, 679–697. https://doi.org/10.1007/978-3-030-66003-1_39.

Yung, K. W. H., & Chiu, M. M. (2023). Secondary school students' enjoyment of English private tutoring: An L2 motivational self perspective. *Language Teaching Research*, 27(4), 907-929. https://doi.org/10.1177/1362168820962139

Yung, Y. (2015). Shadow education in Hong Kong: Private supplementary tutoring and its implications for policy. *International Review of Education*, 61(1), 1-21.

Zhan, S., Bray, M., Wang, D., Lykins, C., & Kwo, O. (2013). The effectiveness of private tutoring: Students' perceptions in comparison with mainstream schooling in Hong Kong. *Asia Pacific Education Review*, 14(4), 495-509. https://doi.org/10.1007/s12564-013-9276-7.

Zhang, W. & Bray, M. (2020). Comparative research on shadow education: Achievements, challenges, and the agenda ahead. *European Journal of Education*, 55(3). https://doi.org/10.1111/ejed.12413.

Zhang, W., & Bray, M. (2020). The effectiveness of private tutoring: Students' perceptions in comparison with mainstream schooling in Hong Kong. *Asia Pacific Education Review*, 21(3), 403-416. https://doi.org/10.1007/s12564-013-9276-7.

Zyzik, E. (2011). Second language idiom learning: The effects of lexical knowledge and pedagogical sequencing. *Language Teaching Research*, 15(4), 313–433. https://doi.org/ 10.1177/ 1362168811412025

Appendices

Appendix 1. Ethics approval

Dear Sihan,

Research Ethics Approval

Research Title: Processing Collocations in the L2: Do L2 Proficiency and L1 Congruency Affect Collocation Learning among Chinese High School EFL Learners?

Research Ethics Reference: EDUC_C1A_24_079

The above application has been considered on behalf of the Education Departmental Research Ethics Committee (DREC) in accordance with the University's procedures for ethical approval of all research involving human participants.

I am pleased to confirm that, on the basis of the information provided to the DREC, ethics approval has now been granted for this study.

Please note the following:

- **Personal data**: It is the responsibility of the PI to ensure that all personal data collected during the project is managed in accordance with the University's guidance and legal requirements.
- **In-person activities**: Any data collection involving in-person interactions with participants must have an up-to-date fieldwork risk assessment in place; further guidance is available from the Safety Office's website.
- **Amendments**: Please notify the committee if you intend to make any amendments to the information in your ethics application as submitted at date of this approval, as all changes must receive ethical approval prior to implementation. The amendment form is available on the SSH IDREC webpage.

We welcome feedback on your experience of the ethical review process and suggestions for improvement. Please email any comments you might have to staff.curec@education.ox.ac.uk / student.curec@education.ox.ac.uk or ethics@socsci.ox.ac.uk.

Yours sincerely,

Dr Faidra Faitaki, AFHEA
Departmental Lecturer in Applied Linguistics
DREC Member

Appendix 2. Participant Recruitment Email (Director)

Participant Recruitment Emails

Director of Reach Top Education,

I am a student at the Department of Education at the University of Oxford, pursuing an MSc in Applied Linguistics and Second Language Acquisition. For my dissertation, I am conducting a study on how L2 learners of English process collocations with a focus on L1 Chinese high school students. Therefore, I am writing to ask you if it would be possible to undertake data collection in your centre, as your students fit the criteria of my study.

The study has been approved by the Education Department Research Ethics Committee (Reference No.XXX). The data collection will take place online via Gorilla and the tasks will take approximately 20 minutes to finish. Participation is voluntary and the research data will be anonymous and confidential. I am attaching a PDF file which contains detailed participant information sheet for your information.

I would be glad to talk to your students and teachers about my research after the data has been collected. If you have any further questions, please do not hesitate to contact me (sihan.fang@education.ox.ac.uk) or my supervisor, Dr. Anna-Maria Ramezanzadeh (anna-maria.ramezanzadeh@education.ox.ac.uk). We would be glad to hear from you. Thank you for your time. I look forward to hearing from you.

Kind regards,

Sihan Fang

Appendix 3. Participant Recruitment Email (Teachers)

Participant Recruitment Emails

To whom it may concern,

I am a student at the Department of Education at the University of Oxford, pursuing an MSc in Applied Linguistics and Second Language Acquisition. For my dissertation, I am conducting a study on how L2 learners of English process collocations with a focus on L1 Chinese speakers. Therefore, I am writing to ask you if it would be possible to undertake data collection in your centre, as your students fit the criteria of my study.

The study has been approved by the Education Department Research Ethics Committee (Reference No.XXX) and by the director of Reach Top Education. The data collection will take place online via Gorilla and the tasks will take approximately 20 minutes to finish. Participation is voluntary and the research data will be anonymous and confidential.

Upon publication, a summary of the results will be made available to those who are interested. Anyone that agrees to participate and completes all the tasks will be offered a PDF of the phrases included in the test.

