The Relationship between Word Knowledge Components and L2 Writing in Chinese University Learners of English

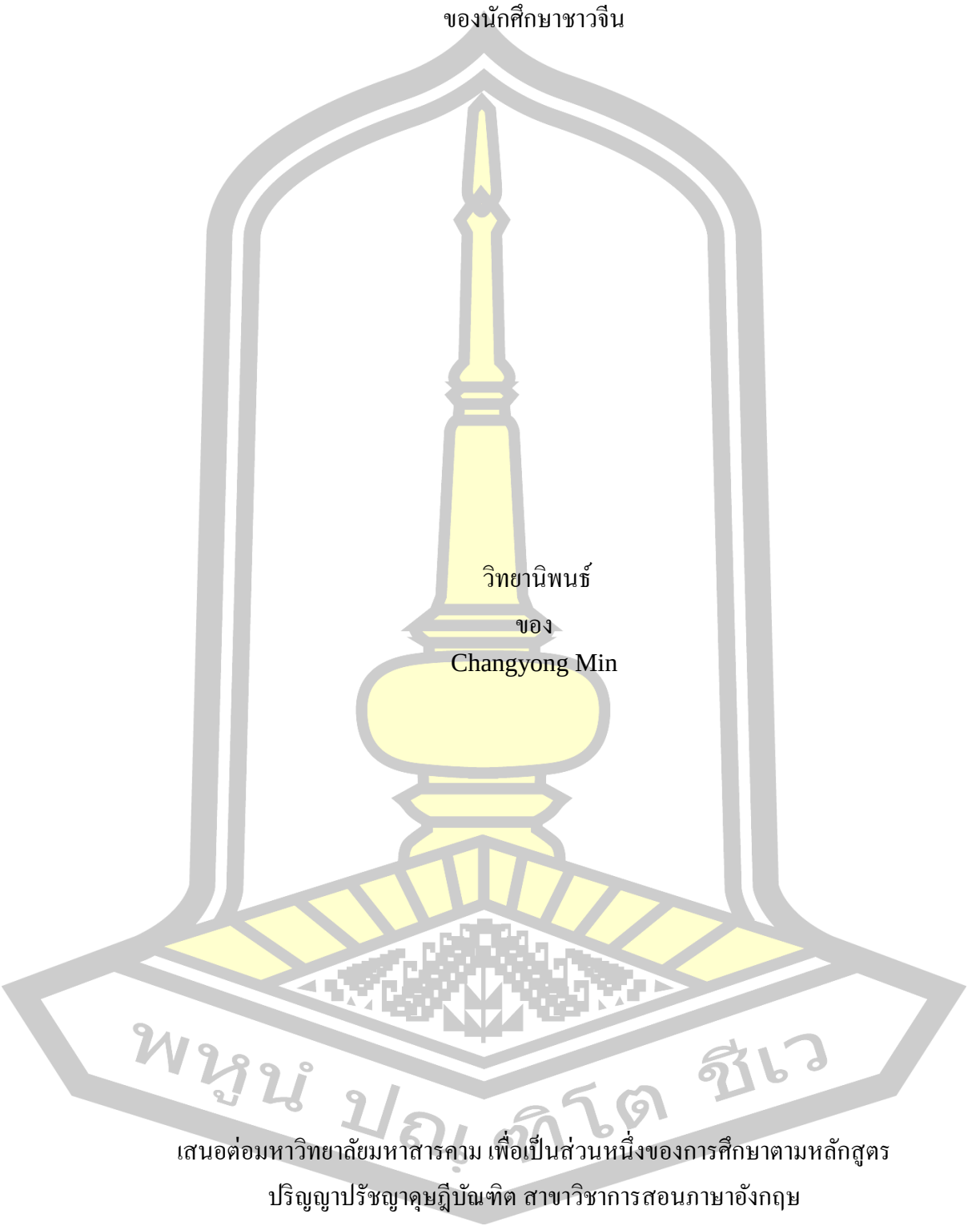
Changyong Min

A Thesis Submitted in Partial Fulfillment of Requirements for

degree of Doctor of Philosophy in English Language Teaching

April 2024

ความสัมพันธ์ระหว่างองค์ประกอบความรู้คำศัพท์และการเขียนภาษาที่สองของผู้เรียนภาษาอังกฤษของนักศึกษาชาวจีน



วิทยานิพนธ์

ของ

Changyong Min

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาการสอนภาษาอังกฤษ

เมษายน 2567

The Relationship between Word Knowledge Components and L2 Writing in Chinese University Learners of English

Changyong Min

A Thesis Submitted in Partial Fulfillment of Requirements

for Doctor of Philosophy (English Language Teaching)

April 2024

The examining committee has unanimously approved this Thesis, submitted by Mr. Changyong Min , as a partial fulfillment of the requirements for the Doctor of Philosophy English Language Teaching at Mahasarakham University

Examining Committee

.....Chairman
(Assoc. Prof. Supakorn Phoocharoensil , Ph.D.)

.....Advisor
(Asst. Prof. Apisak Sukying , Ph.D.)

.....Committee
(Pilanut Phusawisot , Ph.D.)

.....Committee
(Eric A. Ambele , Ph.D.)

.....External Committee
(Prof. Jessie S. Barrot , Ph.D.)

Mahasarakham University has granted approval to accept this Thesis as a partial fulfillment of the requirements for the Doctor of Philosophy English Language Teaching

.....
(Assoc. Prof. Nittaya Wannakit , Ph.D.)
Dean of The Faculty of Humanities and Social Sciences

.....
(Assoc. Prof. Krit Chaimoon , Ph.D.)
Dean of Graduate School

TITLE The Relationship between Word Knowledge Components and L2 Writing in Chinese University Learners of English

AUTHOR Changyong Min

ADVISORS Assistant Professor Apisak Sukying , Ph.D.

DEGREE Doctor of Philosophy **MAJOR** English Language Teaching

UNIVERSITY Mahasarakham University **YEAR** 2024

ABSTRACT

Previous vocabulary research has explored the multi-component nature of a word and provided comprehensive frameworks on what types of word knowledge are needed to learn a word and use the word in writing. Yet, previous studies primarily focused on form and meaning aspects and the nature of multiple word knowledge components associated with L2 writing proficiency remains unclear. The current study, therefore, examined the interrelations between multiple word knowledge components and explored the relationships between these word components and productive word use and L2 writing ability. The study adopted a multi-task approach guided by the word knowledge framework proposed by Nation (2022) and the model of word knowledge components required in writing proposed by Coxhead (2007).

Participants were 147 third-year university EFL learners sampled from a regional university in Mainland China. The study adopted a cross-sectional setting, in which participants did six vocabulary tests, including one receptive vocabulary size test and five productive depth knowledge tests (productive form and meaning, association, productive derivative and collocation). The participants were also assigned to write two argumentative essays and integrate the twenty target words in their writing tasks. These measures were designed and validated to capture multiple word knowledge components relative to lexical use and argumentative writing ability. The principle objective of these tests is to investigate to what extent these different types of word knowledge can relate and contribute to L2 argumentative writing quality and word use in actual contexts.

The correlation and regression results confirmed that vocabulary knowledge can be a network of multiple interrelated word components. The interrelatedness, however, is only limited to the five internal depth knowledge aspects of the same target words. The depth components and overall depth knowledge merely have little relationship with learners' receptive vocabulary size. This result cautions that depth and size are two distinct and separate constructs whose relationship may vary in different contexts. Overall, the hierarchical regression models demonstrate that vocabulary knowledge composed of the six receptive and productive word aspects accounts for 61% variance explained in L2 writing proficiency. This finding strongly

supports previous literature documenting that vocabulary can be one the most essential factors in determining L2 writing quality.

Moreover, the findings in the current study also indicate that productive form- meaning links are mainly related to L2 writing ability and word use, as opposed to the receptive form-meaning knowledge. The receptive size measured by the VLT has weak correlations with L2 writing and cannot significantly predict the writing and lexical proficiency. Association and collocation contribute little to the productive performances, yet still highly correlate with them and significantly improve the regression models in L2 writing. These results have theoretical and pedagogical implications, empirically demonstrating that multi-component word tests can be applied to reveal the multidimensional construct of vocabulary knowledge. The current study provides empirical evidence to the theoretical word knowledge models and yields nuanced ideas regarding the smallest lexical predictors of L2 writing. Pedagogically, receptive form-meaning links are insufficient in EFL vocabulary learning and teaching. Learners and instructors need to focus more on the deeper components of vocabulary knowledge if the learning goal is to improve productive written skills.

Keyword : Chinese EFL university learners, L2 writing ability, multiple word tests, word knowledge components, word size and depth

ACKNOWLEDGEMENTS

This thesis project would not have been possible without the support and help of many people during the whole academic journey of these years. Looking back, the many details in my memory always remind me of not only the academic advancement I have achieved but, more importantly, the gratitude I should forever feel deeply in my heart.

I am most deeply indebted to my supervisor, Dr. Apisak Sukying, for the tremendous supervision he provided me all the way forward, the constant support he gave me both in life and academic study, the inspiring encouragement he offered me whenever I was in trouble, and above all, the trust he places in me to grow into an independent researcher. Dr. Apisak led my way in academia, and his academic enthusiasm kindled my research interest. He sowed the academic seed in me who otherwise would have never engaged in any academic work. My academic world has become bigger and more colourful under his guidance and mentoring. I would also like to express my heartfelt thanks to Dr. Pilanut Phusawisot and Dr. Eric Ambele for their informative courses and fundamental insights into methodologies and research trends. Dr. Pilanut also assisted me in preparing many documents, which are critical to me and make my PhD journey easier. Sincere thanks also go to all the staff and teachers in the English Language Teaching programme and the university dormitory administration for their assistance in any respect.

I would hereby acknowledge the cooperation and support provided by all the participants in the current study, my lovely and hard-working students in the 2020 and 2021 academic years. Without their patience in doing the multiple tests and providing constructive suggestions, the current thesis project would be impossible. In addition, I would like to express my special thanks to my colleague and friend, Ms Yali Yu, who used her expertise in L2 writing to offer me considerable assistance in the inter-rating work without any repayment.

I owe special thanks to the ELT programme and Mahasarakham University for the best scholarship (Postmaster Scholarship) they granted me. The financial support helped me so much in paying tuition fees, publishing papers, and attending international conferences.

To my friends, Thai and Chinese alike, Panassanan Kitichaidateanan, Jiraporn Noipa, Zheng Yuan, Yuxian Hou, Dan Li, Qian Zhao, Qi Liu and Yali Yu, I deeply thank

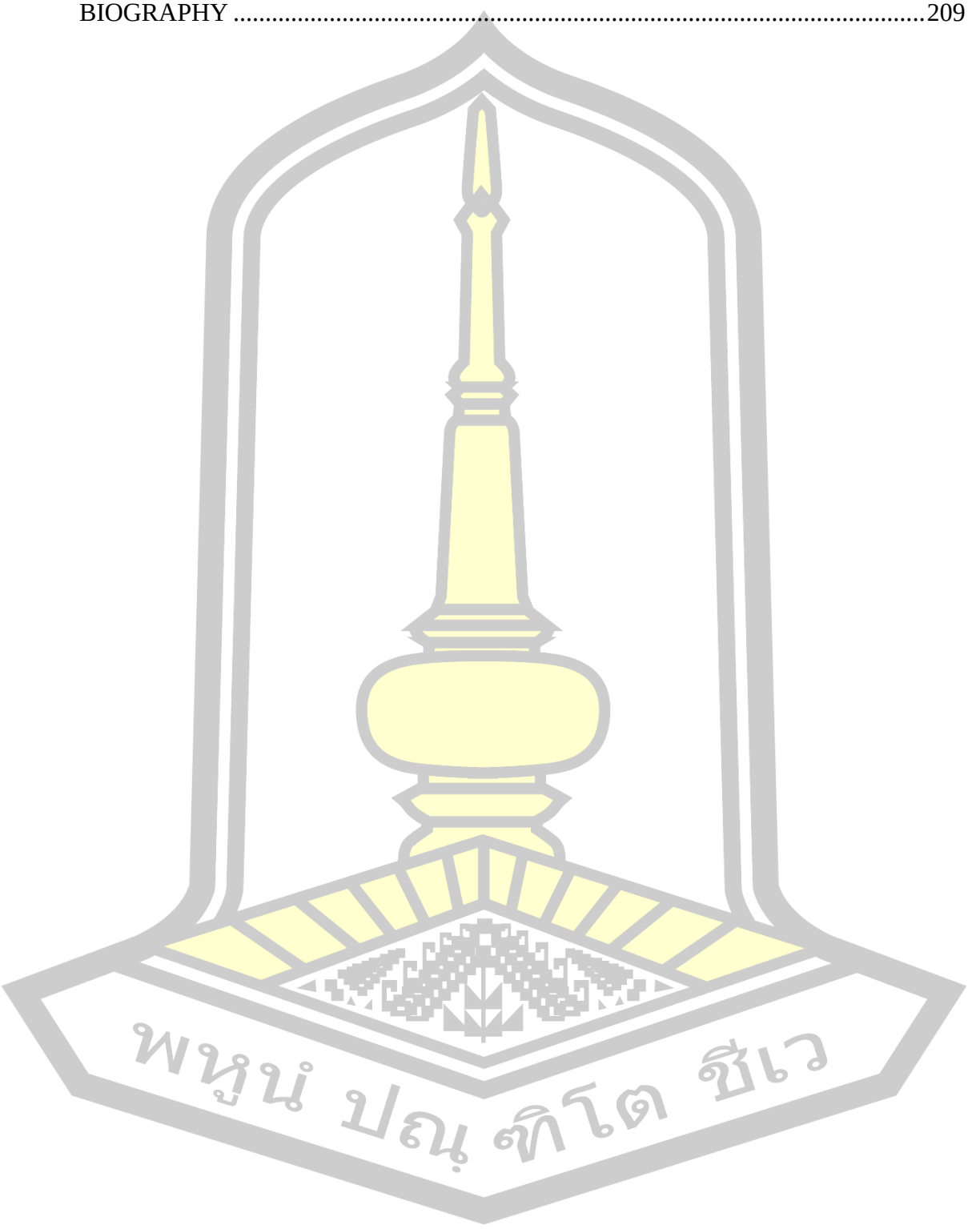
them for the inspiring talk and kind support and encouragement. P Mui, P Ao, and other Thai friends helped me complete much of the work that I could not do myself during the Covid-19. The warm welcome and kindness they extended made me feel at home in Thailand, which will be part of my most beautiful memory. The talk with Zheng Yuan about our research interest and statistics methods largely stimulated my desire to do better in my thesis writing. Qian Zhao, Qi Liu and Yali Yu carefully helped me check the test instruments and offered their valuable suggestions. In addition, the Chinese friends in Parichart Dormitory enriched my life and enlivened my days in Thailand because of their presence and mutual support.

Finally, and most importantly, I want to thank my parents and my family for all kinds of support they provided me in the course of my studies. My parents have always had faith in me and care about my PhD study and the exotic life in Thailand. They love me in their own ways and allow me to be as ambitious as I can to pursue my academic achievement and fulfill my dream. They have been doing their utmost to provide me with financial and spiritual support during these years, which have been a major drive for me to move forward. My gratitude for their unconditional love and support is beyond words to express.

Changyong Min

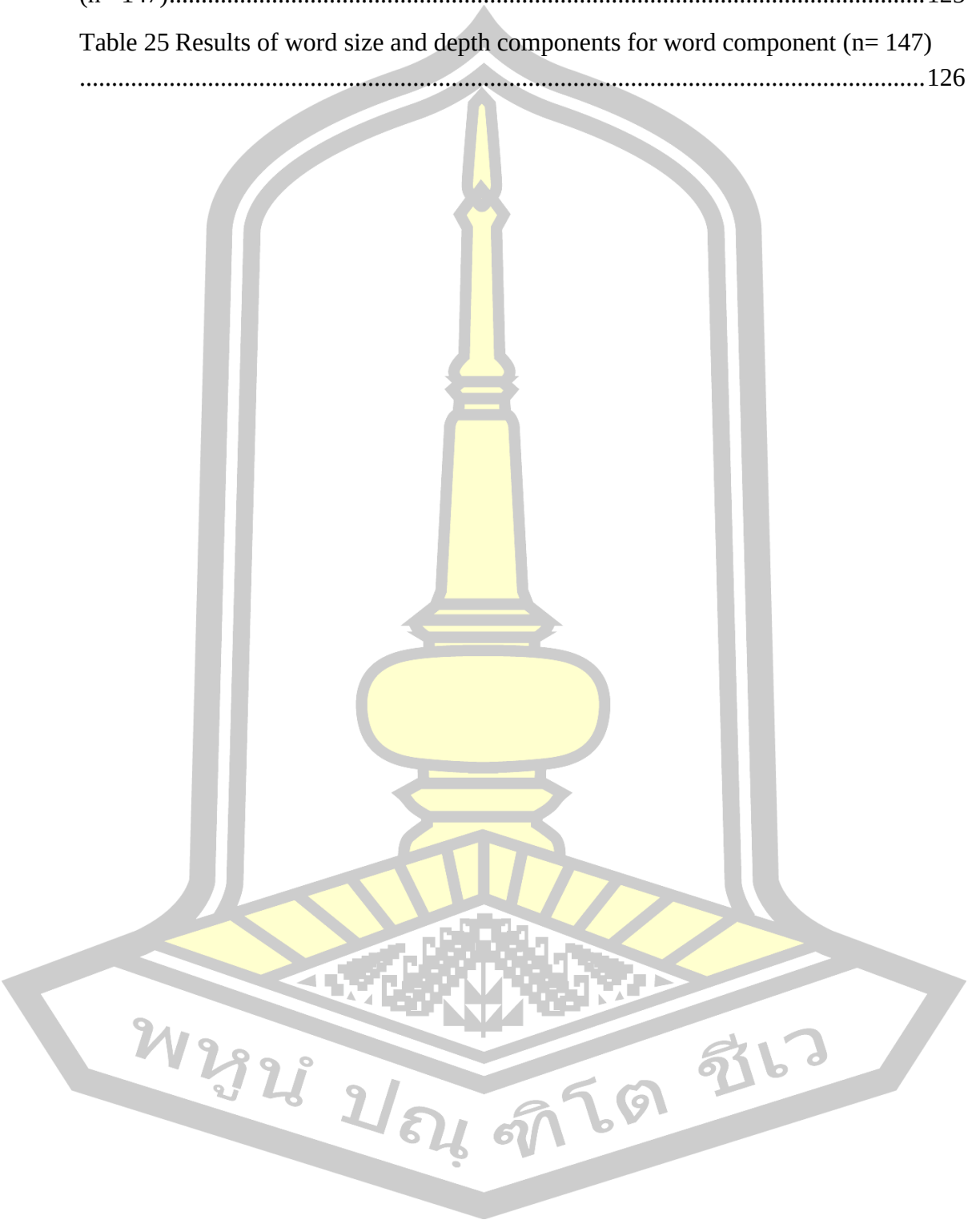
TABLE OF CONTENTS

Page



LIST OF TABLES

Page



LIST OF FIGURES

Page

CHAPTER I
INTRODUCTION

This chapter provides a brief introduction to the study area into which the current study ventures, including the background of the study and the problem statement. It also introduces the aims and significance of the study and raises two research questions based on the review of previous literature. This chapter ends with an explanation of the key terms involved in the current study.

1.1 Background of the Study

Vocabulary knowledge in different dimensions plays an indispensable role in lexical and language proficiency (Coxhead, 2021; Nation, 2022; Qian & Lin, 2020; Schmitt, 2010). This knowledge constitutes the rudimentary units of English acquisition in the English as a Foreign Language (EFL) context as words are the "basic building blocks for language use and development" (Albrechtsen, Henriksen, & Haastrup, 2008, p.22). Vocabulary knowledge allows EFL learners to create sentences, paragraphs and the whole texts, such that some researchers even argue that lexical learning represents English language learning (Gass & Selinker, 2008; Schmitt, 2008), and some EFL learners also regard L2 acquisition as a matter of learning vocabulary (Read, 2000). It is true that sufficient word knowledge precedes effective language use, and empirical evidence supports the close relationship between vocabulary size and reading comprehension (Coxhead, Nation & Sim, 2015; Laufer, 1992; Qian, 1999, 2002; Zhang & Yang, 2017), vocabulary knowledge and listening ability (Wang & Treffers-Daller, 2017), productive word knowledge and oral proficiency (Uchihara & Saito, 2016), and vocabulary size and depth and writing performance (Albrechtsen et al., 2008; Qian & Lin, 2020). It can be conclusive that vocabulary knowledge increases the possibility of EFL learners becoming more proficient in a variety of language skills (Milton & Fitzpatrick, 2014; Stæhr, 2008, 2009).

Vocabulary knowledge can be defined as the knowledge of multiple word components (Coxhead, 2007; González-Fernández, 2022; Laufer et al., 2004; Nation, 2001, 2022; Sukying, 2018). These components are manifold and involve multi-faceted dimensions, embracing, among other components, word form, meaning, collocation,

frequency of use, and association (González-Fernández & Schmitt, 2020; Nation, 2022; Nagy & Scott, 2000; Milton & Fitzpatrick, 2014; Richards, 1976). To date, Nation (2022) has tabulated the most comprehensive framework of word knowledge components, mainly featuring word form, meaning, and constraints of use in receptive and productive dimensions. This theoretical model has provided a profound theory for a considerable volume of studies conducted from a multi-component perspective. Schmitt and Meara (1997) and Schmitt (1998) made the first attempts to break down the complex vocabulary knowledge construct into smaller components and yielded a fine-grained description of this knowledge.

The component approach to vocabulary knowledge has engaged research attention to unravel the nature of acquiring and using a word. For example, research has parsed the grammatical knowledge of word parts, i.e., inflectional and derivative affix knowledge (Chui, 2006; Schmitt & Meara, 1997; Sukying, 2017, 2018, 2022) and receptive and productive mastery of multiple word components (Nontasee & Sukying, 2021; Schmitt, 1998; Webb, 2005, 2007a). These studies found that EFL learners have uneven knowledge of different components, and the acquisition rate of these components differs markedly. This attracts further research attention to explicate the acquisition order of the various word components in contexts (Daskalovska, 2015; Gonzalez-Fernandez & Schmitt, 2020; Sukying, 2022).

Moreover, many studies have revealed that word components are interrelated, particularly between vocabulary size (form and meaning links) and multiple depth components (other components such as collocation and association) (Chui, 2006; Chen & Truscott, 2010; Gonzalez-Fernandez & Schmitt, 2020; Koizumi & In'nami, 2020). However, the most research attention has been paid to the interrelationships between word components and reading comprehension. Some studies examined the effects of different word components on reading proficiency (Kieffer & Lesaux, 2012; Laurence et al., 2018; Li & Kirby, 2015; McLean et al., 2020; Qian, 2002) and generally found that form-meaning connections correlate best with reading ability; many other researchers have investigated the instructional effects of reading practice on word components acquisition (Coxhead & Byrd, 2007; Lin & Hirsh, 2012; Lee & Muncie, 2006; Pichette et al., 2012; Webb, 2007b, 2009).

Since knowledge of different word components contributes to the performance and use of a word in language skills (Coxhead, 2007; Meara, 1996; Zhong, 2016), multiple tasks might serve to capture each component of vocabulary knowledge (Webb, 2005, 2007; Zareva, 2005). This may help to extract detailed information about word knowledge in language use, from which we can make inferences to reveal the mastery level of these word components. Coxhead (2007) drew on Nation's (2001, 2022) component approach to determine which components are necessary to reach the appropriate depth level for writing purposes. The findings suggested that beyond the basic form and meaning knowledge, many components, such as collocation, grammar, and register, are of particular essence in the writing process. It has been documented that vocabulary knowledge in different dimensions impacts the quality of L2 writing (Brun-Mercer & Zimmerman, 2015; Coxhead, 2007, 2012; Johnson, 2016; Laufer & Nation, 1995; Paquot, 2010; Santos, 1988). A growing number of studies have focused on the relationships between various word knowledge components and L2 writing proficiency within the multi-component framework (Choi, 2017; Crossley et al., 2011, 2015; Dabbagh & Janebi Enayat, 2019; Kilic, 2019; Laufer, 1994, 1998; Suyking, 2023; Olinghouse & Leaird, 2009; Olinghouse & Wilson, 2013; Wu, Dixon, Sun, & Zhang, 2019; Zhong, 2016).

Albrechtsen et al. (2008) found that the receptive size of vocabulary and associative depth knowledge can predict the scores of L2 writing. Zhong (2016) and Wu et al. (2019) produced similar results, indicating that receptive form and meaning connections are more critical than other word components. This is consistent with Coxhead (2007) that receptive size knowledge was reported as the most significant lexical element in writing. However, further research has shown the centrality of in-depth word components in predicting L2 writing scores, such as academic register (Coxhead, 2012; Brun-Mercer & Zimmerman, 2015), collocation accuracy (Choi, 2017; Crossley et al., 2015), productive L1-L2 word pairs and adjective synonyms (Wu et al., 2019) and free productive knowledge (Kilic, 2019; Shi & Qian, 2012). Thus, there is a growing need to combine word-specific (depth components) knowledge and word-general (receptive size) knowledge to perceive individual and group differences of vocabulary knowledge in language production (Kieffer & Lesaux, 2012).

Multi-component research has also inquired into the relationship between different features of lexical richness and L2 writing quality. For example, empirical studies have shown high relevance between L2 writing and lexical diversity (Olinghouse & Leaird, 2009; Yu, 2009), lexical sophistication (Crossley, 2020; Ha, 2019) and low-frequency words (Dabbagh & Janebi Enayat, 2019; Johnson et al., 2016; Olinghouse & Wilson, 2013). Accordingly, researchers have developed various measures to assess lexical richness in L2 writing. Laufer and Nation (1995, 1999) identified the flaws inherent in previous productive vocabulary tests and proposed the LFP (Lexical Frequency Profile) to reflect vocabulary size in EFL essays. Fitzpatrick (2007) and Fitzpatrick and Clenton (2010, 2017) focused on the assessment of associative vocabulary knowledge and developed measures to capture this knowledge. In addition, Meara and Bell (2001) created the *P_Lex*, highlighting low-proficient L2 learners' productive knowledge. Taken together, it has become clear that all word components, beyond word form and meaning, can contribute to productive word use and L2 writing performance to some degree (Coxhead, 2007; Zhong, 2016). L2 writers reported that they have to mobilize knowledge of various lexical components during the writing process. For example, they would consider academic register in their lexical choice so that the word they used can reveal their membership in the academic community (Coxhead, 2012). This provides a solid theoretical foundation for the current study probing the relationship between different word components and L2 writing proficiency.

1.2 Problem Statement

Although Nation (2022) and Coxhead (2007) have provided comprehensive models of multiple word components that navigate the way towards a sound understanding of vocabulary knowledge, they have yet to touch upon the interrelationships between these knowledge components (Gonzalez-Fernandez & Schmitt, 2020). Nation (2020) pointed out that his convenient word knowledge model simply describes a set of hypothesized vocabulary features, which do not serve to explain how these different types of knowledge relate to one another and how they behave in actual language use. Consequently, the correlations between various word components remain unclear. This raises questions about which components contribute most to the word knowledge construct and the extent each component (e.g., collocation, derivative) can assist other

component knowledge or foster L2 skills, especially speaking and writing. Despite the ample evidence that vocabulary knowledge is a multidimensional construct (Coxhead, 2007; Kieffer & Lesaux, 2012; Nation, 2022; Gonzalez-Fernandez, 2022), and that multiple components of word knowledge need to be integrated to receptively know the word and productively use it, the bulk of research focused on word components in a piecemeal manner. Vocabulary size or word form and meaning knowledge have drawn the most attention (Laufer, 1998; Laufer & Nation, 1995; Webb, 2008; Webb & Chang, 2012). This is especially evident in studies conducted with Chinese EFL learners. China's EFL vocabulary research primarily focuses on practical methods of vocabulary acquisition, mainly to enlarge vocabulary size (Huang & Shu, 2020). Jiang (2002) demonstrated that Chinese EFL learners tend to achieve higher rating scores and faster reaction times towards L2 words that carry the same semantic content in their L1 language translations. This may explain their attention paid to form and meaning associations and their inability to use these words appropriately in productive situations (Gan, Humphreys & Hamp-Lyons, 2004).

However, vocabulary knowledge is more than knowledge of the form-meaning connections (i.e., vocabulary size); it also requires the acquisition of multiple shades of meaning, semantic associations, and grammatical functions in contexts (i.e., vocabulary depth) (Kieffer & Lesaux, 2012; Read & Dang, 2022; Schmitt, 2010). This construct is so rich and complicated that any single measure merely "gives a minimal impression of the overall lexical knowledge constellation" (Schmitt, 2010, p.166). Kremmel and Schmitt (2016) also stressed that "knowing" a word necessitates a more profound and deeper knowledge, such as derivatives and collocations, in order to use the word well in one or more language skills. Therefore, size alone cannot extract much information about the interwoven relationships between word components. Instead, multiple measures need to be in place to assess other depth components as a whole so that we can paint a clearer picture of the overall state of EFL learners' word knowledge (Cheng & Matthews, 2018; Lemmouh, 2010; Webb, 2013).

The focus on vocabulary form-meaning links is also true in studies on the relationship between word knowledge components and L2 writing proficiency. A large number of studies devoted attention to the lexical richness in L2 writing by measuring lexical

features such as lexical diversity, frequency and sophistication (Choi, 2017; Crossley, 2020; Enbger, 1995; Ha, 2019; Laufer & Nation, 1995). However, these lexical features are more closely related to vocabulary size and are seldom associated with depth knowledge components (Crossley et al., 2011, 2015). In other words, lexical richness in L2 writing mainly captures how many words learners know, not how deep learners know a word. The limited number of previous studies that have attempted to probe the relationship between word depth knowledge and L2 language skills have used instruments that capture a relatively constrained range of vocabulary knowledge components. For instance, previous studies have focused on lexical register and collocation (Brun-Mercer & Zimmerman, 2015; Coxhead, 2012; Crossley et al., 2015) or lexical errors in L2 writing (Llach, 2005, 2011). Some studies measured receptive aspects only (Zhong, 2016) or merely captured one or two depth components (Bestgen, 2017; Leontjev et al., 2016; Sukying, 2023). Therefore, there is little empirical data on the nature of the multiple components and how the relationships between these components influence vocabulary knowledge and language skills.

This paucity of research in this area has led to doubt regarding the components of vocabulary knowledge that should be taught and measured by instructors and researchers to foster EFL learners' lexical ability. A sound understanding of the interplay between different components of learners' word knowledge is needed because the learning outcomes of L2 acquisition are characterized by variability (Ellis, 1995). That is, learners show an uneven knowledge of various word components, particularly in L2 writing (Weigle, 2002, 2007), and some might have inadequate knowledge of certain components even at advanced levels (Levitzky-Aviad & Laufer, 2013; Nontasee & Sukying, 2021). As such, research suggests that separate tests should be devised to measure different components of vocabulary knowledge (Read & Dang, 2022; Schmitt, 2010; Webb; 2013). To this end, additional research using multiple measures of vocabulary knowledge is necessary (Cheng & Matthews, 2018; Sukying, 2018, 2022) to better understand how these word components are interrelated and the extent the individual can be predictive of L2 writing.

1.3 The Current Study

1.3.1 The Aim of the Study

The aim of the current study is to investigate the internal nature of vocabulary knowledge and the relationships between size and depth word knowledge and L2 writing ability. Two hypotheses lay the foundation of the current study: a) vocabulary knowledge is a multidimensional construct with interrelated word size and depth knowledge components in different dimensions; b) receptive word size and productive depth components are related to L2 writing and lexical use ability.

The complexity of the vocabulary knowledge construct suggests that examining the multiple aspects of the same target words is necessary and feasible if the research aim is to have a comprehensive understanding of learner word knowledge and use. It is known to date that receptive and productive form-meaning links represent learners' vocabulary size. Learners largely rely on mapping form and meaning at the early stage of learning and using a word (Coxhead, 2007; Elgort, 2011; Jiang, 2002; Zhong, 2014). As vocabulary knowledge develops, more word depth components would be included in learners' actual use (McLean et al., 2020; Zareva, 2005). Recent studies (González-Fernández, 2022; González-Fernández & Schmitt, 2020) have cautioned that vocabulary knowledge may be more of a unidimensional construct, given the strong correlations among different word components. Yet, the current study suggests that the multidimensional conceptualization of vocabulary knowledge is valuable, if the aim is to understand how these word components behave and contribute to writing proficiency. To our knowledge, the unidimensionality of vocabulary knowledge can hardly be empirically demonstrated or employed in vocabulary research in actual word use. As such, a multi-task approach to vocabulary knowledge is needed for the current study to reach its research aim.

Schmitt (2010) pointed out three difficulties inherent in multi-component tests: a) measuring all word knowledge components is unfeasible; b) the measurement would be cumbersome due to the considerable time and effort needed; c) the cross-test effects may be unavoidable because of the interactions of different word knowledge aspects. In order to achieve the aim of exploring the nature of vocabulary components in actual contexts, the current study devised a battery of tests guided by Nation's (2022)

comprehensive framework of word components and Coxhead's (2007) model of lexical components required in L2 writing. The multi-component tests were based on the existing research on the relationship between different measures of vocabulary knowledge and L2 writing ability. Six word size and depth components that underpin the word knowledge construct were examined. These tests explored the extent to which word components in different formats are differentially associated with and predictive of L2 writing quality in Chinese university learners of English. In addition, the current study also carefully examined the construct measured by each instrument and the arrangement of test administration to ensure the practicality and feasibility of the vocabulary and writing tests.

1.3.2 Research Questions

The current study explored the relationships between the various word components and the extent to which they are related to L2 writing and lexical ability. The following two questions will guide the study:

1. What are the correlations among the six word components, namely, vocabulary size, word pair, form recall, association, productive derivative and collocation, and their relationship with L2 writing and word use?
2. To what extent do these discrete word knowledge components contribute to university EFL learners' word use and overall L2 writing ability?

The first question attempts to determine how the word components interrelate and how well they relate to the target word use and L2 writing proficiency. The second question examines to what extent each of the six components of word knowledge contributes to the essay writing quality.

1.3.3 Scope of the Study

The current study is a cross-sectional quantitative study relative to English language teaching and vocabulary acquisition. Because the study focuses on multi-component vocabulary knowledge and L2 writing, a range of contextual variables should be controlled for a practical and feasible research scope. Thus, participants' differences such as L1 proficiency, gender, cultural background, learning context and other cognitive and psycholinguistic elements in writing were excluded from the current study. As such, a homogeneous group of Chinese university EFL learners were

selected, whose English proficiency were at an intermediate level. Noted, however, that the research setting was limited to a second-class regional university in Mainland China, where the 147 participants were sampled. Twenty academic words at different frequency levels were selected from the Academic Word List (AWL, Coxhead, 2000) and the L2 writing tests were solely limited to argumentative writing tasks with similar topics. In addition, the word knowledge components included receptive and productive types of knowledge, yet productive formats were the focus of vocabulary tests. Only overall word size and association were in receptive formats. Future studies may extend to a broader research paradigm and include a wider range of variables and contexts.

Moreover, only quantitative methods were used to answer the two research questions in the current study. Pearson correlation and hierarchical regression analyses were conducted to determine the interrelations among word components and their relationships with L2 writing and lexical use. Three regression models were built to capture the contributions of the six word components to L2 writing ability, target word use and vocabulary score in writing.

1.3.4 Significance of the Study

The significance of the study is twofold in terms of the theoretical and pedagogical contributions. Theoretically, the current study offers empirical evidence to the hypothesized word knowledge models proposed by Nation (2001, 2022) and Coxhead (2007). The quantitative results demonstrate that vocabulary knowledge can be a network composed of a wide variety of interrelated knowledge components. In addition, the current project fills the gap in the existing literature by examining a wide range of word knowledge components associated with L2 writing proficiency. The study empirically confirms Coxhead's (2007) descriptive framework regarding the multiple types of word knowledge required in writing. Productive form and meaning, productive derivative, association and collocation can significantly contribute to L2 writing ability, albeit in varied degrees.

Based on the theoretical significance, the current study also yields pedagogical implications. It makes sense for EFL learners and instructors to focus on vocabulary form and meaning knowledge since productive form-meaning links explain most of the variance in L2 writing and word use. However, the findings in this study suggest

that only productive or indirect types of form and meaning relationships are the major contributors to written communicating skills. The direct links of form and meaning measured by the VLT cannot significantly predict vocabulary use and L2 writing ability. This indicates that Chinese EFL learners and teachers need to change the rote learning of direct form-meaning links. For example, the derivative knowledge, as part of form knowledge, should be paid special attention to since this knowledge reveals robust prediction to L2 writing and lexical use. Association and collocation should not be overlooked in vocabulary acquisition for writing purposes, even though they contribute relatively less variance than other word components.

1.4 Definitions of Key Terms

The key terms in this research are defined as follows:

Multi-component word knowledge: what is required to know a word, ranging from word form and meaning, association, collocation, derivation, and syntactic behavior to constraints of word use in actual lexical production, as tabulated in the comprehensive word knowledge component frameworks proposed by Nation (2022) and Coxhead (2007).

Receptive vocabulary knowledge: the metalinguistic word knowledge associated with vocabulary size that learners can recognize and understand the word form-meaning links without a context (Greidanus & Nienhuis, 2001; Nation, 2022).

Productive vocabulary knowledge: multiple word knowledge components that learners can correctly and appropriately recall and use in a sentence context and IELTS argumentative tasks (Nation, 2022; Zhong, 2016).

Word depth knowledge: "more than just a superficial understanding of the meaning, but also a rich and specific meaning representation as well as knowledge of the word format features, syntactic functioning, collocation possibilities, and so on" in actual contexts (Read, 2004, p.155).

Word size knowledge: the number or quantity of learners' vocabulary, which is quantifiable and limited to receptive form and meaning connections. A large vocabulary size primarily means more receptive word knowledge (Schmitt, 2014).

L2 Writing Ability: the ability to produce IELTS argumentative written texts by Chinese third-year university EFL learners within a given time with particular attention paid to multiple word knowledge recall and use, such as word meaning, spelling, collocation and various derived forms. (Becker, 2018; Hyland, 2003; Yi, 2009).

1.5 Overview of the Study

This study is composed of six chapters. Chapter One Introduction provides the background of the study with a focus on previous studies examining the relationship between vocabulary knowledge components and L2 writing. This chapter briefly outlines the main theories concerning the multidimensional features of word knowledge. It also describes previous studies on multi-component research and L2 writing and specifies the problem statement, research aim, questions, scope and significance as well as key definitions used in the current study.

Chapter Two Literature Review describes the relevant theoretical frameworks for the current study. It introduces vocabulary size and depth concepts, focusing on three main conceptualizations of word depth knowledge and the multi-component theories. This chapter then describes the main concepts of receptive and productive word knowledge and critically reviews the models form receptive word knowledge to productive word use. Combining L2 writing, the review moves to the relationship between word knowledge and L2 writing and the assessment of word knowledge in L2 writing. The last part of this chapter reviews the relevant multidimensional word studies and research on the multi-component word knowledge in L2 writing.

Chapter Three Methodology details the research paradigm and design of the current study. This chapter describes the research context, participants and the quantitative methods used. Then it introduces the test instruments selected for the receptive and productive word knowledge components and L2 writing tasks. Details including how these instruments were borrowed or adapted and how they were validated to be useful and practical are also described in this chapter. It also presents the scoring scales for vocabulary and writing tests, data collection procedure, and data analysis processes.

Chapter Four Results first presents the descriptive statistics reporting participants' performances on the multiple word knowledge tests and L2 writing and word use

tasks. Then Pearson correlation coefficients and the regression R^2 values were used to determine the interrelations among different word components and between these components and L2 writing and word use scores. After reporting the linear relationship between vocabulary knowledge and L2 writing, this chapter presents the hierarchical regression models to describe the contribution of each word component to L2 writing and word use.

Chapter Five Discussion describes the main findings of the current study associated with relevant previous literature. The discussion includes the multi-component construct of vocabulary knowledge and the relationships between the receptive and productive word components. It also compares with the previous findings related to the strength of correlations among these word components. This chapter also discusses the contributions made by each word component to L2 writing and word use, as well as the plus points and drawbacks emerged from these results.

Chapter Six Conclusion summarizes this thesis project and specifies the contributions of the current study to vocabulary and writing research. It describes the significance of the study theoretically and pedagogically, also including the limitations in research design and test administration. This chapter concludes the study by discussing the implications for future studies from the perspective of the limitations and results generated in the current study.