I am attaching a PDF file which contains detailed participant information sheet for your information. Any student interested in participating can access the test at their convenience via an internet-connected computer. The link to the test is XXXXXXXXXXXX.

I would be glad to talk to your students about my research after the data has been collected. If you have any further questions, please do not hesitate to contact me (sihan.fang@education.ox.ac.uk) or my supervisor, Dr. Anna-Maria Ramezanzadeh (anna-maria.ramezanzadeh@education.ox.ac.uk). We would be glad to hear from you. Thank you for your time. I look forward to hearing from you.

Kind regards,
Sihan Fang

Appendix 4. Participant information sheet

PARTICIPANT INFORMATION SHEET

Central University Research Ethics Committee Approval Reference: [EDUC_C1A_24_079]

1. Introductory paragraph

You are being invited to take part in a research project. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether you wish to take part.

2. Why is this research being conducted?

I am conducting a research study at the University of Oxford looking at how individuals process word combinations. I am interested particularly in how L1 speakers of Chinese learning English process different types of word combinations.

3. Why have I been invited to take part?

You have been recommended to take part since you are a high school student and a native Chinese speaker learning English. Your responses will help us to understand how word combinations are recognised and processed by language users.

4. Do I have to take part?

No. It is up to you to decide whether to take part. You can withdraw yourself from the research, without giving a reason by advising me of this decision. Your personal information will not be stored if you exit before or during the experiment. However, after submitting the results and completing the experiment, the data will be stored and automatically anonymised, at which point it will not be possible to withdraw your data.

5. What will happen to me if I take part in the research?

You will be asked to complete a questionnaire regarding your linguistic background and a brief test to assess your proficiency in the English language. The test will have no impact on your grades or school learning. Following this, you will be asked to complete a simple word recognition exercise. During this exercise, you will be presented with word combinations in English (e.g., EAT DINNER) and asked to determine whether the combination is considered ‘acceptable’ in English. The entire process is expected to take approximately 20 minutes and does not require any prior knowledge. Detailed instructions will be provided to assist you throughout the process.

6. What are the disadvantages and risks in taking part?

The entire study can be accessed and completed online with your personal computer connected to the Internet. You won’t have to engage in any physical activity. You may experience fatigue from prolonged screen time and continuous task completion. To address this, the test duration will be limited to around 20 minutes, there will be a 15 second break halfway during the test.

7. Are there any benefits in taking part?

While there are no financial benefits for those people participating in the research, it is hoped that the study will bring to your attention the importance of learning collocation as a dimension of vocabulary knowledge. This is of great interest to you since knowledge of collocations, and formulaic language in general, has been found to play an instrumental role in second language proficiency, despite often being overlooked in English curricula.

8. What information will be collected and why is the collection of this information relevant for achieving the research objectives?

Before the test starts, your name will be collected to confirm the authenticity of data. You will be asked to fill out a questionnaire about your age and English learning background. I am interested in the duration of your English studies, your self-rated proficiency level, your most recent standardised test score, and your experience of learning English abroad.

9. Will the research be published? Could I be identified from any publications or other research outputs?

The findings from the research will be written up for an MSc. degree in Applied Linguistics and Second Language Acquisition.

Data will be anonymised by replacing students names with ID numbers, and the data containing the personal identifiers (participants' names), and any linkage to the ID numbers will be stored separately and deleted at the end of the project. The study will not report any identifiable information.

10. Data Protection

Your responses on the tests will be stored in a password-protected electronic file. Research data will be stored for three years after publication or public release.

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the research. The University will process your personal data for the research outlined above. Research is a task that is performed in the public interest. Further information about your rights with respect to your personal data is available from the University's Information Compliance web site at https://compliance.admin.ox.ac.uk/individual-rights.

11. Who has reviewed this research?

This research has received ethics approval from a subcommittee of the University of Oxford Central University Research Ethics Committee. (Ethics reference: xxxxx).

12. Who do I contact if I have a concern about the research or I wish to complain?

If you have a concern about any aspect of this research, please contact Sihan Fang (sihan.fang@education.ox.ac.uk) or Dr. Anna-Maria Ramezanzadeh (anna-maria.ramezanzadeh@education.ox.ac.uk), and we will do our best to answer your query. I will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:

The Chair, Social Sciences & Humanities Interdivisional Research Ethics Committee;
Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Boundary Brook House, Churchill Drive, Headington, Oxford OX3 7GB

13. Further Information and Contact Details

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

Sihan Fang
Department of Education
University of Oxford
15 Norham Gardens
Oxford
OX2 6PY

University tel: + +44 (0) 1865 274024
University email: sihan.fang@education.ox.ac.uk

Appendix 5. Consent form

Consent to take part in 'Processing Collocations in the L2: Do L2 Proficiency and L1 Congruency Affect Collocation Learning among Chinese High School EFL Learners?'

Central University Research Ethics Committee (CUREC) approval reference: EDUC_C1A_24_079

Purpose of Study: Understanding how high school EFL learners process word combinations

Please initial each box if you agree with the statement

I confirm that I have read and understand the information sheet for the above research. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

I understand that my participation is voluntary and that I am free to withdraw at any point before and during the experiment without giving any reason. After submitting the results and completing the experiment, I will no longer be able to withdraw any data.

I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project.

I understand that I will not be identifiable from any publications, presentations, videos or websites.

I understand how to raise a concern or make a complaint.

I agree to take part.

I agree that my personal contact details can be retained in a secure database so that the researchers can contact me about incentives. YES / NO

_____	dd / mm / yyyy	_____
Name of participant	Date	Signature

_____	dd / mm / yyyy	_____
Name of person taking consent	Date	Signature

Appendix 6. Questionnaire

Questionnaire

Please answer the following questions about yourself.

Age: _____

Gender: □ Female □ Male □ Non-binary □ Other (please specify)

Handedness: □ Left □ Right □ Ambidextrous (both)

First (native) language: □ Chinese (Mandarin) □ English □ Other (please specify)

Please answer the following questions about your English learning abilities and background.

Age when you started learning English: _____

Where did you start learning English?

□ Home □ School □ Online (e.g. Duolingo) □ Language School □ Other (please specify)

Which variety of English did/do you primarily learn?

□ American English □ British English □ Both □ Don't know □ Other (please specify)

Did/Do you study at a school where English is/was the medium of instruction?

□Yes □No

Have you lived (or do you live) or spent an extended period of time (more than 3 months) in a majority English-speaking country?

□Yes □No

If yes, for how long?

□Less than 6 months □6 months to 1year □More than 1 year □ 1 to 3 years □ More than 3 years

Have you ever learned English in a study abroad context?

□Yes □No

On a scale of 1 (beginner) to 5 (proficient), how would you rate your English proficiency in the following areas:

Reading: 1□ 2□ 3□ 4□ 5□

Listening: 1□ 2□ 3□ 4□ 5□

Writing: 1□ 2□ 3□ 4□ 5□

Speaking: 1□ 2□ 3□ 4□ 5□

On a scale of 1 (none at all) to 5 (all the time), how would you rate your use of English on an average day?

1□ 2□ 3□ 4□ 5□

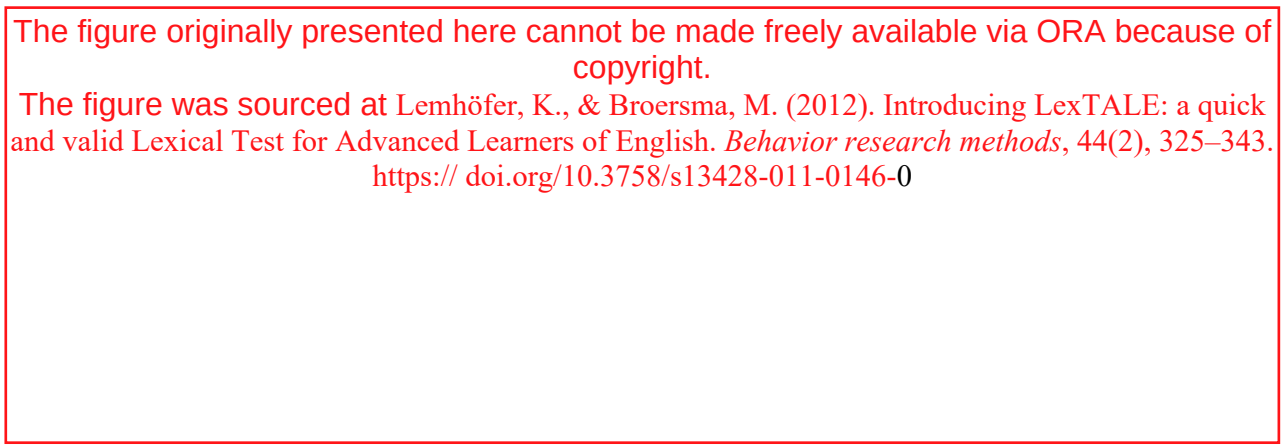
Have you taken any standardised English proficiency test (e.g. IELTS/TOEFL)?