CHAPTER II
LITERATURE REVIEW

This chapter reviews the literature with regard to three major parts: word knowledge constructs, word knowledge and L2 writing and relevant studies on multi-component research. The first part reviews vocabulary size and depth and receptive-productive distinction and the measures hitherto used. By critically comparing various vocabulary dimensions, this part determines that multi-components of word knowledge, including six-word components, are required to assess in the current study. The second part reviews the role vocabulary plays in L2 writing and the relationship between word knowledge and L2 writing ability. There follow the measures of L2 writing and vocabulary knowledge in L2 writing. The third part offers an overall review of multi-component vocabulary research and multi-component studies on vocabulary and L2 writing. It is worth noting that the depth of vocabulary knowledge in the component approach will be focused on in this chapter and thus devoted more space.

2.1 Conceptual Framework of Vocabulary Knowledge

Vocabulary knowledge has been often defined as a construct in the separate trait model, in which various sub-components of a single word can be labeled as distinct types or dimensions (Coxhead, 2007; Gass & Selinker; 2001; González-Fernández, 2022; Nation, 2001; 2022; Richard, 1976). By contrast, the global trait model posits that vocabulary knowledge is an incremental degree of lexical knowledge within a continuum (Henriksen, 1999; Henriksen & Haastrup, 2000; Melka, 1997; Zhong, 2016). Instead of a knowledge cline, another global trait model likens vocabulary knowledge to a metaphorical web of lexical associations (Meara, 1990, 1996, 1997; Meara & Wolter, 2004; Wolter, 2001, 2005). These multifarious dimensions and components make it hard for learners to become proficient in any aspect of language skills without acquiring sufficient lexical knowledge in each dimension. It is not surprising since a word holds myriad internal and external links within the lexical network (Schmitt, 2014), also known as the mental lexicon (Meara, 2009; Singleton, 1999; Wolter, 2001). Despite the complexity of vocabulary knowledge, researchers have theorized a number of descriptive dimensions to better comprehend this construct. Chief among these are size/breadth-depth and reception-production

distinctions which are the best acknowledged global dimensions to describe the overall state of vocabulary knowledge (Anderson & Freebody, 1981; Daller, Milton &Treffers-Daller, 2007; Greidanus & Nienhuis, 2001; Milton, 2009; Nation, 2022; Qian, 2002; Schmitt, 2010, 2014; Zareva et al., 2005). According to Anderson & Freebody (1981), the size of vocabulary knowledge denotes the number of words learners could identify, while depth is the richness of word knowledge or how well learners know about the words. The latter is far more complicated and perplexing than the former. Therefore, size is relatively easy and straightforward to conceptualize and operationalize in assessments because it is typically characterized by counting lexical numbers and closely connecting to word form and meaning (Schmitt, 2010, 2014). On the other hand, depth goes deeper and wider than form and meaning, involving the mastery of more shades of meaning and associations needed to productively use in different contexts (González-Fernández, 2022; Kieffer & Lesaux, 2012). Researchers interpret the depth construct in a wide variety of ways. For example, Read (2000) describes vocabulary depth from a dimension or component approach (see Nation, 2022) by breaking down vocabulary knowledge into various lexical components; Meara and Wolter (2004) argue that words do not stand isolated, but rather closely related to one another. Thus, vocabulary depth refers to every possible link of words in the lexical organization, which is indistinguishable from the size; Schmitt (2014) theorizes that receptive and productive mastery can be a simpler way to understand and operationalize the depth construct. In addition, Daller, Milton & Treffers-Daller (2007) regard the fluency or automaticity of using a word in real time as part of depth knowledge, though independent of the size and depth.

2.1.1 The Size of Vocabulary Knowledge

Vocabulary size entails the number of words learners know or how many words can be recognized (Coxhead, Nation & Sim, 2015; Read, 2004). Typically, vocabulary size measures the ability to fast map the word form to its meaning, thus making form-meaning link the single dimension in size tests (Lin, 2015; Milton, 2009). Receptive size tests primarily assess the lexical knowledge in listening and reading when test takers encounter the word form and match its meaning. In contrast, productive tests focus on how many words test takers can use in speaking and writing when they have

to provide the corresponding word form to express the meaning (Coxhead et al., 2015). Size tests feature prominently in vocabulary assessment research (Vermeer, 2001; Zhong, 2016) due to their single-dimensional character and ease of operationalization. According to Nation (2022), there have been two fundamental methods to measure vocabulary size: dictionary-based sampling and corpus or corpus-based frequency list sampling. Many vocabulary size tests, such as the *Eurocentres Vocabulary Size Test* (the EVST, also known as the Yes/No test) (Meara & Jones, 1990) and the receptive *Vocabulary Levels Test* (the VLT) (Nation, 1983, 1990) are based on the sampling from corpus frequency lists.

Of all the receptive size tests, the VLT (also known as the RVLT), first developed by Nation (1983) for diagnostic assessment purposes, has been widely accepted as the "*de facto* standard" (Meara & Olmos Alcoy, 2010. p. 222). Indeed, many researchers have unanimously acknowledged the validity of the VLT (Read, 2000; Schmitt, 2010; Meara, 1996). Coupled with its simplicity and practicality, the VLT has become the most prominent development in vocabulary research in the last decades (Meara & Olmos Alcoy, 2010; Read, 2020). As such, it has been universally used to test both native and non-native learners' vocabulary size and other language abilities, such as reading, where vocabulary size is a critical variable. A sample cluster is as follows:

1. Apply
2. elect
3. jump ___2___ chose by voting
4. manufacture ___5___ become like water
5. melt ___4___ make
6. threaten

The original VLT test is a form-recognition mapping approach including five sections based on five levels of word frequency: the first 2000 words, 3000 words, 5000 words, the *University Word Level* (UWL beyond 5000 words) and the 10,000 words. Word items sampled in each section represent the corresponding frequency level, i.e. the times a word appears in written products. Regarding the test format, each cluster has six target words and three definitions. Six clusters constitute one frequency level, making the total number of words and definitions at each level 36 and 18, respectively.

Schmitt et al. (2001) revised the test and expanded the target words at each level from 36 to 60 and definitions from 18 to 30. They sufficiently validated the modified VLT through rasch, item, and factor analysis. The results indicate that the new versions of the VLT produce a similar equivalent effect to the original one, pointing to the high prediction of learners' vocabulary size. Moreover, the somewhat outdated UWL (Xue & Nation, 1984) has been supplanted by the *Academic Word List* (the AWL) (Coxhead, 2000), as the AWL is synonymous with academic vocabulary knowledge (Coxhead, 2011; Coxhead & Hirsh, 2007), has drawn growing attention in recent years.

Therefore, the validity and reliability of the VLT have justified its presence in a large number of empirical studies as a vocabulary size measure. For example, Stæhr (2008) compared the scores EFL learners achieved in the VLT with the grades in their listening, reading and writing exams. The testing results suggest that the size knowledge assessed by the VLT is highly relevant to learners' language proficiency, particularly reading. This is unsurprising since the ability to quickly bridge word form and meaning is fundamental to a wide variety of language skills, which is more so for L2 learners (Laufer & Goldstein, 2004). Webb and Chang (2012) substantiated that the VLT is useful for measuring the overall vocabulary growth for L2 words. They conducted a longitudinal VLT test over four years to observe L2 vocabulary acquisition rates. They also found that groups that received the most English instruction on high-frequency words accomplished the highest scores. This result was corroborated by Coxhead and Boutorwick (2018), who also deployed extensive VLT tests among international students with varying backgrounds. The VLT was in a favorable position to measure vocabulary size in EFL contexts. When compared with the Word Associates Test (WAT) (Read, 2000, 2007), the VLT was found to win favor in measuring learners' language proficiency, while the WAT provided little prediction (Janebi Enayat & Amirian, 2016).

The VLT plays a critical role in yielding a large amount of empirical data though it is not without pitfalls. According to McLean and Kramer (2015), the first drawback is that the VLT does not design a section to test the first 1000 levels of word families. It is hard to overstate the importance of this frequency level as it accounts for more than 80% of spoken and written English (Webb et al., 2017), thus is the most valuable part

of word families for learning English. Another weakness is that many of the target words are sampled from old word lists compiled more than 50 years before and cannot represent the words currently used today. In addition, the VLT was initially designed in accordance with the assumption that learners necessarily acquire vocabulary from high-frequency to low-frequency words. Lemmouh (2010) argues otherwise since EFL learners might gain some relatively low frequency, advanced words in language acquisition at school, whereas lacking the basic high-frequency words common in everyday use. Furthermore, Read (2000) criticizes the VLT for its inability to reveal whether learners have genuinely known the word because it simply remains at a low and superficial level. As pointed out by Kremmel and Schimtt (2016), correctly matching the form and meaning is not reflective of lexical employability in real word use since no reading materials provide any prompt multiple definitions for meaning mapping. Targeting the limitations cited above, Webb et al. (2017) improved upon the initial version of the VLT and re-developed two equivalent forms featuring three changes: an added 1000 levels of words, a new item source from Nation's (2012) BNC (British National Corpus) frequency list reflecting contemporary English, and different presentations of matching format.

Admittedly, the VLT is not perfect despite the attempts to promote all facets of this test. It boasts sound validity and reliability and has achieved a superior position like no other size test (Lin, 2015; Meara & Olmos Alcoy, 2010; Schmitt et el., 2001). Specifically, the VLT can be ideally placed to measure high-frequency words and diagnose what level learners have achieved in their learning. In addition, this test is simple to administer and easy to mark and interpret (Nation, 2013), insofar as teachers and researchers can do it to identify learners' overall vocabulary knowledge. The current study uses the VLT modified by Schmitt et al. (2001) to reap these benefits due to the ample test usefulness. This test version would be the best fit for capturing Chinese EFL learners' receptive word size.

In addition to the VLT, the size construct can also be measured by the Yes/No test (Meara & Jones, 1990) and the Vocabulary Size Test (VST) (Nation & Beglar, 2007). The Yes/No test is easy to administer by simply eliciting test takers' responses from their recognition of the target words. This points to a large sample of words that can

be measured in this format, thus suitable for placement purposes (Read, 2000). Regarding the VST, a multiple-choice test for a crude estimate of the overall vocabulary size, it has a 14,000 version and a parallel 20,000 version of frequency levels, making it possible to measure the written receptive size at any level. However, as important as they are, vocabulary size tests simply end with a "superficial treatment of each item" (Laufer & Goldstein, 2004, p. 400). Singleton (1999) seriously doubts treating vocabulary knowledge as separate lexical items, and he argues that the scope of teaching and measuring vocabulary needs to be expanded beyond individual content words. Read and Chapelle (2001) agree to suggest that many vocabulary size tests investigate vocabulary knowledge as separate components "without reference to the functions of words in grammatical structures, texts or discourses" (p.2). Hence, more tests should be designed to determine whether learners know the word and know about the word, meaning the depth of vocabulary knowledge.

2.1.2 The Depth of Vocabulary Knowledge

Vocabulary knowledge is not only about knowledge of the form-meaning connections (vocabulary size) but also about acquiring multiple shades of meaning, semantic associations and grammatical functions in different contexts (vocabulary depth) (Kieffer & Lesaux, 2012; Schmitt, 2014). Accordingly, measuring the size cannot suffice to paint an overall picture of vocabulary knowledge, merely a part of the entirety at best (Ishii & Schmitt, 2009; Read & Dang, 2022), as suggested by Read (2004):

"Learners need to have more than just a superficial understanding of the meaning [of a word]; they should develop a rich and specific meaning representation as well as knowledge of the word's format features, syntactic functioning, collocational possibilities, register characteristics, and so on" (p.155).

This is endorsed by Anderson & Freebody (1981) that "for most purposes, a person has a sufficiently deep understanding of a word if it conveys to him or her all of the distinctions that would be understood" (p.92). These statements draw attention to the fact that deepening vocabulary knowledge is as much a part of vocabulary acquisition as expanding the number of lexical items. In other words, it is not how many (vocabulary size) but also how well (depth or quality) of the lexis needs to be

highlighted in depth (Qian & Lin, 2020; Schmitt et al., 2011). Nevertheless, there has been a sprawling array of different conceptualizations of the construct of vocabulary depth as different researchers have their own disparate understandings, to the point where this concept has become confusing and elusive. When it comes to depth, researchers in practice are coping with a construct that is "inherently ill-defined, multidimensional, variable and thus resistant to neat classification" (Read, 2004, p. 224). Consequently, assessing vocabulary depth has become a mammoth task for researchers (Meara, 1996).

Against the odds, previous studies have categorized the relative depth concepts and provided insights into this construct (Read, 2000; 2004; Schmitt, 2010). Read (2004, p.211) encompassed different paths of developing and operationalizing the depth construct as three distinct theoretical lines: **precision of meaning** refers to a degree of word meaning from a limited, vague idea of the word to a more precise, accurate knowledge of the meaning; **comprehensive word knowledge** is defined as various word knowledge components more than semantic features acquired by learners, including "orthographic, phonological, morphological, syntactic, collocational and pragmatic features", which is also known as the component approach; **network knowledge** points towards the integration of the word into the lexical organization or network in learners' mental lexicon, in which learners connect and distinguish the word from the related words. The three categories summarized by Read (2004) can be categorized to be declarative knowledge, i.e. the depth of vocabulary knowledge learners can have access to and report consciously in tests, as opposed to procedural knowledge, i.e. the implicit depth of vocabulary knowledge, which is more related to vocabulary use and fluency in productive language skills.

Following Read's (2004) categorization, Schmitt (2010, 2014) postulates that the reception-production dichotomy can be a more straightforward way to conceptualize the depth construct. Vocabulary knowledge gradually increases its depth from the receptive to the productive ends. Similarly, based on Read's (2000) developmental concept, Schmitt (2010, p.38) believes that the "development scale" within a continuum can be used to perceive and operationalize the depth, not unlike Read's (2004) precision of meaning. The difference is that the development approach contains

more components than meaning, including, among other components, spoken and written form, grammar, collocation, register and associations. In this scale, the overall proficiency of the word ranges from no knowledge at all to full mastery. The two scholars' (2010) categorizations of the depth concept make this construct more accessible to researchers. Based on these, the current study attempts to provide a clearer picture of the depth, as the below chart indicates.

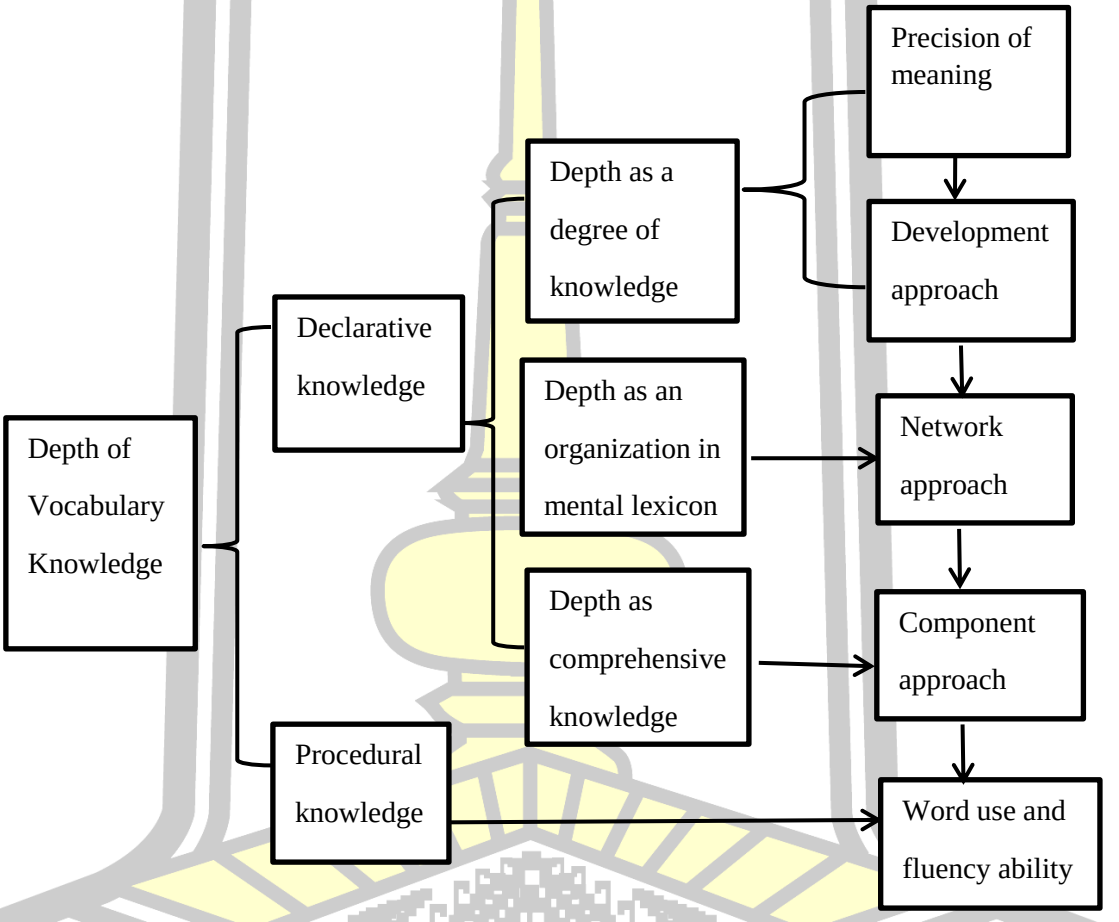


Figure 1 Different conceptualizations of vocabulary depth based on Read's (2004) and Schmitt's (2010) categorizations.

This chart above simply makes a rough and oversimplified summary of the depth since Read (2004) admits that these approaches overlap considerably due to the interwoven relationships of different knowledge components. The following sections will review vocabulary depth in accordance with the outline of the figure. The depth approaches featured in the declarative knowledge will be examined separately, and the procedural knowledge (word use models) will follow suit.

2.1.2.1 Depth as an Incremental Degree of Knowledge

Despite being popular as a way to conceptualize depth, the development approach could hardly be operationalized in empirical tests as little has been known regarding the process of this progression (Schmitt, 2010). This section narrows down and reviews the precision of the meaning approach. According to Read (2004), the precision of meaning refers to a degree or scale of vocabulary knowledge from a low level of mastery to a more precise and accurate understanding. For example, if the need arises to convey the meaning in a sound and comprehensive manner, learners need to sufficiently perceive the word meaning. This is often difficult to obtain, especially when learners encounter high-frequency words with no context. They also find it perplexing to distinguish polysemous words that carry multiple shades of meaning. Complicated still, there are distinctions between the meanings of a word in everyday use and specialized and technical contexts. All these simply elude L2 learners from linking and discerning the word meanings in actual use. Therefore, the precision of word meaning aims to measure whether learners achieve the exact meanings and what learners have mastered or missed. The most commonly used test of this kind is the Vocabulary Knowledge Scale (VKS), developed by Paribakht and Wesche (1996, 1997).

The earliest example of the VKS is Dale's (1965, cited in Read, 2000) four-stage scale, measuring knowledge development along four steps. Following this line, Paribakht and Wesche (1996, 1997) expanded the scales to five stages and rendered the VKS today (see the example below). This test instrument is referred to as a self-report test by Read (2000, 2004) and a self-assessment test by Milton (2009). The VKS was developed with the initial interest in investigating the effects of extensive reading on learners' incidental acquisition of word meaning. This means that the test elicits lexical knowledge from the early stages of recognition to self-assessment and verifiable demonstrations of word use: to provide a synonym, L1 translation and write a sentence with the target word at the last stages (Read, 2007). In other words, the VKS was designed to reveal learners' development from receptive knowledge to basic productive word use (Schmitt, 2010).

I. I don't remember having seen this word before.
II. I have seen this word before, but I don't know what it means.
III. I have seen this word before, and I think it means_____. (synonym or translation)
IV. I know this word. It means _____. (synonym or translation)
V. I can use this word in a sentence: _____. (Write a sentence.) (If you do this section, please also do Section IV.)

(Wesche & Paribakht, 1996: 30)

Paribakht and Wesche (1996) managed to justify this test in measuring incidental word acquisition. They conducted a test and post-test using the VKS with a two-week interval in between and yielded a high level of reliability at 0.89. Test-takers' scores were highly aligned (correlation was 0.92 to 0.97) with the self-reported interviews as to how much they knew about the words, indicating a sound test validity. Paribakht (2005) and Wesche and Paribakht (2009) also used the VKS to assess the ESL learners' retention of new vocabulary. The VKS played a crucial role in the operation of these studies. Joe's (1998) study also provided evidence for the test's usefulness when he used the VKS to measure how and to what extent the target words had been acquired during the text-retelling process. Moreover, the VKS has been adapted by researchers to look into learners' lexical quality. For example, Zareva et al. (2005) revised the VKS to examine six variables related to quantity, quality and metacognitive awareness of vocabulary knowledge. They replaced the sentence writing task at Scale V with a productive elicitation task, requiring learners to associate three related words. It was concluded that the VKS is sensitive to the overall mastery of word knowledge and the development of language proficiency. However, despite their modifications, these allegedly receptive and productive items still seem to measure the same construct (Zhong, 2016). This stresses the need for designing separate tests to measure the multiple components of word knowledge if the depth is the testing goal (Read, 2000; Schmitt, 2014; Webb, 2013).

While the VKS has been touted as a traditional depth measure, it is not above criticism and limitations. First and foremost, the VKS is not straightforward to understand because it simply represents a mixture of receptive and productive word knowledge.

Stages I to IV focus on receptive form recognition and meaning recall, whereas stage V jumps to a much higher level of ability to productively use the word in a sentence. Simply put, the test addresses inconsistent word knowledge constructs integrated into a single test instrument (Schmitt, 2010), which causes trouble for its scoring and reliability. This is the major reason why the current study does not use the VKS in its test battery. The attempt to encapsulate multiple components within one test might be problematic once and for all.

Furthermore, researchers also criticized the VKS for its inability to measure a wide range of deep vocabulary knowledge (Milton, 2009; Read, 2000; Schmitt, 2010; Wolter, 2005). For example, Read (2000) and Wolter (2005) argued that the VKS could not touch on multiple shades of meaning in that knowing one definition of the word suffices to correctly complete all items. It is no better regarding its measuring power of productive components either. As many researchers (Milton, 2009; Milton & Fitzpatrick, 2014; Read, 2000; Schmitt, 2010) pointed out, a prompted sentence writing task on stage V might not well serve the purpose of measuring learners' productive ability to use the word in context. Some L2 learners seemingly could produce reasonable, even complicated sentences with the target words without a real understanding. Taken together, the VKS as a specific measure of vocabulary knowledge degree is flawed in operationalizing the complex depth construct.

2.1.2.2 Depth as a Lexical Network

Unlike the development scale, a host of researchers compared vocabulary depth to a process of building a lexical network (Henriksen, 1999; Henriksen et al., 2008; Hills et al., 2009; Read, 1998, 2000, 2020), also known as the mental lexicon in which words are stored and linked (Meara, 1996; Meara & Wolter, 2004; Wolter, 2001, 2005). The approaches to conceptualizing the network concept are varied (Zareva et al., 2005). For instance, the lexical network can be perceived as the way how words are orchestrated and stored in the mental lexicon, i.e. paradigmatic associations (such as synonyms), syntagmatic associations (such as collocates) and analytic links (the key meaning of the word) (Greidanus & Nienhuis, 2001; Read, 2000; Wilks & Meara, 2002). Another way of interpreting the network approach is the lexical organization in which words are characterized as interconnected nodes (Henriksen, 1999; Henriksen

et al., 2008; Meara, 1996; Meara & Wolter, 2004; Wolter, 2005). The number of nodes in the organization corresponds to the vocabulary size, and the links of these nodes represent the depth of vocabulary. With each new word (node) added to the organization come a larger vocabulary size and new links of deeper knowledge. Learners who are more proficient in vocabulary depth tend to have denser and stronger links of these nodes. The current study reviews the former way of the association as it has been put in the spotlight of research in recent years.

The most widely used association test is the Word Associates Format (WAF), created by Read (1993, 1998, 2000) as a typical measure for associative patterns of word knowledge (Schmitt, 2010; Yanagisawa & Webb, 2020). Having modified several versions, Read (2000) decided that assessing to what extent learners have known the meanings of a word should be the major target of the WAF since eliciting all components of vocabulary knowledge is as problematic as it is impractical. Thus, the WAF primarily measures such three components as form and meaning, concept and referents, and collocation (Webb, 2013), as the below example indicates (Read, 2000, p.184).

common

complete light ordinary shared	boundary circle name party

In the WAF, selective adjectives alone are measured to ensure more consistency. Test-takers need to first build paradigmatic associations by choosing synonyms or words that share at least one semantic feature with the target word from the left box. Likewise, they establish syntagmatic associations by selecting words from the right box, meaning collocates or words that typically appear adjacent to the target word. The keys assigned to the two types of associates could be 2-2, 3-1 or 1-3 in order for a minimum potential of successful random guessing. This effort notwithstanding, Read (2000) acknowledges that random guessing, in many cases, remains a lingering problem in the WAF. That said, Schmitt et al. (2011) argue that the WAF could make a useful tool to measure learners' depth of word knowledge as it delves into the uncertain nature of collocational combinations. Moreover, considering that many traditional word association measures are somewhat unwieldy in practice, Fitzpatrick

and Thwaites (2020) advise that the WAF could be easier and simpler to operationalize in empirical studies due mainly to its high construct validity, clear instructions and limited options. Yanagisawa and Webb (2020) pointed out that the WAF has been the most popular depth measure in research to facilitate the feedback on learner' vocabulary knowledge progress and unravel the nature of word knowledge development.

Indeed, the WAF has been used by a number of researchers to operationalize their studies (Alsahafi, 2023; Batty, 2012; Dabbagh & Janebi Enayat, 2022; Janebi Enayat & Amirian, 2020; Nassaji, 2006; Qian, 2002; Qian & Schedl, 2004). Qian (2002) and Qian and Schedl (2004) adapted the WAF format to what they called the DVK (the depth of vocabulary knowledge), mainly measuring three components of vocabulary knowledge: synonymy, polysemy and collocation. They have substantiated that the WAF suffices to assess these components of vocabulary depth with reliability as high as over 0.88. The WAF test can reveal the significant correlations between the depth components and word size in predicting academic reading performance. Qian and Schedl's (2004) study also indicated that the modified WAF (the DVK) could produce a positive and meaningful washback effect for ESL vocabulary teaching and acquisition practices. Nonetheless, it is worth noting that these encouraging results could hardly be generalized since merely meaning and collocation were embraced in the empirical tests without regard to the bulk of other depth components. Webb (2013) noted that the WAF is primarily restricted to meanings and collocations since Read's (1993) initial testing goal is meaning-centered. For this reason, Webb and Sasao (2013) call for more separate tests of multiple components to measure vocabulary depth so that each component, such as synonymy or collocation, could be put under the microscope for a nuanced understanding.

Aiming to validate the WAF, Dabbagh and Janebi Enayat (2022) and Janebi Enayat and Amirian (2020) compared it with other depth and size tests such as the VKS and the VST. Their comparisons suggest that the WAF is more predictive of EFL learners' vocabulary size and has a higher correlation with scores of highly proficient EFL learners. This points towards the WAF deserving more attention along with the boost of English proficiency. The same holds true in Nassaji's (2006) study, where the WAF

was used to examine how well the depth knowledge measured could predict learners' lexical inferencing in context. Learners with stronger associative capacity were found to apply more effective inferencing strategies in the reading task. This result is unsurprising because L2 learners need profound word knowledge to build lexical associations. Those who have stronger and more integrated lexical representations of lexis are in a more robust position to derive words from the rich lexical knowledge base. Accordingly, L1 language users tend to have a richer and denser lexical network than L2 learners (Jiang, 2000).

Nevertheless, the WAF also draws a chorus of criticism as to whether it has the quality as a depth measure. Wolter (2005) holds a less optimistic attitude towards the WAF being successful in assessing the depth knowledge in that word size accounts for a large part of the test scores. Wolter (2005) and Meara and Wolter (2004) believe that the clumsy attempt to measure depth by breaking down vocabulary into different components is "unfortunate" (p. 86). Instead, the size and organization might be a more productive way to construe vocabulary knowledge. Vermeer (2001) endorses this view and points out that an extensive network of a word means knowing as many associated words as possible so that learners can deepen and strengthen their knowledge. In the same token, size and depth are conceptually the same and inter-dependent lexical components; if the former is expanded, the latter will be deepened simultaneously. However, regarding this whole organization concept, they neither provide specific ways to operationalize it nor shed light on the features of those links that characterize vocabulary depth. Moreover, it seems true only when depth is narrowly perceived in the single sense of word meaning, disregarding other more productive components such as collocation, derivatives and register, among others (Schmitt, 2014).

Read (2000) further indicates that all words are treated independently in the WAF for a broader coverage of the target words, isolated from any context. Context should be paid due regard in assessing vocabulary depth for productive use. More meanings of a word might be expressed and conveyed by the surrounding words in a context than the target word per se (Ruhl, 1989). Richard's (2011) newly-developed depth measure in multiple, contextualized sentences say much about the importance of context,

especially for high-frequency words. Recently, Zhang and Koda (2017) did a fine-grained review on the WAF vis-a-vis four major areas---design features of the test, test administration conditions, the scoring system of the test, and characteristics of the test takers. After critically discussing the relevant studies and findings, they concluded that the psychometric properties of the WAF have been plagued by problems inherent in variables of the four areas. Therefore, the WAF warrants more validation evidence and refinements for future studies. Based on the aforementioned discussions, the current research adapts the WAF to a single synonym test in a sentence context as one of the multiple tests to look into the synonymous component of word knowledge.

2.1.2.3 Depth as Component Approach

Despite the limitations found in the WAF, it shows every sign of being a popular type of depth measure. Nonetheless, more components than meaning and collocation about a word need to be measured to delve into word knowledge in productive contexts. In this regard, Read (2004, p.223) calls attention to "a broader range of measures" for a comprehensive understanding of the depth knowledge. Under such circumstances, separate and multi-component tests focusing on individual knowledge components might be a more precise and comprehensive way to measure depth (Read & Dang, 2022; Webb, 2013). This is also called the strength of vocabulary knowledge, meaning to what extent a single word component is known (Levitzky-Aviad & Laufer, 2013; Milton, 2009, 2013; Webb, 2013).

The notion of the component approach stems from Richard's (1976) eight assumptions about what it involves to know a word, ranging from word form and meaning, association, collocation, derivation, and syntactic behavior to restrictions of use. Richard's (1976) assumptions are somewhat a bewildering and unorganized array of word knowledge components (Milton & Fitzpatrick, 2014). In addition, it seems strange that the most fundamental meaning component was ranked the lowest in his list (Meara, 1996). Nonetheless, these added lexical elements inspired Nation (2001, 2022) to encompass more word components that needed to be acquired to use the word well. Although there has never been a list that encapsulates all components of word knowledge, Nation's (2001, 2022) framework of word knowledge is the "nearest thing we have to a definitive list of what it means to know a word" (Milton, 2013,

p.59). Nation codifies vocabulary knowledge into three major dimensions, namely form, meaning and use, as the general levels of knowledge. Each dimension comprises three components, and each component is further subdivided into receptive and productive uses, making a total of eighteen components, as shown in the table below (Nation, 2001, p.49).

Table 1 What is involved in knowing a word

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognizable in this word?
		P	What word parts are needed to express the meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this?
	constraints on use	R	Where, when and how often would we expect to meet this word
		P	Where, when and how often can we use this word?

Note: R: receptive knowledge, P: productive knowledge

Nation (2001, 2022) makes a clear distinction between different elements of vocabulary knowledge and provides us with a comprehensive understanding of these lexical components. The first-dimension **form** illuminates the pronunciation, spelling and the inflected and derived forms of the word; the **meaning** dimension features the form-meaning link, the denotations and connotations, and the semantic links, such as polysemy and synonymy; the **use** dimension consists of the grammatical components of the word, formulaic patterns such as collocations and idioms, and the sociolinguistic constraints like occasions and registers. Moreover, a clear line is drawn between receptive and productive distinctions, constituting the depth construct in its entirety. This taxonomy provides convenience for different types of research engaging in receptive or productive or both combined contexts. It mainly holds true in the current study, which concentrates on productive components of vocabulary depth in L2 writing as it makes possible to capture learners' vocabulary knowledge in a wide range of qualities. Nonetheless, it is noted that this knowledge component list merely serves as a standard work of reference instead of all the lexical components learners typically know in their mental lexicon. While Nation (2001, 2022) does not specify the depth construct in the taxonomy, his tabulation of the word components closely connects to

and reveals vocabulary depth characterized by the development scale and network theory.

Unlike the development approach, the component approach construes vocabulary knowledge from a more micro-level perspective, looking into the word's internal structure. It appears that the former is closely tied in with the latter. Milton and Fitzpatrick (2014) believe that some components, say, the form and meaning, apparently precede the others in an acquisition, like the collocation and grammatical functions. Laufer and Goldstein (2004) have also demonstrated that form recognition is the most straightforward in vocabulary acquisition and the initial stage of retrieving the word. It is unthinkable if learners can connect a word with others and use it in L2 writing before they can recognize the form and match the meaning. This can be further corroborated by Henriksen's (1999) model, which spells out the incremental processes from vocabulary size to depth and use with her triple continua: partial-precise knowledge, depth of knowledge, receptive-productive competence (see more in Section 2.2.4). These development stages correspond precisely to the knowledge components articulated in Nation's (2022) framework if they are examined in a fine-grained manner. It follows from this point that Nation's tabulation of word components can possibly make both a component clarification and a developmental scale --- all knowledge components progress from word form to meaning and to use (Milton & Fitzpatrick, 2014). For example, the VKS (Paribakht & Wesche, 1996), often used as a developmental test, follows exactly the same sequence as Nation's layout in Table 2 from form recognition to meaning mapping (synonym and translation) and to word use (sentence production). In this regard, the component approach can be more of a practical, feasible and, above all, comprehensive method to empirically operationalize the development of the vocabulary components in the list.

Likewise, the network approach to vocabulary depth (Henriksen, 1999; Meara, 1996) also draws heavily on the component approach. Henriksen (1999) and Meara (1996) argue that vocabulary knowledge can be deepened by strengthening the links between words in the mental lexicon. Those links, however, can be made when many words have been acquired (Milton, 2009; Milton & Fitzpatrick, 2014). More importantly, word knowledge concerning certain components of the associated words needs to be

known beforehand (Zhong, 2014). Chief among these components are form and meaning because they are the most fundamental elements to cement the relationships between words. Indeed, Miller (1999) points out that the vocabulary links in the mental lexicon can simply be seen as the form-meaning associations. If the form is connected to different meanings, the word has polysemous links; by contrast, if multiple forms can convey the meaning, then the words make synonymous links (Miller, 1999). Lexical links go beyond the rudimentary form and meaning associations. For example, when learners build connections between the words *socialize* and *peers,* they first have to identify the forms and meanings of the two words. There follows the realization of the word parts, i.e. *socialize* is a family member of *social* and *peers* is a plural, grammatical functions as a verb and a noun and finally, the associations and constraints of word use in a particular context. Taken together, it is logical to assume that both the development approach and the network metaphor can be perceived as the conceptualization of various components depicted in the component approach. However different the former two approaches claim to be, they eventually come back to the internal structure of a single word guided by Nation's (2022) framework. The corollary of this analysis is that the component approach includes and exceeds the other two counterparts and can better guide the measures to capture vocabulary depth knowledge. As Fitzpatrick and Milton (2014) concluded, delving into the complex and dynamic components of the lexicon facilitates a deeper investigation of the lexicon as a whole.

Notably, the comprehensiveness of this framework also becomes one of its disadvantages because measuring all the eighteen components is unthinkable and unattainable (Milton, 2009; Read, 2000, 2004; Schmitt, 2010). A typical method is to assess several measurable components of dozens of specific words (Schmitt, 2010, 2014). Nonetheless, researchers (Read, 2004; Wolter, 2005) express concerns about the time- and effort-consuming nature of the multi-component tests for both researchers and test-takers alike. Doubts were also voiced about the assessment purpose of eliciting a large amount of data on a small number of words (Read, 2004), such that it might miss the woods for the trees (Merea, 1996). However, multi-component tests merit attention since few studies have empirically scrutinized the nature of these vocabulary components at the level of this detail (Kieffer & Lesau,

2012; Schmitt, 2010). Consequently, little has been known regarding the development and interrelationships between different lexical components (Gonzalez-Fernandez & Schmitt, 2020). This is especially so when certain word components feature heavily in the productive performance of L2 writing. For instance, it remains a mystery how various components are orchestrated in function, and if the growth of one component, say, the association knowledge heightens the component of collocation in the essay writing process. Coxhead (2007, 2011, 2012) paid special attention to the relationship between these word components and L2 writing. Coxhead (2007) capitalized on Nation's (2022) framework of vocabulary knowledge to introduce her model of word components required for producing an L2 essay. With this model, conducted a case study involving seven L2 university learners. These L2 writers reported that they had to mobilize different word components during the writing process, chief among which are written form, form and meaning links and academic register (see more in Section 2.3). Before unraveling word components in language production, it is necessary to first explicate the concepts of receptive and productive vocabulary knowledge.

2.1.3 Receptive versus Productive Vocabulary Knowledge

Receptive and productive vocabulary knowledge provides a clearer picture regarding the conceptualization of vocabulary depth, focusing on how learners could handle and control the lexical items in language skills (Schmitt, 2014). This distinction has long existed regarding learners' different abilities to access and use the word. Schmitt (2010, 2014) maintains that the receptive-productive dichotomy should be regarded as valid because, in both EFL and native language acquisition, some words might be known and understood in some contexts yet cannot be used productively in others. Milton (2009) also supports the validity of this distinction, and he holds that there are moments when learners encounter words they perceive to know but can never call into mind to use smoothly and naturally. Thus, it is more than necessary to spell out what receptive and productive vocabulary knowledge means and what is involved in vocabulary use (Melka, 1997; Schmitt, 2010).

As simple as this distinction seems, researchers have yet to reach a consensus on conceptualizing and operationalizing receptive and productive vocabulary knowledge. Receptive vocabulary knowledge has usually been defined as the capacity of either

recognizing the word form (Laufer et al., 2004) or discerning the form and meaning (Sukying, 2017, 2018, 2022; Webb, 2008a) or producing L2-L1 translation of word pairs (Webb, 2008b, 2009). Productive vocabulary knowledge has often been defined as the ability to activate and retrieve word form and meaning for free language production in context (Laufer et al., 2004; Sukying, 2017, 2018, 2022; Webb, 2008a), or recalling L1-L2 translation of word pairs in contextualized learning (Webb, 2008b, 2009). Laufer and Goldstein (2004) prefer passive and active dichotomy to that of reception and production. Their computer-adaptive vocabulary size test drew a distinction between active and passive word meanings with four categories of word-from link: passive recognition, active recognition, passive recall, and active recall (p.399). Among the four components of vocabulary knowledge, passive recognition is the easiest and active recall is the hardest for learners to acquire. In contrast, passive recall proves to be the best predictor of lexical performance (Laufer & Goldstein, 2004). Nevertheless, it is confusing to construe the four terminologies that refer to supplying the word form or meaning for active and passive recalling or recognizing tests (Schmitt, 2010). Laufer and Goldstein (2019) have recently developed a new computerized test of vocabulary size, measuring vocabulary knowledge in four receptive and productive modalities: productive recall, receptive recall, productive recognition, and receptive recognition.

Based on Laufer and Goldstein's (2004) studies, Schmitt (2010) adapted and clarified the four receptive and productive vocabulary knowledge components. Four clearer terms were coined in his definitions: meaning recognition and form recognition for receptive knowledge and meaning recall and form recall for productive knowledge.

Table 2 Schmitt's (2010, p. 86) four stages of degrees of word *knowledge*

Word Knowledge	Word-knowledge tested	
Given	Recall	Recognition
Meaning	Form Recall (supply the L2 item)	Form reception (select the L2 item)
Form	Meaning recall (supply definition/ L1 translation)	Meaning recognition (select definition/ L1 translation)

Supporting Schmitt's (2010) categorization, Nation (2022) follows to use the four stages of vocabulary knowledge to elucidate this distinction. Specifically, Nation puts the definitions of the receptive-productive distinction within contexts of actual language use. Nation (2022) theorizes that learners could recognize and comprehend the word form and meaning for receptive use in listening and reading while correctly and appropriately producing words in speaking and writing for productive mastery. In other words, receptive use involves perceiving the spelling and sound form of the word and recalling its meaning. Productive use needs to retrieve and produce the spoken and written form of the word to express the meaning learners have in mind.