□Yes □No

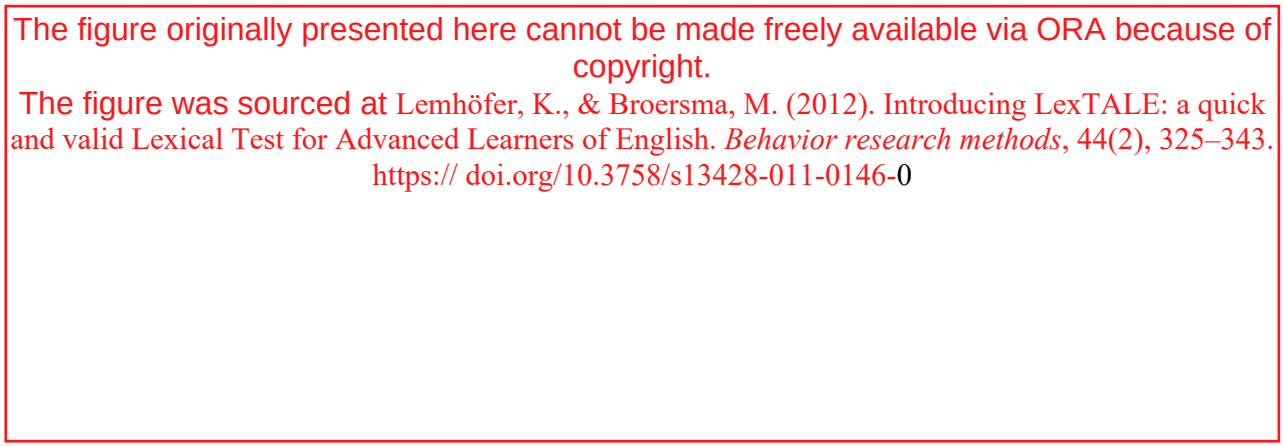
If yes, what is your most recent test score? _____

Appendix 7. Task screens

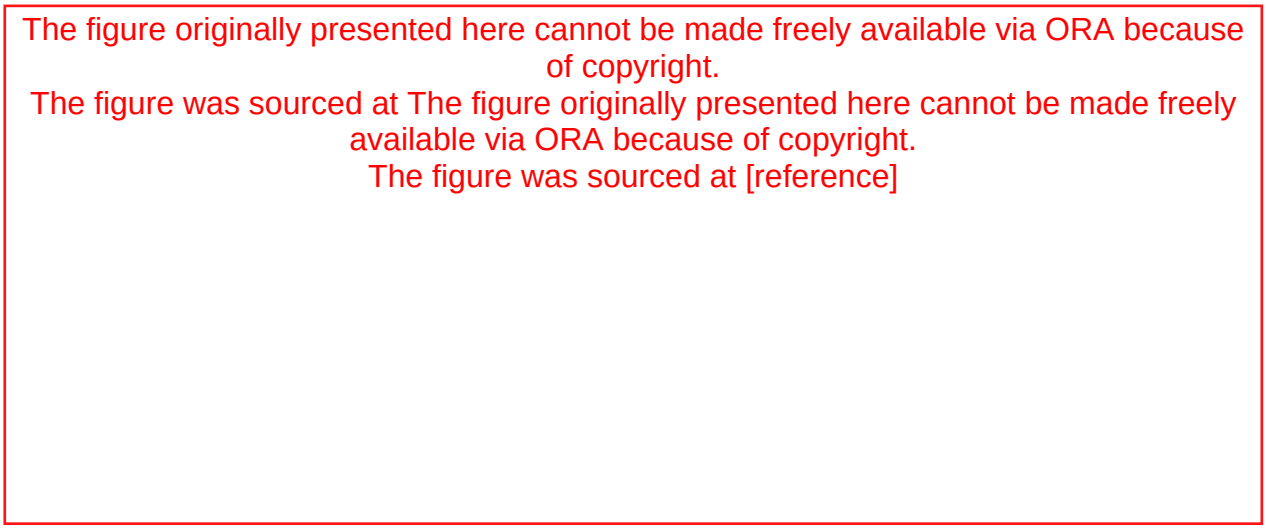
Appendix 7.1. LexTALE instructions



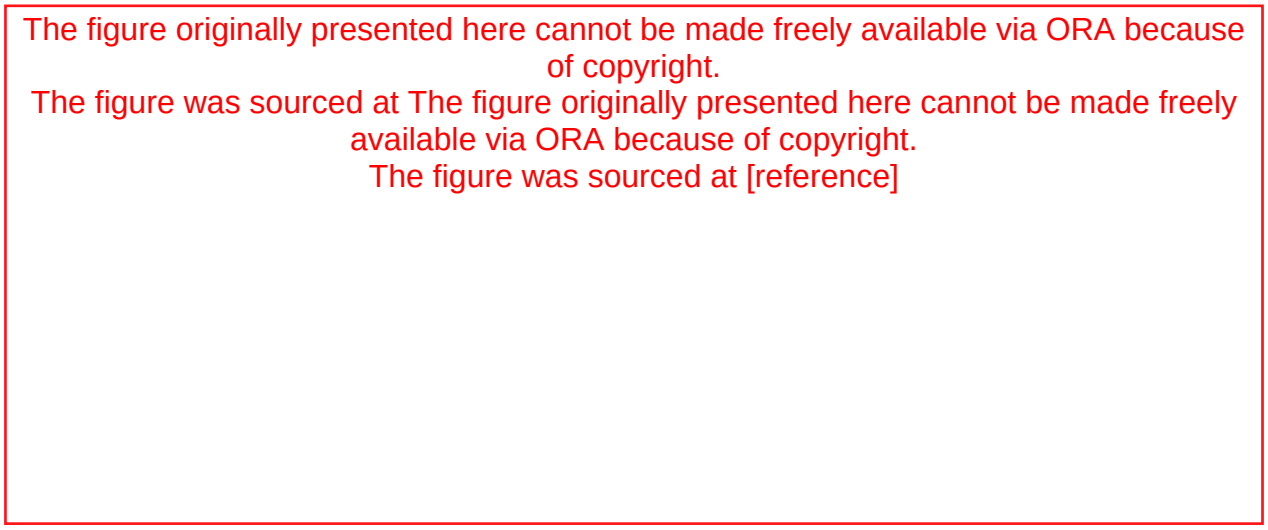
Appendix 7.2. LexTALE task screen



Appendix 7.3. Collocation judgment acceptability test instructions



Appendix 7.4. Collocation judgment acceptability test screen



Appendix 8. Debrief form

Debrief Form

Did you encounter any unfamiliar words while completing the tasks? □Yes □No

Did you experience any distractions while completing the tasks? □Yes □No

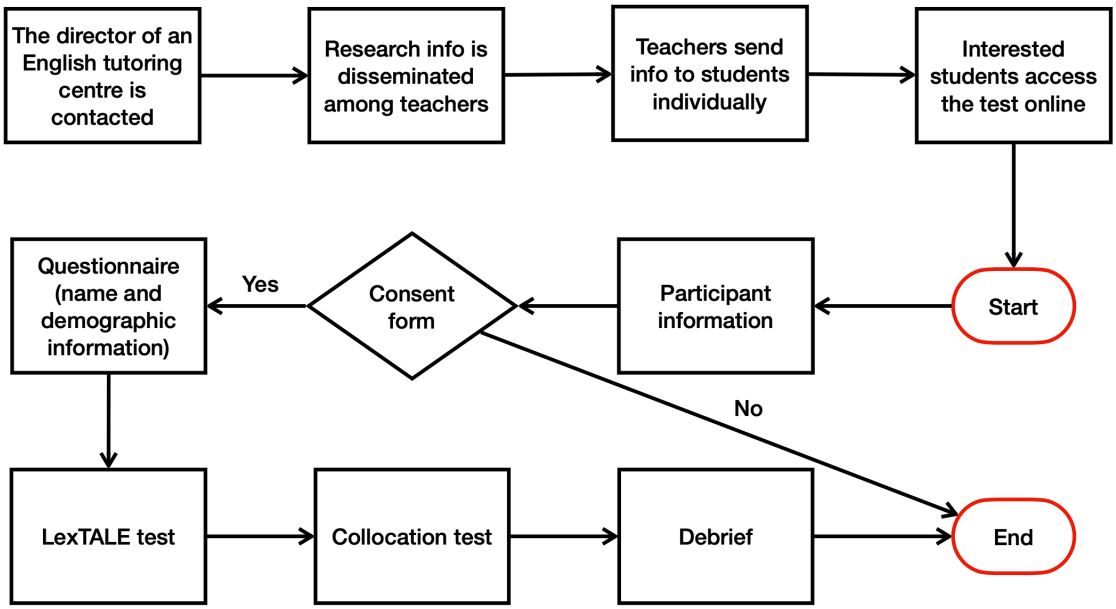
Did you experience any internet connectivity issues while completing the tasks? □Yes □No

Did you experience any delay or failure of response while completing the tasks? □Yes □No

Thank you again for your participation!

If you have a concern about any aspect of this study, please contact XXX, and we will do our best to answer your query.

Appendix 9. Research flowchart



Appendix 10. Collocation items

No.	Item	Type	Freq	Mil	MI
1	play a game	EM	1291	1.31	4.52
2	wear a hat	EM	227	0.23	6.13
3	light a candle	EM	250	0.25	9.10
4	watch a video	EM	308	0.31	3.86
5	sign a contract	EM	409	0.41	6.27
6	pay a price	EM	493	0.50	4.44
7	draw a picture	EM	253	0.26	3.37
8	tell a story	EM	1427	1.45	3.55
9	sing a song	EM	615	0.62	6.71
10	fight a war	EM	332	0.34	4.05
11	hear a voice	EM	199	0.20	4.14
12	solve a problem	EM	616	0.62	6.96
13	drive a car	EM	653	0.66	4.65
14	ride a horse	EM	222	0.23	6.21
15	attend a meeting	EM	151	0.15	5.71
16	leave a message	EM	2046	2.07	3.16
17	press a button	EM	228	0.23	7.74
18	win a race	EM	113	0.11	3.68
19	pass a law	EM	465	0.47	3.65
20	open a window	EM	320	0.32	4.17
21	pick a fight	EO	385	0.39	3.53
22	keep a diary	EO	121	0.12	3.66
23	cut a deal	EO	911	0.92	3.21
24	throw a party	EO	359	0.36	3.14
25	take a pill	EO	273	0.27	3.62
26	place an order	EO	205	0.21	3.55
27	catch a cold	EO	170	0.17	3.11
28	pay a visit	EO	369	0.37	3.89
29	make a difference	EO	8701	8.76	3.21
30	raise a question	EO	105	0.11	4.33
31	deal a blow	EO	109	0.11	5.42
32	cast a vote	EO	294	0.30	6.83
33	draw a conclusion	EO	105	0.11	6.18
34	grab a bite	EO	289	0.29	5.42
35	serve a purpose	EO	217	0.22	4.85
36	file a report	EO	320	0.32	4.82
37	fly a plane	EO	148	0.15	5.99
38	give a speech	EO	525	0.53	3.06
39	lend a hand	EO	482	0.49	3.89
40	wear a beard	EO	140	0.14	4.46
41	eat a drug	MO	0	/	/
42	touch a goal	MO	0	/	/
43	close a radio	MO	0	/	/
44	sit a bus	MO	0	/	/
45	open a computer	MO	0	/	/
46	pull a violin	MO	0	/	/
47	release a heart	MO	0	/	/