All the definitions above seem to restrict receptive and productive vocabulary knowledge to measuring word form and meaning. As Webb (2009) noted, there is more to mastering a word than form and meaning. For this reason, Read (2000) proposed a twofold distinction to provide a more comprehensive understanding of measuring receptive and productive knowledge: recognition versus recall and comprehension versus use. The two pairs are parallel and can be used interchangeably for different assessment purposes. **Recognition** is the receptive response ability in which test-takers are required to show their understanding of the meaning of a target word. **Recall** means the productive ability to retrieve the target word from their mental lexicon stimulated by some designed word. This distinction can be illustrated by the word pair experiment in which L2-L1 word translation points to recognition, and the reverse process refers to recall. By contrast, Read (2000) holds that the comprehension-use distinction is a more complex way to distinguish and investigate reception and production in context. **Comprehension** refers to the receptive ability to understand a word in listening and reading, while **use** alludes to the actual process of productively using words in speech and writing.

Recognition and recall are more likely to be used in discrete measures to assess selective and context-independent lexical items. In contrast, use and comprehension tend to be embedded and context-dependent measures, entailing more tasks with different lexical abilities. Either pair alone cannot shake off problems if merely one pair of them is used in the assessment. On the one hand, recognition-recall primarily measures the isolated words in decontextualized situations (Read, 2000). Nonetheless,

any word becomes meaningful only when it is placed in a natural context, which brings vocabulary tests closer to language use and produces positive washback effects for classroom implications (Qian, 2008). Chapelle (1994) and Read and Chapelle (2001) concur with this point and argue that vocabulary knowledge should be defined with an interactionalist approach, meaning that receptive and productive word use are subject to contextual constraints. Van Zeeland (2013) compared vocabulary assessments with and without contexts and identified 35% discrepancies between word knowledge in the two ways of testing. This indicates that context impacts vocabulary knowledge, and vocabulary tests with no context may not truly reflect lexical mastery.

On the other hand, comprehension-use might be time-consuming to operationalize and difficult to control as the target words might be inferred from the context. Worse still, test-takers may avoid using the target words, throwing the test validity into question. Allowing for the pitfalls inherent in the two types of receptive and productive assessment, the twofold distinction has to be combined to capture both general decontextualized word knowledge and context-specific knowledge in actual use (Pearson, Hiebert, & Kamil, 2007; Read, 2000). Following this idea, the current study conflates the two pairs of definitions to operationalize receptive and productive word knowledge tests. First, the recognition test (the VLT) captures learners' general vocabulary size. Then productive recall tests and limited sentence contexts are deployed to measure various components of productive knowledge. Finally, the target words are embedded in L2 writing to assess this knowledge in language performance.

2.1.3.1 Relationship between Receptive and Productive Vocabulary Knowledge

There is a growing consensus that receptive and productive vocabulary knowledge should not be regarded as an either-or dichotomy. Instead, the two constructs are deemed to stand on the two ends of a spectrum of word knowledge, also known as the incremental continuum (Henriksen, 1999; Meara, 1990, 1997; Melka, 1997). Melka (1997) agrees that the reception-production distinction is more of a gradual cline in the mental lexicon than a clear-cut division. Receptive knowledge of a word is inadequate and incomplete information to activate the word in contexts. This receptive knowledge varies in degrees, yet all gradually evolve to productive levels (Henriksen,

1999; Haastrup & Henriksen, 2000; Melka, 1997). Some components of the word may be ripe for production, while others may remain receptive. Only when receptive knowledge reaches a certain threshold of richness can it be retrieved and produced naturally. However, Read (2000) doubts the possibility of locating the point where receptive knowledge grows into productive mastery in the development process. Meara (1990, 1997) holds that the relationship between receptive and productive vocabulary knowledge resides in the differences of lexical links in the mental lexicon. In other words, receptive and productive words are stored in the lexical organization composed of different types of interconnected words. Productive words can be easily accessed and linked anywhere in the mental lexicon (Meara, 1990) or lighted up by internal stimuli (Schmitt, 2010; Zhong, 2014). By contrast, receptive words can merely be activated by outside stimuli, such as the words seen in texts or heard in speeches. Accordingly, there is no need to draw a clear demarcation between receptive and productive word knowledge. A receptively known word can also be activated for productive use if there are rich links in the particular context.

Ample research evidence has shown that learners generally have a larger receptive vocabulary knowledge than productive vocabulary (Fan, 2000; Laufer, 1998; Laufer & Paribakht, 1998; Read, 2000; Waring, 1997a; Webb, 2008a; Zhong & Hirsh, 2009). The former grows faster than the latter (Laufer, 1998; Schmitt, 2010; Webb, 2008a). This is hardly surprising because word knowledge of the productive character is more difficult to come by at a higher level than the receptive kind (Schmitt, 2010). Productive mastery is usually found in varied contexts and involves more lexical components. These components take more time and effort to develop (Schmitt, 2014). Furthermore, learners have to access productive word knowledge automatically, subject to time restrictions and reacting speed (Milton, 2009). Consequently, EFL learners often demonstrate less productive than receptive word knowledge (Laufer & Paribakht, 1998; Milton, 2009; Schmitt, 2010, 2014; Waring, 1998). One exception is Zhong and Hirsh's (2009) finding that after a four-month classroom instruction among a group of intermediate Chinese EFL learners, they revealed more productive word knowledge than receptive. This means that the learning needs and specific learning tasks may be the major drives for vocabulary acquisition. In this respect, Webb (2005, 2009) has demonstrated the superiority and efficacy of productive word learning for

gaining more both receptive and productive word knowledge. This parallels the previous findings in Griffin (1992) and Ellis and Beaton (1993) that productive learning of L1-L2 word pairs, though more difficult, is an effective method for EFL vocabulary development.

2.1.3.2 Importance of Productive Vocabulary Knowledge

Productive vocabulary knowledge is more advanced than receptive kind insofar as the former embraces and exceeds the latter. Sustained studies have been here to support the importance of productive vocabulary knowledge in different language performances (Cheng & Matthews, 2018; Choi, 2017; Kilic, 2019; Koizumi & In'nami, 2013; Kim, Crossley & Kim, 2022; Milton, 2013; Shi & Qian, 2012; Pellicer-Sánchez, 2019; Yu, 2009; Warnby, 2023).

It is hard to overstate the indispensability of word knowledge in English speaking and writing abilities. For instance, Yu (2009) and Kilic (2019) offered empirical insights into vocabulary knowledge being an essential indicator of general EFL writing and speaking skills. Yu (2009) examined 200 compositions with different topics and 25 interviews focusing on oral performances. From the data analysis, she established significant correlations between vocabulary diversity, one type of productive word knowledge, and participants' overall language proficiency. Their vocabulary knowledge profoundly impacted the judgment of writing quality and spontaneous speaking abilities. In addition, Choi (2017) provided SEM (Statistical Equation Modelling) data pointing towards the direct contributions made by productive word knowledge to L2 writing, while receptive knowledge merely plays an indirect role at best. Karakoç and Köse (2017) empirically showed that the more productive knowledge learners have, the more successful they will be in L2 writing performance due to the significant impact this knowledge has on writing.

Similarly, Kilic (2019) evidenced that productive size and depth of word knowledge strongly predict EFL learners' writing and speaking performances, particularly writing (26%). In contrast, receptive size adds relatively little to the variance (2% for writing and 1.4% for speaking). Productive word knowledge has also been posited to play a critical role in speaking proficiency. A number of studies say a lot about the contributions made by productive knowledge to speaking, such as the L1-L2 word

translations (Koizumi & In'nami, 2013), productive associative knowledge (Uchihara & Saito, 2016), and the form-recall knowledge elicited by the PVLT (Laufer & Nation, 1999) (De Jong et al., 2012). Nevertheless, it should be noted that the need to make productive vocabulary a high priority in L2 learning is not only because of its importance in productive skills. It carries equally similar weight in receptive reading and listening skills. For example, Cheng and Matthews (2018) found that productive orthographic and phonological knowledge accounts for the most variances in reading and listening scores. The result contradicts the previous finding by Qian (2002), who postulated that receptive size substantially explains (54%) the EFL reading ability. This further reiterates the significance of focusing more on productive components of the word in not only productive but also receptive performances.

Despite the merit found in productive word knowledge, it has not yet been paid as much attention as receptive size in assessment (Nizonkiza & van den Berg, 2014; Pearson, Hiebert & Kamil, 2007). Although Schmitt (2010) advocates that vocabulary knowledge in production should be the best way to assess vocabulary depth, he also acknowledges that little has been known and done regarding this knowledge. There are some factors at play for this paucity of research. The major reason is that productive word knowledge involves more lexical components, more complex measuring processes, and less controllable contexts and results (Read, 2000; Schmitt, 2014). Nonetheless, the existing problems should not justify productive knowledge being disregarded or even discarded in research and assessment. As Laufer and Nation (1999, p.45) pointed out, assessing productive word knowledge in addition to receptive size allows us to inquire into such questions as a) the development of different components of word knowledge over time; b) the interrelationships among various word components; c) correlations between different word components and vocabulary acquisition and teaching.

In addition to receptive size, the current study scrutinizes various components of productive knowledge of a word vis-a-vis language production. In other words, word usability is central to this study instead of metalinguistic word knowledge. Some researchers have formulated an overarching model in an attempt to articulate the mechanism and development from vocabulary knowledge to lexical ability.

2.2 Models from Word Knowledge to Word Use

Vocabulary knowledge components have been elaborated in different dimensions, such as vocabulary size, depth and receptive-productive distinction. All these components of word knowledge have been established to closely tie in with language skills and can be instrumental to successful language use. However, the relationship between vocabulary knowledge and competence is such that the line between the two constructs is relatively blurry, and competence sometimes is viewed as part of knowledge (Milton & Fitzpatrick, 2014). It is tempting to conclude that accumulating as many components of the word as possible into the mental lexicon is the ultimate goal of acquiring vocabulary. However, this knowledge, be it in the form of lexical information stored in memory or a well-organized lexical network, is declarative word knowledge (Albrechtsen et al., 2008). This conceptualized model is sometimes categorized as the "traditional impoverished approaches that concentrate on building a store of definitional meanings" (Robinson, 1988, p.1). Simply put, the knowledge components are information stored in the mental lexicon, descriptive and static in nature, rather than procedural knowledge that learners know how to readily use and perform implicitly and unconsciously. This declarative knowledge constitutes the prerequisite, whereas the procedures to use it serve to realize and achieve this knowledge. Thus, probing into word knowledge in such procedures is necessary because lexis should become closely attuned to users' manipulation instead of analytic descriptions isolated from any context. Indeed, there is a distance between the "know that" and "know how" of a word (Daller, Milton & Treffers-Dallar, 2007; Meara, 1996; Milton, 2013). The particular essence of mastering a word is using it in communication rather than knowing the metalinguistic word knowledge (Milton, 2009; Schmitt, 2014).

2.2.1 Chapelle's Model of Vocabulary Ability

Vocabulary has been conceptualized as a kind of linguistic ability by Chapelle (1994), revolving around the "capacity for language use in context" (p.163). In this definition, Chapelle (1994) draws on the "interactionalist construct definition of communicative language ability" by Bachman (1990, p.108), suggesting that a) vocabulary use deeply resonates with learners' lexical characteristics (vocabulary knowledge, processes to use the words, and communicative strategies), contextual factors and the interactions

between them; b) vocabulary ability includes vocabulary knowledge and the capability to use the knowledge in natural production. Therefore, three components are featured in Chapelle's (1994, p.164) definition of vocabulary ability: (1) the context of language use; (2) vocabulary knowledge and processes; and (3) the metacognitive strategies required for vocabulary use in context.

The context of language use

According to Chapelle (1994), the context from an interactionalist perspective refers to the linguistic situation in which language production is carried out and the broader social and cultural contexts constraining the words used by language users. More specifically, lexical context comprises three major elements: field, tenor, and mode (Halliday & Hasan, 1989). Field denotes the subject matters and linguistic actions occurring in a specific setting; Tenor means the particular speakers or writers and their relationships and objects; Mode points to the linguistic channel and genre of the language selected in the context. The differences of field, tenor and mode affect the word meaning and lexical choice. For example, the word "propaganda" means the false or incomplete information used by the government or political groups in the English context. In contrast, the corresponding translation of this word in the Chinese context has no such negative referent. Another example is that the words chosen by learners for writing a report of an English novel may differ markedly from those used for academic writing purposes.

The current study is set in the context of Chinese EFL learners' productive word knowledge in L2 writing. For starters, academic words are solely investigated as opposed to other general and specialized words, in that academic vocabulary is highly relevant to the comprehension and production of L2 university compositions (Nation, 2013; Skjelde & Coxhead, 2020). Thus, this study needs to emphasize academic word learning and teaching in the Chinese tertiary context. In addition, the present study capitalizes on authentic L2 writing topics selected from the writing section of the International English Language Testing System (the IELTS), putting the measures of academic words in actual writing contexts.

Vocabulary knowledge and fundamental processes

The components of vocabulary knowledge and fundamental processes shed light on

the declarative and procedural stages of vocabulary development, respectively. These stages all center around a particular context. The three declarative stages consist of vocabulary size, knowledge of word characteristics and lexical organization, whereas the procedural stage is vocabulary access.

Chapelle (1994) defines vocabulary size as how many content words learners know within a specific context. Contrary to the popular belief that vocabulary size can be estimated with an absolute number, the interactionalist definition specifies that the size of learners' vocabularies is dynamic in growth; thus, it should be defined and measured according to the particular context where the vocabulary is used. The second dimension refers to the multiple components of vocabulary knowledge similar to Nation's (2001, 2022) taxonomy, including phonological, orthographic, morphological, semantic, syntactic, pragmatic and collocational characteristics. Some of these components are usually acquired in an incorrect, incomplete and unanalyzed manner in the early stages. For example, learners might mistake the lexical representations of the word and make errors in mapping word form and meaning. In some cases, they may not have an in-depth understanding of the word and use it smoothly across contexts. Thus, the knowledge of word characteristics should be construed within and underpinned by a particular context. Regarding lexical organization, Chapelle (1994) theorizes that it is well placed to describe the dynamic state of morphemes and words in the mental lexicon and how they are related to one another. Such a vocabulary organization is not definite and fixed at different stages of development. Moreover, different contextual factors significantly impact language users' lexical connections between words.

The last dimension specifies the fundamental procedural processes to access the word: a series of cognitive activities of using the word receptively and productively. For example, learners may first input and encode the form and meaning knowledge of a word in their short memory, then access its structural and semantic property from the mental lexicon, integrate its lexical-semantic content in the text and parse and compose its morphological components. These cognitive activities of using a word are also deeply rooted in a certain context, such as reading a text.

The metacognitive strategies required for vocabulary use in context

In addition to the cognitive knowledge and activities needed to manipulate word knowledge, learners' metacognitive strategies are also of importance to use the word in real situations. This component tends towards linguistic and sociocultural lexical strategies language users adopt in different communication, which in Bachman's (1990) definition is strategic competence. Specifically, this ability might include setting communicative goals, planning language use such as word choice or paraphrasing, controlling linguistic execution, changing topics, and avoiding meanings. According to Chapelle (1994), these metacognitive strategies are needed, especially when learners with relatively limited vocabulary wish to function well and achieve goals in communication. As learners' vocabulary size enlarges and word knowledge expands, such strategies would be used without notice and become less of an issue. Nonetheless, lexical strategies always play a significant role in the successful control of word performance.

2.2.2 Robinson's Model from Vocabulary Knowledge to Performance

Robinson (1988) is more concerned with procedural word knowledge and how learners could mobilize various components of word knowledge to perform in discourse negotiation and rise above lexical problems in communication. He (1988, p.4) maintains that the preoccupation with a lexical repository that is static and declarative while disregarding the "actional nature" and "fluid meanings" of words is inadequate, mistaken and even misrepresented. Instead, learners need to know how to establish and realize the relations and meanings of different word components amid actual contextualized use. Moreover, the categorizations of vocabulary components fabricated by Nation (2022) and others merely represent an idealized, static taxonomy. There are some grounds for concern, notably regarding the pragmatic ability and communicative competence brought by such lists of vocabulary knowledge. Therefore, based on Canale and Swain's (1980) four areas of communicative approaches, namely grammatical competence, discourse competence, sociolinguistic competence, and communication strategies (p.27), Robinson (1988) proposed a model of vocabulary competence, integrating word knowledge of declarative components and the procedural abilities which learners could draw upon to move their vocabulary

knowledge to word use (see table 3).

Table 3 Vocabulary knowledge and competence-based on Robinson's sketch

Lexical communicative competence			
Vocabulary knowledge			Vocabulary skills
That (declarative)		How (procedural)	
linguistic	social	cognitive	
knowledge		using and learning	
strategic		reception	
discoursal		production	
sociolinguistic		communication	
grammatical			
potential communication			actual communication
non-time constrained			time-constrained
static			dynamic
competence			performance
idealized			realized
capacity			manifestation

Robinson (1988) takes issue with the divorce of static vocabulary knowledge from actual use. In this model, he postulates what word competence involves, conjoining various real conditions such as time constraints, actual contacts and lexical strategies. Furthermore, nuanced ways of converting knowledge to procedures are also illuminated under the guidance of Canale and Swain's (1980) communicative framework. Nonetheless, many of the arguments and methods provided are merely a descriptive solution, devoid of empirical experiments. He never touches upon any reasonable measure or assessment criterion teachers and researchers could use to assess the procedural components of lexical competence. Although Robinson calls attention to the fact that lexical knowledge should be acquired in tandem with performance, his model is merely represented in crude terms. All this warrants more clarification and validation in vocabulary assessment.

2.2.3 Daller et al.'s Three-dimensional Lexical Space

Daller et al. (2007) resort to an analogy or metaphor of three-dimensional lexical space to drive home the complex and multifarious nature of vocabulary knowledge and abilities. Vocabulary knowledge is not merely perceived as a size and depth dichotomy. Instead, the fluency or automaticity of this knowledge represents yet

another dimension (see Figure 2).

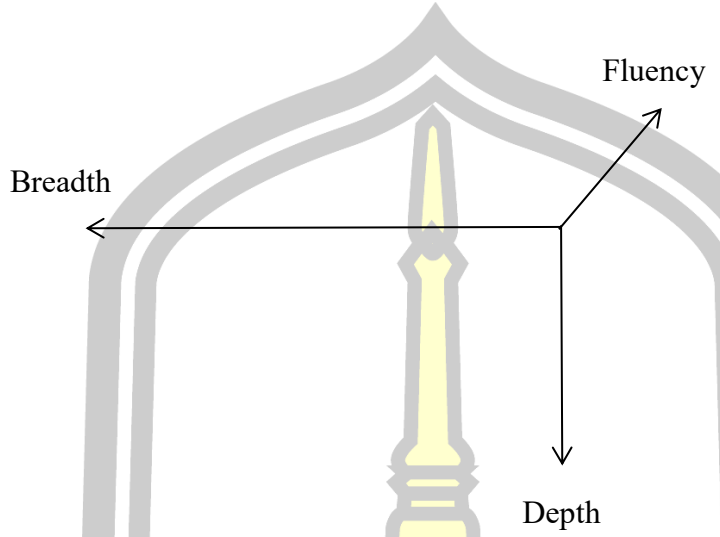


Figure 2 The lexical space: dimensions of word knowledge and abilities (Daller et al., 2007, p.8).

Each dimension in this cubic space model signifies a component of knowing or using the word. It is well placed to distinguish learners' different types of word knowledge: words that can be quantifiable at the form-meaning level, words that have been known more deeply and comprehensively or connected with others, and words that can be recalled readily to use in speaking and writing. In other words, the size or depth of words is not merely the element that counts decisively in knowing a word. The speed or fluency to retrieve the words in real scenarios is just as important (Qian & Lin, 2020). This crystal-clear demarcation of abstract word knowledge makes it easier to locate where learners are in knowing a word, not least the EFL learners. Indeed, it particularly holds true in Chinese EFL word learning since numerous Chinese students seemingly have a large vocabulary size but a limited capacity to use these words in communication. While others with a relatively smaller vocabulary yet more exposure to real English might possess considerable fluency to use these words idiomatically and appropriately.

However, the lexical space could merely emerge as a metaphor to scaffold a thorough perception of word knowledge which is inherently difficult to pin down. It implies that each dimension in this space could foster and function along one direction without regard to the other two directions. Such a lexical space can hardly exist within

learners' mental lexicon, and it is impossible to develop and operationalize the interwoven word knowledge components categorically separate. If learners have substantial knowledge or capability in one direction, it is unthinkable that the other dimensions would not be influenced.

2.2.4 Henriksen's Three Dimensions of Vocabulary Development

Henriksen (1999) attempted to paint a clearer picture of vocabulary development along three interrelated dimensions: two declarative dimensions --- the partial to precise word knowledge and depth of vocabulary knowledge; and one procedural dimension --- the receptive to productive competence. The triple-dimension model is built on an established consensus that an incremental continuum exists in learners' mental lexicon. Vocabulary knowledge progresses from incomplete, shallow knowledge to in-depth comprehension and word production. On the part of the first continuum, the partial to precise dimension refers to a process of word meaning intake and form-meaning mapping. Learners at this stage build extensional links by labeling and packaging different word meanings, developing from a vague recognition to an accurate mastery. With new items added in extension, learners deeply manipulate the new words and achieve depth of knowledge by reordering and restructuring their lexicon to create intensional links, also known as the network building process. This associative network typically develops along with three phases coined as "notice-analyze-integrate" (p.225). In this cognitive process, the new word gradually finds its place and achieves depth in the lexical field (Haastrup & Henriksen, 2000).

The third continuum is the control and procedural dimension as opposed to the first two declarative dimensions. Both receptive and productive language skills are needed in this continuum to readily mobilize and access the precise word meaning and connected knowledge in the metal lexicon. Henriksen (1999) concurs with Melka (1997) and Meara (1990, 1997) that the distinction between reception and production needs to be regarded as a continuum rather than a dichotomy in that many words are initially acquired receptively and subsequently grow into production. Moreover, the automaticity by Meara (1996) should become a hidden yet critical component of receptive and productive competences. It has been argued that the three dimensions in question correlate closely with one another. Nevertheless, because the current study

taps into vocabulary knowledge under Nation's (2022) multi-component framework, the reception-production division is considered a dichotomy for separate, multiple tests. Moreover, the automaticity or fluency of word knowledge is not included as a knowledge component in the present study. This will be controlled by allowing as much time as the participants need to complete these tests.

Henriksen's (1999) seminal three-dimension of vocabulary knowledge helps to cast light on the complex process of vocabulary acquisition and development. However, two questions merit more attention about her model. For one thing, the first two dimensions are termed an "interlanguage semantization process" (p.308), which is simultaneously developed on an ongoing basis. Since the word learning is concomitant with the system-changing processes, it is hard to see the point of drawing a clear line between the two continua assuming that such a line could be drawn in practice. For another, it seems that more space has been devoted to the declarative components of meaning understanding and network building. This begs the question of how the benefits of size and depth knowledge are felt in the procedural continuum of receptive and productive use. As the author realizes herself (1999), more empirical studies need to be done to operationalize these continua of vocabulary development. Few such studies, however, have capitalized on this model to clarify and validate all these components (Schmitt, 2010).

In retrospect, several models by previous researchers venturing to outline the nuanced components and phases from vocabulary knowledge to lexical performance have been identified and discussed in this section. Robinson (1988) underlined the dynamic and fluid nature of vocabulary knowledge and argued that lexical knowledge components need to be conceptualized and defined during real procedures in use. Unlike the other models, Daller et al. (2007) provided us with a cubic space to liken different dimensions of vocabulary knowledge and abilities, suggesting that the three constructs in question are developed towards their own axis, independent from one another. In comparison, Henriksen (1999) and Chapelle (1994) emphasize the interconnections among lexical knowledge and competence components. Nevertheless, contextual factors are at the heart of Chapelle's model, going through all vocabulary development stages, and the context in Henriksen's model is not touched on as a particular concern.

In addition, whereas Henriksen conceptualizes vocabulary depth as a lexical organization, Chapelle perceives word depth as a construct combining both the concept of lexical network and the various components of word knowledge. Qian (2002) agrees that Chaplle's word characteristics are similar to the multiple word components in Nation's (2022) comprehensive list.

These models serve to navigate the way through the perplexing, confusing constructs of vocabulary knowledge and competence. They also have made it more accessible for researchers to conceptualize and measure how vocabulary knowledge relates to word use and to what extent lexical knowledge has been mastered and used by learners.

2.3 Vocabulary Components Operationalized in the Current Study

Drawing on Nation's (2022) comprehensive framework of vocabulary knowledge, Coxhead (2007) specified what components of a word are necessary to be involved in producing a piece of an essay (See Table 4). Coxhead (2007, 2011, 2012) conducted a series of empirical studies under the multi-component model. Her findings suggest that many components, such as collocation, association, and register, are also critical factors in using a word in addition to the basic form and meaning connections. L2 university learners reported that they have to mobilize knowledge of various lexical components during the writing process. According to Coxhead (2007), all the components tabulated in the table, including form, meaning and use, might be related to word use and could contribute to L2 writing. Nation (2022) also suggested that it is of high relevance to look into how the various components of a word fit into the language production process. This provides a solid theoretical foundation for this study probing into the relationship between different word components and L2 writing proficiency. As such, the current study measures six word knowledge components, including written form, receptive and productive form-meaning links, word associations, productive derivatives and collocations.

Table 4 Knowledge required for the production of a word in writing (Coxhead, 2007, p. 332).

Form		How is the word written and spelled?
Meaning	Form and meaning	What word form can be used to express meaning?
	Concepts and referents	What item can this concept refer to?
	Association	What other words can we use instead of this one?
Use	Grammatical function	In what patterns must we use this word?
	Collocations	What words or types of words must we use with it?
	Constraints of use	Where, when and how often can we use this word?

2.3.1 Form and Meaning

Coxhead (2007) argued that the written form of a word is one of the components capturing the most attention from L2 writers. That is, how to correctly spell a particular word is of essence in the writing process. Indeed, according to Nation (2022), spelling skills not only reflect the issues arising from other word components but also impinge on learners' writing strategies. Poor spelling causes damage to word use, such as using limited vocabularies or favoring frequent words and avoiding words that are difficult to spell. Laufer and Nation (1999) developed the Productive Vocabulary Levels Test (PVLT) under the idea that L2 learners may know a certain word and use it in a sentence only when required. However, they might avoid using it in their own right in a free writing task mainly because of incomplete knowledge of word form. Schmitt (2010) deployed the PVLT in a writing context and suggested that the PVLT can be a form-recall measure, focusing on form-meaning connections.

Form-meaning connections were reported as one of the decisive factors L2 writers consider in L2 writing (Coxhead, 2007). This is hardly surprising since form and meaning have long been identified as the most fundamental elements in any mode of language production (Laufer & Goldstein, 2004; Levitzky-Aviad & Laufer, 2013). Laufer (1994, 1998) noted that vocabulary size (form-meaning links) is the most striking difference between native and nonnative speakers because it signals the number of words they can control in free language production. Moreover, a larger vocabulary, in a sense, also contributes to a higher level of lexical richness (Crossley et al., 2011, 2015; Laufer & Nation, 1995). However, form-meaning connections in productive contexts can sometimes be operationalized as the L1-L2 word pairs knowledge (Waring, 1997; Webb, 2009). This is more so in L2 writing because L2

writers primarily map an L2 word form to its L1 translation, not directly to the word meaning (Jiang, 2000, 2002; 2004). In other words, when learners wish to use an L2 word, they usually resort to the L1 word translation before recalling the corresponding word form. Together, form-recall knowledge, vocabulary size, and L1-L2 translations are all of value to measure in terms of form-meaning links in L2 writing. Concepts and referents in the "meaning" category will be excluded because measuring the multiple shades of meaning is complicated, if not impossible, in lexical production. Alireza (2017) substantiated that participants find it challenging to provide multiple meanings even after explicit instructions. Therefore, the current study will take up the constructs of form-meaning links: form-recall word knowledge, vocabulary size and L1-L2 translations, and limit to measuring the core meanings which L2 learners are more likely to know and provide (Nation, 2013).

2.3.2 Word Associations

Generally, vocabulary associations come in three different relationships between words: paradigmatic (e.g., synonymy), syntagmatic (e.g., collocation) and analytic (e.g., words bearing a key meaning of the target word) (See Section 2.1.2.2). Nation (2001) and Coxhead (2007) limited the association knowledge in language production to synonymy, that is, "what other words can we use instead of this one"? Indeed, Nation (2013, 2022) asserted that synonymy represents the most pervasive and important relationship between words. L2 writers in Coxhead's (2007) case study also described association knowledge as a key element to enriching their vocabulary, so they could select the most appropriate word to fit into the context. Thus, comparing a bunch of semantically related words and selecting the best one to convey the meaning can be one of the main focuses of the L2 writing process. Zhong (2016) found that association knowledge shares much variance with word meaning in predicting productive word use in sentence writing. It can be an indispensable process to compare synonyms to express the intended meaning, even if the retrieval and comparison process cannot be revealed in the produced sentence. Moreover, research also shows the high correlation between L2 writing proficiency and synonymy used as one of the cohesive devices. The higher-level learners used more synonyms in their writing to promote writing cohesion than the lower proficiency learners (Ferris, 1994; Liu & Braine, 2005). However, Schmitt and Meara (1997) cautioned that L2 learners paint

an unpromising picture about this knowledge since they could merely supply half of the associative knowledge even for words they rated as known. Thus, word association, mainly the synonymy as one of the major components of vocabulary depth knowledge in meaning comprehension, warrants more research attention, especially in language production (Meara, 2009; Schmitt & Meara,1997).

2.3.3 Productive Derivatives

The grammatical functions refer to the patterns of a word L2 learners use in language production. These grammatical patterns embrace derivative and inflectional forms in actual use (Nation, 2013; Sukying, 2022). The derivative words, also known as word class, are of particular concern because these derived forms make up the word family of a word (e.g., happy, happily, happiness, unhappy). Comparatively, the inflectional forms point towards the grammatical rules used with a noun or verb, mainly referring to the person, tense and number (Sukying, 2022). The inflectional forms will be excluded from this study because this knowledge barely adjusts the syntactic category of the word in the sentence it is used, thus merely making small changes to the word (Sukying, 2022). Schmitt and Meara (1997) indicated that inflectional scores ranked high and stable in their tests mainly because inflected knowledge is grammatically rule-based and can be generalized. On the contrary, derivative forms are idiosyncratic and involve the prefix, base word and suffix, leading to the word formation. Identifying and manipulating this morphological knowledge not only helps learners acquire a rounded understanding of the internal structure of a word but also facilitates their language communication (Sukying, 2022). Nation (2013, 2022) stressed that the choice of derivative forms, particularly of a verb, can determine the grammatical structure of the whole sentence. Zhong (2014) also agrees that derivative knowledge in context determines where to put the word in a sentence and what other words should be used with it. That is, the collocation knowledge is highly relevant to the choice of the derivatives. As such, productive derivative knowledge in L2 writing should be one of the main testing goals in the current study.

2.3.4 Collocations

Collocations or multi-word units have been singled out as a subcategory of word use, which reiterates its productive attribute of vocabulary since the target word and the

immediate neighbouring words make up a local context (Nation, 2013, 2022). Learners may know a word but cannot use it properly in context mainly due to their lack of collocation knowledge (Webb & Sasao, 2013). Nation (2013, 2022) pointed out that collocation is essential in written and oral production because it makes learners sound more native and fluent. Storing a large number of ready-made lexical sequences can reduce their learning burdens and shorten their choosing time in actual use. Previous empirical studies have yielded compelling evidence as to the critical role collocations play in L2 writing. For example, collocations are deeply related to more sophisticated word use to promote L2 writing quality. Yoon (2018) and Kyle and Crossley (2015) unraveled the centrality of multi-word units (frequent bi-grams and tri-grams) in L2 argumentative essays. Proficient L2 writers would prefer to use a wide variety of phrases common in native writing, while low proficiency writers produce far less of them (Crossley, 2020; Kyle & Crossley, 2015). If the collocations can be used properly and accurately, they would, to a large extent, contribute to the overall writing quality and rating scores. Crossley, Salsbury, and McNamara (2015) found that collocation accuracy predicted a considerable amount of variance, as much as 84% of the holistic scores of L2 written samples. Therefore, it can be safe to conclude that native-like collocations, as a critical lexical component for high lexical proficiency and L2 writing quality, are worth assessing in a multi-component vocabulary study.

However, the constraints of use was excluded in the current study because this subcategory mainly embraces sociolinguistic and cultural factors in vocabulary use, such as the word register (Nation, 2022). To date, there have been no validated instruments to measure these factors (Schmitt, 2010; Zhong, 2014).

2.4 Vocabulary and L2 Writing

2.4.1 The Role of Vocabulary Knowledge in L2 Writing

That vocabulary knowledge and vocabulary use are different constructs that need to be organically integrated into L2 lexical acquisition have been discussed, as is amply demonstrated in the above models. The former constitutes various types of lexical knowledge stored in the mental lexicon, while the latter points towards the actual manifestation of this knowledge in natural production (Laufer, 2005; Laufer &

Goldstein, 2004; Levitzky-Aviad & Laufer, 2013). One of the key arenas where such vocabulary use plays a critical, if not the most decisive role, is L2 writing. Indeed, numerous researchers have empirically evidenced the centrality of vocabulary knowledge in L2 writing (Astika,1993; Bestgen, 2017; Johnson et al., 2016; Kim, Crossley & Kim, 2022; Li, 2022; Morris & Cobb, 2004; Staehr, 2008; Sukying, 2023). Many factors are indeed at play in the complicated writing process; a rich and diverse vocabulary can be the most fundamental component of a quality and effective L2 essay (Kim et al., 2022; Qian & Lin, 2020; Olinghouse & Leaird, 2009; peng et a;., 2023). For instance, among the variables that contribute to the holistic quality of L2 writing (content, organization, vocabulary, language use and mechanics), vocabulary was found to be the most striking predictor, accounting for more than 83% of the variance even if it merely has a relative weighting of 20 points to the total score (Astika, 1993). Similarly, Kim et al. (2022) found that L2 vocabulary knowledge predicts L2 writing proficiency better than the other variables: L1 reading and writing abilities and educational background. They further indicated that robust vocabulary knowledge enables learners to readily apply writing skills acquired in L1 writing in their L2 writing practices. Supporting this finding in one of their recent studies, Johnson et al. (2016) revealed that the accurate use of high-frequency words plus a repertoire of less frequent vocabulary was associated with stronger L2 writing performance.

Qualitatively, Leki and Carson (1994) surveyed how to better evaluate and assist ESL learners in their academic writing. It was reported that language issues, vocabulary and grammar, are of particular concern in their writing. As articulated by one of the students, in many cases, “I simply run out of words necessary to express what I am actually thinking, the word (chosen) would only be semi-fitting with my thoughts” (p. 91). This urgent need for vocabulary and grammar signals their demand for a speedy process of the language in timed writing. It is more so in L2 writing because L2 learners have a relatively smaller working memory capacity for retrieving and coordinating cognitive resources than their L1 peers (Llach, 2011; Weigle, 2002). Their short working memory makes it hard to translate their ideas, images, and experiences into linguistic forms stored in their long-term memory. This constraint, coupled with the lack of linguistic knowledge, hampers L2 writers from paying as

much attention to higher-order issues such as the content and organization (Schoonen et al., 2003; Wu, Dixon, Sun, & Zhang, 2019). L2 writers tend to plan less, organize fewer ideas, and edit less for content. Instead, they have to make “lengthy searches” for the right word and syntactic forms (Weigle, 2002, p.36). The corollary of this focus on language is that L2 writers often fail to truly represent their ideas in text production, reducing the written product to something that is not well suited to their initial intention.

Nevertheless, vocabulary figures even more prominently in L2 writing than grammar (Qian & Lin, 2020) because grammar is more related to the text structure, whereas vocabulary can make or break the whole communication (Gass & Selinker, 2008). Without vocabulary, grammar becomes meaningless and abstract linguistic rules that express nothing in communication. Hence, a rising tide of questions previously explained by grammar is now moving towards vocabulary (Bates & Goodman, 1997). More specifically, vocabulary is a major force driving grammatical structures (Bates & Goodman, 1997), which is depicted by Levelt (1989, p.181) as the “lexically driven” procedures. In this process, vocabulary serves as a critical mediator between the representations and grammatical encoding. That is, a sentence can be produced only when there is a set of proper lexis behind the triggering message (ideas) and grammatical properties. Thus, the chosen word determines a sentence's grammar, morphology and phonology (Nation, 2022). For example, Vögelin et al. (2021) and Peng and colleagues (2023) found that a good manipulation of lexical sophistication and diversity in L2 argumentative writing positively impacted teachers’ assessment of grammatical knowledge in analytic scoring. Accordingly, lexical errors made by L2 writers are many and varied, causing more severe and disruptive results than grammar. According to Llach (2005, 2011), these errors vary in type and can be a strong predictor and assessment criteria for the holistic quality of L2 writing. Chief among these are semantic lexical errors, namely, the wrong word choice, which most undermines the intelligibility of the meaning (Llach, 2005). Santos (1988) also evidenced that native professors could, by and large, comprehend the EFL writing while frowning upon the lexical errors as unacceptable. Being the driving force of language production (Gass & Selinker, 2008), vocabulary knowledge, as such, has been widely acknowledged as the single most crucial component of linguistic

competence in L2 writing (Bestgen, 2017; Hyland, 1997; Qian & Lin, 2020; Schoonen et al., 2003).

There is good reason to draw attention to vocabulary in L2 writing, notably for those advanced EFL learners because many of them have acquired a good knowledge of grammatical rules without a rich vocabulary. They need to become as fluent and automatic as possible at lexical use so that more time and cognitive resources can be devoted to higher intellectual demands (Leki & Carson, 1994). This speedy and accurate response to word use in timed writing is challenging as it takes a good command of vocabulary depth knowledge. L2 writers need to first thoroughly understand the components of a single lexical item, as listed by Laufer (1990) and Nation (2022). These individual words also need to be connected in every possible way by activating other related words in the mental lexicon (Meara, 1996). Only when these knowledge components are intensely mobilized and organized in real-time can a quality sentence in L2 writing be generated smoothly. However, the picture seems not encouraging regarding EFL/ESL learners effectively using such depth knowledge in their L2 writing. For example, it was reported that nonnative English users are increasingly subject to the same strict writing criterion as their native peers (Angelova & Riazantseva,1999). Unfortunately, their vocabulary size is much smaller than previously thought, and lexical variation is much less than is usually reported, which impairs their language comprehension and production (Treffers-Daller & Milton, 2013). This places them at a disadvantage and fuels their difficulties in meeting academic standards. Such a survey as this was also administered by Hyland (1997) and Evans and Green (2007), and the data collected indicates the central role vocabulary knowledge plays in L2 writing, hence the importance of a rich and fluent vocabulary for L2 writers to break down the most demanding barrier to academic success.

2.4.2 Relationship between Vocabulary Knowledge and L2 Writing

As has been well demonstrated above, vocabulary knowledge, by proxy, can represent the critical linguistic competence in L2 writing and determine the quality of the writing. L2 writing, in turn, facilitates acquiring a depth of vocabulary knowledge and word retention. The strong two-way relationship between vocabulary knowledge and

L2 writing is multidimensional and mutually supportive (Laufer, 2013; Lin, 2015; Llach, 2011).

Vocabulary knowledge of different components contributes to L2 writing. First and foremost, it has been well-documented that the size of vocabulary can predict L2 writing performance (Albrechtsen et al., 2008; Kim et al., 2022; Qian, 2002; Staehr, 2008). Staehr (2008) provided data that the correlation between writing proficiency and vocabulary size is as high as 73%. This means that learners' large vocabulary size can explain more than half of the variance in L2 writing capability. Perhaps more convincing are the two lexical characteristics in writing reflected or affected by vocabulary size, namely, the lexical frequency and lexical variation, also coined as the productive size (Crossley, Salsbury & McNamara, 2015; Laufer, 2013). L2 learners with extensive vocabulary could use more infrequent and less frequent words. Similarly, a large vocabulary size also means that learners can use more diverse words instead of repeated words in writing as measured by lexical diversity (variation). Engber (1995) established the relationship between L2 writing quality and lexical richness and found the highest correlation between error-free lexical variation and writing scores. Moreover, Engber (1995) also stressed that lexical correctness or accuracy is more essential for advanced learners. Their writing may carry less diversity but is still high quality, with the exact word being precisely used. Lexical accuracy, therefore, means more to advanced EFL learners and emerges as a major writing assessment criterion both quantitatively and qualitatively (Hyland, 2003; Llach, 2011; Polio, 1997, 2001; Peng et al., 2023; Weigle, 2002).

Notably, accurate word use in writing induces deeper and more sound word knowledge. Because many EFL learners lack a thorough mastery of the word, they sometimes attempt to shun difficult words so that they might avoid lexical errors. EFL writers reported that they were not trying to find the precise words in the writing process; instead, they were more inclined to settle for words that express vague, similar meanings (Albrechtsen et al., 2008). Levitzky-Aviad and Laufer (2013) evidenced that the rate of productively using words lags far behind that of vocabulary knowledge growth. EFL learners still refrain from using low-frequent words and collocations until university years or later. It can be concluded that enlarging

vocabulary size is the prerequisite for L2 writing purposes, whereas the accurate retrieving of these resources in timed writing is simply as important. Therefore, it can be an urgent need to implement lexical instructions focusing on more components of vocabulary and use in L2 writing.

A growing body of research has inquired into pre-writing vocabulary instructions and the effects thereof on L2 writing (Coxhead & Byrd, 2007; Lin & Hirsh, 2012; Lee, 2003; Lee & Muncie, 2006; Webb, 2009). Researchers virtually agree that lexical instructions, implicit and explicit, are necessary to promote the quality of L2 writing (Csomay & Prades, 2018). For instance, Lee (2003) and Lee and Muncie (2006) observed significant improvement in the writing quality with more low-frequency and newly learned words being used after the vocabulary instruction. Native speakers judged post-instruction writing to have better sentences and richer content because of using high-level words. Moreover, Lin and Hirsh (2012) compared the effects of incidental reading and explicit instruction on vocabulary intake vis-a-vis L2 writing. They found that direct vocabulary teaching can better help L2 writers correctly use the new words than incidental reading activities. Webb (2009) supported this finding that EFL learners who received productive pre-learning of L1-L2 word pairs can use more of the target words in writing picture descriptions. These results make sense since words in receptive knowledge can hardly flash through the mind in actual use for not being profoundly processed or encountered enough. It seems reasonable, if not conclusive, that focus-oriented teaching is more effective than context-oriented learning for productivity (Laufer & Shmueli, 1997). Following this line, researchers have advocated that language instructors should be encouraged to provide more opportunities to focus on the lexical items in classroom teaching (Coxhead & Byrd, 2007) or intensify teacher elicitation, explanation, discussion and negotiation, and above all, the word manipulation in L2 writing (Lee & Muncie, 2006).

On the other hand, writing can be a sound facilitating tool for gaining more vocabulary knowledge and profound knowledge in different dimensions (William, 2012). Such knowledge acquired in the writing process is likely to be engraved in the learner's memory (Laufer, 2013). Manchón, Murphy and Roca (2007) attributed this effect to the lexical retrieval strategies adopted by L2 writers for fluent and efficient language

production. According to some researchers (Cumming, 2001; Hyland, 2011; Manchó n, 2011; Manchón et al., 2007), L2 writing is an intensive problem-solving activity, including planning, formulation and revision stages. These phases necessitate the access and choice of extensive lexis, particularly at the formulation stage, where lexical retrieval entails a deep cognitive process. To solve lexical problems, L2 writers usually resort to three retrieval strategies: the use of L1, backtracking and use of dictionary. These aiding techniques help to remember the target words even if they are not intended.

The problem-solving nature of composing a piece of writing increases the possibility of L2 writers consolidating and exercising more control over their lexical knowledge and expanding new word knowledge (Cumming, 1990; Swain & Lapkin, 1995). According to William (2012), this learning outcome is due in part to the three features peculiar to L2 writing: its slow pace, the permanent record it leaves and the high demand for lexical precision in use. Thus, writing can be one of the most effective pushed outputs (Swain, 1985) that take up plenty of cognitive resources to encode form-meaning connections while composing (Ellis, 2003; William, 2012). However, these writing features' side-effect is that they are more time-consuming than other learning modes. For example, Webb (2005) indicated that reading tasks yielded better results in vocabulary acquisition than writing if the same amount of time was allowed. Yet if given longer time as learners needed, they could do better in word learning and retention in writing tasks.

In addition, William (2012), as well as other researchers (Ellis, 2003; Schoonen et al., 2009; Wolff, 2000), contended that writing takes a much higher level of linguistic accuracy than it does in other language skills. The demand for precision in writing can, to a large extent, activate lexical awareness and drive word knowledge development. This heightened accuracy of words “calls for constant interaction” between the word knowledge stored in learners’ minds and the newly learned words (Wolff, 2000, p.219). In so doing, they might have to restructure and reconnect lexical items in the best possible way to render their ideas to an ideal prose. In other words, this interactive activity encourages learners to reflect on the word knowledge they already have and negotiate and intake more knowledge. Following this line, the

current study not only sets sights on what words L2 learners can use in writing but also the extent to which they can use them precisely related to the multi-components of knowing a word, such as the meaning in a particular context, spelling, grammatical patterns, and word parts, to name but a few (Coxhead & Byrd, 2007).

The Involvement Load Hypothesis (ILH) (Hulstijn & Laufer, 2001; Laufer & Hulstijn, 2001) can be another theoretical explanation for writing being an instrumental factor of vocabulary knowledge growth. The ILH claims that any real-life task or communicative situation can induce a certain degree of involvement. Three components characterize this involvement load: **need** (the external-imposed or self-imposed need arising to know a word), **search** (the information processing to find the right word) and **evaluation** (the comparison with other words to assess if the word selected is the best fit). Craik and Lockhart (1972) theorized that whether a new word can be imprinted in the long-term memory resides not in how long it has been known in the short-term memory, but rather in the depth with which the word has been initially processed. This theory has been well placed to explain why vocabulary learning in L2 writing could engender a deeper level of cognitive processing (Keating, 2008; Kim, 2008; Hulstijn & Laufer, 2001; Pichette, Serres & LaFontaine, 2012). For instance, Kim (2008) compared reading comprehension and writing tasks and found that writing (sentence writing and composition writing) is more effective in initial word learning and better retention. Keating (2008) and Pichette et al. (2012) produced similar results, suggesting that sentence writing yielded superior word recall ability over sentence reading tasks even though this plus point weakened with time. Therefore, it can be safe to conclude that writing can be a sensible way to learn how to use a new word in context (Laufer, 2013).

2.5 Assessment of L2 Writing

2.5.1 How to Assess L2 Writing

The assessment of L2 writing can be a daunting task since writing ability is a broad, complex construct that involves multifarious elements in different dimensions (Bachman, 1990; Knoch and Chapelle, 2017; Llach, 2011; Slomp, 2012, Weigle, 2002). The product-based and process-based definitions of this construct have been the most frequently cited and aroused heated debate (Llach, 2011; Yi, 2009). The

product-based approach focuses on the final written text with particular attention paid to language structures and accuracy in grammar, vocabulary and choices of syntactic patterns (Becker, 2018; Hyland, 2003; Yi, 2009). Thus, grammatical and lexical errors are intolerable, and their existence tarnishes the writing quality (Biria & Karimi, 2015; Llach, 2011). In Contrast, some researchers prefer to treat writing ability as a series of cognitive activities that are incremental and interactive with each other such as planning, drafting, revising and editing (Barkaoui, 2019; Grabe, 2001; Hyland, 2003; Slomp, 2012). Because these stages are difficult to pin down and resistant to a monolithic construct (Slomp, 2012), the process-based approach is operationalized far less than the product-based one in assessment (Biria & Karimi, 2015; Cho, 2003).

Accordingly, there have been three main approaches to measuring writing: indirect multiple-choice test, timed impromptu writing test and portfolio assessment (Llach, 2011, p.52). The summary of the three kinds of writing measures is displayed in Table 5. The multiple-choice test is used to test writing ability by inferring from their choice of grammatical and lexical items, thus also called the indirect writing test (Kim & Schatschneider, 2017; Llach, 2011; Hyland, 2003; Weigle, 2002). This approach highlights the grammar and vocabulary proficiency that underpins L2 writing capabilities. In this regard, it can be more statistically reliable than any other method (Hyland, 2003) and easy to mark and interpret scores. Nevertheless, test validity pays the price as it targets not directly towards learner writing proficiency but rather the lower-level language abilities.

Table 5 Main features of writing testing methods (Llach, 2011, p. 53)

Multiple-choice test	*Timed impromptu writing test*	*Portfolio assessment*
Objective	Subjective	Subjective
Easy to administer	Easy to administer	Time-consuming
Easy to mark	Requires rater training	Requires rater training
High statistical reliability	Lower reliability	Lower reliability
Low validity	High validity	Highest validity

Direct measures, including timed impromptu tests (product-based approach) and portfolio assessment (process-based approach), are the most widely used tests in writing assessment. Because both methods measure actual writing and students' compositions, they share higher validity than indirect assessment. The timed impromptu writing test features heavily in the L2 writing test (Biria & Karimi, 2015).

This approach is simple to administer as it sees writing as a written product and assesses merely the final text regarding its multidimensional aspects. L2 writers usually are required to perform an essay within a given time after they read some prompt materials or instructions. The single essay plays a critical role as the major test instrument in this assessment approach. Instead of a final composition, the portfolio assessment embodies the writing-as-a-process notion and assumes that measuring the writing process can be more effective than a written product (Cho, 2003; Harmer, 2001). While essay test is also indispensable in portfolio assessment, a collection of student essays would be written from the first draft to the final version. Instructors provide help in this process and observe the development of student writing ability. Notwithstanding its merits in promoting writing instruction and idea generation, this approach can be time-consuming and error insensitive to grammar and vocabulary (Ferris & Hedgcock, 2005; Yi, 2009).

As such, the current study takes the product approach, which might be more stable and measurable and has long been the centerpiece of assessment on writing (Jeffery, 2009; Slomp, 2012). L2 learners would be assigned a single writing question selected from the IELTS (Task 2 in academic module) to perform an essay within the time constraint. Nonetheless, the essay rating might be an issue more than necessary to be spelled out to ensure the reliability of the direct assessment of writing quality.

2.5.2 Scoring Procedures and Rubrics

Since a quality essay is an orchestration of knowledge and skills in multi-dimensions (Schoonen, 2012), rating scales of a direct writing test are essential because they serve to realize theoretical constructs and indicate what elements need to be assessed about writing (Hamp-Lyons, 2011; Weigle, 2002). A set of reliable evaluating methods has been devised: holistic, analytic, and primary-trait scoring. The current study merely reviews the former two methods as holistic and analytic scoring have sparked the most debate (Li & He, 2015; Neumann, 2012; Olinghouse & Santangelo, 2010; Schipolowski & Böhme, 2016). Overall, the two types of scales mainly differ in three respects: scoring methods, assumptions about the whole-parts relationship and decision-making processes (Li & He, 2015).

Table 6 Differences between holistic and analytic scoring

	Holistic scale	Analytic scale
Scoring method	A single score for each script	Multiple scores for each script
Whole-parts relations	The whole equals all parts	Parts add up to the whole
Decision-making	Combining traits together	Rating multiple traits separately

Holistic Scoring

Many writing assessments, particularly large-scale tests, rely on holistic scoring, in which a single score assigned to the script reflects the rater's overall impression of the writing quality (Olinghouse & Santangelo, 2010; Weigle, 2002). The score is given against a set of rating scales or rubrics specifying the scoring criteria at each level. Raters adhere to the rubric and provide an overall score ranking the assessed script to a certain level. The IELTS writing test is a well-known example of a holistic scoring rubric. There are nine bands of scores with detailed scoring criteria in four dimensions at each band.

Indeed, holistic scoring is quick and economical for the direct writing test. Moreover, this scoring can engender better validity since it represents a close, natural reaction to the script instead of distracting attention from separate elements (Li & He, 2015; Llach, 2011). Holistic scoring also shows high validity and reliability in the tests of different genres of writing tasks in empirical studies (Olinghouse, Santangelo, & Wilson, 2012; Schipolowski & Böhme, 2016). However, holistic scoring is criticized as neither transparent nor informative (Llach, 2011) because the single score cannot inform which aspect accounts for the dominant part of a learner's writing ability. Also, worth mentioning is that the holistic score might be correlative with superficial features such as handwriting and text length, weakening its validity (Weigle, 2002).

Analytic Scoring

Instead of a single score, analytic scoring assigns separate scores to different aspects of the assessed script. Each aspect has detailed criteria, against which raters give sub-scores to content, vocabulary, grammar or mechanics, depending on the testing purpose. The sub-scores can finally be added up or averaged to an overall score. Thus, the analytic scale provides more detailed and instructional information concerning

different dimensions of writing ability. The best example of analytic scoring should be the one created by Jacobs et al. (1981) (See more in Section 3.3.3). This typical analytic scoring covers five aspects: content, vocabulary, grammar, language use and mechanics.

Analytic scoring has been favored and widely employed by a number of writing test developers (Ghalib & Al-Hattami, 2015; González et al., 2017; Troitschanskaia et al., 2019; Wiseman, 2012). The first advantage is that analytic scoring is more user-friendly, especially for inexperienced raters, because it is easier to understand and clearer to apply to each area than holistic scales (Weigle, 2002, 2007). In addition, analytic scoring is in a position to evaluate L2 writing performances since L2 writers might have markedly varying capacities for different aspects of writing (Ghalib & Al-Hattami, 2015; Wiseman, 2012). For example, some learners may be strong in structure organization but weak in grammar, while others may have a large vocabulary but lack ideas. The diagnostic information extracted from analytic scoring makes all this clear about L2 learners' strengths and weaknesses. Moreover, analytic scoring has also been found more reliable than its holistic counterpart due mainly to the detailed descriptors and additional items (Knoch, 2009; Weigle, 2007; Wiseman, 2012).

Taken together, analytic scoring, though more time-consuming and costly, might be better placed to achieve the testing goal of the current study. With this scoring, the researcher can yield a nuanced understanding of Chinese L2 learners writing ability by tapping the target writing variables (Brown, 2004; Llach, 2011; Weigle, 2002).

2.5.3 Measures of Vocabulary Knowledge in L2 Writing

It is paramount that L2 learners acquire as much knowledge as possible about various word components in L2 writing. This knowledge in language production or lexical proficiency is deeply related to the judgment of writing quality, and the manifestations thereof in L2 writing, need to be well assessed (Engber, 1995; Laufer & Nation, 1995; Lin, 2015). Vocabulary knowledge in L2 writing sometimes emerges as lexical richness as an umbrella term with several typical lexical features in assessment, including among other features, lexical diversity (variation), sophistication, frequency profile, density and originality (Engber, 1995; Ha, 2019; Kojima & Yamashita, 2014; Morris & Cobb, 2004; Read, 2000). Crossley, et al. (2015, p. 571) referred to these

features as the “behavioral constructs”, driving the measures of the theory-based constructs such as the size and depth at the language production level. These features of lexical proficiency scaffold an overall perception of how L2 learners produce and process vocabulary knowledge (Laufer & Nation 1995; Qian 1999; Schmitt 1998). Of all these lexical features, lexical diversity, sophistication, and frequency are the most normally used as lexical measures in L2 writing.

2.5.3.1 Lexical Diversity

Lexical diversity, also known as lexical variation, means the proportion of different words in written or oral production that do not repeat the words already encountered (Jarvis, 2013). It calculates the “type-token ratio” (the TTP) between the number of different words (types) and the total number of running words (tokens) in a text. Lexical diversity reflects lexical proficiency, assuming that the more proficient L2 writers are, the wider variety of words they can recall and produce in their writing. However, the TTP has drawn a chorus of criticism because the differing text length can distort it, hence flawed reliability (McCarthy & Jarvis, 2007). To overcome this limitation, a collection of new measures were developed to assess lexical diversity, such as the D measure (Malvern & Richards, 1997; Malvern, Richards, Chipere, & Durán, 2004) and the MTLD (Measure of Textual Lexical Diversity) (McCarthy, 2005).

Lexical diversity has been found to correlate with language proficiency (Daller & Phelan, 2007). For instance, error-free lexical variation contributes best to L2 writing scores (Engber, 1995) and writers’ word diversity can be a stable and consistent prediction of L2 writing quality (Olinghouse & Leaird, 2009) and one of the most predictive indices of native essay quality (McNamara, Crossley, & McCarthy, 2010). However, research also suggests that the correlations might be varied, as was demonstrated by Filipino and Chinese English learners whose writing scores bear little correlations with vocabulary diversity assessed by the D measure (Yu, 2009). Similarly, Akbari (2017) observed no significant difference in EFL essays written in year 1 and year 2, focusing on the lexical diversity of academic words. They blamed this on lexical features as a measure of how varied the words are in a text, instead of how well learners can use the words. Akbari (2017), as such, suggested that L2 writers

have to bring the various word components listed in Nation's (2001) framework into productive knowledge to precisely use the words in L2 writing.

2.5.3.2 Lexical Sophistication

Lexical sophistication is another measure used to manifest lexical richness in language production, mainly referring to "the proportion of relatively unusual or advanced words in L2 learners' text" (Read, 2000, p. 203). The "advanced" words can be perceived and operationalized in a wide variety of ways, such as the low-frequency words beyond the most frequent 2000 words assessed by the LFP (Laufer & Nation, 1995), the most typical academic words (Coxhead, 2000) and the less concrete and familiar words (Crossley, & McNamara, 2011; Saito et al., 2016), among other lexical features. Moreover, collocations or multi-units in writing have been found to be a striking property of lexical sophistication, endowed with significant advantages (Crossley, 2020; Ellis, 2012; Siyanova-Chantura & Martinez, 2015).

Lexical sophistication can be one of the most effective indexes of L2 essay quality (Crossley, 2020), such that the higher quality L2 writing has, the more sophisticated words it contains. For example, regression analysis in Ha's (2019) study suggested that writing quality improved significantly with lexical sophistication characterized by using the second 1000 most frequent GSL words (General Service Word List, West, 1953). This parallels Laufer and Nation's (1995) and Meara and Bell's (2001) claims that low-frequency words are the best predictor of lexical proficiency. Meara and Bell (2001) developed the *P_Lex*, a measure to calculate the difficult words or words beyond the most frequent 2000 words. This measure calculates the *lambda* value, counting the number of infrequent words in every ten-word segment in a text. The *P_Lex* shows sound test-retest reliability and concurrent validity, assisting EFL/ESL lexical proficiency assessment (Lemmouh, 2010; Lin, 2015).

However, Kyle and Crossley's (2015) and Yoon (2018) implied that other word properties added more prediction to lexical richness scores, such as the frequent n-grams and context-specific words used in argumentative essays. Moreover, Zhang et al. (2021) found that lexical sophistication featuring the range and frequency of academic words significantly correlated with argumentative and letter writing. While the construct of lexical sophistication is complex with multidimensional indices, it

generally contributes to and is associated with higher L2 writing quality. Learners are supposed to use more sophisticated words, be it academic words, infrequent words or frequent collocations.

2.5.3.3 Lexical Frequency Profile

The Lexical frequency profile (LFP) developed by Laufer and Nation (1995) draws heavily on various word lists to reflect vocabulary size knowledge in L2 writing. It mainly measures to what extent learners can bring vocabulary size to free productive use in their writing. Details of this information are elicited from the percentage of words in an L2 essay at four frequency levels: the first 1000 most frequent words, the second 1000 most frequent words, the AWL words and the "not-in-the-list" words, i.e. words not in any word list mentioned above. The proportion of word frequency is calculated by counting how many word families are used at each frequency level. The AWL and "not-in-the-list" words are deemed low-frequency words. The notion behind the LFP is that the more low-frequency words are used in an essay, the more lexical richness it has.

Laufer and Nation (1995) claimed that the LFP is more valid and reliable for assessing productive word knowledge in L2 writing than any other measure, such as lexical diversity and sophistication. Research also considerably employs the LFP as a useful diagnostic and sensitive testing tool for word use (Douglas, 2015; Min & Qian, 2012; Morris & Cobb, 2004; Munice, 2002). For example, Min and Qian (2012) found that the free productive word knowledge assessed by the LFP correlated best with Chinese web-based learners' writing scores. The LFP results also showed a significant relevance to a higher lexical proficiency rated by human judgments of writing quality (Douglas, 2015). Moreover, Morris and Cobb (2004) used the LFP to predict ESL learners' academic performance and found that vocabulary profiles correlated well with high proficiency learners' academic writing skills.

However, Meara (2005) suggested that the LFP is not sensitive to the modest changes in the size of productive word knowledge, and it merely works well when the samples compared have disparate word sizes. In addition, Meara and Bell (2001) compared the *P_Lex* and the LFP and concluded that the LFP is unsuitable for short texts produced by low proficiency learners. Consequently, it seems that no single measure can suffice

to assess productive vocabulary in L2 writing and it is more than necessary to introduce multiple measures to do it (Lin, 2015; Nation, 2007). Moreover, the above-mentioned measures seem to be more related to vocabulary size (knowledge of the lexicon) and seldom associated with vocabulary depth (knowledge of individual words) (Crossley et al., 2015). For instance, Crossley et al. (2011, 2015) argue that lexical diversity and frequency profile mainly measure the breadth of knowledge, assuming that learners with a large vocabulary tend to produce more varied and low-frequency words. Therefore, research calls for multiple measures assessing the productive components of word-depth knowledge in different dimensions.

2.6 Relevant Studies

This section mainly summarizes and reviews the relevant studies associated with the multi-component research of vocabulary knowledge, vocabulary knowledge and L2 writing within multi-component frameworks.

2.6.1 Multi-component Studies on Vocabulary Knowledge

It is still unclear how various word components are acquired and developed in lexical acquisition. In order to establish the development of two types of word knowledge, Schmitt and Meara (1997) made the first multi-component attempt to investigate the suffix and associative knowledge correlating with vocabulary size and language proficiency. They found considerable variability in EFL learners' vocabulary knowledge characterized by a poor mastery of these word knowledge components within an academic year. These findings are significant and serve as a reminder that some components of word knowledge lag far behind in EFL acquisition. For this reason, another longitudinal research involving more knowledge components was administered by Schmitt (1998), who tracked the acquisition of three international postgraduates regarding four-word components: form recall, meaning recall, grammatical knowledge and association knowledge. He found that participants had varying mastery of the four components, and with each component increased, the others interrelated to rise in the meantime. Because the study was conducted in response and one-on-one interview format, the results could hardly be absolved of subjectivity. However, the two studies manifest that multi-component tests, though time-consuming and limited in word number, are worthwhile since "such research

produces a very detailed description of vocabulary knowledge, making it well worth the effort" (Schmitt, 1998, p. 286). As such, the multi-component approach makes it easier to break down the complex construct and ascertain the interrelations between different word components.

Word component knowledge is essential, and many L2 learners may be short of this knowledge. For example, Chui (2006) examined Hong Kong university learners' vocabulary size and components of depth knowledge, including grammatical functions, derivatives, collocations, and word meanings. Results showed that neither the size nor the depth knowledge was satisfactorily mastered, particularly collocation and derivative knowledge. Therefore, the need arises to provide more lexical instructions on various word components both receptively and productively. Webb (2005, 2007a, 2007b, 2009) attempted to investigate receptive and productive learning of five-word components: grammar, syntax, orthography, association, form and meaning. The series of studies (2005, 2007a, 2009) compared receptive sentence reading and productive sentence writing and examined receptive and productive word-pair learning. These studies reveal that productive learning can be superior and more effective than receptive learning in vocabulary acquisition. In addition, Webb (2007b) also explored incidental vocabulary acquisition through reading tasks and found that the more times learners encountered these words, the more knowledge they gained in these word components. However, the learning context used in Webb's studies cannot be generalized since incidental reading contexts may vary remarkably. The single glossed sentences can hardly represent an ideal context for vocabulary knowledge learning. Chen and Truscott (2010) followed Webb to expand the measured word knowledge to seven lexical components, examining Chinese EFL learners' vocabulary acquisition and retention. They introduced richer reading contexts in passage reading rather than single sentences. Their findings largely parallel Webb's study, except that some components such as meaning, spelling and word class develop at different rates. However, a relaxing control on participants' reading tasks might be difficult to establish causal links between variables.

It has become clear that the various components are acquired unevenly, and different learning methods may lead to varying mastery of these components. However, the

order of acquisition and the interactions between word components remain to be explored. In recent studies, Koizumi and In'nami (2020) and Gonzalez-Fernandez and Schmitt (2020) employed structural equation modelling (SEM) to determine various components of word size and depth knowledge. Koizumi and In'nami (2020) devised five tests for 255 low to intermediate-level EFL learners: one size test, four depths tests (one association test, two polysemy tests and one collocation test), all in multiple-choice format under the guidance of Nation's (2022) multidimensional approach. They used a single-factor model and a correlated two-factor model to examine the data. The results revealed that the correlated two-factor model fits best with the data, suggesting that the size and depth are two separate yet highly correlated constructs. This is consistent with previous studies in this aspect (Qian, 2002; Qian & Schedl, 2004; Schmitt, 1998; Schmitt & Meara, 1997; Vermeer, 2001). However, four multiple-choice tests focusing on meaning recognition were used to measure depth component knowledge, which is no different from the size test in character. This raises the question about the validity of the results they obtained from these receptive measures.

Consequently, Gonzalez-Fernandez and Schmitt (2020) used more productive tests and closely looked into the interrelationships of word components and their acquisition order. In their study, eight tests were administered to measure the receptive recognition and productive recall of four components of vocabulary knowledge: form-meaning link, derivatives, multiple meanings, and collocations. The SEM analyses indicated that all assessed components are highly correlated, ranging from .70 to .94. These word components also have a high factor loading to contribute to vocabulary knowledge. Moreover, this study attested that the acquisition order varies markedly for different components and word knowledge recall usually lags behind recognition. This is aligned with Laufer and Goldstein's (2004) findings that form recall is the hardest to acquire. Gonzalez-Fernandez and Schmitt's (2020) study merits attention because it has involved the most extensive multi-component tests to date. However, this study was merely conducted in a cross-sectional setting, making the results less conclusive than a longitudinal study which might otherwise provide more fruitful insights into vocabulary knowledge acquisition and development.

Similar to Gonzalez-Fernandez and Schmitt's (2020) study, Nontasee and Sukying (2021) also attempted to establish the interactions between multiple word knowledge components and their learnability, that is, how easy or difficult a certain component is for learners to master. This study was conducted in a Thai high school context where samples from three grades sat six receptive and productive (respectively three) tests. Three-word components were tapped: word parts, form meaning links and collocation knowledge. Participants did the receptive and productive word tests one by one without instruction. The findings of this study confirm that the three components of word knowledge develop at an incremental continuum from receptive to productive ends. Knowledge of word parts scored highest, while the collocation tests gained the lowest scores. This signals that morphological knowledge precedes form-meaning and collocation knowledge in receptive and productive contexts when considering the degree of acquiring difficulty. However, it is worth mentioning that the researchers measured collocation knowledge with the PVLT, which is primarily regarded as a form-recall test (Schmitt, 2010). It is unclear whether learners' form recall ability or collocation knowledge helped them gain the score. Thus, the test validity remains uncertain on the part of this test. In addition, the researchers deployed a set of cross-sectional tests without experimental treatment to determine the learnability of word components characterized by a longitudinal process. It might still be hard to pin down the difficulty degree of acquiring the component knowledge if no instructions or self-learning is allowed.

Reading proficiency has primarily drawn the most research attention from the multi-component perspective. Previous research looked into the interactions between various components of word knowledge and reading comprehension. Qian (1999) investigated the relationship between word size, depth and reading ability, measuring seventy-four Chinese and Korean ESL learners' word size with the VLT, depth knowledge, including synonymy, collocation and morphological knowledge with the DVK model. The findings revealed a positive and significant correlation between the size and depth components. Nevertheless, receptive size knowledge accounted for 60% of the variance of reading comprehension, and the depth knowledge of collocation and association added another 11%. Qian's (1999) study identified moderate to positive correlations between different word components and receptive reading ability.

Similarly, Laurence et al. (2018) delved into academic vocabulary and reading performance. They created four novel measures focusing on associated depth: formulaic patterns, topical associations, superordinate words and definition tasks. Using the factor approach, they determined the relationships between each component of the academic words sampled from the AWL (Coxhead, 2000) and native speakers' abilities to answer forty-eight items based on reading a passage. They found that the definition task explained the most variance in reading comprehension and the formulaic patterns and topical associations also showed unique correlations. This finding supports Qian's (1999) study because the definition test is not unlike the VLT test. It requires test takers to choose one right meaning to match the target word from four options with three distractors.

However, McLean et al. (2020) suggested otherwise in their four-modality tests to measure 103 Japanese university students' form and meaning knowledge and their reading comprehension. They respectively used Yes/No test, form recall, meaning recall and meaning recognition and correlated their scores in these tests with their reading performance in the Test of English for International Communication (TOEIC). They found that form recall was the strongest predictor of reading comprehension, followed by meaning recall, meaning recognition and Yes/No test. This is inconsistent with the previous assumptions in Qian (1999) that receptive size in form-meaning mapping is more related to reading skills. Nevertheless, the meaning and form tests overlapped considerably to the point where it is impossible to draw a clear line between the data engendered from the multiple tests.

2.6.2 Multi-component Studies on Vocabulary and L2 Writing

Despite being critical, there has been limited research on the relationship between various components of vocabulary knowledge and productive language production, such as L2 writing. Recent exceptions include Zhong's (2016) study on receptive word components and word use in sentence writing tasks and Kilic's (2019) research into multiple word components and speaking and writing proficiency. Guided by Nation (2022) and Coxhead's (2007) frameworks of vocabulary knowledge, Zhong (2016) investigated the interrelationships among five-word components in a receptive format, including form recognition, meaning comprehension, word class, and multiple

choice of association and collocation. She adapted the VKS and WAT to devise word knowledge and sentence writing tests. Data analyses show that meaning and form figure prominently (74.1% variance) in the prompted sentence writing. At the same time, word class, association, and collocation predict starkly less yet still improve and align with word use ability. Thus, these components need to be paid attention to in vocabulary acquisition for productive use. Nevertheless, Zhong (2016) used sentence writing tasks to predict EFL productive word use ability. The question is that sentence writing has been proven ineffective in measuring contextualized vocabulary depth (Read, 2000; Schmitt, 2010), hence the need to design more productive tests in natural and authentic contexts.

Contrary to Zhong's (2016) study, Kilic (2019) introduced a more comprehensive battery of tests featuring the RVLT for receptive vocabulary size, the PVLT for form recall knowledge, and Qian's (2002) DVK (the depth vocabulary knowledge) model, i.e. synonymy, polysemy, and collocation tests for vocabulary depth. This study indicated that depth tests and form recall significantly explained both the writing and speaking tasks alike, whereas receptive size contributed relatively little. This result means that productive depth components of word knowledge account more for the quality of L2 production ability than vocabulary size. This contradicts Zhong's (2016) finding, which is unsurprising because L2 writing bears a much higher involvement load than sentence writing used in Zhong's (2016) tests (Laufer & Hulstijn, 2001). Ren and Liu (2020) found that using the target word in L2 writing stimulates a more profound cognitive process than sentence writing, involving more word knowledge than form and meaning. This is also true in Shi and Qian's (2012) study on web-based Chinese English learners' writing quality. They tested passive (receptive) size, controlled active form-recall knowledge, and free active word knowledge in essay writing. Their findings further determine that free active (productive) knowledge is more related to writing quality than the other two components.

Productive word knowledge exerts a stronger hold on L2 writing performance. However, as Laufer (1994) and Laufer and Nation (1999) pointed out, vocabulary form and meaning knowledge should always play an indispensable role in composing an essay. For example, Dabbagh and Janebi Enayat (2019) determined that vocabulary

breadth knowledge elicited from the VLT better predicted descriptive writing scored holistically and analytically while depth knowledge from the WAF predicted little variance. Moreover, Wu et al. (2019) identified that receptive and productive word size contributed more significantly to L2 writing quality than other depth components: adjective synonyms and morphological awareness. However, it should be noted that they respectively used L2-L1 and L1-L2 word pairs to measure receptive and productive size with a base of sixty words. The question is whether this small sample can fully represent L2 learners' word size. More important is that productive L1-L2 translation knowledge was the best predictor and adjective synonyms predicted more variance in higher-level learners. This means that productive word components still are the best contributors to L2 writing ability. Similarly, Choi (2017) measured L2 writing vis-a-vis various receptive and productive word components: receptive size with the VST, meaning association and collocation with the WAF, and form-recall knowledge. She observed a direct role played by productive knowledge in L2 writing while receptive knowledge had an indirect bearing mediated by productive knowledge.

Perhaps more convincing about vocabulary knowledge in L2 writing are the multiple components listed in Nation's (2022) taxonomy of constraints on word use, such as word frequency and collocation. Crossley et al. (2015) examined lexical proficiency components: collocation accuracy, word frequency and lexical diversity in 240 L2 written texts. They found that the depth of vocabulary knowledge characterized by collocation accuracy is the strongest predictor of lexical proficiency in L2 writing. The accurate use of the multi-word units explained 84% of the variance of holistic writing scores, while lexical diversity and frequency, usually seen as indices of breadth knowledge (Crossley et al. 2011), merely accounted for 3% and 1%, respectively. On the contrary, Olinghouse and Leaird (2009) compared four lexical features: vocabulary diversity, less frequent vocabulary, mean syllable length, and a number of polysyllabic words in two L2 writing tasks. They found that vocabulary diversity and less frequent lexis explained the most variance of L2 writing quality, while lexical diversity was the only stable and consistent variable across the two writing tests. Nevertheless, Olinghouse and Wilson (2013) and Csomay and Prades (2018) cautioned that the genre of L2 writing could be a decisive factor in determining which word component plays the most critical role. For example, among the components they examined,

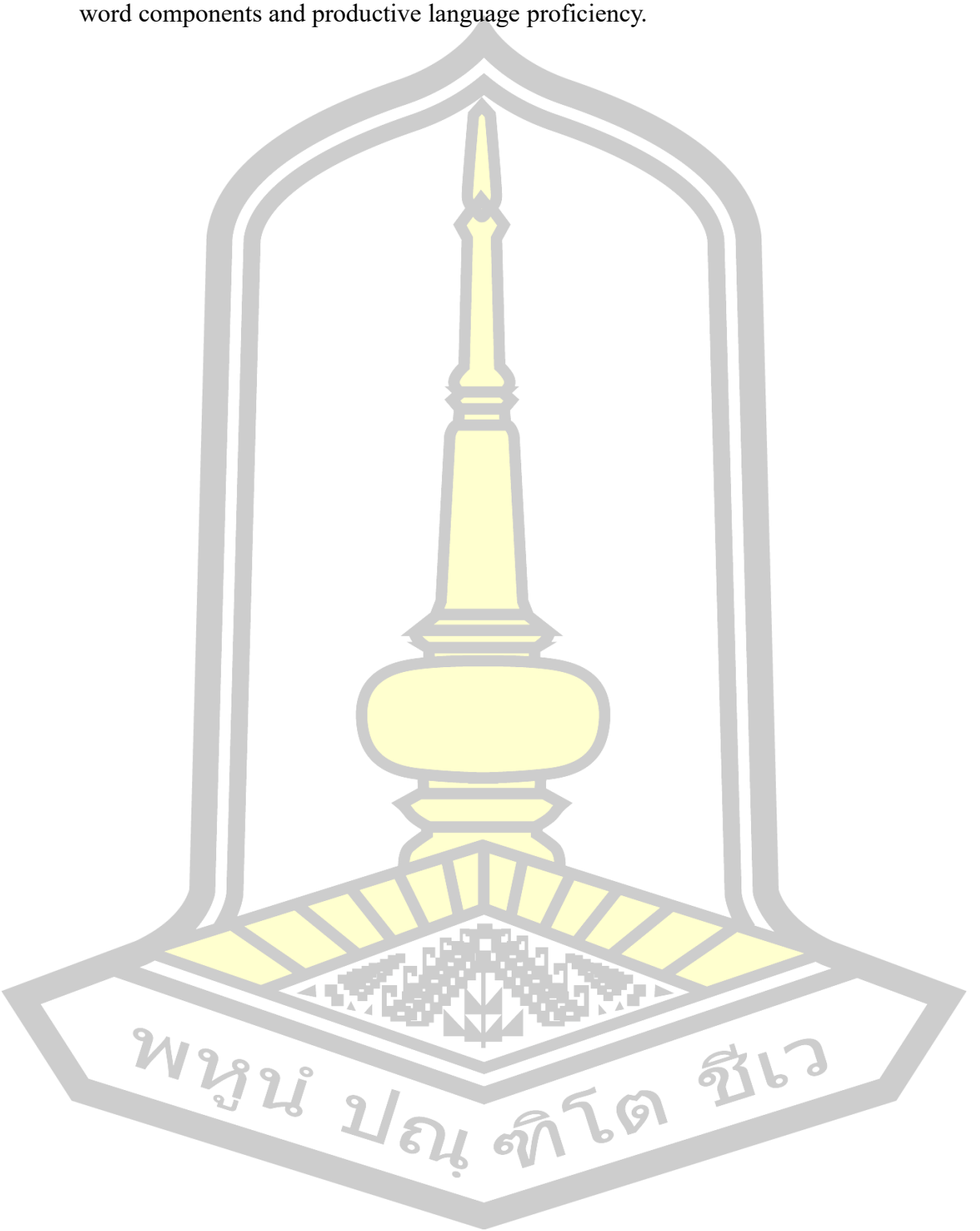
vocabulary diversity predicted story texts best, register and content words uniquely explained persuasive texts, and content words accounted for the most variance of informative texts. This result confirms the high relevance of lexical diversity in Olinghouse and Leaird (2009). Also, it points towards the central role of content words which appear to be low-frequency words in L2 writing, particularly in academic vocabulary (Olinghouse & Wilson, 2013).

L2 writers seem to be expected to use more low-frequency lexis to enhance writing quality. Karakoç and Köse (2017) tested receptive size, controlled form-recall and full productive knowledge in LFP associated with L2 writing. They established that learners with a sound knowledge of the low frequency words (above the most frequent 2000 words) could produce well-developed essays even if they did not have a large word size. Corroborating this finding, Johnson et al. (2016) examined knowledge of receptive and productive word components and confirmed that unduly using high-frequency words in writing negatively correlated with writing performance. As such, they called attention to a balance struck between accurate use of high-frequency word families and a repertoire of mid- to low-frequency (from 2K to 2K-3K word lists) word families. Dabbagh and Janebi Enayat (2019) also provided empirical data showing that mid- and low-frequency words were a strong predictor of the overall assessment of L2 writing. This is consistent with previous arguments made by Meara and Bell (2001) and Vermeer (2004) that a larger vocabulary enables L2 writers to use more infrequent words. The need of less frequent lexis for stronger L2 writing performance is closely attuned to lexical sophistication (See details in Section 2.5.3.2).

2.7 Chapter Summary

In retrospect, this chapter reviews the global concepts of vocabulary knowledge, such as size and depth, receptive and productive knowledge, including models combining word knowledge and competence. There follows the relationship between vocabulary knowledge and L2 writing and the measures of these constructs. This review takes a multi-component approach to inquiring about multiple word components in language skills such as L2 writing proficiency. We wake up to the fact that previous studies on this research area, though insightful, are not compelling enough to unravel the nature of multi-component word knowledge in language production. Therefore, the current

study ventures into this research gap and delves into the interplay between various word components and productive language proficiency.



CHAPTER III
RESEARCH METHODS

This chapter presents the overall research design and the operationalizations of receptive and productive vocabulary tests and EFL argumentative writing tasks. This chapter also gives a full account of the participants, instruments, procedures, rubrics, and data analysis processes. The description includes all the statistical tools that validate the vocabulary and writing tests.

3.1 Research Paradigm and Design

The study focused on vocabulary size and depth of knowledge and L2 writing within a multi-component framework. It adopted a quantitative research approach to provide a numerical description of the interactions between various word components and their relationships with L2 writing and lexical ability. This quantitative research aimed to examine the linear relationship between six vocabulary knowledge variables. It also attempted to deploy these variables to predict the feasibility of L2 writing ability and to determine to what extent such variables could be used to explain learners' L2 writing ability.

The study employed a cross-sectional design to investigate Chinese intermediate-level EFL learners' vocabulary knowledge and L2 writing proficiency. Cross-sectional research involves a snapshot of data collection; that is, research in which quantitative data is garnered from one or more cohorts of participants within a short period of time or at a single point in time (Phakiti, 2015). As such, the current study used a battery of multi-component vocabulary knowledge tests and L2 writing tasks to examine EFL learners' word size, depth knowledge and L2 writing ability. The measures comprised six vocabulary tests: one receptive test to measure the overall vocabulary size and five productive tests to measure word depth knowledge underpinned by five vocabulary components. Moreover, this study also devised two writing tasks with an interval of one week. All these tests will be expanded on in detail in the following sections.

3.2 Participants and Context

Research Context

Since China implemented the reform and opening-up policy, English has become an important foreign language for primary, secondary and tertiary students (Qi, 2016). According to the curriculum guidelines of China's Ministry of Education (MoE) (2003), students in Mainland China are required to start learning English in Primary Three. Accompanying China's rise in areas like world commerce and culture exchange, the whole society now attaches high importance to learning English (Bolton & Graddol, 2012).