48	obtain a profit	MO	0	/	/
49	catch a chance	MO	0	/	/
50	wash a photo	MO	1	/	/
51	pay an effort	MO	0	/	/
52	think a solution	MO	0	/	/
53	grow a habit	MO	0	/	/
54	exercise a skill	MO	1	/	/
55	hit a videogame	MO	0	/	/
56	raise an example	MO	0	/	/
57	return a debt	MO	0	/	/
58	cross a birthday	MO	0	/	/
59	talk a language	MO	0	/	/
60	call a takeout	MO	0	/	/
61	eat a party	BL	0	/	/
62	touch a speech	BL	0	/	/
63	wash a question	BL	0	/	/
64	catch a solution	BL	0	/	/
65	cross a blow	BL	0	/	/
66	pull a difference	BL	0	/	/
67	cut a radio	BL	0	/	/
68	grab a language	BL	0	/	/
69	draw a debt	BL	0	/	/
70	release a bite	BL	0	/	/
71	sit a vote	BL	0	/	/
72	raise a report	BL	0	/	/
73	talk a beard	BL	0	/	/
74	take a habit	BL	0	/	/
75	make a cold	BL	0	/	/
76	fly a profit	BL	0	/	/
77	lend an example	BL	0	/	/
78	exercise a visit	BL	0	/	/
79	wear an effort	BL	0	/	/
80	hit a purpose	BL	0	/	/

Appendix 11. AR and RT of individual items

No.	Item	Type	AR	RT (milliseconds)
1	play a game	EM	0.91	1446.010
2	wear a hat	EM	0.97	1680.954
3	light a candle	EM	0.65	2194.804
4	watch a video	EM	0.97	1293.433
5	sign a contract	EM	0.65	2495.827
6	pay a price	EM	0.59	2088.711
7	draw a picture	EM	0.94	1446.081
8	tell a story	EM	0.88	1198.743
9	sing a song	EM	0.94	1243.705
10	fight a war	EM	0.59	1857.909
11	hear a voice	EM	0.91	1594.981
12	solve a problem	EM	0.97	1113.632
13	drive a car	EM	0.94	1295.078
14	ride a horse	EM	1.00	1656.290
15	attend a meeting	EM	0.79	1811.048
16	leave a message	EM	0.79	1781.666
17	press a button	EM	0.82	1789.396
18	win a race	EM	0.94	1478.382
19	pass a law	EM	0.68	1718.293
20	open a window	EM	0.94	1430.258
21	pick a fight	EO	0.44	2475.357
22	keep a diary	EO	0.76	1765.891
23	cut a deal	EO	0.53	2020.864
24	throw a party	EO	0.24	2105.652
25	take a pill	EO	0.71	1700.546
26	place an order	EO	0.47	2270.203
27	catch a cold	EO	0.65	1794.406
28	pay a visit	EO	0.47	1781.797
29	make a difference	EO	0.94	1447.814
30	raise a question	EO	0.74	1516.363
31	deal a blow	EO	0.09	1971.188
32	cast a vote	EO	0.44	2229.663
33	draw a conclusion	EO	0.74	2207.203
34	grab a bite	EO	0.38	2505.334
35	serve a purpose	EO	0.59	2464.311
36	file a report	EO	0.59	2376.763
37	fly a plane	EO	0.56	2074.614
38	give a speech	EO	0.88	1494.160
39	lend a hand	EO	0.53	2097.339
40	wear a beard	EO	0.35	2282.316
41	eat a drug	MO	0.44	2002.777
42	touch a goal	MO	0.35	1780.606
43	close a radio	MO	0.35	2220.893
44	sit a bus	MO	0.50	1577.896
45	open a computer	MO	0.09	1994.620
46	pull a violin	MO	0.71	2011.636
47	release a heart	MO	0.76	1937.038