These quantitative tests were conducted at Shanxi Normal University, Taiyuan, China, a regional higher degree institute in Mainland China. The researcher has been teaching English for five years at this university. The major course the researcher taught in the given semester was titled *Understanding Contemporary China: Reading and Writing Course.* All university students majoring in English Education must enroll in this compulsory course. One of the main teaching objectives of this course was to acquire more depth of vocabulary knowledge and prepare students for their forthcoming academic writing in the last academic year, which parallels the research aims of the current study.

Participant Profiles and Sample Criteria

The current study adopted convenience sampling to select potential participants enrolled in four intact classes. The participants sampled were 147 third-year students majoring in English Education, and the researcher was familiar with their English proficiency levels. The sample was composed of 140 females and 12 males. This uneven gender distribution reflects the proportion of China's male and female English learners in higher education, which leads to a limitation of the sample that may not be mediated. They were between 19 and 21 years old with 10-13 years of English learning experience from primary school. During the semester, when they sat for the tests, they attended eight courses related to the English language, making up nearly 14 hours of exposure to English every week. In addition, many students would spare more time to prepare for the Graduate School Entrance English Examination (the GSEEE). The participants' vocabulary level was estimated by the scores of the 2000

(96.5%), 3000 (92.1%), 5000 (76%) and 10000 (30.46%) sections of the Vocabulary Levels Test (Schmitt et al., 2001), with a compound score of 73.5% at the four frequency bands.

Before admission to the university, all participants were encouraged to prepare for one of the most influential English tests in China: The National Matriculation English Test (NMET). The NMET is a nationwide English test with other subjects held annually for secondary students to gain university admission (Cheng, 2008). Participants in this study prepared for three years for the NMET, focusing mainly on English reading and grammatical and lexical knowledge (Hu, 2003; You, 2010). After enrolling at the university, all participants completed a credit-bearing English writing course and participated in the TEM4 (the Test for English Majors Band 4) at the end of the second year. The TME4 is another annually-held English test to measure the English proficiency of Chinese undergraduate English majors, mainly tapping into vocabulary and grammar (Treffers-Daller & Huang, 2020). It is a criterion-referenced test requiring all English majors to take to examine whether they have met the requirements of the foundation stage. Therefore, the TME4 has been proven valid and reliable as an intermediate-level test (Jin & Fan, 2011).

The participants were sampled as per the following criteria: (1) All participants had attended the writing course and obtained the credit so that they have mastered basic knowledge and skills of L2 writing; (2) All participants had at least passed the TME4 (60-69 scores) so that they have reached the intermediate level of English proficiency.

3.3 Instruments

The test instruments in the current study subsumed sub-tests designed to measure Chinese university EFL learners' overall size of vocabulary, vocabulary depth knowledge characterized by lexical components which were deemed as critical lexical elements in L2 writing and practical to be tested: productive form and meaning links, productive derivative knowledge, synonymous knowledge, and collocation production ability. In addition to these lexical tests, two L2 writing tests followed suit to measure participants' ability to use the target words in their writing and their overall writing proficiency.

3.3.1 Test for Receptive Vocabulary Size

The Vocabulary Levels Test (VLT) was used to capture the receptive size of vocabulary in the current study since this test has been universally accepted as "the closest thing we have to measure vocabulary knowledge" (Schmitt, Schmitt & Clapham 2001, p.60). Invariably, the VLT and the later defined versions stem from four incremental word levels representing separate frequency bands. These frequency bands (noted that the academic section is not frequency-based) help gauge learners' mastery of each frequency level, thereby providing an estimation of the overall vocabulary size (Cameron, 2002, see Section 2.1.1).

The VLT developed by Schmitt et al. (2001) was adopted in the current study to yield an overall picture of Chinese EFL learners' vocabulary size. It is noted that the 10,000-frequency level was also included, even though this frequency band has been substantiated to be barely associated with L2 writing proficiency (Lin, 2015; Laufer & Ravenhorst-Kalovski, 2010; Schmitt et al., 2004). Indeed, most EFL students find it difficult to employ words at the 10,000 level in their writing or other language performances. However, the 10,000 band represents an important part of vocabulary knowledge that assists in estimating EFL learners' overall vocabulary size. Thus, it makes sense to include this level of words in the vocabulary knowledge test. On the other hand, the most frequent 1000-level words, though carrying as much weight as 80% in English (Webb, Sasao & Balance, 2017), were regarded as too fundamental to be measured in the recognition-matching format for EFL university learners (Lin, 2015; Schmitt et al., 2004). The current study, as such, mainly targeted words at 2000, 3000, AWL (not frequency-based), 5000 and 10,000 levels when deploying the VLT.

Scoring of the VLT

The scoring of the VLT was calculated at both the mastery of each frequency level and the total vocabulary size. Cameron (2002) presented the scores of the VLT merely reflecting the command of each frequency level separately. She also defined the mastery threshold of a certain level at 89%; that is, test takers have to correctly answer a minimum of 26 words out of 30 to prove a good command of that frequency level. With regard to the overall vocabulary size, there has been no universally accepted calculating method. In this study, the way created by Schmitt and Meara (1997) will

be used because of its simplicity and practicality. They postulated that the proportion of the correct answers at each level could represent the proportion of total words known at that level. Then, adding up the total words known at the five levels leads to the estimation of the overall vocabulary size. For example, if a learner correctly answers 15 words out of 30 at the 3000 level, s/he knows 50% of the words at this level in crude terms, namely, 500/1000 words. In this way, the words known at other levels can be calculated, hence the whole vocabulary size. Therefore, in the current study, there would be a total of 150 items in the VLT test (3 items x 10 clusters at each level x 5 levels = 150 items). Each correct item gains one point, making a maximum of 150 points.

Test Usefulness of the VLT

Regarding the test's usefulness, Schmitt and colleagues (2001) made every attempt to validate that the VLT can be reliable and practical. They conducted a native-speaker pilot study, suggesting that the VLT is answerable for proficient English learners. Also central to the VLT is its high practicality because test-takers can complete it reasonably (31 minutes). The usability of this test also includes its ease of scoring and interpretation without resorting to any special equipment. The results reveal the whole state of vocabulary size more than most of its counterparts (Schmitt et al., 2001).

Schmitt et al. (2001) and other researchers (Read, 1988; Beglar & Hunt, 1999) have amply demonstrated the high validity of the VLT. Read (1988) found significant scales between scores in different sections, which means that the frequency levels in the test are highly scalable and that the mastery of the lower frequency levels signals the mastery of the higher ones. Schmitt et al. (2001) assumed that the validity of the VLT could be established if it could reflect the learning sequence from high-frequency to low-frequency levels. They corroborated Read's (1988) finding and provided a coefficient over .90 in the scalability analysis. Moreover, Beglar and Hunt (1999) focused on the 2000-word level and UWL level and found the two tests bear high concurrent validity when correlating with the TOEFL scores. Based on this, Schmitt et al. (2001) used the factor analysis to lend support to the assertion that all levels in the VLT are unidimensional, and vocabulary knowledge is its major measuring trait.

As with the reliability of the VLT, Schmitt et al. (2001) used Cronbach's alpha to validate the internal reliability indices. They argued that the VLT was designed to test how many words learners know at a frequency level, and the sample of items in each section needed to be large enough to cover more words at a certain level. This encouraged them to revise the original 18 items (6 clusters) in each section to 30 items (10 clusters) to ensure better reliability. They found that 30 items are sufficient to represent the words at each frequency level, which has also been validated by other empirical studies (Laufer & Goldstein, 2004; Lemmouh, 2010).

Table 7 Reliability of the levels section (Cronbach's alpha) in Schmitt et al. (2001)

Level	Number of items per version	Version 1	Version 2
2000	30	0.921	0.922
3000	30	0.929	0.927
5000	30	0.927	0.927
Academic	30	0.958	0.96

3.3.2 Tests for Depth of Vocabulary Knowledge

The current study measured the productive depth of vocabulary knowledge, as tabulated as lexical components in Nation's framework (Nation, 2022). It should be noted that the depth construct lacks a standard measure available. Therefore, all the components were counted among the most essential elements in L2 writing in this study. These vocabulary tests were administered independently to test EFL learners' capacity to manipulate word form and meaning, derivative knowledge, synonyms, and collocation during the writing tasks.

3.3.2.1 Selecting Target Words

Building the Corpus for Target Words

The current study collected an overall group of words consisting of altogether 100 words. This small corpus was used as a word list for participants to learn during the given semester. In other words, the corpus of 100 words is the lexical base on which all depth tests would build and from which the final target words would be selected. First, all the 100 words were selected from the textbook used by third-year university students whose major was English Education. The textbook was a classic textbook designed for Chinese intermediate-level English majors. During the semester, students

were supposed to learn six to eight book passages and master at least 200 words productively. Moreover, the researcher was the teacher of this course for all the participants in the given semester. The rationale for choosing the 100 words from this course was that all target words could have the opportunities to be touched on in classroom teaching and students' self-directed learning. This would ensure that all the words would receive different degrees of instruction and learning. In addition, these words sourced from the textbook would guarantee a high motivation for learning among the participants. They have to sit the course exam at the end of the semester, in which vocabulary would account for a large share of their grades.

Second, all the 100 words selected from the textbook belonged to academic words. To this end, the Academic Word List (AWL, Coxhead, 2000), which is regarded as one of the most preferred collections of academic vocabulary, was used as a reference. Academic vocabulary from the AWL can be a key factor for successful university writing (Durrant, 2016) and determines EFL/ESL learners' academic performance and success (Daller & Xue, 2009; Morris & Cobb, 2004). The AWL (Coxhead, 2000) was compiled based on the word frequency, range and uniformity of occurrence in a corpus of 3.5 million running words. It comprises 570-word families, accounting for 10% of the total words in the written academic texts of four disciplines. This word list provides guidance on the most useful and valuable words for university students with academic goals. It has been widely used in vocabulary teaching and research all over the globe, such that the AWL is now synonymous with academic vocabulary knowledge (Coxhead, 2011; Coxhead & Hirsh, 2007). Thus, the AWL could serve the purpose of the current study to sample the generic academic vocabulary.

Additionally, the AWL was divided into ten rank-order sub-lists to indicate the word frequency in each sub-list. The first three sub-lists represent the most frequent academic words, while those in sub-list 10 are at the lowest frequency. Even though sub-lists 5-10 add little to the coverage of the whole list, they are still worth learning as these less frequent words can be found in a wide range of texts (Coxhead, 2000). Targeting these words pays high dividends for university learners in their vocabulary acquisition and use in reading and L2 writing. Therefore, the words in this corpus were selected according to the following criteria:

a. The 100 words were all academic words from the AWL, and meanwhile, they were the words in the textbook that participants would learn during the semester.
b. The words should be mastered productively within the given semester according to the curriculum of the textbook.
c. The words should have different derived forms, synonyms and collocations in actual use.
d. The words should frequently occur in written English and can be readily and commonly used in students' writing.

Following the above criteria, the present study selected the target words for the overall lexical base. The 100 words selected from both the textbook and the AWL are presented alphabetically as follows:

Table 8 The corpus of 100 words from both the textbook and the AWL

Verbs	achieve, acknowledge, assign, acquire, assure, anticipate, consume, consult, commit, compound, credit, confine, conceive, detect, deprive, determine, dispose, emerge, expand, eliminate, equip, encounter, enable, extract, establish, feature, facilitate, highlight, initiate, indicate, impose, imply, invoke, issue, justify, label, manipulate, modify, norm, negate, precede, persist, promote, rely, resolve, relax, reinforce, recover, retreat, restore, reside, reverse, restrain, stimulate, undergo, vary
Adjectives	adequate, arbitrary, considerable, discrete, devoted, evident, ethnic, enormous, exposed, grave, intrinsic, identical, integral, inevitable, incentive, inherent, inclined, legal, mutual, odd, prior, primary, relevant, straightforward, voluntary
Nouns	approach, access, alternative, attempt, bulk, capacity, equivalent, function, intelligence, insight, injure, lecture, motive, margin, norm, reluctance, revenue, schedule, vision

This small corpus mainly comprised verbs, nouns and adjectives, in which verbs constituted the bulk of the list. Adverbs were excluded because, on the one hand, the AWL was compiled as word families, suggesting that the headwords (mainly stem nouns and verb forms) can represent knowledge of other morphologically distinct forms (Coxhead, 2000). On the other hand, learners who command the knowledge of adjectives are usually assumed to know adverbs, too (Ishii, 2005). In addition, some words can be both a verb and a noun, such as *access* and *schedule,* which would be

used subject to specific tests. From the 100 words, the final target words were narrowed down for the following word depth tests.

Selecting the Final Target Words

The selection of the final target words needed to revolve around the two research questions. The words chosen should be well placed to address the various word knowledge components: productive form and meaning, synonyms, derivatives and collocation. Thus, the final target words were selected purposefully according to the lexical attributes of a word. Because participants had to randomly use the final target words in their L2 writing, the current study suggests that the number of the target words needs to be controlled at twenty. Twenty target words might be reasonable for the multiple depth tests and L2 writing tasks. Moreover, the final target words had to fit in with the contexts of the L2 writing topics. This makes it possible for participants to integrate the target words into their essay writing tasks. As such, half (10 words) of the target words were selected according to the context of one writing question and the other half for another question. The current study decided that the following words would be in a good position for the tests:

a. Words with definite, distinct meanings in the corresponding Chinese translations so that productive form-meaning connections would be easily made.
b. Words that have at least two synonyms so that the target words can be linked with other words in the association test.
c. Words with a variety of derivative forms in a natural productive context.
d. Words with definite and commonly used collocations are also common in Chinese and can possibly be prompted by a Chinese sentence.
e. Words that are highly frequent in Chinese EFL learners' argumentative writing so that these words can possibly be used in the writing task.
f. In order to control the difficulty of the target words, two words from each sub-list from rank 1 to rank 10 were selected, making a total of twenty.

According to the above six criteria, twenty final target words were generated in the following table. The pilot study used these words in the following productive tests of vocabulary depth knowledge. Ten third-year English learners who did not attend the

main study piloted study by doing all depth tests. After some modifications, some words were replaced in the actual administration of the tests, and the final target words in the main study are presented in Table 9.

Table 9 The 20 final target words for depth tests

Word class	Numbers	Final target words
verbs	10	achieve, consume, detect, expand, indicate, justify, rely, restrain, stimulate, persist
nouns	5	approach, access, lecture, intelligence, schedule
adjectives	5	devoted, exposed, inclined, prior, voluntary

Among the twenty target words, there were ten verbs, five nouns and five adjectives since verbs account for approximately 50% of the 570 headwords on the AWL. All the target words and their derived forms (verb, noun and adjective) were measured in depth tests and assigned to use in the following two writing tasks.

3.3.2.2 The Productive Form Test

The PVLT (Laufer & Nation, 1999) was deployed to measure productive form knowledge in the current study. Initially, the PVLT, also referred to as a productive size measure, was created as the productive counterpart parallel to the receptive VLT. The PVLT has yet to achieve the same level of validity as its receptive version (VLT). However, Laufer and Nation (1999) claimed that it is ideally placed to test how many words learners can readily use productively in writing. Yet, Read (2000) and Schmitt (2010) raised doubts as to what the PVLT measures. In this respect, Schmitt (2010) had participants do the PVLT before using the target words in a writing context. He confirmed that it is more of a form-recall test, focusing on the productive form-meaning link. Lemmouh (2010) has also substantiated the PVLT as a valid form-recall test in his empirical study. Moreover, Laufer and Nation (1995) found a moderate correlation between the scores of the PVLT and the LFP, suggesting that the PVLT is related to the ability to produce words in L2 writing.

Following this line, the current study borrowed the format of the PVLT to measure participants' productive recall ability of the word form in a particular context of syntactic levels. It should be noted that the adapted version in this study was not frequency-based, but solely limited to the target words. The initial letters provided as cues would be restricted to minimum letters according to the word length. Since the

PVLT was regarded as a productive test, Laufer and Nation (1999) claimed that it is reasonable to provide the minimum number of letters as long as test-takers could eliminate the possibility of choosing alternative words. If the first two letters were not enough to disambiguate the cue, then one more letter would be added. Moreover, all the sample sentences containing the target words were selected from two widely-used corpora: the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). These corpora have a wide coverage of words, including more than one hundred million words. The main function of the corpora is to indicate the frequency and context of a word. Noted, however, that the contexts of the sentences should be familiar to participants, and therefore, sentences with difficult words or unfamiliar topics were excluded to control the contextual effects.

Laufer and Nation (1999) developed four versions of the PVLT at each frequency level and found that they correlated well and led to similar decision-making. This indicates that the PVLT is reliable enough to achieve satisfactory equivalence among parallel versions. In addition, this test can be as practical as the VLT in that it is easy to operate and interpret. However, the scoring criterion of the PVLT is somewhat unclear because it is a little subjective to judge spelling mistakes. According to Laufer (1998), grammatical errors and minor spelling mistakes were not marked as incorrect. Yet, in the current study, the grammatical errors were ignored, but the spelling mistakes of the stem word were marked incorrect in order for better reliability. Each correct spelling of the word in the sentence achieved one point, making a total of twenty points in this part. Zero points would be given for blanks, non-existent words and words that did not match the syntactic contexts.

Instructions

Complete the underlined words. An example has been done for you.

Example The government imposed a ban on the sale of wild animals.

1. The supermarket wants to ex_____ its business by adding two more stores.

3.3.2.3 The Productive Meaning Test

Learning word pairs has always been a popular and effective way of learning word meaning (Nation, 2008, 2022; Webb, 2009). Receptive learning of word pairs means memorizing the corresponding L1 translation of the L2 word form, while productive learning involves recalling the L2 word form when meeting a decontextualized L1 item (Webb, 2009). Laufer and Goldstein (2004) referred to the productive word pairs as active recall of an L2 word for the given L1 translation. In the current study, the active (productive) recall test format in Laufer and Goldstein (2004) was adapted to measure Chinese learners' productive form-meaning knowledge. As Laufer and Goldstein (2004) did, the first letter of the target word was provided as a prompt in case participants would use other words with the same meaning.

In some cases, more than one Chinese translation of the target words was presented, so participants were more likely to know the target words. The Chinese translations were selected from the definitions explained in the *Oxford Advanced Learner's English-Chinese Dictionary, ninth edition*. This dictionary is the most widely-used English-Chinese dictionary for teaching and learning in the Chinese EFL context. Regarding the scoring, the correct mapping of the target word gained one point, and any spelling mistake was intolerable and marked incorrect.

Instructions

Complete the word according to the Chinese translation. An example has been done for you.

Example 洞察力 --- insight

1. 强加/ 征税 --- i_____

The current study adopted the L1-L2 word-pair test because it has been validated as a better learning method for EFL learners' productive performance than receptive L2-L1 learning. Previous empirical studies have evidenced that productive L1-L2 word learning can be a more effective approach to vocabulary knowledge if the learning goal is productive word use (Griffin & Harley, 1996; Mondria & Wiersma, 2004; Waring, 1997; Webb, 2009).

Another theoretical support for the content validity of the test stems from the results of psycholinguistic vocabulary research (Jiang, 2000, 2002; 2004). In Jiang's (2000) model of word form-meaning mapping, an L2 word is initially mapped to its L1 translation, not the meaning itself. In other words, the L2 word is acquired by borrowing or copying the L1 lemma information (i.e., meanings and concepts). Consequently, the L1 translation mediates the use of the L2 word in productive contexts. Even proficient EFL learners still rely on their L1 semantic system in using an L2 word (Jiang, 2004). Many EFL advanced students depend on the translation of a word to search and eventually bring out the word itself in their writing. Because this process is opposite to the direction in which they usually establish the mapping, active recall is the hardest for learners to handle in vocabulary tests (Jiang, 2000; Laufer & Goldstein, 2004).

3.3.2.4 The Productive Derivative Test

The current study combined the test formats developed by Sukying (2018) and González-Fernández and Schmitt (2020) in the productive derivative test. Sukying's (2018) MPAK (the more-controlled productive affix knowledge) was devised to measure EFL learners' affix knowledge. The MPAK was divided into two parts: a) participants must write as many inflected and derived words as possible in part A, and b) they had to select the appropriate form from the word family they built to complete a sentence in part B. The MPAK was proved to boast sound internal consistency with the Cronbach Alfa coefficient as high as .92. Moreover, Sukying (2018) also used a Likert scale done by seven raters in the EFL context and confirmed the content validity of this test. However, the testing goal of the MPAK is to look into how much affix knowledge EFL learners can produce. This differs from this study, which aims to explore EFL derivative knowledge in terms of L2 writing ability. Thus, the current study solely measured the productive knowledge of word class and merely employed part B in the MPAK, namely, the sentence completion task.

On the other hand, González-Fernández and Schmitt (2020) put all tests in a sentence context and measured four types of productive derivative knowledge of each word: verb, noun, adjective and adverb. Some word types may be non-existent in their tests. Building on the above two studies, the current study measured two types of word classes in sentence completion tasks. Because the target words selected in this study

involved three parts of speech: verb, noun and adjective, the derived forms measured for each word were restricted to two word class. Take the word *initial* as an example; it is an adjective and thus would be tested in terms of its verb and noun. Adverbs were excluded from this test since most adjective forms have already signaled the derived forms of adverb knowledge. Learners likely know the adverb knowledge if they can produce the adjective form of a word (Zhong, 2014; Ishii, 2005). Sukying (2022) provided further evidence that EFL performances in adjective and adverb knowledge acquisition are strikingly similar (adjective was 36.98% and adverb 36.35% for upper secondary students). No significant difference was found between an adjective and adverb performance in correlation analysis. It can be concluded that EFL learners of all vocabulary sizes can recall and use a word's adverbial knowledge (Sukying, 2022). Thus, there is little point in measuring the productive knowledge of both adjective and adverb forms in the same test.

It also makes sense to measure each word according to the most frequent derivatives that EFL learners are likely to use in their writing. According to Bauer and Nation (1993), the word family was categorized into seven levels in affix acquisition based on four criteria: frequency, productivity, regularity, and predictability. Among these levels, level 3 is hypothesized to be the most frequent and regular derivational affixes (Sukying, 2022). Those derived forms that are rarely met in productive contexts were removed. Regarding the syntactic contexts in this test, all sentences were sourced from the COCA and the BNC. Still, the sentences should be readily understood by participants, and those with difficult words or unfamiliar topics were excluded to control the contextual effects. Based on these, the productive derivative test in this study was designed as follows:

Instructions

Fill in the blanks with the appropriate forms of the prompt words. Some of the forms may not exist. As for the non-existent forms, just leave it blank. Wild guessing will cause point loss.

initial

Verb: They have decided to _____ a public debate on these issues.

Noun: The government took the_____ in the fight against corruption.

As with the scoring of this test, the appropriate form of the word with its right spelling was awarded one point. Grammatical errors would cause no point loss. For instance, if the correct answer "*stimulation*" is written as "*stimulations*", one point would be given because it signals that the participant might have acquired the derivational knowledge. Blank spaces and wrong word forms gained no points, and the total score in this part was forty altogether.

3.3.2.5 The Association Test

The association test in the current study solely focused on synonymy, that is, the paradigmatic associations. The WAF created by Read (1998, 2000) was modified to measure participants' synonymy knowledge. The WAF (see more details in Section 2.1.2.2) has been validated as a reliable and valid measure for the lexical network test (Schmitt et al., 2011; Qian, 2002). It has also been regarded as practical and straightforward in actual administration (Fitzpatrick & Thwaites, 2020). However, the WAF is not without its limitations, the most obvious of which are context-independent (Read, 2000) and confusion in scoring (Schmitt, 2010). Targeting the two drawbacks, the current study adapted the WAF into a synonym test in sentence contexts indicated in the below example:

Instructions

Tick off words in the below box with the similar meanings as the underlined word in the sentence. There may be 2-3 keys in the box. **DO NOT** choose more than 3 words.

1.The church was carefully restored after the war.

repair	restart	fix	reinvent	renew

First, the original version, combining synonymy and collocation tests, was split into a single synonymy test to address the scoring problem. Second, the target words without context were put in a sentence as a contextual prompt. In so doing, despite the test still being a receptive format, it has, to some extent, become more productive in measuring depth knowledge of a word. The syntactic contexts should be as accessible as possible to participants so that they can understand each sentence. Moreover, among the five words in the box, there were 2-3 keys in order to minimize the possibility of correct guessing. The participants would notice that the words they selected should be well suitable to replace the underlined words in the sentence. The criteria for selecting the synonym associations include:

a. The associates were selected mainly based on similar word meanings, which can replace the target words in the sentence. All the synonyms of the target words came from the thesauruses of the *Longman Dictionary of Contemporary English* and *Merriam-Webster Dictionary.* These dictionaries also provide the most frequent example sentences for using the words.
b. The synonyms in the box belonged to the same word class as the target words.
c. If possible, the words in the box, including the distractors, should be at the same or lower-frequency levels.

The distractors should be semantically unrelated to the target word, while some had a similar morphological resemblance to the target words. This might better test whether participants knew the target synonym. Choosing each correct synonym gained one point, and the missed synonym subtracted one point. The choice of distractors also

reduced one point. Ticking off more than three words would be regarded as wild guessing and zero points would be given to that item.

3.3.2.6 The Collocation Production Test

The collocation production test was developed based on the previous research on productive collocation by González-Fernández and Schmitt (2015, 2020). They took a form-recall approach to assess productive collocation knowledge in a sentence context. The first letters of the target words in the collocations were given as cues in the sentence gap. This may help to constrain the range of collocates with similar meanings. In addition, an extra L1 sentence was written beforehand to provide a summary of the contextual information. González-Fernández and Schmitt (2015) conducted a series of three pilot studies to ensure the validity and reliability of the test. Native speakers were asked to do the test to make sure the items were answerable, clear and simple. However, the original version provided no direct prompt in the L1 sentences in case participants could correctly answer the target words without true knowledge of the collocations.

The current study adapted this test, borrowing the form-recall format in natural sentences. The difference from the original version was that the Chinese sentences offered equivalent translations of the target collocations, which may help participants find the collocations in question. The contexts of the Chinese sentences should be as familiar as possible to the participants. Likewise, the English sentences containing the collocations were selected from the sample sentences in the *Longman Dictionary* and *Macmillan Dictionary*. The phrase banks of the two dictionaries provide a rich corpus of the most commonly-used collocations attached with formal example sentences. All the target collocations and sentences were also searched in the BNC to ensure a high frequency. It should be noted that sentences with difficult words and unfamiliar topics, Chinese and English sentences alike, were discarded from this test. The modified collocation test in the pilot study is as follows:

Instructions

Complete the sentences with an appropriate collocation (习惯搭配). Collocation means phrases in which the word given always appear with other words in sentences. The first letters of the words (including prepositions) you have to use to make the collocation has been provided. These target collocations you have to complete include different types: noun + preposition, adjective + noun, verb + noun, among other combinations. The Chinese sentence has prompted the target collocation.

Example

上班通勤对很多人来说都很头疼，尤其是在早晚高峰期。

Many commuters have to bear the congestion during peak hour.

In the above example, the test measured whether participants knew the frequent collocations of the target words in a syntactic context. As González-Fernández and Schmitt (2015, 2020) did, the first letters of words in the target collocations were provided. The above Chinese sentence served as another prompt, providing more information concerning the collocation tested, meaning "*It can be a hard time for many people who have to commute to work, especially during peak hour*". Participants must identify the collocation and complete the "peak hour" target words. They were told that collocations in this test would include different types of collocates: noun + preposition, adjective + noun, verb + noun, among other combinations. The collocations selected in this study should meet the following criteria:

a. Collocation is defined in this study as the lexical patterns of two or more words that occur together in texts (Schmitt, 2000). The target words may appear before or after the co-occurring words in the collocation.
b. The target collocations came from the most frequent phrase banks of the *Longman dictionary* and *Macmillan dictionary* and also showed high frequency in the BNC.
c. The collocations should be common in formal L2 writing and have equivalents in Chinese translations. Those collocations that are rare for Chinese learners to encounter and contain difficult words were excluded.

The scoring of the collocation production test was marked as correct or incorrect. Grammatical errors were ignored and would not cause point loss. Minor mistakes, such as spelling errors, would not be deducted points because participants may have partial knowledge of the collocation. Collocates that did not match the context or did not use the target words were marked incorrect.

3.3.2.7 The Usefulness of the Depth Tests

Criteria for Developing the Depth Tests

The five depth tests were developed in accordance with four criteria, which were designed to ensure that all depth tests were well-placed to capture the productive word components in L2 writing. These criteria were as follows:

a. Since the current study associated vocabulary knowledge with lexical and L2 writing ability, the depth tests were designed in productive instead of receptive formats. Productive word knowledge was considered more critical in this study. It is noted that the association test was an exception in that the current study attempted to measure as many associations of the word as possible. Thus, the original receptive recognition-mapping format of the WAF was maintained.

b. The five depth tests were all measured in a sentence context, which means that the word knowledge should be activated in a single sentence rather than any other context.

c. All five depth tests were devised to capture vocabulary depth instead of size knowledge. Association and collocation have been the two depth components measured in the WAF (Read, 2000, 2020), and derivative knowledge has also been regarded as another depth component worth measuring (Schmitt, 2010; Zhong, 2014). Form-recall and word pair tests, as two productive relationships of form-meaning links, were also captured as depth components according to the definition of word depth knowledge in the current study.

d. Only one type or aspect of word depth knowledge, which was determined by the previous literature, was designed to be captured by each of the five depth tests.

Validity and Reliability of the Depth Tests

Regarding the construct validity of the depth tests, sufficient evidence had been provided to validate that the test scores reflected the lexical ability we intended to measure and very little else (Bachman & Palmer 1996). This was achieved in the current study through a comprehensive description of the relevant depth knowledge constructs and a sound outline of the validity of the respective test instruments in chapter two (see Section 2.1.2). In addition, in order to ensure the validity of the depth tests, the current study borrowed and adapted the tests that had been sufficiently validated in previous studies. Despite some limitations, the PVLT (form-recall), L1-L2 word pair (productive meaning), morphology test (productive derivative), and the WAF (association and collocation) have been substantiated to be valid and empirically used by a large number of studies (Laufer & Nation, 1999; Laufer & Goldstein, 2004; Gonzalez-Fernandez & Schmitt, 2015; Sukying, 2017, 2018; Read, 2020; Yanagisawa & Webb, 2020).

As mentioned in the previous sections, the PVLT has been validated to be a form recall measure that can be used to assess productive form knowledge in context (Schmitt, 2010; Lemmouh, 2010). L1-L2 word translation test also has a high content validity to measure meaning knowledge in context, which has been theoretically and empirically validated in previous literature (Laufer & Goldstein, 2004; Jiang, 2000, 2002). Thus, these two measures in the current study were valid to capture the productive types of form-meaning links, as opposed to the receptive VLT size.

As with the association, productive derivative and productive collocation tests, the test formats have been proven suitable for capturing the target word components in previous studies (Gonzalez-Fernandez & Schmitt, 2015; Sukying, 2017, 2018; Read, 2020). Moreover, three experienced EFL instructors reviewed the depth tests and ensured whether the target words and sentences were clear, simple and natural to measure the intended vocabulary knowledge. The teachers did the depth tests until they could correctly answer all items without misunderstanding and confusion, suggesting that all items were appropriate to be used. Afterwards, ten English majors with similar English levels to the participants did the depth tests before a focus group interview. Based on the pilot study and feedback, some collocations, derivatives and sample sentences were removed and replaced.

The reliability of the depth tests was also checked. The internal consistency of the

vocabulary depth tests measures to what extent test-takers' performances on different parts of these tests are consistent. The internal reliability coefficient is often computed with formulas such as Kuder-Richardson formulae (KR-20 and KR-21) and Cronbach's alpha. The current study used the method of Cronbach's alpha because, in contrast to KR-20 and KR-21, it could be deployed to calculate the internal reliability of test items that were dichotomously scored. Cronbach's alpha indicates the homogeneity or unidimensionality of a measure, meaning whether the test items measure a single latent trait or construct (Cho, 2016). The coefficient ranged from 0-1, and the higher the value, the higher the internal consistency of the test items presented.

Table 10 Reliability (Cronbach's alpha) for all vocabulary depth tests

Form recall	Word pair	Derivative	Association	Collocation
.830	.855	.842	.862	.860

The reliability for all vocabulary depth tests is presented in Table 10. The coefficient values Cronbach's alpha > .70 can be interpreted as respectable, > .80 is very good, and >.90 is excellent. All Cronbach's alpha values for the depth tests in the current study were over .80, showing that the depth instruments achieved high internal consistency.

The current study also validated the practicality of the depth tests by conducting a pilot study and a follow-up focus group discussion. Ten third-year English majors who were not among the participants did the vocabulary depth tests. They were not assigned to do the writing tasks because they did not receive any compensation for the pilot study. During the focus group interview, the ten students gave feedback on the five depth vocabulary tests about word difficulty, test administration and time allotment. In addition, these tests were also checked by the three EFL instructors who have taught English writing, reading, grammar, and vocabulary for more than fifteen years. Based on the suggestions offered by the students and teachers, the researcher replaced five out of the twenty target words in the main study.

3.3.3 The L2 Writing Test

Selecting the Writing Questions

The L2 writing test included two writing questions selected from the writing section of the IELTS (International English Language Testing System), task 2 in the academic

module. As one of the most widely-used, high-stakes English language tests, the IELTS plays a critical role in many people's studies all over the globe (Uysal, 2009; Shaw, 2004). Task 2 of the IELTS writing component boasts sound validity and reliability, requiring test-takers to write a 250-word essay on an idea, argument or question (Weigle, 2002; Shaw & Falvey, 2008). This section carries more weight in the writing test, which is similar to the genre of the most typical university essay writing. IELTS writing (task 2) also shares the same rhetorical focus on "evaluation" with university writing assignments (Moore & Morton, 2005;1999; Uysal, 2009). Moreover, the IELTS constantly endeavors to achieve a high construct validity and relevance of the writing tasks through expert judgement and empirical approaches. All these lead to a good match between the test and target domain tasks, thus being accessible to university learners. As such, task 2 of the IELTS writing test (academic module) might be in a good position to measure L2 writing ability in the current study.

Topic knowledge in the writing test is a salient point that needs to be spelled out since writing is always about a certain subject (Weigle, 2002). In order to control the topical effects, the current study adopted the first option for constructing a definition of topic knowledge proposed by Bachman and Palmer (2010). That is, topic knowledge was excluded from the construct definition of the L2 writing test of this study. To this end, two writing items sharing the same education category in the IELTS writing test (task 2) were selected.

Writing task one: Competition for places at university is increasing. Why do more and more people want to study at university? Is this a positive or negative development?

Writing task two: Today, millions of university students have to enroll in online learning for higher education. Colleges and universities offer e-learning programs and courses. Do you agree or disagree with the popularity of online learning?

The reason for choosing these two questions was that the two writing topics belong to the category of university education: one is about the trend of university learning, and the other is about university online learning; both are familiar to third-year Chinese university students. This might minimize the differences in topical knowledge and

ensure that most participants would have something to offer in writing due to their first-hand experiences. The above questions required participants to write two argumentative essays because, for one thing, argumentative structure plays a critical role in task 2 of IELTS writing (Coffin, 2004). Argumentation is defined as "a process of establishing a position which is then defended through the use of evidence, negotiation, logic, etc." (Coffin, 2004, p.4). For another, argumentative writing can be the most common and useful genre of academic writing that learners and scholars use to defend and convince ideas (Soleymanzadeh & Gholami, 2014). Therefore, L2 writers in the current study were more likely to use examples, evidence and reasons to discuss a proposition or question in argumentative writing.

Scoring Rubrics of L2 Writing

Regarding the assessment criteria and rating scales of L2 writing, the current study employed analytic scoring instead of holistic scoring. As such, the analytic rating scale developed by Jacobs et al. (1981) in rating L2 writing quality was used.

The four writing assessment criteria in IELTS, namely, task response, coherence and cohesion, lexical resource and grammar and accuracy, have undergone several phases of the validation programme and boast sound validity, reliability, impact and practicality (the VRIP) (Shaw & Falvey, 2008; Shaw, 2004). However, the IELTS writing criteria represent a typical holistic scoring rubric with one benchmark at each band. Raters assign a single score to each script against the main benchmark, which might be problematic for L2 raters. Thus, the IELTS criteria were discarded in this study because a) L2 raters would find it hard to interpret and stick to the criteria throughout the rating process, and the consistency weakened as a result; b) the 1-9 scoring scale is rarely seen in Chinese EFL context, and it would be difficult for raters and students to generate useful information from this scoring.

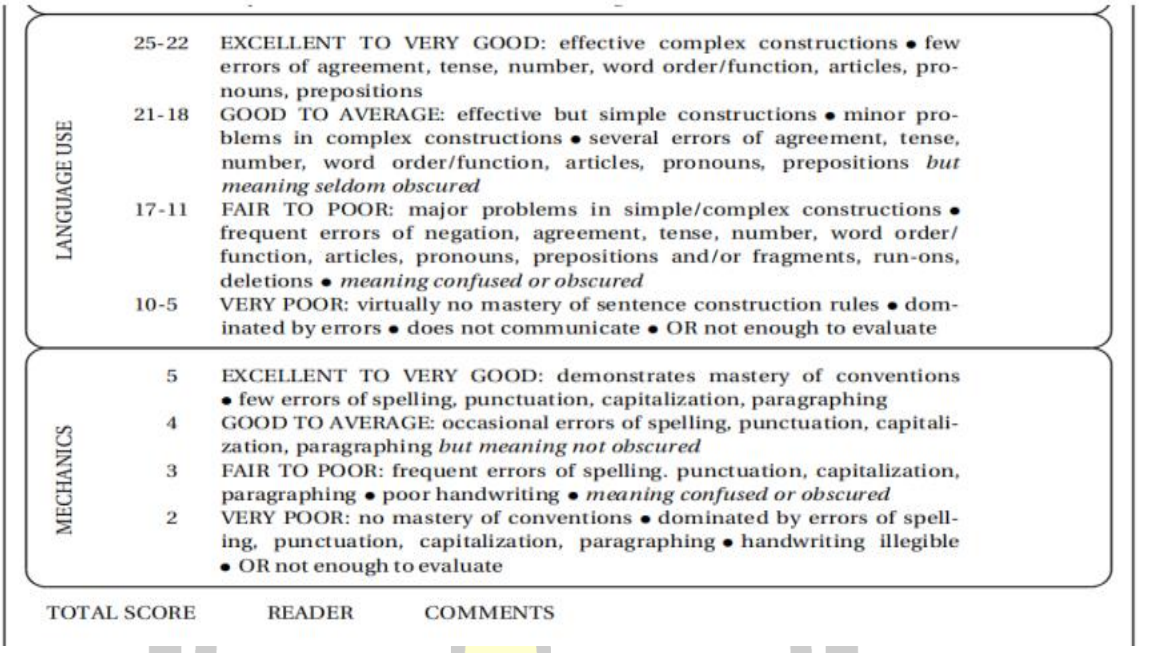
For these reasons, the scoring scale created by Jacobs et al. (1981) was chosen because it has been one of the most widely used analytical scales in EFL/ESL college-level writing programs (Weigle, 2002) and has been deployed in a number of empirical studies (Janssen, Meier & Trace, 2015; Uludag & McDonough, 2022; Winke & Lim, 2015; Wang, 2014). Five dimensions were included in this rating scale, and the weights in each dimension were slightly modified in the present study as the research

focused on vocabulary and language use. The revised version of the current study includes content (20%), organization (15%), language use (30%), vocabulary (30%) and mechanics (5%). Each dimension is attached with detailed scoring descriptors specifying four levels of quality and how many scores can be gained at each level: excellent to very good, good to average, fair to poor and very poor (see Figure 3). This makes the rating process crystal-clear and ensures sound reliability, especially for EFL teachers and researchers. After the multidimensional scoring, an overall score of the script can be obtained by adding up the scores of the five dimensions. This analytic scale can be reliable and valid to capture the subtle differences in multidimensional writing skills (Yoon, 2018).