48	obtain a profit	MO	0.32	2273.124
49	catch a chance	MO	0.15	2018.096
50	wash a photo	MO	0.53	2080.177
51	pay an effort	MO	0.29	2453.104
52	think a solution	MO	0.29	2214.384
53	grow a habit	MO	0.53	2098.036
54	exercise a skill	MO	0.29	2135.678
55	hit a videogame	MO	0.71	2314.815
56	raise an example	MO	0.32	2117.489
57	return a debt	MO	0.32	2236.209
58	cross a birthday	MO	0.91	1855.699
59	talk a language	MO	0.50	1909.953
60	call a takeout	MO	0.74	2146.282
61	eat a party	BL	0.91	1531.282
62	touch a speech	BL	0.74	1810.914
63	wash a question	BL	0.88	1798.000
64	catch a solution	BL	0.44	2446.548
65	cross a blow	BL	0.35	2585.660
66	pull a difference	BL	0.74	2249.217
67	cut a radio	BL	0.62	1795.924
68	grab a language	BL	0.82	1858.263
69	draw a debt	BL	0.50	2270.325
70	release a bite	BL	0.74	2409.690
71	sit a vote	BL	0.59	1895.341
72	raise a report	BL	0.41	1955.407
73	talk a beard	BL	0.82	2350.001
74	take a habit	BL	0.65	2039.859
75	make a cold	BL	0.82	2035.184
76	fly a profit	BL	0.88	1854.001
77	lend an example	BL	0.56	1764.444
78	exercise a visit	BL	0.76	1860.631
79	wear an effort	BL	0.79	1836.752
80	hit a purpose	BL	0.65	2059.869

Appendix 12. Frequency profile of component words

Frequency Type	Percentage	Words
1k types	80.00	a_[75] an_[5] blow_[2] call_[1] car_[1] catch_[3] chance_[1] close_[1] cold_[2] cross_[2] cut_[2] deal_[2] difference_[2] draw_[3] drive_[1] eat_[2] effort_[2] example_[2] exercise_[2] fight_[2] fly_[2] game_[1] give_[1] grow_[1] hand_[1] hear_[1] heart_[1] horse_[1] keep_[1] language_[2] law_[1] leave_[1] light_[1] make_[2] meeting_[1] open_[2] order_[1] party_[2] pass_[1] pay_[3] picture_[1] place_[1] play_[1] press_[1] price_[1] problem_[1] profit_[2] pull_[2] purpose_[2] question_[2] race_[1] raise_[3] report_[2] return_[1] ride_[1] serve_[1] sign_[1] sing_[1] sit_[2] song_[1] speech_[2] story_[1] take_[2] talk_[2] tell_[1] think_[1] throw_[1] touch_[2] visit_[2] voice_[1] vote_[2] war_[1] watch_[1] wear_[3] win_[1] window_[1]
2k types	11.67	attend_[1] beard_[2] birthday_[1] bite_[2] bus_[1] button_[1] debt_[2] habit_[2] hat_[1] hit_[2] lend_[2] message_[1] pick_[1] plane_[1] radio_[2] skill_[1] solution_[2] solve_[1] wash_[2]
Academic Word List (AWL) types	3.33	computer_[1] conclusion_[1] contract_[1] file_[1] goal_[1] obtain_[1] release_[2]
OFF types	5.00	candle_[1] cast_[1] diary_[1] drug_[1] grab_[2] photo_[1] pill_[1] takeout_[1] video_[1] videogame_[1] violin_[1]

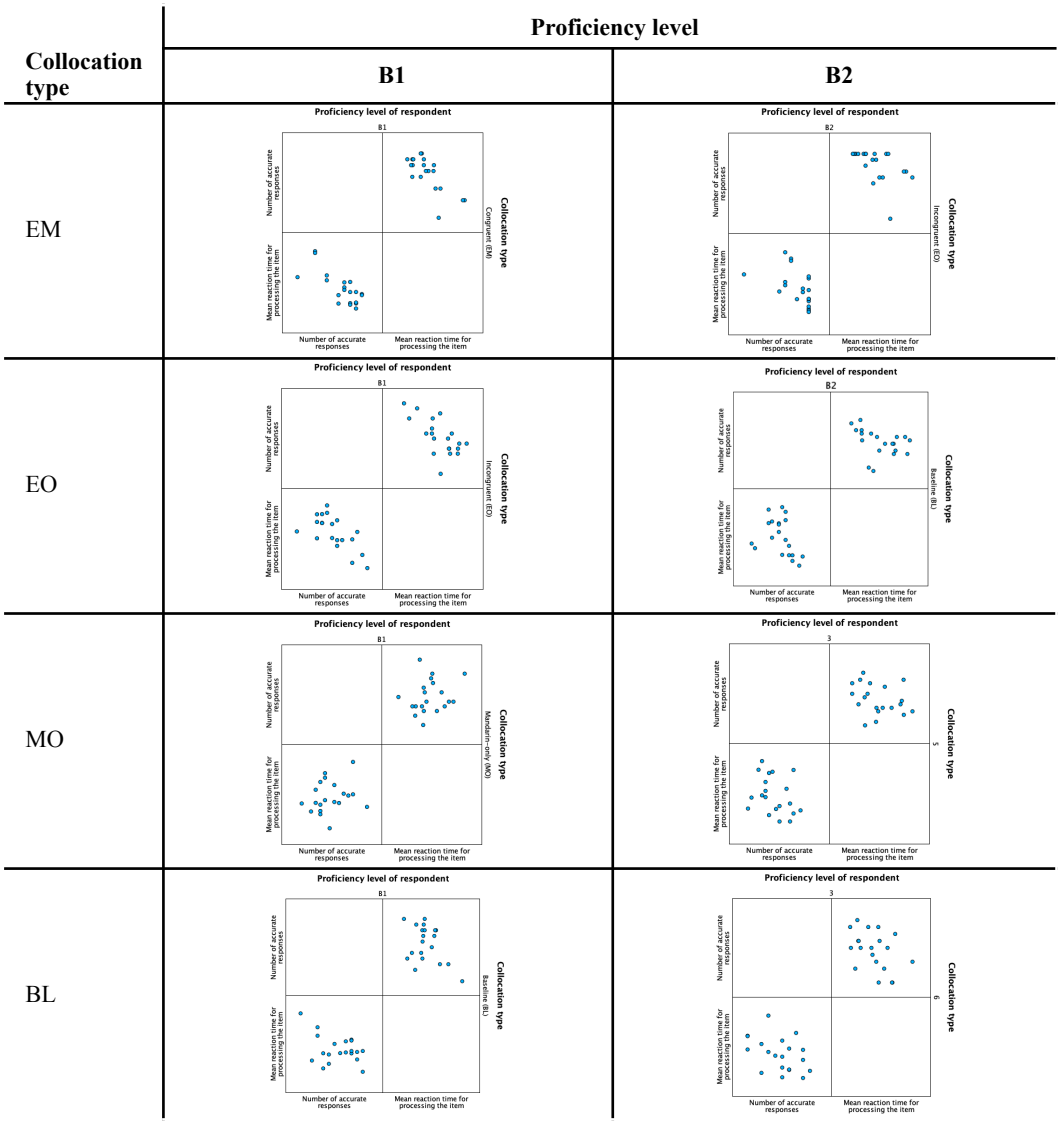
Appendix 13. MANOVA assumptions

Homogeneity of covariance matrices

Box's M	30.799
F	1.404
df1	21
df2	82814.338
Sig.	0.103

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

Linearity scatterplot of dependent variables



Multicollinearity

Collocation type * proficiency level	Dependent variables		Accuracy rate	Reaction time
EM-B1	**Accuracy rate**	Pearson Correlation	1	-.744**
		Sig. (2-tailed)		<.001
		N	20	20
	Reaction time	Pearson Correlation	-.744**	1.00
		Sig. (2-tailed)	<.001	
		N	20	20
EM-B2	**Accuracy rate**	Pearson Correlation	1	-.507*
		Sig. (2-tailed)		0.022
		N	20	20
	Reaction time	Pearson Correlation	-.507*	1.00
		Sig. (2-tailed)	0.022	
		N	20	20
EO-B1	**Accuracy rate**	Pearson Correlation	1	-.665**
		Sig. (2-tailed)		0.001
		N	20	20
	Reaction time	Pearson Correlation	-.665**	1.00
		Sig. (2-tailed)	0.001	
		N	20	20
EO-B2	**Accuracy rate**	Pearson Correlation	1	-0.367
		Sig. (2-tailed)		0.112
		N	20	20
	Reaction time	Pearson Correlation	-0.367	1.00
		Sig. (2-tailed)	0.112	
		N	20	20
MO-B1	**Accuracy rate**	Pearson Correlation	1	0.213
		Sig. (2-tailed)		0.367
		N	20	20
	Reaction time	Pearson Correlation	0.213	1.00
		Sig. (2-tailed)	0.367	
		N	20	20
MO-B2	**Accuracy rate**	Pearson Correlation	1	-0.295
		Sig. (2-tailed)		0.207
		N	20	20
	Reaction time	Pearson Correlation	-0.295	1.00
		Sig. (2-tailed)	0.207	
		N	20	20
BL-B1	**Accuracy rate**	Pearson Correlation	1	-0.400
		Sig. (2-tailed)		0.081
		N	20	20
	Reaction time	Pearson Correlation	-0.400	1.00
		Sig. (2-tailed)	0.081	
		N	20	20
BL-B2	**Accuracy rate**	Pearson Correlation	1	-0.430
		Sig. (2-tailed)		0.059
		N	20	20
	Reaction time	Pearson Correlation	-0.430	1.00
		Sig. (2-tailed)	0.059	
		N	20	20