ESL COMPOSITION PROFILE

STUDENT DATE TOPIC

	SCORE	LEVEL	CRITERIA	COMMENTS
CONTENT		30-27	EXCELLENT TO VERY GOOD: knowledgeable • substantive • thorough development of thesis • relevant to assigned topic	
		26-22	GOOD TO AVERAGE: some knowledge of subject • adequate range • limited development of thesis • mostly relevant to topic, but lacks detail	
		21-17	FAIR TO POOR: limited knowledge of subject • little substance • inadequate development of topic	
		16-13	VERY POOR: does not show knowledge of subject • non-substantive • not pertinent • OR not enough to evaluate	
ORGANIZATION		20-18	EXCELLENT TO VERY GOOD: fluent expression • ideas clearly stated/supported • succinct • well-organized • logical sequencing • cohesive	
		17-14	GOOD TO AVERAGE: somewhat choppy • loosely organized but main ideas stand out • limited support • logical but incomplete sequencing	
		13-10	FAIR TO POOR: non-fluent • ideas confused or disconnected • lacks logical sequencing and development	
		9-7	VERY POOR: does not communicate • no organization • OR not enough to evaluate	
VOCABULARY		20-18	EXCELLENT TO VERY GOOD: sophisticated range • effective word/idiom choice and usage • word form mastery • appropriate register	
		17-14	GOOD TO AVERAGE: adequate range • occasional errors of word/idiom form, choice, usage *but meaning not obscured*	
		13-10	FAIR TO POOR: limited range • frequent errors of word/idiom form, choice, usage • *meaning confused or obscured*	
		9-7	VERY POOR: essentially translation • little knowledge of English vocabulary, idioms, word form • OR not enough to evaluate	



LANGUAGE USE	25-22	EXCELLENT TO VERY GOOD: effective complex constructions • few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions
	21-18	GOOD TO AVERAGE: effective but simple constructions • minor problems in complex constructions • several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions *but meaning seldom obscured*
	17-11	FAIR TO POOR: major problems in simple/complex constructions • frequent errors of negation, agreement, tense, number, word order/function, articles, pronouns, prepositions and/or fragments, run-ons, deletions • *meaning confused or obscured*
	10-5	VERY POOR: virtually no mastery of sentence construction rules • dominated by errors • does not communicate • OR not enough to evaluate
MECHANICS	5	EXCELLENT TO VERY GOOD: demonstrates mastery of conventions • few errors of spelling, punctuation, capitalization, paragraphing
	4	GOOD TO AVERAGE: occasional errors of spelling, punctuation, capitalization, paragraphing *but meaning not obscured*
	3	FAIR TO POOR: frequent errors of spelling. punctuation, capitalization, paragraphing • poor handwriting • *meaning confused or obscured*
	2	VERY POOR: no mastery of conventions • dominated by errors of spelling, punctuation, capitalization, paragraphing • handwriting illegible • OR not enough to evaluate

TOTAL SCORE READER COMMENTS

Figure 3 Jacobs et al.'s (1981) scoring profile

Scoring Rubrics of the Target Words in L2 Writing

The twenty target words were provided for participants to use in their L2 writing test. To this end, the target words were selected early to match the contexts of the two writing topics. Thus, the target words were assigned to two groups according to the core meaning of each word, making two semantic sets and other lexical features of these words were not considered. As such, the core meanings of ten words, including *indicate, achieve, justify, prior, exposed, intelligence, stimulate, lecture, devoted, and persist,* were deemed closer to the context of writing task one, in which Chinese participants were more likely to use these ten words. The other ten words (*approach, consume, rely, access, expand, detect, voluntary, schedule, restrain, and inclined*) fit more neatly into writing task two semantically. Participants were required to integrate at least five out of the ten words in writing each of the two items. They were informed beforehand that they could use any derived form of the target words in their writing.

The scoring of target words used in the L2 writing was adapted from Zhong's (2016) 3-point scoring scale. This scale was devised in her study to assess learners' receptive word knowledge and sentence writing ability. The scoring criteria in this scale embraced such multiple word components as appropriateness, spelling, meaning

accuracy, lexical grammar and collocation accuracy. Three points were awarded for words used correctly and appropriately in L2 writing. Two points were given for words with minor mistakes. These mistakes were limited to one-letter spelling errors (such as adding or missing a letter, misspelling a letter, or mistakenly placing two adjacent letters), grammatical errors (such as mistakes in word parts) and inappropriate collocation (such as "root reason" or "rely in"). One point was given if the word did not fit the context without interfering with understanding the intended meaning. No point was given if the word was misused in meaning and inappropriate in context. The maximum score for the target words would be 30 for each writing task. Inter-rater reliability was also examined in the scoring of the target words. The other rater was trained on the 3-point scale and discussed with the researcher about the uncertain word appropriateness. The maximum score for the target words was 60 points for the two writing tasks. The total scores of all vocabulary tests and the calculation of these scores are summarized in Table 11.

Table 11 Total scores of all vocabulary knowledge tests

Tests	Total scores	Calculation
The VLT	150	3 items × 10 clusters at each level × 5 levels = 150
Form recall	20	20 items × 1 point each = 20
Word pair	20	20 items × 1 point each = 20
Association	57	20 items with max 57 keys = 57
Productive derivative	40	20 items × 2 points each = 40
Productive collocation	20	20 items × 1 point each = 20
Target words	60	20 items (10 items in each writing task) × max 3 points = 60

Inter-rater reliability was verified for L2 writing and target word scores (see Table 12) (Schoonen et al., 2011). The second rater was an experienced English instructor who has taught L2 writing for more than 15 years and was also working on her PhD program. She first examined the assessment criteria and band descriptors prior to the rating work. Then, the scoring profile was explained to her, and she was trained by marking five scripts under the researcher's guidance before her independent rating. The scores awarded by the two raters for each essay and target words were compared, and if the difference was within five scores, they were considered acceptable. If the difference were greater than five scores, the two raters would re-rate the essay and target words to ensure better scoring consistency. The correlation coefficients of all

scores between the two raters ranged from .60 to .90, suggesting high consistency in scoring.

Table 12 Pearson correlation coefficients between the two raters

Writing Task	Content	Organization	Vocabulary	Language	Mechanics	Overall score	Target words
I	.867	.804	.850	.833	.658	.975	.965
II	.833	.789	.874	.776	.642	.936	.976

3.4 Data Collection Procedure

The six vocabulary component tests were administered on two consecutive days to avoid test fatigue, and the testing sequence is shown in Table 10. The VLT is independent of the other depth tests and was the first one to be administered. Schmitt (2010) cautioned that the depth tests designed to capture the multiple components of a word may cause a cross-test effect, which is hard to eliminate in this type of measure. The process of taking one test may provide learning opportunities for tests administered later. In order to minimize the cross-test effect, the depth tests were arranged according to the difficulty of multiple word tests, namely, from the hardest to the easiest tests. The difficulty levels of each depth test were determined with reference to the mean scores of multiple tests in previous literature (Read & Dang, 2022; Schmitt & Zimmerman, 2002; Webb, 2005, 2009; Zhong, 2016). According to previous studies, the difficulty sequence of depth tests in the current study was productive collocation, productive derivative, association, form-recall and word pair. As such, the ordering of the five depth tests was arranged accordingly, as shown in the below table.

The participants did not notice early on that there would be another set of tests on the following day using the same target words. They were informed that the tests were designed to measure how much they had mastered these words, and the test scores would not count in their final scores for that semester, so the tests were low-stakes. The time given for each word test was decided according to the pilot study, yet they were allowed more time if some of them could not complete a certain test. This may reduce the influence of time pressure. During the vocabulary tests, they were not allowed to look up dictionaries or other materials. The researcher and another teacher

monitored the tests and collected each test before handing out the next one so that participants had no chance to seek answers from the previous tests.

Table 13 Vocabulary Tests Sequence

Sequence	Tests	Time allotment (min)
Day 1	The VLT Test	30
	Collocation Test	20
	Derivative Test	25
Day 2	Association Test	20
	Form-recall Test	20
	Word-pair Test	20

One week after the vocabulary tests, the participants were assigned to do the first writing task and use at least five of the ten target words provided in their writing. They were told to write the second argumentative essay and integrate at least five of the other ten target words in their writing the following week. Similarly, they were not allowed to look up any dictionary or material during the writing process. Participants were given one hour to complete the writing tasks, yet another ten to fifteen-minute extension was allowed if they could not conclude the tasks. The current study suggests that the time extensions in vocabulary and writing tests would not significantly impact participants' performances since most of the tests are productive and low-stakes in nature. The extensions may help to minimize wild guessing and incomplete test samples, thereby reducing missing data.

3.5 Data Analysis

When data was garnered from the word tests and writing tasks, the current study used the Pearson product-moment correlation and regression analyses to answer the first research question: What are the correlations among the six word components, namely, vocabulary size, word pair, form recall, association, productive derivative and collocation, and their relationship with L2 writing and word use? The correlation analyses indicated the relationships and the strength of correlations among all word knowledge components and their relationships with L2 writing and word use scores.

In light of the close relationships between various word components, some of the depth test instruments may capture more than one-word knowledge component. Most notably, productive form and meaning are two components that are inseparable parts of each other and featured in other depth components. Nation (2022) and Coxhead (2007) also categorized form and meaning in the same Meaning aspect. In addition,

derivative knowledge is regarded as part of form, and association is a subcategory of meaning in their models. Collocation has been substantiated to be closely related to meaning (Qian, 1999), and it was measured in the same test format as form recall in the current study. As such, partial correlation analyses were conducted to check the variances uniquely explained by association, derivatives and collocation in L2 writing and word use that were also shared with productive form and meaning. This calculation may help to indicate the unique variance shared between two variables (Field, 2009). The productive form was first controlled during the partial correlation analyses, and the partial correlations between other depth components and L2 writing and word use were calculated. The same calculation was done when productive meaning was controlled. Then, the partial correlation coefficients were compared with the standard correlation results. If the partial correlations decreased, the controlled variable (productive form or meaning) shared a large portion of the variance. For example, it was found that when form recall was controlled, the variance explained by collocation in L2 writing and word use dropped dramatically to an insignificant level. This indicates that the productive collocation test also captured a considerable amount of form recall knowledge.

Then hierarchical regression analyses were used to answer the second research question: To what extent do these discrete word knowledge components contribute to university EFL learners' word use and overall L2 writing ability? The hierarchical regression models were conducted with each step entering one independent variable, and all word components were entered one after another. This method indicates the variance that can be explained by each individual predictor variable. The current study suggests that hierarchical regression models might be more insightful and informative than the simultaneous or stepwise counterparts, particularly for studies with multivariate regression analyses. The simultaneous or stepwise regression models can also offer the overall variance accounted for by all variables, yet can hardly reveal the predictive level of each variable. With the hierarchical methods, the R^2 Change and B values in each step would demonstrate to what extent the respective word component can predict and contribute to L2 writing and word use (Field, 2009; Keith, 2015).

It is noted that the entry order of these word components is critical in hierarchical

regression models since the sequence of entry significantly impacts the predictive effect of each predictor variable (Field, 2009; Keith, 2015). Different entry orders may yield varying regression results. The current study determined the entry order based on previous studies and literature in multidimensional vocabulary research. It has been documented that form-meaning links are lexical knowledge that is usually acquired by L2 learners at the very beginning. That is, receptive and productive vocabulary sizes are often learned at the early stage and then proceed to word depth knowledge (Jiang, 2002; Laufer & Goldstein, 2004; Nation, 2022; Webb, 2008, 2009). As such, receptive and productive form and meaning components were entered into the models first in the current study. The derivative knowledge followed form and meaning aspects to be entered as derivative knowledge is also regarded as form knowledge. Because of its metalinguistic nature (Schmitt, 2010), this knowledge might also be acquired at an early stage, even though learners find it difficult to master all word family members (Schmitt, 2010; Schmitt & Meara, 1997; Schmitt & Zimmerman, 2002).

In terms of the sequence of association and collocation, the current study suggests that association should be more readily mastered than collocation. Association, captured as synonyms in the current study, has been categorized as a kind of meaning knowledge (Nation, 2022) that appears in the partially known stage (Whitmore et al., 2004). By contrast, collocation has been empirically proven to be the most difficult word depth knowledge in vocabulary acquisition. EFL learners at all proficiency levels are found to struggle with collocation, especially in language production (Laufer & Waldman, 2011; Nguyen & Webb, 2017). As such, it may be reasonable to enter the predictor variables into the regression models as the following sequence: receptive size (the VLT), productive form and meaning, productive derivatives, association and collocation.

In addition to the null hypothesis significance testing (*p*-values), effect sizes were also calculated as a critical yardstick in the current study. The significant tests may indicate whether there is an effect between two means, correlations and predictions, while effect sizes report the strength of the impact (Aberson, 2010; Zhong, 2014). In other words, it measures the magnitude of the effect that has practical significance in the real world (Cohen, 1988; Selya, Rose, Dierker, Hedeker, & Mermelstein, 2012). The

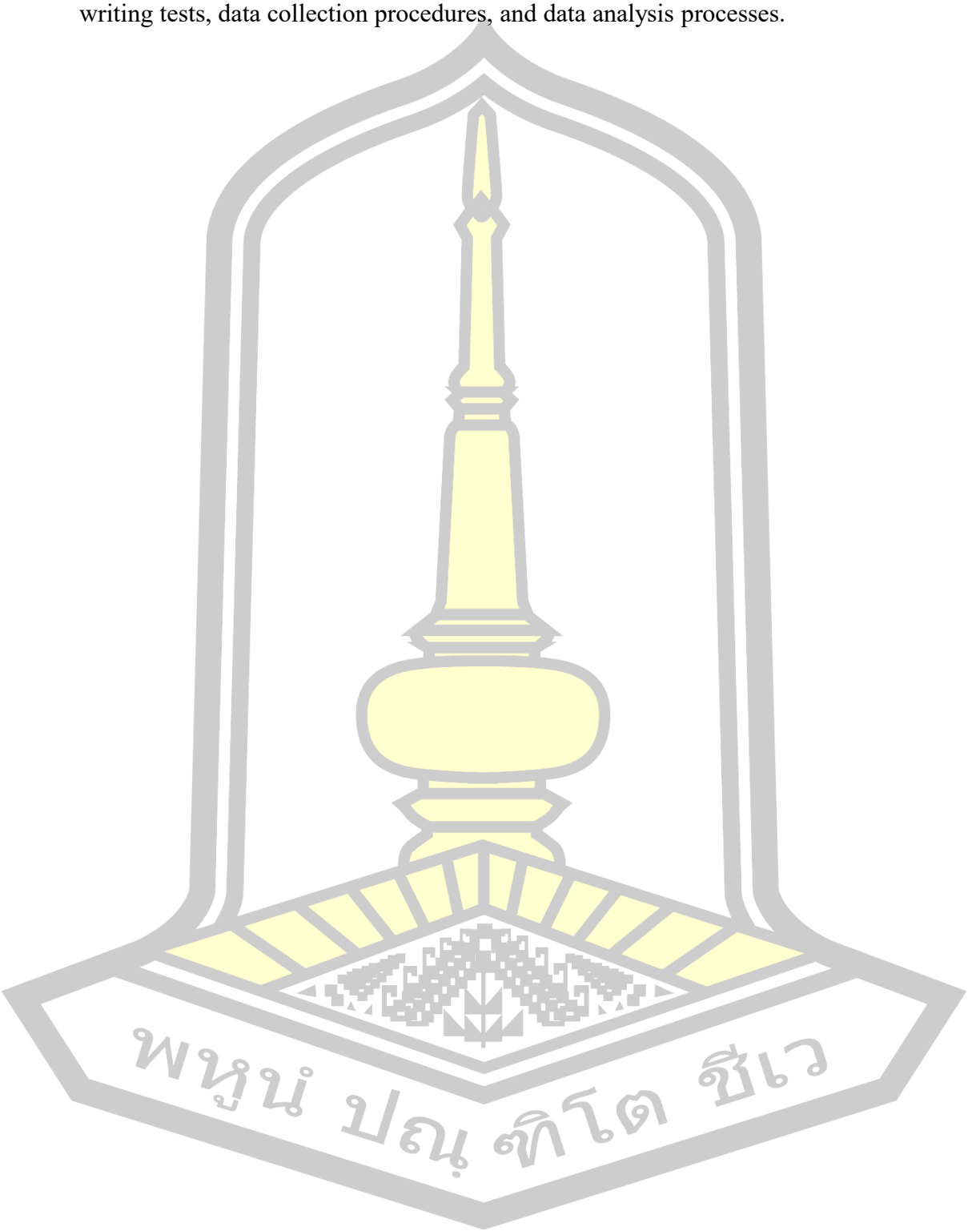
current study squared the correlation coefficient *r* to calculate the effect size of correlation, namely, the R^2 values. The R^2 in a linear relationship indicates the portion of variance from one variable that is also accounted for by the other variable (Cohen, 1988). According to Cohen (1988), 0.01 R^2 represents a small effect, 0.09 a medium effect and 0.25 a large effect in correlations.

The Cohen's f^2 (Cohen, 1988) was calculated in the regression models to determine the effect sizes of predictions. Cohen's f^2 is appropriate for use within a multiple regression model, and the calculating formula for global effect size is presented as $f^2 = \frac{R^2}{1-R^2}$ (Selys et al., 2012). The global effect size shows the overall effect of the variance that can be explained by all variables together. However, measuring the local effect size might be more relevant to the individual effect of each word component in the current study, as shown in the formula $f^2 = \frac{R^2\mathrm{AB} - R^2A}{1-R^2AB}$. In this formula, *B* is the variable of interest, and *A* represents all other variables. Accordingly, R^2A is the variance accounted for by variable *A,* and R^2AB is the variance accounted for by *A* and *B* together. The numerator of the formula reflects the variance that is uniquely explained by *B,* and the denominator calculates the variance that cannot be explained by variables *A* and *B* in the dependent variable (Selys et al., 2012). The current study not only calculated the global effect sizes presented by a set of word components as predictor variables in L2 writing and word use, but also the local effect sizes to determine the strength of effect for individual word knowledge components. Cohen (1988) suggests that an effect size of f^2= 0.02 is small, 0.15 is medium and 0.35 is large.

3.6 Chapter Summary

This chapter details the research paradigm and design of the current study, including the research context, participants and the quantitative methods used. Then, it introduces the test instruments selected for the receptive and productive word knowledge components and L2 writing tasks. Details, including how these instruments were borrowed or adapted and how they were validated to be useful and practical, are

also described in this chapter. It also presents the scoring scales for vocabulary and writing tests, data collection procedures, and data analysis processes.



CHAPTER IV
RESULTS

This chapter presents the quantitative results garnered from the tests of word size and depth knowledge components and L2 writing tasks and word use in writing. The chapter comprises three sections. The first section reports the descriptive statistics, summarizing participants' performance on multiple vocabulary knowledge tests and L2 writing and word use scores. The second section uses the Pearson correlation coefficients to address the first research question regarding the relationships among the six vocabulary knowledge components, L2 writing scores, and word use scores. The third section depicts the contribution of word size and depth in predicting the overall L2 writing performance and productive vocabulary use in L2 writing. Multiple regression analysis and related methods are reported in the final section to cope with the second research question on the contributions made by multiple word knowledge to L2 writing and productive word use.

4.1 Vocabulary and Writing Scores

The descriptive statistics for scores on the vocabulary size and depth knowledge components are presented first, followed by the L2 writing and productive word use scores, including the maximum and minimum scores, mean, standard deviation, skewness and kurtosis. The percentages of the total scores calculated by dividing the total score of each test by the mean were also included. The scores gained on the VLT represent Chinese university EFL learners' vocabulary size as the baseline word knowledge, while the combined score of all word depth components amounts to the overall depth word knowledge in a Chinese university context. The scores on the two writing tasks and vocabulary scores in L2 writing were also combined and calculated into percentages for a sound understanding. Because the total scores for each test are different, all raw scores were calculated into percentages for the ease of comparison and analysis. The summary of descriptive statistics of vocabulary knowledge is shown in Table 14.

Table 14 Scores on vocabulary knowledge size and depth (n = 147)

Test	Total	Min.	Max.	Mean (%)	Std. Deviation	Skewness	Kurtosis
The VLT	150	70	149	116.2 (77.5)	14.498	-0.156	0.295
2000	30	23	30	28.95 (96.5)	1.416	-2.301	6.061
3000	30	16	30	27.32 (92.1)	2.401	-1.822	4.924
Academic	30	17	30	28 (93.3)	2.531	-1.974	4.470
50000	30	2	30	22.80 (76)	5.662	-0.869	0.544
10,000	30	0	30	9.14 (30.46)	8.288	0.794	-0.401
Form recall	20	3	20	14.32 (71.6)	3.324	-0.646	0.499
Association	57	23	49	38.42 (67.4)	6.136	-0.599	-0.044
Collocation	20	4	19	13.20 (66)	3.111	-0.234	-0.480
Derivative	40	14	39	30.01 (75.02)	5.380	-0.534	-0.293
Word pair	20	8	20	17.58 (87.9)	2.666	-1.365	1.299
Depth	157	55	143	113.54 (72.3)	17.086	-0.800	0.555

As illustrated in Table 4.1, participants achieved the highest percentage in the mean of the total score in word pair at 87,9% (M = 17.58, SD = 2.666), followed by 77.5% (M = 116.2, SD = 14.498) in the VLT and 75.02% (M = 30.01, SD = 6.136) in the derivative test. It can be seen that the other mean scores that are over 70% of the total score are form recall (M = 14.32, SD = 3.324) at 71.6% and the overall depth score (M = 113.54, SD = 17.086) at 72.3%, while association (M = 38.42, SD = 6.136) and collocation (M = 13.2, SD = 3.111) have similar percentage scores with collocation being the lowest at 66% among depth component tests. The overall word size and depth tests were performed fairly well, and word size knowledge by the VLT was a little higher than the overall depth percentage score (M = 113.54, SD = 17.068).

Further examination was checked for the normality of all the test scores. The skewness and kurtosis for the majority of tests were found to be within normal range except that the VLT had a kurtosis higher than three at the 2000 level, 3000 level and academic level frequency, hence a high skewness on these score distributions. This may be caused by the homogeneity of participants' remarkable ability to recognize high-frequency words receptively. Moreover, the scores of all depth component tests and the overall depth score are negatively skewed, suggesting that participants' scores on these tests clustered towards higher scores.

The summary of the descriptive statistics about participants' L2 writing (W) scores and the scores they acquired in the target word use (TW) and vocabulary component (VC) scores in L2 writing are presented in Table 15. The highest percentage in the mean of

the total score was found in the vocabulary component scores participants achieved in writing task one (VC one) at 80.4% (M = 24.13, SD = 1.664) and in writing task two (VC two) at 81.5% (M = 24.46, SD = 1.142), respectively. In terms of L2 writing, participants gained similar scores in the two writing tasks, with writing task one (W one) at 76.52% (M = 76.52, SD = 5.589) and writing task two (W two) at 79.09% (M = 79.09, SD = 4.307). The lowest percentages came at the twenty target words participants chose to use in their two writing tasks. The target word use in writing task one (TW one) only achieved 65.73% (M = 19.72, SD = 4.796) and in writing task two (TW two), at 69.9%. The combined percentage scores of L2 writing (CW), target word use (CTW) and vocabulary component score (CVC) are also shown in the table at 77.8%, 67.8% and 80.98%, respectively.

Table 15 Scores on L2 writing and word use in writing (n = 147)

Test	Total	Min.	Max.	Mean (%)	Std. deviation	Skewness	Kurtosis
W one	100	58	90	76.52 (76.52)	5.859	-0.783	1.306
W two	100	63	90	79.07 (79.07)	4.307	-0.628	0.891
TW one	30	8	30	19.72 (65.73)	4.796	-0.001	-0.894
TW two	30	10	30	20.97 (69.9)	4.315	-0.220	-0.686
VC one	30	20	28	24.13 (80.4)	1.664	-0.362	0.260
VC two	30	20	27	24.46 (81.5)	1.142	-0.424	1.039
CW	200	125	179	155.59 (77.8)	8.425	-.640	1.431
CTW	60	24	58	40.68 (67.8)	7.517	.084	-.580
CVC	60	40	55	48.59 (80.98)	2.310	-.220	1.031

The distribution of scores on L2 writing and productive word use was also checked for normality. Skewness and kurtosis were found to be within the acceptable range of ±3, suggesting a normal distribution of test scores. The skewness values of almost all tests in this part are negative, meaning that these tests clustered closer towards higher scores. The kurtosis values were found negative in the scores of the target word use in writing tasks, which indicates that participants tended to obtain homogeneous scores on the target word use. The positive kurtosis for scores of L2 writing and vocabulary component in writing scores means that more participants acquired scores at the high and low areas than when it is a normal bell curve distribution.

4.2 Relationships between Vocabulary Components and L2 Writing

This section presents the findings regarding the first research question: What are the relationships among the six vocabulary knowledge components, and how are they related to the overall writing performance and productive word use in L2 writing

tasks? Correlation analysis was used to see the strength of correlations between all these vocabulary variables and L2 writing variables. Multiple regression was also deployed with the specific *r* square values to examine the variance in L2 writing and productive word use that can be explained by each word size and depth component and the combined vocabulary depth score.

Figure 4 and Table 16 show the boxplot distributions and statistics for the five word depth components scores gained by Chinese university EFL learners. The mean and median of these word depth components are comparatively close, suggesting that the data distribution is relatively symmetric. The median lines in the middle of the boxplots indicate that except for word pair, which has the highest median (90), the other four word depth components present close median, decreasing from derivative (75) to form recall (70) and to the association (68.42) and collocation (65). Note that the word pair shows no whiskers above the third quartile (also known as the 75% percentile), and the weighted average of the word pair at the 75% percentile is 100. This means many Chinese participants could obtain a full score (100) on the word pair test. Together, the data distributions of the five depth components suggest that Chinese university EFL learners have a progressive acquisition of word depth knowledge from productive meaning and form links to association and collocation knowledge.

The boxplot also presents the case numbers of outliers. Four outliers were identified in word pair, two in form recall and one in association, while the data for collocation and derivative are normally distributed. It is noted that the top line of the box is the 75% percentile (Q3), and the bottom line of the box is the 25% percentile (Q1). Outliers that are respectively higher and lower than the 1.5 interquartile range (IQR) from the Q3 and Q1 percentiles are considered mild outliers, whereas those that are respectively higher and lower than 3 IQR from the Q3 and Q1 are deemed extreme outliers (Field, 2009; Keith, 2015). By calculation, all the outliers below the Q1 cut-offs in word pair, form recall and association are mild with circle marks. Thus, these data points were also included as they do not critically impact the overall data analysis.

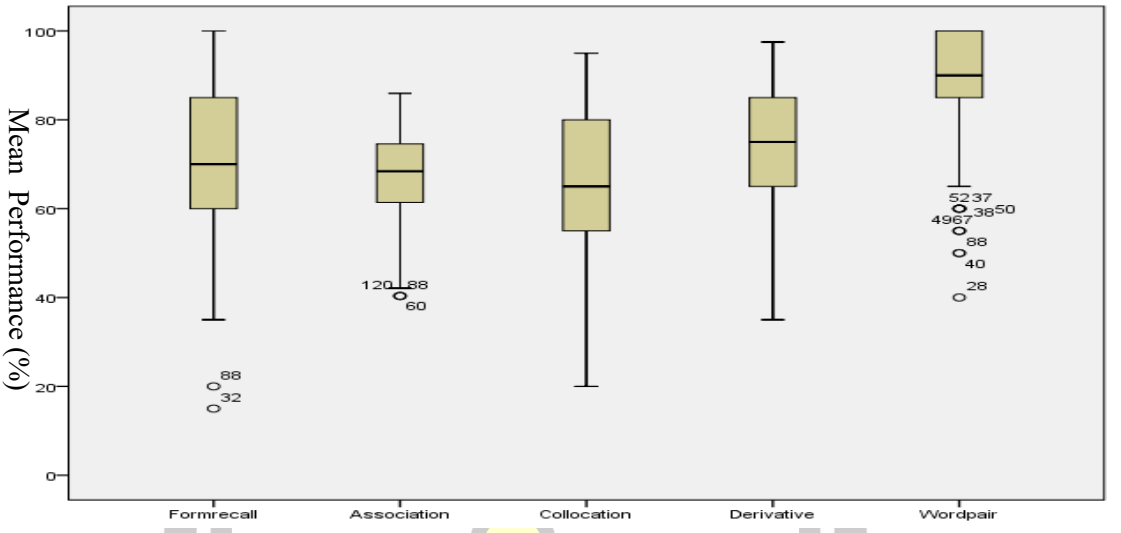


Figure 4 Boxplot for word depth knowledge components

Table 16 Descriptive statistics for word knowledge components

	From Recall	Association	Collocation	Derivative	Word pair
Mean	71.59	67.41	66.02	75.03	87.89
Median	70	68.42	65	75	90
St. Deviation	16.62	10.77	15.56	13.45	13.33
IQR	25	14.04	25	20	15
Skewness	-.646	-.599	-.234	-.534	-1.366

Note: N = 147

From Figure 5 and Table 17, it can be seen that both word size by the VLT and the overall depth knowledge have close mean and median, indicating that the two sets of data are normally distributed. The whiskers of the VLT and overall depth score also present a symmetric distribution. The mean and median of the VLT are a little higher than those of the depth knowledge, meaning that Chinese university EFL learners have more receptive word knowledge than productive depth knowledge. Moreover, both data sets are negatively skewed, with the majority of participants obtaining scores that concentrate towards the higher ends of the scales.

Separately, four outliers in the depth score and two in the VLT score were found, all below the bottom line of the boxes. According to the case-wise diagnostic, these outliers are all considered mild and marked with circles. No extreme outliers in the datasets that would significantly influence the analysis results. Therefore, the data analysis process included all outliers in the VLT and depth knowledge.

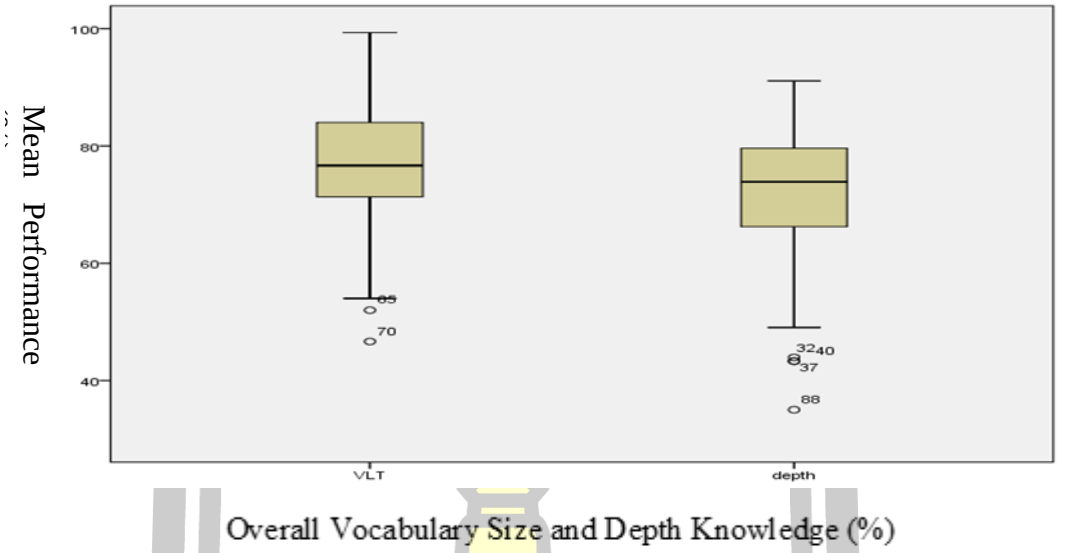


Figure 5 Fig Boxplot for word size and depth knowledge

Table 17 Descriptive statistics for word size and depth knowledge

	Word size (VLT)	Word depth knowledge
Mean	78.48	72.32
Median	76.67	73.89
Std. deviation	9.665	10.88
IQR	12.67	13.38
Skewness	-.156	-.800

Note: N = 147

4.2.1 Relationships among Multiple Vocabulary Components

Vocabulary knowledge is a complex construct of multiple components (Nation, 2022; Suyking, 2018; Zhong, 2016). The correlation matrix with regard to the six receptive and productive word size and depth components and the overall word depth score are shown in Table 18. All the correlation coefficients between the receptive size and productive depth word knowledge components are positive and significant at $p < 0.01$ level, particularly among the word depth components, ranging from .52 to .86. However, all pairs of correlations between the VLT and depth component variables are relatively weak, ranging from small ($r = 0.10$) to medium ($r = 0.30$) effects (Cohen, 1988; Field, 2015).

The strongest correlations among vocabulary depth components were found between form recall and depth ($r = 0.86$, $R^2 = 0.73$) and derivative and depth ($r = 0.86$, $R^2 =$

0.74), followed by association and depth (r = 0.85, R^2= 0.72). Form recall and derivative have the highest correlation among the individual depth components at r = 0.73, R^2= 0.52. Form recall also correlates significantly with collocation (r = 0.67, R^2 = 0.44), which may be caused by the similar test format of the two tests, both requiring participants to complete the word in a sentence with the first word letters being given as cues. Other pairs of word depth components that bear strong correlations are derivative and word pair (r = 0.62, R^2=0.39), form recall and association (r = 0.61, R^2 = 0.37), and form recall and word pair (r = 0.60, R^2= 0.36).

Table 18 Correlation coefficients among multiple word components (n = 147)

Test	1	2	3	4	5	6	7
1. VLT	-						
2. Form recall	.17**	-					
3. Word pair	.12**	.60**	-				
4. Association	.18**	.61**	.58**	-			
5. Derivative	.17**	.73**	.62**	.59**	-		
6. Collocation	.18**	.67**	.53**	.52**	.54**	-	
7. Word depth	.20**	.86**	.77**	.85**	.86**	.75**	-

*P*** < .01; *p** < .05 (two-tailed)

In summary, vocabulary size, as measured by the VLT is not significantly correlated with word depth knowledge in Chinese university EFL learners. However, the five depth components measured have positive and large correlations with one another and all of them are highly associated with the overall depth knowledge. This suggests that the word depth construct is composed of various interrelated sub-knowledge components as tabulated in Nation's (2022) proposal.

4.2.2 Relationships between Word Knowledge and L2 writing and Word Use

Globally, this section examines the correlations between the overall vocabulary size and depth scores and L2 writing and productive word use scores in L2 writing. The correlation coefficients of the VLT and the overall depth score associated with scores of two writing tasks (W) and the target words (TW) and word components (VC) are shown in Table 19. As illustrated, vocabulary size by the VLT barely correlates with L2 writing and productive word use, according to Cohen (1988). All pairs of correlation coefficients between the VLT and L2 writing scores and word use scores

are lower than 0.20 at $p < 0.01$ level (two-tailed). This points towards the little relationship between word size by the VLT and L2 writing and productive word use performances. By contrast, the overall depth knowledge score presents moderate to strong correlations with the two writing scores and the target words and vocabulary component scores. It is apparent that word depth knowledge is more associated with writing task one ($r = 0.63$, $R^2 = 0.40$) as well as its target words ($r = 0.67$, $R^2 = 0.45$) and vocabulary component ($r = 0.59$, $R^2 = 0.35$) than writing task two. Together, the overall depth score has statistically linear and large correlations with the combined L2 writing (CW) score at $r = 0.72$, $R^2 = 0.52$, and the combined target word (CTW) score at $r = 0.76$, $R^2 = 0.58$, and the combined vocabulary component (CVC) score at $r = 0.67$, $R^2 = 0.45$.

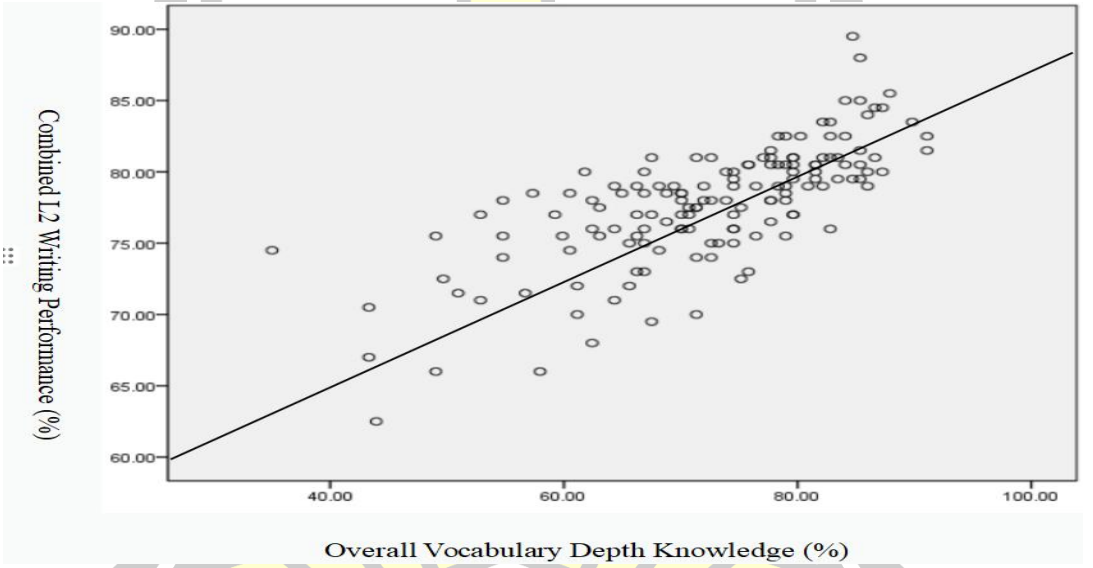


Figure 6 Scatterplot of depth knowledge and L2 writing performance

The scatterplot of the overall vocabulary depth knowledge and combined L2 writing score shows that the scores between the two performances illustrate a linear and positive correlation in Chinese university EFL learners despite not being a perfect line. That is, the L2 writing score grows along with increased vocabulary depth knowledge. Chinese EFL learners with higher word depth knowledge scores tend to fare better in L2 writing tasks and vice versa. This suggests that L2 writing proficiency has a linear and positive relationship with the amount of word depth knowledge they can manipulate.

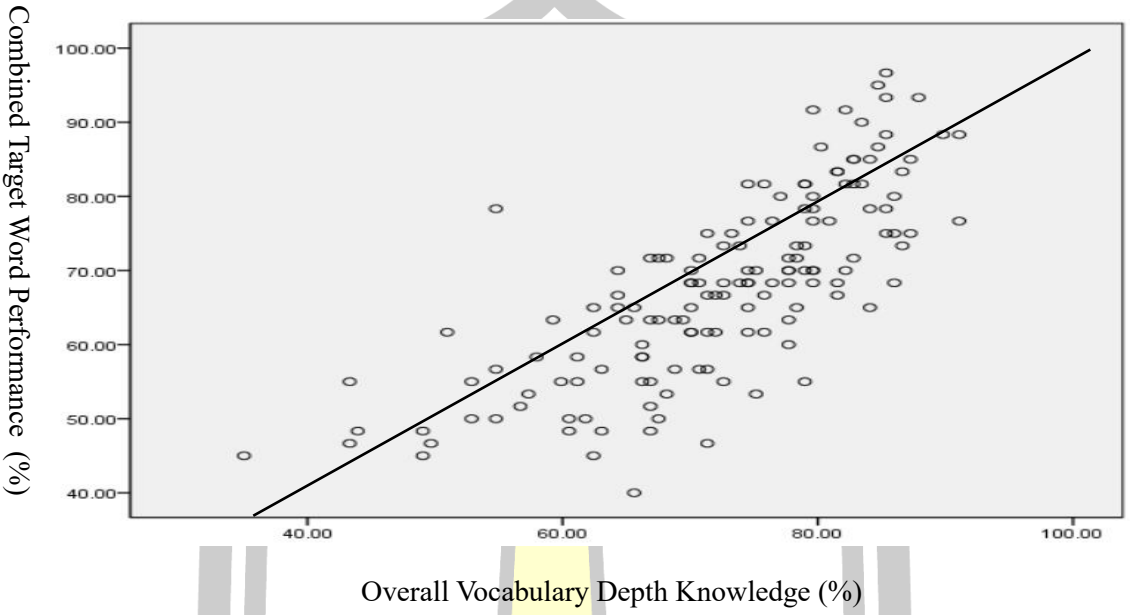


Figure 7 Scatterplot of depth knowledge and target word performance

The scatterplots in Figure 7 and Figure 8 also illustrate positive and linear relationships between the overall vocabulary depth knowledge and the target word performance and vocabulary component score in L2 writing. Nevertheless, neither figure shows a perfect line between the related variables. The scores of the target words are generally lower than the overall word depth score. Despite Chinese university EFL learners' good knowledge of word depth tests, they may still have difficulty using the target words in authentic contexts. A homogeneous quality characterizes the vocabulary component scores obtained by Chinese EFL learners; many participants obtained the same scores on this part. This is probably due to the analytic criteria used for rating vocabulary component scores in L2 writing. Taken together, the target words performance and general vocabulary use in L2 writing are in positive proportion to vocabulary depth knowledge. In other words, with more word depth knowledge, Chinese participants performed better in productive word use in L2 writing.

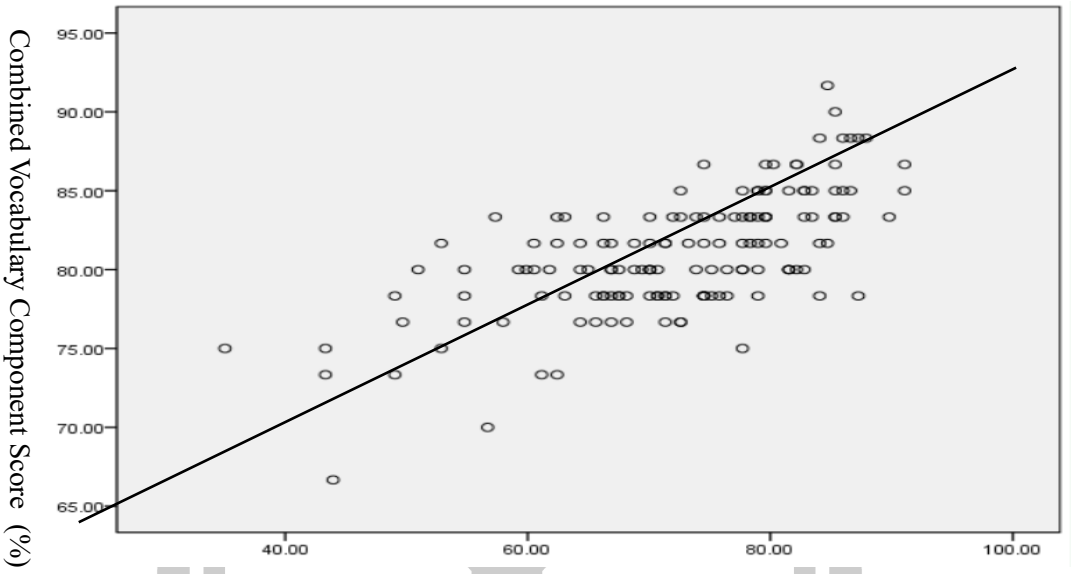


Figure 8 Scatterplot of depth knowledge and vocabulary component score

With a fine-grained examination, the variables relative to L2 writing and productive word use within each task also bear positive and significant correlations. The highest coefficient came at writing task one (W one) and its vocabulary component (VC one) (r = 0.84, R^2= 0.70), followed by writing task two (W two) and its vocabulary component (VC two) (r = 0.78, R^2= 0.60). Moreover, strong correlations were also found between two writing tasks and their target word scores, with writing task one (W one) and its target word score (TW one) at r = 0.61, R^2= 0.37 and writing task two and its target words score (TW two) at r = 0.63, R^2= 0.40. However, the correlations between W one and W two and their vocabulary scores are rather weak with reference to Cohen (1988), which means that productive word use is highly tied in with specific writing topics.

Table 19 Correlations between size and depth and L2 writing and word use (n = 147)

Test	1	2	3	4	5	6	7	8	9	10	11
1. VLT	-										
2. Depth	.20**	-									
3. W one	.07**	.63**	-								
4. W two	.19**	.56**	.36**	-							
5. TW one	.12**	.67**	.61**	.23**	-						
6. TW two	.14**	.59**	.36**	.63**	.36**	-					
7. VC one	.03**	.59**	.84**	.31**	.61**	.32**	-				
8. VC two	.15**	.49**	.31**	.78**	24**	.54**	.33**	-			
9. CW	.15**	.72**	.88**	.76**	.54**	.57**	.74**	.61**	-		
10. CTW	.15**	.76**	.59**	.51**	.84**	.80**	.57**	.46**	.67**	-	
11. CWC	.09**	.67**	.76**	.61**	.56**	.50**	.89**	.73**	.84**	.64**	-

*P*** < .01; *p** < .05 (two-tailed)

It is clear that the overall depth knowledge has large correlations with L2 writing and productive word use. Individually, the correlations between the specific word depth knowledge components and L2 writing and word use were also checked, and the results are shown in Table 20. Looking separately, all depth components measured have moderate to large correlation coefficients with the two writing tasks and their vocabulary scores ranging from r = 0.33 to r = 0.66. However, with the combined writing and vocabulary scores, the depth components present even more positive and significant correlations with L2 writing and word use performances. The highest correlation coefficients were found between derivative and combined writing score (CW) (r = 0.76, R^2= 0.57) and combined target words score (CTW) (r = 0.77, R^2= 0.59), followed by form recall and CW (r = 0.62, R^2= 0.38) and CTW (r = 0.74, R^2= 0.55). The lowest correlation coefficients are between collocation and CW (r = 0.49, R^2= 0.24) and combined vocabulary component (CVC) score (r = 0.48, R^2= 0.23). Moreover, word pair and association also have significant and strong correlations with the combined writing score (CW) at r = 0.61 and 0.50, respectively. The combined target words (CTW) score is more correlated with word depth components than the combined word component (CVC) scores. This is perhaps because the twenty target

words that participants chose to use in their writing tasks are the same words used in all the five depth tests.

Table 20 Correlations between depth components and L2 writing and word use (n = 147)

Test	Form recall	Association	Collocation	Derivative	Word pair
W one	.54	.44	.43	.64	.56
W two	.47	.39	.38	.61	.43
TW one	.66	.44	.52	.64	.54
TW two	.55	.41	.35	.63	.47
VC one	.52	.43	.44	.55	.54
VC two	.37	.38	.33	.55	.39
CW	.62	.50	.49	.76	.61
CTW	.74	.52	.53	.77	.62
CVC	.55	.50	.48	.67	.53

$P^{**} < 0.01$; $P^{**} < 0.05$ (two-tailed)

In summary, this section presents the findings concerning the first research question about the correlations among the multiple vocabulary size and depth components and their relationships between L2 writing and productive word use. Results from the data analysis show that although the five depth components are highly correlated with one another, the VLT size component is not significantly related to the other depth word components. This may be caused by the variability of Chinese EFL learners' mastery of word size and depth knowledge. Moreover, word size by the VLT was not linearly related to L2 writing performance and merely had meagre correlation with productive word use in writing tasks. The reasons will be discussed in detail in chapter 5. By contrast, word depth knowledge presents a linear and significant correlation with L2 writing and word use. All five depth components in question have moderate to strong correlations with two writing tasks and productive vocabulary use and their combined scores. Chief among these components are derivative, form recall and word pair. Association and collocation have relatively small correlations with L2 writing and productive word use tests. This further corroborates the idea that multiple components of productive word knowledge play a substantial part in language production in a real and natural context and that receptive knowledge is not enough for productive purposes.

Although separate tests were used to capture different word knowledge components, they are closely related to one another in real contexts (Gonzalez-Fernandez, 2022; Gonzalez-Fernandez & Schmitt, 2022; Sukying, 2022), particularly in terms of form and meaning (Nation, 2022; Schmitt, 2010). For example, association is regarded as part of meaning knowledge and morphological form is part of form knowledge (Nation, 2022; Schmitt, 2010). Thus, partial correlations were also examined to look into how much unique variance between certain depth components (association, derivative and collocation) and L2 writing and word use is also shared by productive form and meaning knowledge. It was found that when meaning (word pair) was controlled, the partial correlations between association and L2 writing and word use had a significant drop of 54% and 51%, respectively, from the standard correlations. More importantly, when form (form recall) was controlled, the partial correlations between collocation and L2 writing and word use reduced dramatically (approximately 71% and 87%, respectively). Partial correlations between derivative and L2 writing and word use also reduced about 30% when form recall was controlled. This suggests that a large portion of the unique variances between these depth components and L2 writing and word use were also shared by productive form and meaning.

4.3 The Contribution of Vocabulary Knowledge to L2 Writing and Word Use

This section describes the results addressing research question two: To what extent do vocabulary knowledge components (i.e., receptive vocabulary size, L1-L2 word pair, form recall, association, productive derivative and collocation production) contribute to Chinese university EFL learners' L2 writing and productive word use? Hierarchical regression models and related statistical methods were used to analyse the data.

4.3.1 The Contribution of Vocabulary Knowledge to L2 Writing

Table 21 below illustrates the results from multiple regression analysis regarding the prediction of multiple word depth components on L2 writing performance. The R^2 indicates the overall variance of L2 writing that can be accounted for by each model composed of the corresponding predictor word component and the component(s) in the previous model(s). For example, in Model 1, the R^2 indicates the total variance in L2 writing that can be explained by form recall only, while the R^2 in Model 2 shows the proportion of L2 writing variance explained by form recall and word pair. Following this calculation, Model 5 includes all of the five word depth components in explaining

the variance of L2 writing. The R^2 change represents the addition of variance contributed by the respective word depth component(s) to L2 writing performance. The local effect size for each model was calculated using Cohen's f^2 to present the variance in L2 writing that can be explained by each variable predictor. According to Cohen (1988), the effect is small at $f^2 \geq 0.02$, medium at $f^2 \geq 0.15$, and large at $f^2 \geq 0.35$. Moreover, the B values, standard errors (SE B) and Beta (β) extracted from model 5, which includes all depth components, are also reported here. Specifically, the B value (slope) denotes the relationship between predictor variables and the dependent variable, which in the current study indicates the predicted increase in L2 writing for each unit of increase in the depth word components. The Beta (β) is the standardized value for B, suggesting the contribution of each word component to L2 writing.

Table 21 Regression results of word depth components for L2 writing (n= 147)

Model & Predictors	B	SE B	Beta (β)	Sig.	R^2	R^2 Change	effect size
Model 1				.000	.380	.380	0.969
Form recall	.016	.023	.064	.000			
Model 2				.000	.469	.089	0.227
Word pair	.062	.023	.197	.000			
Model 3				.000	.607	.138	0.352
Derivative	.179	.026	.571	.000			
Model 4				.000	.607	.000	.000
Association	-.003	.028	-.008	.970			
Model 5				.000	.608	.001	0.002
Collocation	.010	.020	.038	.600			

$p < 0.001$ (2-tailed); dependent variable = L2 writing

As illustrated in Table 21, form recall alone (F (1,145) = 88.992) in Model 1 is statistically significant at $p < 0.001$ in explaining 38% variance of L2 writing with a large effect size at $f^2 = 0.969$. With the word pair in Model 2, the R^2 increases by 0.089 to 0.469 with a medium effect size ($f^2 = 0.227$). The R^2 change is dramatically significant with the addition of derivative in Model 3 (F (3,143) = 73.586, $p < 0.001$), increasing the variance explained in L2 writing to 60.7% and the effect size is also large ($f^2 = 0.352$). The model with form recall, word pair and derivative accounts for more than half of the variance of L2 writing. When entering association in Model 4 and collocation in Model 5, the R^2s barely change with association increasing zero and collocation adding 0.1% of the overall variance accounted for in L2 writing. However, all of the five models, including Model 4 and Model 5, are statistically significant at $p < 0.001$ level, and the F values are significant in all five models. This means that all

five regression models fit well with the data in predicting the variance of L2 writing. Although association and collocation are insignificant in explaining the variance of L2 writing, they are also important word components in L2 writing performance.

More specifically, the R^2 change and the local effect size indicate the contribution of each vocabulary depth component to L2 writing. Form recall has the largest value of R^2 change and effect size, followed by the derivative knowledge, which ranks second in effect size and considerably increases the R^2. Word pair is the third strongest predictor in the variance explained and effect size values. The association suggests no addition to the variance explained and effect size in L2 writing, and the contribution by collocation is apparently small in effect (f^2= 0.002). However, the global effect size calculated for Model 5, which includes all predictors, is large (f^2 = 1.551). This indicates that the five productive word depth components together have a large effect in predicting the variance of L2 writing.

Apart from R^2 change and effect size, the B value and its standardized Beta (β) are also critical yardsticks for judging the effect strength of word depth components on L2 writing. The highest B value came in derivative at 0.179, meaning that with one unit increase in derivative, the L2 writing score would increase by 0.179 units with other variables being held constant. As one unit of word pair increases, 0.062 units would increase in the L2 writing score, followed by form recall (B = 0.016) and collocation (B = 0.010). Yet, association has a negative B value, which may be caused by the fact that participants applied little association knowledge directly in L2 writing tasks. Correspondingly, derivative (β = 0.571) and word pair (β = 0.197) significantly contribute the most variance to L2 writing, followed by form recall (β = 0.064) at the p < 0.001 level. Neither association (β = -0.008) nor collocation (β = 0.038) are statistically significant.

It is noted that the sequence of entering the predictor variables in hierarchical regression models influences the variance accounted for by the individual variables (Keith, 2015). The entry order in the current study is built upon the previous literature, and the word components EFL learners would mainly use in the L2 writing process are form and meaning knowledge (Jiang, 2002; Laufer & Goldstein, 2004; Zhong, 2014; Nation, 2022). Moreover, because the current study looks into the internal structure of

productive vocabulary knowledge, the multiple components are highly related and may interact with one another. The large correlations among these word components may cause collinearity. Thus, collinearity diagnostics were reported here, as shown in Table 22.

Table 22 Collinearity diagnostics for word depth components

Depth components	Tolerance	Variance inflation factor (VIF)
Form recall	.350	2.589
Word pair	.511	1.955
Derivative	.408	2.449
Association	.535	1.870
Collocation	.518	1.929

The results show that all tolerance values are within the range of 0.3-0.6, and the VIF values are below 6 (the higher the tolerance and the lower the VIF is, the more independent the depth component is). Tolerance values above 0.1 and VIF values below 10 are considered free of concern (Field, 2009; Keith, 2015). This suggests that there are no issues regarding collinearity among any pair of word depth components.

Globally, Table 22 shows the results when the vocabulary depth knowledge combining all five depth components and the receptive size knowledge elicited from the VLT were entered the regression models with L2 writing as the dependent variable. With depth alone, the Model explains more than half of the variance in L2 writing ($R^2 = 0.522$) with a large effect size ($f^2 = 1.092$) at the significant level of $p < 0.001$. The L2 writing score presents a positive, linear relationship with the depth knowledge, and each unit increase in depth knowledge would result in 0.280 (B value) units of increase in L2 writing. The contribution of the overall depth knowledge to L2 writing is 0.723 ($\beta = .723$), which is a large effect. However, with the addition of the VLT in Model 2, R^2 change and effect size has no change. The B value and Beta are also negative, indicating that the VLT is negatively correlated with L2 writing in the regression model and thus not statistically significant at $p < 0.001$ level. This means that the VLT cannot significantly predict L2 writing performance.

Table 23 Regression results of combined depth knowledge and size for L2 writing (n=147)

Model & Predictor	B	SE B	Beta (β)	Sig.	R^2	R^2 change	effect size
Model 1				.000	.522	.522	1.092
Vocabulary depth	.280	.022	.723	.000			
Model 2				.000	.522	.000	.000
VLT	-.001	.026	-.002	.970			

$p < 0.001$ (2-tailed); dependent variable =L2 writing;

In summary, vocabulary depth knowledge can significantly predict L2 writing, while receptive size by the VLT makes no contribution to L2 writing performance in Chinese university EFL learners. More specifically, among the depth knowledge components, form recall, word pair and derivative explain 60.7% of the variance in L2 writing, and derivative contributes the most, followed by word pair, while form recall has the largest $R^2$and effect size. This indicates that productive form and meaning knowledge play the most critical part in Chinese learners' L2 writing. More importantly, understanding morphological knowledge and L1-L2 word meaning draws the most attention in L2 writing at the Chinese university level. Association and collocation together explain merely 0.1% of the total variance, suggesting that Chinese university EFL learners rely little on association and collocation knowledge in their L2 writing performance. Nevertheless, all the regression models, including association and collocation, are statistically significant, meaning that even though association and collocation are not individually significant in predicting L2 writing, they are still necessary and important word knowledge components in L2 writing proficiency.

4.3.2 The Contribution of Word Knowledge to Productive Word Use

Hierarchical regression models were examined to determine the contribution of word size and depth components to the target word score in Table 24 and the vocabulary component score in Table 25. The R^2s indicating the total variance of productive word use that each model can explain are presented. The R^2 change represents the addition of variance contributed by the respective word component(s) to productive word use. The local effect size in each model has been calculated using Cohen's f^2 to present the variance in productive word use that can be explained by all individual predictor in the corresponding model. Moreover, the two tables below also report the B values, standard errors (SE B) and Beta (β) extracted from the models in the last step, which includes all

word size and depth components. The entry order of building the regression models for productive word use reflects the typical sequence of word knowledge acquisition. It is documented in the previous literature that vocabulary knowledge is often acquired from receptive form and meaning links to productive retrieval and use (Nation, 2022; Read, 2000).

As demonstrated in Table 24, the VLT alone in model 1 cannot significantly predict the target word use and the effect size is small (F (1, 145) = 3.455; f^2= 0.071). The addition of form recall in Model 2 significantly brings about the largest R^2 change, explaining 55% of the total variance in the target word use score at the $p < 0.001$ level with the largest effect size at f^2= 1.622. Word pair and derivative respectively add 4.5% in Model 3 (F (3,143) = 70.137; $p < 0.05$) and 7.8% in Model 4 (F (4,142) = 73.171, $p < 0.05$) to the overall variance explained. The R^2 changes are statistically significant in these two models, resulting in a medium effect size (f^2= 0.138 and 0.240, respectively). However, with association and collocation, neither the group models at the last two steps nor the added predictors are statistically significant in predicting the target word use. The R^2 change brought by association and collocation together is only 0.1%. This means that word components together in Model 5 and Model 6 cannot significantly predict the target word use.

Specifically, only form recall, word pair and derivative are significant predictors of target word use. Chief among these is derivative, of which each unit of increase leads to 0.410 (B value) units of increase in the target word use (β = 0.441). Form recall is the second largest contributor (β = 0.365) with each unit of increase resulting in 0.275 units of increase in the target word use, holding the other variables constant. Although the word pair has a smaller contribution to explaining the target word use variance (β = 0.145), it is still statistically significant in predicting the target word use at $p < 0.001$ level.

Table 24 Regression results of word size and depth components for target word score (n= 147)

Model & Predictors	B	SE B	Beta (β)	Sig.	R^2	R^2 Change	effect size
Model 1				.000	.023	.023	0.071
VLT	.007	.064	.005	.065			
Model 2				.000	.550	.527	1.622
Form recall	.275	.061	.365	.000			
Model 3				.045	.595	.045	0.138
Word pair	.145	.063.	.154	.000			
Model 4				.022	.673	.078	0.240
Derivative	.410	.070	.441	.000			
Model 5				.176	.675	.001	0.003
Association	.055	.077	.047	.457			
Model 6				.179	.675	.000	0.000
Collocation	.009	.054	.011	.875			

$p < 0.001$; $p < 0.05$ (2-tailed); dependent variable = target word score

Chinese university EFL learners may rely more on morphological knowledge (derivative), productive form recall and L1-L2 word meaning (word pair) knowledge in using the twenty target words in their writing tasks. The derived forms of the words and word form spelling caught the most attention when participants manipulated the target words in real contexts. The mapping of L1 to L2 meaning was another focus for them to use these words. The VLT, association and collocation cannot significantly predict the target word use, meaning that Chinese university EFL learners relied little on the receptive recognition of form and meaning knowledge and the deeper productive knowledge of association and collocation.

From the results presented in Table 25, the variance of vocabulary component score explained by word size and depth components gradually progresses from $R^2 = 0.009$ in the one-component Model 1 to $R^2 = 0.483$ in the six-component Model 6. All six group models are statistically significant in predicting the vocabulary component score in L2 writing at $p < 0.001$ level. Thus, each increase in the percentage of variance in vocabulary components is significant, which means that all word components together can significantly help explain the variance of vocabulary component in L2 writing. The largest R^2 change was found in form recall, with 29.7% of the variance in the vocabulary component of L2 writing being explained. Form recall also has the largest effect size among all the word components ($f^2 = 0.574$). Form recall and word pair together in

Model 3, that is, productive form and meaning knowledge account for 36.8% of the overall variance that can be explained in the vocabulary component score. With the addition of derivative in model 4, the R^2 change obtained the second largest increase of 0.104, and the total variance explained added up to 47.2%. Although the VLT, association, and collocation also contribute small variance, they cannot individually predict the vocabulary component in L2 writing at a significant level.

When observed specifically, derivative contributes most to the vocabulary component as each unit of derivative increase brings about 0.136 (B = 0.136, β = 0.475) units of increase in vocabulary component score in L2 writing. The effect size brought by derivative is at the medium level with f^2 = 0.201. What follows is word pair whose B value is 0.035 (β = 0.120), meaning that each unit of increase in word pair knowledge produces 0.035 units of increase in vocabulary component score. The effect size by word pair to explain the variance of vocabulary component is relatively small (f^2 = 0.120). Despite the large proportion of variance explained by form recall, it contributes little to the vocabulary component with reference to the B values (B = 0.005, β = 0.021). As for the VLT, the B value, Beta are negative, which means that the VLT is not positively associated with vocabulary components in L2 writing. Association and collocation are not significant predictors individually, although their additions contribute a small amount of variance explained in general vocabulary component in L2 writing.

Table 25 Results of word size and depth components for word component (n= 147)

Model & Predictor	B	SE B	Beta (β)	Sig.	R^2	R^2 Change	effect size
Model 1				.000	.009	.009	0.017
VLT	-.017	.025	-.043	.266			
Model 2				.000	.306	.297	0.574
Form recall	.005	.024	.021	.000			
Model 3				.000	.368	.062	0.120
Word pair	.035	.025	.120	.000			
Model 4				.000	.472	.104	0.201
Derivative	.136	.027	.475	.000			
Model 5				.000	.478	.006	0.012
Association	.034	.030	.094	.197			
Model 6				.000	.483	.005	0.010
Collocation	.025	.021	.101	.234			

$p < 0.001$; $p < 0.1$; $p < 0.5$ (2-tailed); dependent variable = vocabulary component

In summary, word depth components, particularly productive form and meaning aspects, can robustly predict productive vocabulary use in L2 writing. In contrast, vocabulary size, as measured by the VLT has little impact on productive word use in an authentic context. These word size and depth components appear to have better prediction (67.5%) on the twenty target words that Chinese participants integrated into their writing than on the vocabulary component scores (48.3%) they gained in L2 writing. This may be because the five depth component tests used the same target words as those participants used in their writing tasks, hence a higher correlation between the two sets of data. More specifically, productive form knowledge (derivative and form recall) contribute most to the variance of the target word use in L2 writing, whereas productive form and meaning knowledge (derivative and word pair) can best predict the variance of vocabulary component in L2 writing. Association and collocation can neither significantly predict the target word use nor the vocabulary component in L2 writing. Yet, the regression models, including association and collocation, are statistically significant in predicting the vocabulary component. This means that the association and collocation knowledge of the words used in L2 writing still play critical roles in the overall word use score.

4.4 Chapter Summary

This chapter first presents the descriptive statistics reporting participants' performances on the multiple word knowledge tests and L2 writing and word use tasks. Then Pearson correlation coefficients and the regression R^2 values were used to determine the interrelations among different word components and between these components and L2 writing and word use scores. After reporting the linear relationship between vocabulary knowledge and L2 writing, this chapter presents the hierarchical regression models to describe the contribution of each word component to L2 writing and word use.

CHAPTER V
DISCUSSION

Fruitful quantitative results are obtained in the previous chapter to answer the two research questions regarding the relationship between receptive and productive word knowledge components as well as their relations with, and contributions to, L2 writing and productive word use. This chapter will discuss these results with reference to the findings in the previous literature under the multi-component frameworks of word knowledge. It mainly includes three sections of discussion: the statistical description, the correlation analysis and the contribution of each word component to L2 writing and word use in writing.

5.1 Descriptive Statistics

Judging from the statistical description, Chinese university EFL learners show varied degrees of mastery in vocabulary size, depth knowledge and L2 writing ability. Taken together, they could achieve better performances in receptive and productive form-meaning link tests, namely, the VLT and the word pair test, than other productive depth tests. Vocabulary form and meaning mapping knowledge is the part of word knowledge that Chinese university EFL learners know best. This is no surprise since form and meaning knowledge are the most fundamental and essential word components and the central concern of a word in context (Laufer, 1994, 1998; Laufer & Goldstein, 2004; Levitzky-Aviad & Laufer, 2013). Vocabulary semantic and form entities are usually learned together and constitute the very first step of learning a word. It makes sense that EFL learners are more likely to know the form-meaning associations in vocabulary acquisition and obtain higher scores on these tests (Jiang, 2002; Miller, 1999).

Notably, the L1-L2 word pair gained the highest score, pointing to the importance of this productive form and meaning association in EFL vocabulary learning (Waring, 1997; Webb, 2009). This result is relatively surprising since many participants in the current study could achieve a full score on this test, which could be reflected by the highest mean score (M = 87.5%) among all vocabulary tests. This is inconsistent with Laufer and Goldstein (2004) who identified that L1-L2 word pair translation (active recall) is actually the most difficult form-meaning link that learners know worst. The discrepancy may be caused by the fact that Chinese EFL learners mainly focus on the

L1 translation equivalent when learning a word, and thus, the L1 translation becomes the main connecting point between the L2 word in lexical production (Jiang, 2000, 2002). The typical learning method by Chinese EFL learners may partly explain the highest score in the word pair test.

On the other hand, the receptive size measured by the VLT gained similar and high percentages at the first three levels (academic words are not frequency-based), followed by the 5000 and 10,000 levels. The decrease in scores with frequency levels is expected because vocabulary knowledge is more likely to be acquired at high frequency levels than low frequency levels, supporting the acquisition order of knowing a word (Nation, 2022). The overall VLT score is also higher than the overall depth score, meaning that the participants have more receptive word knowledge than productive depth knowledge. The mastery of word knowledge is consistent with the previous literature documenting that receptive vocabulary size is usually larger than productive word knowledge, and depth usually lags behind size in growth rate (Laufer, 1998; Schmitt, 2010; Webb, 2008).

Moreover, the scores of other productive depth components have similar mean scores, especially among form recall, collocation, association, derivatives and the overall depth score. The cross-effect or mutual interference between different depth tests may be responsible, even though the five depth tests were arranged from the hardest to the easiest to minimize the effects of the previous tests. Schmitt (2010, 2014) and Read (2004) have cautioned that the cross-effect might be one of the limitations common in multi-component word knowledge tests in all studies. The similar mean scores among depth tests may also be due to the intrinsic associations featured in these vocabulary components since the internal components of word knowledge are interrelated and overlap to a large degree (Gonzalez-Fernandez, 2022; Gonzalez-Fernandez & Schmitt, 2020; Schmitt, 1998).

Regarding the L2 writing and productive word use scores, Chinese university EFL learners achieved similar performances on the two essay writing tasks and lexical proficiency. This may be partly due to the analytic rating criteria used for L2 writing tasks, which is more apparent for L2 raters to award scores more than the holistic criteria. There may be no big difference of scores given for each part in writing

according to the descriptors in each scoring scale. The similar scores on the two writing tasks also suggest that Chinese university EFL learners' L2 writing proficiency and vocabulary competence are constant in real and similar contexts.

However, the scores participants gained on the target words (M = 67.8%) they randomly chose to integrate into their writing tasks are obviously lower than vocabulary component scores (M = 80.98%). Three reasons may be accountable for this score difference. First. it may indicate that using a word in a controlled context may cause more difficulty for EFL learners than in a free context. This suggests that different productive tasks may activate and call for different lexical competence (Zhong, 2014). Second, the difference may also be explained by the specific writing topics, which may exert influence on the choice and use of a word. Topics indeed would constrain the use of the target words in context (Weigle, 2002). Third, the requirement of the writing tasks may also serve to explicate the lower scores of the target words. Because at least five prompt words (the target words) were required to be used in each of the writing tasks, many participants only used five of the target words in their writing, even though they gained a high score in the essay writing. They did not use more target words perhaps to avoid extra lexical errors when using them in writing. This may lead to a relatively lower score of the target words than the vocabulary component.

The descriptive results show that Chinese university EFL learners have a varying mastery of different word knowledge components. Generally speaking, they have more receptive word knowledge than productive with a larger vocabulary size than depth knowledge. Word form and meaning links are the best mastered aspects, while depth components like association and collocation relatively lag behind.

5.2 Multi-component Word Knowledge and L2 Writing and Word Use

This section will discuss the first research question regarding the nature of vocabulary knowledge associated with L2 writing ability and lexical use proficiency. Under the multi-component models, this section discusses and compares the results of the current study with previous findings, centering around the correlations among multiple receptive and productive word knowledge components and the relationship between these word components and L2 writing and word use in writing.

5.2.1 Multi-component Word Knowledge Construct

Previous research has generally agreed that vocabulary knowledge is a complex construct that is composed of various receptive size and productive depth components or dimensions (Coxhead, 2007; Nation, 2020, 2022; Read, 2000; Schmitt, 2010, 2014). Due to the time-consuming nature and limited target items of multi-component tests, few studies have attempted test batteries that encapsulate a comprehensive range of word components concurrently (Read, 2000; Schmitt, 2014). Consequently, the nature of the internal structure of word knowledge remains to be explored (Gonzalez-Fernandez & Schmitt, 2019). The current study is one of these attempts to look into the internal relationship between these word knowledge types. What is known to date is that some word components may be interrelated, and some may be less related (Schmitt, 1998). For example, word components such as form, meaning, association and derivatives may be closely connected (Schmitt, 1998), and receptive and productive form-meaning links, derivatives, multiple meanings and collocations are strongly inter-correlated (Gonzalez-Fernandez & Schmitt, 2020). These different types of word knowledge have also been put in a receptive-productive developmental continuum and positive and significant relations were found between the size and depth word components structurally and functionally in context (Qian, 1999, 2002; Zareva, 2005; Zhong, 2016).

The correlation analysis in the current study also reveals positive and significant relationships between the receptive and productive word components, while the five productive depth components show large correlations with one another. These results provide empirical evidence confirming that vocabulary knowledge is a network of interrelated word knowledge components. This supports a number of previous studies (Gonzalez-Fernandez , 2022; Gonzalez-Fernandez & Schmitt, 2020; Schmitt & Meara, 1997; Schmitt, 1998; Zhong, 2016) that explored the internal structure of word knowledge and demonstrated the interrelatedness between different word components. Specifically, the five productive depth components, namely, productive form and meaning, association, productive derivatives and collocation, are highly correlated, ranging from $r = 0.525$ - 0.725 at $p < 0.001$ level. This means that the various knowledge types of a word do not stand independent of other components, but strongly interact with one another. The increase in one word component might facilitate the

learning of other components (Schmitt & Meara, 1997) and no one component is used alone in language production (Gonzalez-Fernandez & Schmitt, 2020).

The correlation matrix also shows that all five depth components are highly associated with the overall word depth construct, which is also a major finding in Gonzalez-Fernandez and Schmitt (2020). In addition, the current study found that derivatives or morphological knowledge correlates most with the depth score (r = .86), even higher than productive form and meaning. This echoes previous studies demonstrating that morphological awareness can be a key contributor to the word depth construct (Lemmouh, 2010; Li & Kirby, 2015). This knowledge, therefore, in a productive context may better facilitate vocabulary acquisition and use (Sukying, 2020). Overall, productive form and meaning (derivatives, form recall and word pair) constitute the main lexical elements in the depth construct, followed by association and collocation.

Compared with the first attempts of multi-component tests by Schmitt and Meara (1997) and Schmitt (1998), the findings of the current study quantitatively support their studies. The positive and significant correlations in this study are aligned with the moderate correlations between word parts (suffix knowledge) and word association in Schmitt and Meara (1997). Beyond word parts and association, the current study extends to the relations among more depth components, such as productive form-meaning connections and collocation production. The significant inter-correlations in this study also lend quantitative evidence to the qualitative findings by Schmitt (1998), who conducted one-to-one interviews across one academic year. Four vocabulary components in Schmitt (1998), including form, meaning, derivatives and association, are interrelated, and meaning seems to be more closely linked to association and derivative knowledge.

Moreover, our results of the positively significant correlations between the various word components add to Zhong's (2016) receptive tests of multiple word knowledge aspects. Zhong (2016) focused on the receptive aspects of word knowledge and devised five receptive recognition measures to examine the receptive-productive incremental process. The current study attempts to go a step further to explore the same word components as Zhong (2016), but all vocabulary tests in this study were in productive formats. The two studies obtained similarly strong relations between the

receptive and productive word components. In addition, Zhong (2016) tapped into Chinese junior high school students, whereas the current study targeted Chinese university EFL learners. It suggests that the internal structure of vocabulary knowledge may be similar between the lower proficiency learners and the more advanced university EFL learners in the same English-learning context.

Vocabulary knowledge indeed can be conceptualized and operationalized as multiple distinct word knowledge components. The current study and the previous similar studies provide the necessary salience to the interrelated nature of these word components. Yet, the multiple word components interact with one another in distinct magnitudes, and the next section will discuss the strength of correlations between these word components.

5.2.2 Strength of Correlations between Word Knowledge Components

It is clear that the various word knowledge components tabulated in Nation's (2022) model are suggested to be interconnected (Schmitt and Meara, 1997; Webb, 2008). Yet, the strength of correlations between different word components may vary markedly, and some may be more related and some less or unrelated (Schmitt, 1998). This also holds true in the results of the current study.

The receptive form-meaning link (vocabulary size) assessed by the VLT has positive yet small correlations with the other depth components. All pairs of correlation between the VLT and other word tests were below $r = 0.20$. This is inconsistent with the strong relationships (ranging from $r = 0.760$ to 0.895) between the VLT and other depth components in Gonzalez-Fernandez and Schmitt (2020). The insignificant correlation also contrasts with a number of previous studies showing that vocabulary size is closely related to depth knowledge (Qian, 1999, 2002; Schmitt, 2014; Tseng & Schmitt, 2008; Vermeer, 2001). However, despite being closely related, size and depth were also found to be separate and independent components of vocabulary knowledge (Gonzalez-Fernandez & Schmitt, 2020; Qian, 2002; Tseng, 2011; Zhang & Lu, 2015). For example, Qian (2002) found that the VLT size and the Depth of Vocabulary Knowledge (DVK) have fairly large correlations. Qian (2002) also spelled out that the small shared variance between the two constructs points to the fact that size and depth tap into different types of word knowledge. That is, size and depth are different

constructs involving varying word knowledge aspects, even though they are overlapped to a certain extent. This might explain that the recognition knowledge in receptive size and recall ability in productive depth are inherently distinct competence and need different measures to assess. The awareness of this distinction can be critical to the conceptualization of vocabulary knowledge (Gonzalez-Fernandez & Schmitt, 2020).

Apart from the different conceptualizations of size and depth constructs, the meager relationship between vocabulary size and depth in the current study may be closely related to the vocabulary learning strategies peculiar to Mainland China university students. Because of the fierce competition in many tests, most notably, the National Matriculation English Test (NMET), Chinese English learners primarily regard learning English as a simple process of memorizing vocabulary (You, 2010). They accumulate vocabulary knowledge by rote through a variety of word lists characterized by word spelling and form-meaning links. Consequently, most attention is paid to mapping word form and meaning knowledge, ignoring many other types of deeper knowledge. For many, rote learning of the form-meaning links may be the most effective way for them to enlarge their vocabulary size to cram for tests. This situation may also coincide with the previous claims by Schmitt (2014) that the relationship between size and depth becomes weaker for learners with a large vocabulary size. Although word size has increased quickly, the depth knowledge has lagged behind (see Schmitt, 2014 for a review). This uneven growth rate widens the gap between size and depth knowledge.

Comparatively, all productive depth components of the target words in the current study indicate strong correlations with one another, which aligns with the findings by Gonzalez-Fernandez and Schmitt (2020). Yet, their correlations were clearly stronger and more comprehensive than those in the current study. This may be caused by several differences between the two studies. First, productive depth components caught more attention in this study, whereas Gonzalez-Fernandez and Schmitt (2020) set sights on the receptive and productive types of four vocabulary components and the acquisition order. The research focuses, therefore, are different in the two studies. Second, the participants in Gonzalez-Fernandez and Schmitt (2020) were from varying

backgrounds aged between 18-65 with different learning goals, while the participants in the current study were intermediate-level university EFL learners. Their structure of vocabulary knowledge and productive abilities may be distinctive. Finally, they employed Spearman's coefficients in the correlation analysis, the results obtained from which might be substantially different from the Pearson's correlations used in the current study (Hauke & Kossowski, 2011). The above reasons may help to explain the larger correlation coefficients between the word components in Gonzalez-Fernandez and Schmitt (2020). Even the correlations between the various word components of the same target words may produce varying magnitudes of relationship (Zhong, 2014).

With a fine-grained examination, form recall and productive derivative knowledge in the current study have the highest correlation among the individual depth components (r = 0.73, R^2=0.52). This pair of large correlations supports the findings by Gonzalez-Fernandez and Schmitt (2020) that the strongest correlation among all the productive tests is between form recall and derivative recall at r = .811. Zhong (2016) also identified a large correlation between form knowledge and word class, though in receptive recognition formats. The consistency of the close relationship between form and derivatives is expected since derivative knowledge is the subcategory of the form dimension, as shown in Nation's (2022) comprehensive word component model. The derived forms determine the formation of a word and make up the word family, which may naturally connect to form recall knowledge. In addition, derivative knowledge is also highly correlated with association in the current study. This is partially consistent with the medium to strong correlations between the derivative and association in Schmitt and Meara (1997) and Zhong (2014). Yet, the cross-sectional results of the current study regarding this pair of correlations generally support the previous findings. Zhong (2014) and Schmitt and Meara (1997) employed a longitudinal method and found that the relationship between derivatives and association becomes more correlated as vocabulary knowledge develops.

Moreover, previous studies (Schmitt, 1998; Zareva, 2005) have shown that association is closely related to meaning knowledge, which also finds expression in the current study. Our study found that association bears a fairly strong correlation with

productive meaning and form (r = .58 and .61, respectively). This makes sense since knowing associations, especially synonyms, may help understand the full meaning of a word and recall its form in context. As a result, association can be regarded as part of a broader meaning system that the word fits in (Nation, 2022; Schmitt, 2010). Zareva (2005) also reported consistently large correlations between productive association knowledge and self-reported meaning knowledge. Her study generated much stronger correlations (r > .800) between association and meaning than did the current study. The difference may be attributed to the discrepancy of word knowledge conceptualizations and methodologies in the two studies. The current study sets forth vocabulary knowledge construct under the multi-component model developed by Nation (2022) and used multiple tests to elicit each word component separately. On the contrary, Zareva (2005) sought empirical evidence to explain the vocabulary knowledge construct and attempted to validate Henriksen's (1999) three-dimensional model of partial to precise meaning, network depth and receptive to productive mastery. As such, the instrument used in her study to capture different word knowledge components was the Vocabulary Knowledge Scale (VKS, Wesche & Paribakht, 1996). Combining all the receptive and productive word tests within a single test instrument with four self-reported steps may explain the larger correlations between association and meaning in Zareva (2005) than in the current study. The VKS has been regarded as a measure that could not touch on multiple shades of meaning since knowing one meaning of the word suffices to correctly complete all items (Read, 2000; Wolter, 2005, see more in Section 2.1.2.1).

It is a surprising result that collocation has a relatively large correlation with the other depth components, particularly with form recall (r = .67), productive meaning (r = .53) and derivatives (r = .54). In her qualitative interviews, Huang (2010) also identified similar findings that collocation correlates most with word meaning, followed by grammar, which corresponds to derivative knowledge in this study. Contrarily, Zhong (2014) found that collocation and form have the smallest relation, even though other word aspects generate medium to strong correlations. The large correlation between collocation and form in the current study may be explained by the similar test format of the two tests. The design of the form recall and collocation tests both required participants to complete the word form in a sentence, with the first letter(s) being

given as a cue. The two tests may capture similar lexical knowledge because of the similar test design. In addition, derivative knowledge in context determines where to put the word in a sentence and what other words should be used with it (Nation, 2022; Zhong, 2016). That is, collocation is highly relevant to the choice of the derivatives in a sentence context, which may serve to explain the close relationship between collocation and derivative knowledge.

It can be a challenge to compare and discuss the strength of correlations between the various multiple word components, even though these components have the same names in different studies. The correlation results may vary to a certain degree according to the different definitions of vocabulary knowledge construct and the respective methodologies used in different studies. Some studies captured receptive word components and some productive or both; some used Spearman Rank coefficients and some Pearson's correlations to analyze the strength of correlations. All of these differences may yield varying outcomes of correlation magnitudes even between the same pair of word components. Overall, the current study is relatively similar to Gonzalez-Fernandez and Schmitt (2020) and Zhong (2016) in terms of the word components measured and the quantitative methods used, and therefore, identified consistently high correlations among the depth word components: form, meaning, derivatives, association and collocation.

5.2.3 Relationships between Word Knowledge and L2 Writing and Word Use

As indicated in the correlation analysis, vocabulary size measured by the VLT has a weak relationship with L2 writing and productive word use ($r < .20$, $p < 0.001$), which is insignificant in effect. Conversely, the combined depth knowledge presents statistically linear and large correlations with L2 writing ($r = .72$), target word use ($r = .76$) and word component score in writing ($r = .67$). Globally, this result deviates from the long-established theoretical stance that vocabulary size is primarily the central concern and a better indicator of writing proficiency (Laufer, 1994; Laufer & Nation, 1995) and is highly associated with the four micro-language skills (Laufer & Ravenhorst-Kalovski, 2010; Milton, 2013; Stæhr, 2008; Schoonen & Verhallen, 2008). This surprisingly small correlation between receptive size and L2 writing and word use in the current study is possibly caused by the test administrations. Because the

VLT was administered online due to the restriction of Covid-19, many test-takers may have the chances to find the form-meaning mapping answers readily during the test. This could be reflected by the negative skewness of the VLT scores, meaning that most scores clustered towards high values, especially at the 2000, 3000 and academic levels. On the part of the L2 writing test, they had to complete the argumentative writing task within one hour. This may lead to the gap between the inflated VLT score and the real timed writing score.

Lemmouh (2010) generally obtained similar findings as the current study, showing that vocabulary size has no relationship with the take-home essay writing scores, while productive depth reveals moderate correlations with the lexical richness in L2 writing tasks. These relations increased with time, supporting the previous findings by Schmitt and Meara (1997). Yet, Lemmouh (2010) used the PVLT as a measure to assess the productive size knowledge, whereas the current study borrowed the PVLT format as a word depth component to capture the form recall of the target words. In this sense, the construct of vocabulary size in the two studies may not be exactly the same, and the correlations thereof, might not be comparative directly. In addition, the current study assigned timed writing tasks, while Lemmouh (2010) used take-home essays. The controlled writing tasks and the totally free writing tasks may elicit and activate apparently different word knowledge. Yet, the two studies are consistent that it is lexical quality rather than lexical quantity that is more related to L2 learner writing proficiency.

Our results regarding the correlations of vocabulary size and depth to L2 writing are also roughly consistent with Lin's (2015) findings. In her study, moderate relations were identified between the two word knowledge constructs and essay writing, but depth elicited from the DVK (including collocation, synonym and association) is more associated with writing proficiency. The relationship between the overall depth and L2 writing in the current study is even stronger than that generated in Lin's (2015) study. This may be because the current study required test-takers to integrate the target words in their writing. This creates a closer connection between multiple word components and L2 writing scores, particularly with the score of the target words used in writing. It is noted that the participants from Mainland China in Lin's (2015) study have a

depth knowledge weakly related to their writing proficiency (r = 0.28), compared with their Hong Kong counterparts (r = 0.48). Lin (2015) suggested that specific vocabulary learning contexts may significantly affect the correlations and contributions of vocabulary knowledge to writing proficiency. This will be fully discussed in Section 5.3.1 as the participants in this study also come from Mainland China.

Compared to vocabulary size, the depth of vocabulary knowledge, including all depth components in the current study, reveals close to or above strong relations with L2 writing and word use. This supports a number of studies (Baba, 2009; Lemmouh, 2010; Lin, 2015; Sukying, 2023; Varnaseri & Farvardin, 2016) showing the same or similar findings. The interrelations between depth and L2 writing in this study echoes Laufer's (1994, 2013) statements that lexical quality is interconnected with writing quality, and that the two constructs have a strong two-way relationship. This is unsurprising since the depth of vocabulary knowledge provides a wider range of extensive information about the word in learners' mental lexicon (Nation & Gu, 2007) and significantly improves the fluency and automaticity of lexical production in context (Schmitt, 2008). These factors are all directly related to writing ability. However, since the depth construct is so complicated that many more types of vocabulary knowledge are involved than the form-meaning receptive size. To date, there has been no standardized measure to assess vocabulary depth knowledge (Read, 2004; Schmitt, 2014). Previous studies approaching depth from the component perspective have revealed disparate correlations between different depth components and writing proficiency in different contexts.

In the current study, the strength of correlations between multiple depth components and L2 writing and word use follows such a magnitude sequence as derivatives, form recall, word pair, association and collocation. That is, morphological awareness best correlates with L2 writing, followed by productive form and meaning senses and association networks. Collocation production is the only component that correlates moderately with L2 writing. The correlation strength of multiple word depth components in L2 writing roughly corroborates Zhong (2014), who measured the same word components in receptive format associated with L2 sentence writing. The only difference is that derivatives in the current study are better indicators of L2

writing and word use than productive form and meaning. The discrepancy may be due to the different productive tasks used in the two studies since the controlled writing task may need more morphological changes of a word than the controlled sentence writing task. Essay writing may require richer vocabulary forms than sentence writing.

Lemmouh's (2010) findings are partially aligned with our results since morphological knowledge and association in her study are the two depth components that exhibit a relationship with the lexical richness of the writing performance. The association was found to have the largest correlation with literature-essay writing tasks in Lemmouh's (2010) study. This is reasonable since different writing genres may call for different types of word knowledge. Literature writing is more likely to encourage L2 writers to vary the words used, and thus needs a larger semantic lexical network. This is inherently different from the argumentative IELTS writing tasks assigned in the current study. In addition, Sukying (2023) and Varnaseri and Farvardin (2016) also support the current results that depth components are more closely associated with L2 writing ability than vocabulary size. That said, both studies also captured moderate relations between vocabulary size and L2 writing proficiency. This is probably because the participants they tapped are postgraduate students who may have better lexical ability in L2 writing. Yet, they simply used the WAF or the PVLT to assess a limited range of depth components, which may be insufficient because the WAF has no context (Read, 2000) and the PVLT is merely a form recall measure (Schmitt, 2010).

Collectively, the relationships between vocabulary knowledge and L2 writing ability in the current study are consistent with some previous studies and contrast with some others. When looking into the internal structure of word knowledge, however, we found that the correlations between multiple word components and L2 writing ability are disparate and varied with methodologies, contexts, and construct definitions. A more detailed and fine-grained discussion is needed to examine the relations and predictions of each word component to L2 writing and word use in the following sections.

5.3 Contributions of Multiple Word Components to L2 Writing and Word Use

The current study used hierarchical regression models to analyse the extent to which the various word knowledge components contribute to L2 writing and word use. The results show that the six-component model, including the VLT size, productive form and meaning, association, productive derivative and collocation, explains roughly 61% variance of L2 writing. This portion of explained variance, though lower than the 78% variance explained in sentence writing (Zhong, 2014) and the 84% variance accounted for by vocabulary in ESL composition writing (Astika, 1993), is apparently higher than the 25% variance explained in the IELTS writing task in Lin's (2015) study. It can be conclusive that vocabulary knowledge indeed plays a critical role in successful L2 writing performance.

L2 writing is a complicated process and many factors are at play in addition to vocabulary knowledge. The quality of an L2 essay could be determined by linguistic factors such as grammar, syntax, L1 writing skills, and cohesive devices, among others. Non-linguistic factors such as content, organization and mechanics may also account for a large share of variance in L2 writing. Nevertheless, the current study and many other previous studies all substantiate that a rich and diverse vocabulary can be the most essential factor of a quality and effective L2 essay (Astika, 1993; Kim et al., 2022; Laufer, 1994, 1998; Olinghouse & Leaird, 2009; Peng et al., 2023). Moreover, L2 writers need to become as fluent and automatic as possible at lexical use to devote more time and cognitive resources to higher-order issues, such as content and organization. The speedy and accurate response to word use in timed writing creates a higher demand for word depth knowledge. Coxhead (2007, 2012) suggested that all word components in her model may relate and contribute to writing proficiency, which has been empirically evidenced by the regression results in the current study.

Note, however, that some word components are inherently overlapped and connected to each other, as shown in the above sections. For example, derivative knowledge or word parts are part of form knowledge and association is a subcategory of meaning knowledge (Nation, 2022; Schmitt, 2010). Partial correlation analyses indicate that a large portion of variance between some depth components and L2 writing and word use is also shared with productive form and meaning. As such, the following sections

will discuss the relationships between vocabulary knowledge and L2 writing and word use from the perspective of the contribution made by each word component: receptive size, productive form and meaning, productive derivative, association and collocation.

5.3.1 Contributions of Receptive Size to L2 Writing and Word Use

Vocabulary size (receptive form-meaning links), as assessed by the VLT in the current study, cannot significantly predict L2 writing and productive word use, even though the VLT was the first entered predictor variable in the regression models. The R^2 and effect size, though significant in the model at $p < .001$ level, are also tiny, meaning that the receptive size has little effect on the productive performances. This result is surprising since the receptive vocabulary size has been proven to correlate with, and contribute to, L2 reading comprehension (Qian, 1999, 2002; Stæhr, 2008), listening comprehension (Milton, Wade & Hopkins, 2010; Stæhr, 2009), speaking ability (Janebi Enayat & Derakhshan, 2021) and L2 writing proficiency (Albrechtsen, Haastrup & Henriksen, 2008; Dabbagh & Janebi Enayat, 2019; Milton, Wade & Hopkins, 2010; Stæhr, 2008). Taken together, the receptive size can be a strong contributor to the four language skills, which is opposite to the current study.

The receptive size indeed is fundamental since mapping the form and meaning of a certain number of words is the prerequisite for any unassisted language skill. Yet, this knowledge is merely the very first step of knowing a word and thus only "superficial knowledge" (Qian, 2002, p.515) or "superficial familiarity of the word" (Laufer & Goldstein, 2004, p.400). In this regard, Dabbagh and Janebi Enayat (2019) explained the stronger prediction of the VLT than the WAT in L2 writing performance with reference to the analytic rating criteria of the writing task. They posited that the rating process of the L2 essays may be more inclined towards lexical sophistication and diversity, which are more related to vocabulary size (Crossley et al., 2015). As such, many depth components, such as collocation and association, may be ignored by assessors. The current study also used the same analytic rating criteria in writing, but generated an opposite result. The explanation by Dabbagh and Janebi Enayat (2019) might be partially tenable in the current study since more separate depth tests, instead of a single combined WAT, were devised to capture each depth component. The WAT

was originally developed to be meaning-oriented (Read, 2000; Webb, 2013) and thus may not be sufficient to elicit the depth knowledge.

For three reasons, Stæhr (2008) also argued that the close relationship (r = 0.73, R^2= 0.52) between receptive size knowledge and productive L2 writing proficiency in his study was unexpected. These reasons are also the major differences from the current study. Firstly, only the VLT was used in Stæhr's (2008) study, while more productive measures were employed in the current study. The variance explained in our results may be shared by the other productive measures like the PVLT, also referred to as productive size (Laufer & Nation, 1999). Secondly, the quality of an L2 essay can be determined by a variety of factors. Vocabulary size was simply one of the determinants in the analytic writing criteria used in the current study, and more depth components were focused on in the rating process. The functions of size may be undermined and could not be reflected in the overall writing and word use scores. Thirdly, the L2 writers in Stæhr's (2008) study were allowed to use dictionaries, while participants in the current study were not allowed to consult dictionaries or other materials. This may also lead to the big difference of the two studies in the predictive power of the VLT in L2 writing ability. Notably, the participants in Stæhr (2008) were low-proficiency EFL learners and more than half of them could not master the most frequent 2000 words. By contrast, the participants in the current study were at an intermediate level, many of whom had a much larger vocabulary. This, again, empirically supports Schmitt's (2014) speculation that the receptive size of learners with a small vocabulary may contribute more to L2 writing performance than depth knowledge and vice versa.

This result also relates to Lin's (2015) findings since the participants in the current study are Mainland Chinese university students, not unlike one group of participants in her study. Lin (2015) explained the relatively higher prediction (11%) of receptive size and the smaller contribution (4%) of the DVK depth to L2 writing in terms of the English learning context. As mentioned earlier, English learners from Mainland China are deeply shaped by the unique English learning environment, mainly referring to the high-stakes English tests and vocabulary teaching and learning strategies. The most influential English test is the NMET, which mainly focuses on the capacity to recognize a large size of vocabulary for reading comprehension tests and multiple-

choice questions. Productive skills are largely ignored in the test, speaking ability being excluded, and the writing requirements and rating criteria being not able to measure learners' writing ability (Qi, 2007; You, 2010). Consequently, vocabulary is primarily learned and taught to pass the tests instead of communicating with productive skills. In order to obtain high scores on the tests, Chinese English learners rely too much on various vocabulary books with alphabetically arranged word lists. Students usually memorize by rote the word form and meaning with little or no context and other word depth components. Many even use the respective Chinese meaning as the word acquisition medium, and word-for-word translation has become the major word learning and teaching strategy (Gan, Humphreys & Hamp-Lyons, 2004). Words learned as such may largely remain at receptive levels and cannot be readily used for communicating purposes, such as L2 writing tasks.

Lin (2015) attributed the high prediction of the VLT size in L2 writing to the word learning context and strategies peculiar to Mainland Chinese students. However, the current study may interpret its data differently since our results show that the VLT size has small correlations with L2 writing and word use and cannot significantly predict productive skills. These results may lend more support to the situation in which Chinese university EFL learners have a large vocabulary that does not contribute to productive performances. The high mean score (M = 77%) in the VLT echoes the previous claims that many EFL learners simply "know little about a large number of words" (Schmitt, 2014, p.915) and have "lots of words in their mental lexicon but with poor organization" (Milton, 2009, p.150). In other words, they cannot translate their large vocabulary size to profound vocabulary depth in productive word use and overall L2 writing proficiency.

5.3.2 Contributions of Productive Form and Meaning to L2 Writing and Word Use

Although the receptive size measured by the VLT in the current study cannot predict L2 writing and word use, it cannot be concluded that word form-meaning links have no contribution to actual language production. The VLT size only constitutes a direct link of form and meaning knowledge (Laufer & Goldstein, 2004) and learners likely fail to utilize the full richness of their vocabulary size in writing (Laufer & Nation, 1999).

Therefore, more indirect or productive types of form and meaning relationships are worth investigating to draw on this knowledge fully for productive purposes (Laufer & Nation, 1999; Laufer & Goldstein, 2004; Webb, 2009). Even though form and meaning was measured independently as two word components in the current study, they can hardly be separated in word learning and use (Jiang, 2002; Miller, 1999). In their models, Nation (2022) and Coxhead (2007) categorized form and meaning together in the meaning aspect. As such, the current study will discuss the contributions of productive form and meaning concurrently in one section.

In the current study, productive form (form recall) and meaning (L1-L2 word pair) are strongly correlated with L2 writing and word use scores. The regression models also show that the two productive components account for close to or above 50% variance explained in L2 writing and word use. This means that roughly half of the variance can be explained by productive form and meaning alone. Apart from the large R^2 values, the effect size brought by form recall and word pair in L2 writing ($f^2 = 1.145$) and target word use ($f^2 = 1.76$) are also apparently larger than the other depth components. The heavy weight that productive form and meaning carry in vocabulary use may explain why the two components are primarily used to define receptive and productive vocabulary knowledge (Laufer & Goldstein, 2004; Nation, 2001; Webb, 2008, 2009). This may also point to the fact that most vocabulary tests focus on form and meaning aspects because it can be an effective method to draw out the majority of vocabulary knowledge. The regression results in our study confirm the previous findings (Webb, 2009; Zhong, 2014) that measuring form and meaning alone could be used to evaluate the overall state of learner vocabulary ability in language production.

Moreover, because form and meaning are the initial stages of acquiring a word, learners are more likely to access form and meaning knowledge in real contexts. This may explicate why many general English textbooks at beginning and intermediate levels allocate more than half, if not all, of the lexical excises to form and meaning connections (Brown, 2010). This means that form and meaning are also the word components that EFL teachers and textbooks pay the most attention to in pedagogy. Perhaps it is more expected in the current study since productive form and meaning links, referred to as active (productive) vocabulary by Laufer and Goldstein (2004),

are the most fundamental components in language production (Laufer & Goldstein, 2004; Miller, 1999). The most striking difference between native and nonnative speakers is the number of words they can control in free language production (Laufer, 1994, 1998) and a large active vocabulary also contributes to a higher level of lexical richness in writing (Crossley et al., 2015; Laufer & Nation, 1995).

These quantitative results in the current study are consistent with Coxhead's (2007) qualitative findings and Zhong's (2014) receptive results that form and meaning knowledge plays the most critical part in productive skills. The current study bears more resemblance to Coxhead's (2007) study. University students in her study generally reported that they paid more attention to ensuring the correct match of conceptual expression and checking the accuracy of word form and meaning links. That is, they also relied heavily on the recall and retrieval of form and meaning during their writing as much as university EFL learners in the current study did in the writing tasks. Yet, Zhong (2014) even generated a much stronger prediction (R^2= 0.78) of receptive form and meaning in sentence writing tasks. This appears to be reasonable since sentence writing may not call for as many linguistic and cognitive intellectual elements as essay writing. Form and meaning in her study may naturally account for more importance than the current study.

It is noted that the current study elicited productive form knowledge from the PVLT (Laufer & Nation, 1999) and captured productive meaning knowledge from the L1-L2 word pair test (Laufer & Goldstein, 2004). Although Laufer and Nation (1999) developed the PVLT to assess to what extent learners can use a word in a controlled context, Schmitt (2010) argued that the PVLT is more of a form-recall test, measuring the form-meaning link at a syntactic level. The current study deployed the PVLT format and found that form recall by the PVLT explains the most variance in all regression models with the largest effect size. This empirically echoes the previous suggestions made by Stæhr (2008) that since the VLT is strongly related to L2 writing performance, its productive version (PVLT) may be even more highly associated with writing proficiency. This result also reinforces the findings by Nation (2022) and Coxhead (2007) that the written form of a word is one of the word components capturing the most attention of L2 writers. In other words, L2 writers are likely to

consume most of their cognitive resources on how to correctly spell a particular word in the writing process. Form recall ability in a real context not only impacts the use of other word components but also impinges on learners' writing strategies. Poor spelling in writing damages the real word use, and learners may use limited vocabularies or favor frequent words and avoid words that are difficult to spell (Nation, 2022). However, the large effect size may also stem from the prompted target words in the writing tasks. Because all target words were already provided, participants may not need to recall them, but only to integrate them in writing. In this sense, the high correlation and prediction of productive form in this study may not be conclusive, which can be one of the limitations of the current study.

By contrast, the productive meaning of the word pair test has relatively smaller R^2 values and effect sizes in L2 writing and word use. The rationale for using this test in the current study is that L1-L2 word translation could measure the vocabulary size of word meaning (Laufer & Goldstein, 2004). Yet, Schmitt (2010) postulated that it actually assesses the highest strength of form-meaning links, which also has been established as a productive form-meaning measure (Waring, 1997; Webb, 2007, 2009). This makes sense because L2 writers are likely to directly translate the L1 meaning they have in mind into the corresponding L2 word in the writing process (Jiang, 2002). The deployment of the L1-L2 word pair in the current study has been empirically evidenced by Webb (2009), who showed that productive word pairs contribute more to productive word use and L2 writing ability than their receptive counterparts. Our study also relates to Wu and colleagues (2019), who found that vocabulary size measured by word pairs relates most to L2 writing proficiency. The less variance accounted for by productive word pair in the current study may be because this test overlaps to a certain degree with form recall by the PVLT. As claimed by Schmitt (2010), if learners know the respective L2 word prompted by the L1 meaning, they still have to recall and produce the L2 word form, thus returning to form recall ability. The variance may be already shared by form recall before word pair was entered into the model.

In addition, Laufer and Goldstein (2004) found that passive recall (L2-L1 translation) is associated most with class grades, including writing performance. To a certain degree, their study may also help to explain the relatively smaller contribution of

productive meaning (L1-L2 translation) to L2 writing and word use than form recall in the current study. Like the PVLT, while the word test was a productive L1-L2 word translation, the word pair knowledge elicited in L2 writing tasks was a receptive L2-L1 translation as the target words were already prompted before writing. Participants did not need to recall the form and meaning links but only had to recognize the prompted words and use them in their writing.

5.3.3 Contributions of Productive Derivative to L2 Writing and Word Use

Derivative knowledge is a sub-component of word parts that are subsumed under the category of form dimension in Nation's (2022) word knowledge component model. Derivative or morphology should be a major component of word depth knowledge and has been found to moderately correlate with reading comprehension (Qian, 1999). This depth component may be more related to productive language skills since different derived forms are needed to fit in with different grammatical and contextual situations. In Zhong's (2014) words, derivative knowledge needs to be activated to determine what patterns of the word should be used in context and where the word should be put in the sentence. These functions of derivative knowledge also determine the collocates used with the word and the word class of the surrounding words. This has been strongly supported by the close correlations between the derivative and the multiple depth components in the current study.

The current study revealed the highest correlations between productive derivative and L2 writing and productive word use (all above r = .70). The regression model also showed that productive derivative adds a significant portion of 13.7% variance explained in L2 writing, even though it was the fourth predictor variable to be entered. This suggests that derivative knowledge is still a requisite lexical component apart from productive form and meaning knowledge in L2 writing. The positive and significant result aligns with the previous studies (Leontjev et al., 2016; McCutchen & Stull, 2015; Schmitt & Zimmerman, 2002; Zhong, 2014), documenting the critical role of derivatives in writing. For example, Schmitt and Zimmerman (2002) found that even if L2 learners have uneven mastery of productive derivative knowledge, this knowledge accounts for more than 10% of the lexical errors identified in L2 university writing. In the current study, the correlations between productive derivative and L2

writing and word use are even stronger than Leontjev and colleagues' (2016) findings. They devised the most comprehensive receptive and productive derivative tests and generated moderate to large relations between derivative knowledge and writing proficiency. Yet, not all measures of derivative knowledge in their study were demonstrated to predict writing proficiency.

Interestingly, judging from the Beta (β) values in the regression models, the productive derivative contributes most to both L2 writing and word use. The values are even higher than productive form knowledge even if form recall explains the most R^2 with the largest effect size. This may be due to the shared variance between derivatives and form, as morphological knowledge is regarded as part of form knowledge (Nation, 2022). This implies that knowing more derived changes of the word is as important, if not more important, as form recall for productive use as derivative knowledge can be a good reflection of form in context (Nation, 2022). Given that the target words were prompted in the writing tasks, participants in the current study may reasonably rely more on the morphological changes of the words in their writing. In other words, derivatives, to a large extent, cover the functions of form knowledge and represents the contributions made by form in writing. This further confirms the previous findings (McCutchen & Stull, 2015) that derivative knowledge not only facilitates English learners' spelling in writing, but also assists their word production and text generation. Learners may use morphological awareness to expand vocabulary and enhance their fluency to retrieve words from their mental lexicon. In addition, when L2 writers can manipulate the derived forms in composing a sentence, the sentence generated would be revised accordingly, which is central to essay generation (Berninger, Nagy & Beers, 2011). This may help to explain why derivative knowledge could be a larger contributor than productive form to L2 writing ability when these two variables were entered into the same explanatory regression model.

The current study shows that productive derivative contributes more to L2 writing and word use than all other depth components, including productive form and meaning. However, this knowledge might not be paid sufficient attention to by EFL learners and teachers in pedagogy. This relates to the previous empirical findings (Schmitt, 1998; Schmitt & Meara, 1997; Schmitt & Zimmerman, 2002) that EFL learners at beginning, intermediate and even advanced levels were found to lag behind in derivative

knowledge. General English textbooks have also not attended to derivatives as much as they deserve in lexical exercises (Brown, 2010). The current findings, however, manifest that morphological awareness needs to be raised for EFL lexical learning and teaching, especially for communication purposes. In addition, research on the relationship between morphological knowledge and L2 writing proficiency has been considerably unattended, which calls for more research attention (Yoon, 2017).

5.3.4 Contributions of Depth Network to L2 Writing and Word Use

Association and collocation have been regarded as two types of lexical sense relations to other words in the mental lexicon (Haastrup & Henriksen, 2000). The depth network can be perceived as the way how words are orchestrated and stored, namely, paradigmatic associations (synonyms) and syntagmatic relationships (collocation restrictions), as assessed in the current study (Haastrup & Henriksen, 2000; Read, 2000). Synonymy can be the most pervasive and important relationship between words, while collocation represents the productive attribute of vocabulary since the target word and the immediate neighbouring words make up a local context (Nation, 2022). Following this line, the two types of lexical relationships may naturally contribute to productive contexts, such as L2 writing performance.

However, the regression models in the current study show that neither association nor collocation can significantly predict L2 writing and word use as independent predictor variables. This suggests that association and collocation cannot individually contribute to L2 writing and word use. Chinese university EFL learners mainly relied on productive form, meaning and derivative knowledge during their writing and lexical use. This seems to contradict the widespread belief that association and collocation, as two critical word depth components, are indispensable in context (Coxhead, 2007; Meara, 2009; Nation, 2022; Schmitt, 2010). Yet, with the addition of association and collocation, the five-component and six-component models are statistically significant at $p < 0.001$ in predicting L2 writing and vocabulary component except for the target word use. The R^2 and effect sizes brought by association and collocation to L2 writing and vocabulary component are marginal but significant. The explained variance and effect sizes respectively increased by 0.1% and 0.3% for L2 writing and 1.1% and 2.2% for vocabulary component when association and collocation were entered. This, to a

certain degree, suggests that the two components have a tiny yet significant effect on L2 writing and extensive vocabulary use. Similar to Zhong's (2014) findings, association and collocation also marginally account for the variance of word use in sentence writing for junior high school students in the beginning. After four months of normal classroom instruction, these two depth components were found to have a much larger effect on productive word use. This suggests that association and collocation would become increasingly important for productive word use as vocabulary knowledge develops. Thus, longitudinal research is warranted and recommended for future studies to inquire into the increases in the predictive power of association and collocation to L2 writing and word use.

The non-significant prediction of the target word use by association and collocation may be caused by the limited number of target words used by the participants. In the L2 writing tasks, participants were only required to randomly choose at least five of the target words to use in their writing. Despite being told that the more target words are used, the better, many students still only use five of them. Even though many obtained a high score in the writing and vocabulary component, their target word scores were relatively low. This can be reflected by the mean score of target words (below 70%), which is starkly lower than that of writing score and vocabulary component. This may probably lead to the non-significant contributions made by association and collocation to the target word use.

The small predictions of association and collocation in L2 writing and word use in the current study may also be due to the shared variance with productive form and meaning. Because association and collocation are inherently connected to meaning (Zhong, 2014), the WAF devised to measure association and collocation is regarded to be meaning-centered (Webb, 2013) and semantically-related (Greidanus & Nienhuis, 2001). This will be separately discussed in the following sub-sections.

5.3.4.1 Contributions of Association to L2 Writing and Word Use

In the models of word knowledge component by Nation (2022) and Coxhead (2007), the association is solely limited to synonymy in language production under the meaning dimension, namely, "what other words can we use instead of this one"? The inclusion of association in the meaning category denotes the close relationship between the two

variables. Nation (2022) argued that association is part of the meaning system, suggesting that learners largely rely on meaning senses to link semantically related words. Association and productive meaning in the current study may have mutual interference in predicting L2 writing and word use, which has been reflected in the partial correlation results. When productive meaning was controlled, the standard correlation between association and L2 writing (r = 0.501, R^2=0.251, p < 0.001) dropped 54% to the partial correlation of r = 0.232 (R^2 = 0.034, p < 0.01). The significant decrease in the partial correlation suggests that Chinese university EFL learners may depend largely on form-meaning mapping and have an incomplete system of word associations, on which they might seldom depend during their writing.

The non-significant association in the current study also contrasts with Coxhead's (2007) qualitative case study, in which learners reported that lexical association can enrich their vocabulary and assist their selection of the most appropriate words. They also described that comparing semantically related words is one of the most critical processes that engages their attention in writing. However, this lexical comparison process may not be revealed by quantitatively examining the writing output as it only happens during the writing process. Simply put, the L2 writing score cannot provide insights into how L2 writers associated, compared and retrieved words in their writing. This may help explain the insignificant prediction of association to L2 writing and word use. Zhong (2014), who also used sentence writing tasks, suggested that perhaps a retrospective interview or think-aloud protocol may help garner associative knowledge during the writing process.

In addition, while participants in the current study were instructed that they could use synonyms or words with similar meanings, they may not need or bother to compare any word in their writing as the target words were prompted. In other words, the controlled L2 writing tasks may not be able to activate association altogether. Laufer (2003) pointed out that risk avoidance and time pressure can significantly affect test performances beyond the test-takers' word knowledge. Time pressure may not pose a major challenge to our participants as they were given as much time as they needed to complete the controlled writing tasks. The nature of these tests was also low-stakes, of which participants were informed before the tests. However, risk avoidance may be

common in all tests, particularly in the controlled writing tasks. Since the target words were provided beforehand, they may not take the risk of attempting more associated words in their writing. Even for the target words, many only chose five of them to use as required, even though they could achieve high scores in other tests and the overall writing tasks.

Moreover, the writing tasks assigned in the current study were argumentative essays composed by intermediate university learners. Non-linguistic factors such as writing skills and text genres may also influence the relationship between association and writing ability. For example, previous research shows high correlations between L2 writing proficiency and synonymy used as a major cohesive device. The higher-level learners used more synonyms in their writing to promote writing cohesion than the lower proficiency learners (Ferris, 1994; Liu & Braine, 2005). Our intermediate participants may not have sufficient writing skills to use associations to make their essays more cohesive. The writing training they received mainly focuses on vocabulary and beautiful phrases and may largely ignore text cohesion devices (You, 2010).
Lemmouh (2010) and Baba (2009) found that association contributes most to literature essay writing and summary writing, which differs from the argumentative writing used in the current study. As Lemmouh (2010) used the lexical variation as the main yardstick of literature writing scores, association may naturally associate with, and contribute to, the varied lexical expression needed in this type of writing. Thus, it may not be conclusive that the strong relation between association and lexical variation equals that between association and the overall writing ability. Similarly, Baba (2009) identified that the ability to define words could uniquely contribute to summary writing proficiency. This is reasonable since defining words and summary writing may call for the same lexical associative knowledge. As explained by Baba (2009), defining words requires learners to build a structure of semantic network of words, while summary writing needs to paraphrase a number of similar words with synonyms. The argumentative writing tasks in the current study may not directly impose such a strong demand for association knowledge as literature and summary writing tasks.

5.3.4.2 Contributions of Productive Collocation to L2 Writing and Word Use

Nation (2022) has stressed the importance of collocation in written texts and oral speech because it makes language production sound more native and fluent. A large number of ready-made lexical sequences stored in the mental lexicon can indeed reduce the learning burdens and mediate learners' struggle in actual word use. The results produced in the current study, however, suggest that productive collocation cannot significantly predict L2 writing and word use. Chinese intermediate university learners may have a limited amount of collocation knowledge and rely little on this knowledge in their productive written work. This is inconsistent with previous studies (Crossley, 2020; Crossley, Salsbury, and McNamara, 2015; Kyle & Crossley, 2015) documenting the critical role collocation plays in writing proficiency.

For example, Crossley and colleagues even showed that the accurate use of collocation could predict 84% variance explained in the holistic writing score. While Crossley (2020) and Kyle and Crossley (2015) found that low proficiency learners could barely use collocations, proficient learners could produce a wide variety of them in writing performances. The intermediate Chinese EFL learners in the current study apparently did not reach the proficiency level defined by Crossley and his colleagues. The written samples collected in their study were composed by ESL learners in a university in the United States. Their participants might be able to manipulate more native collocates than participants in the current study who had never studied in an English-speaking context. That said, the findings by Crossley et al. (2015) support the current study that depth indeed can be a stronger indicator of writing proficiency than word diversity and frequency which are related to breadth of knowledge.

The findings by Nguyen and Webb (2017) may partly support the current results that intermediate Vietnamese university EFL learners are not even close to mastery of receptive collocations at the first three 1000 frequency levels. Chinese university EFL learners, like their Vietnamese counterparts, may focus more on single-word items instead of multi-item units and lack awareness of the importance of collocation in communication. Perhaps more directly supportive to the current findings are studies by Laufer and Waldman (2011) and Zhong (2016). Collocation contributes little to sentence writing performance (Zhong, 2016) and L2 learners at all proficiency levels

including advanced learners cannot accurately use verb-noun collocations in their writing (Laufer & Waldman, 2011). Their explanations may also hold for the insignificant prediction of collocation in L2 writing and word use in this study.

Zhong (2016) explained the low prediction of collocation in sentence writing in terms of the avoidance strategy, suggesting that learners may use collocation only when they are confident in using it correctly. The participants in Zhong (2016) were junior high school students who may have insufficient knowledge of collocation. It makes sense that they lack confidence in using collocation in production. However, Laufer and Waldman (2011) showed that even advanced L2 learners who may be confident in using more collocations in writing still cannot use them correctly. Conversely, the more collocations they produced, the more errors they made and these errors would not decrease with proficiency levels. The reasons provided by Laufer and Waldman (2011) are also true for the current study, that is, semantic transparency and L1 transfer of collocations. Because collocations are made up of frequent individual words, L2 learners may tend to ignore the ready-made formulas when they are acquiring the language. They primarily construct collocations from individual words, not from the existing patterns. Moreover, due to the negative L1 transfer, Chinese EFL learners who learn collocations by word-for-word translation may often disregard the restrictions on word combinations. For example, Chinese EFL learners often directly translate the Chinese collocation “学习知识” into English in their writing as “learn knowledge”, while the typical one is “gain or acquire knowledge” (Dodigovic, Ma & Jing, 2017).

In the current study, the non-significant contribution by collocation in L2 writing and word use may also be explained by the shared variance between collocation and other depth components. As mentioned earlier, collocation has a close relationship with meaning as collocation sets the contextual boundary for meaning and the meaning of the single items determines the possibility of collocation (Qian, 1999). The correlation analysis in the current study reveals that productive collocation and meaning have a fairly large correlation (r = 0.533, R^2= 0.284). As such, partial correlation analysis was assessed and found that the partial correlation (r= 0.249, R^2= 0.039, p < 0.01) between collocation and L2 writing decreased approximately half of the standard correlation (r= 0.491, R^2= 0.242, p < 0.001) when productive meaning (word pair) was controlled. Yet,

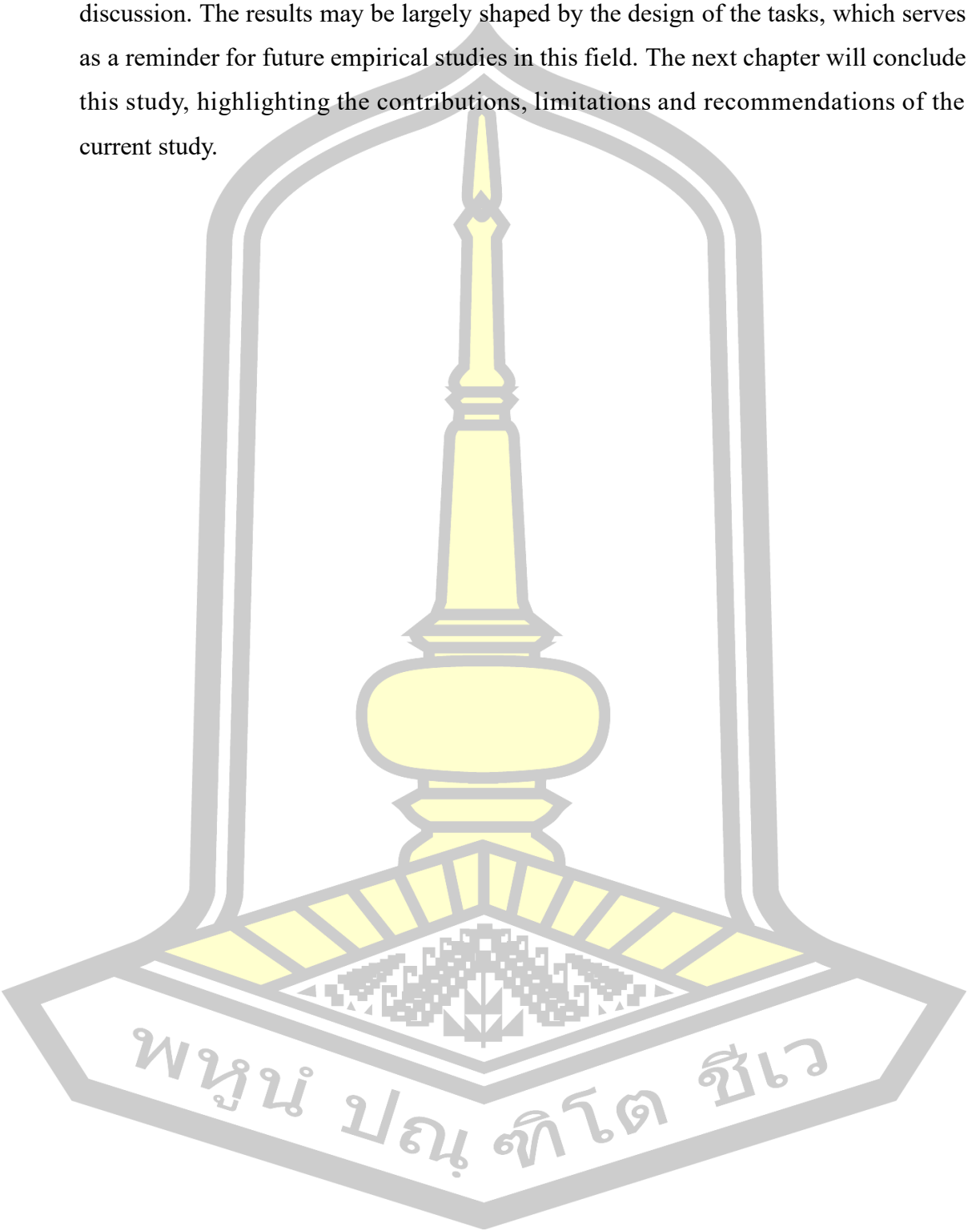
when productive form was controlled, the partial correlation between collocation and L2 writing dropped dramatically to an insignificant level (r = 0.138, R^2= 0.012, p = 0.097). This means that the majority of the variance (74%) that is uniquely explained by collocation in L2 writing is also shared with productive form. When collocation was entered into the model, a large portion of variance had already been accounted for by productive form and meaning. This may cause the non-significant prediction of collocation to L2 writing and word use in the current study.

The strong correlation between collocation and form recall (r = 0.67, R^2= 0.44) may stem from the similar test format, both requiring to complete the target words in a sentence with the first letter(s) being provided (see Section 5.2.2). This close relation between form recall and collocation may also come from the inherent connection between the constructs in a sentence context. As explained by Schmitt (2010, p.202), collocation knowledge actually provides the "transparency" of the answers for form recall in the sentences. For example, the frequent collocation *income tax* serves to make the form knowledge of the target word *income* rather obvious. If test-takers have intuitions about the frequent collocation, form recall is less of an issue. In other words, the contextual defining power of collocation to form recall may lead to a large correlation between the two tests. This may also explain the extensive overlaps between the two constructs and the large mutual interference in the predictive power of collocation in L2 writing and word use.

5.4 Chapter Summary

In summary, this chapter discusses the results generated in the current study with reference to the findings of previous literature revolving around the interrelationships between vocabulary components and their relations and contributions to L2 writing and word use. During the discussion, the strength of correlations between multiple word components varied greatly. The contributions each of these word components makes to L2 writing and word use also differ from one another. The current study suggests that putting the lexical components in a controlled L2 writing context and exploring their relationships can be necessary and significant. A body of fruitful insights into vocabulary knowledge, L2 writing ability and lexical proficiency have been generated. In addition, the limitations of this study concerning the mutual

interference of word tests and the controlled writing tasks also emerged during the discussion. The results may be largely shaped by the design of the tasks, which serves as a reminder for future empirical studies in this field. The next chapter will conclude this study, highlighting the contributions, limitations and recommendations of the current study.



CHAPTER VI
CONCLUSION

This chapter concludes the thesis project by producing a brief summary of the theoretical and pedagogical contributions of the findings obtained in the study. There follows the discussions of the limitations in terms of the test instruments and research design. In accordance with the current limitations and results, relative implications and suggestions will also be provided for future studies looking into the nature of vocabulary knowledge in writing and lexical proficiency.

6.1 Summary of the Study

The current study aims to investigate the the nature of multiple word knowledge components associated with L2 writing and productive word use in writing within a multi-component framework. This study is guided by two research questions regarding the correlations between various word components and their relationships with, and contributions to, word use and L2 writing ability. Accordingly, six separate measures, including one receptive word component (overall word size) and five productive depth components (form recall, L1-L2 word pair, association, productive derivative and collocation), were designed and validated. This empirical study was conducted to tap into Chinese intermediate-level EFL learners' word knowledge relative to their word use and writing ability. Correlation and regression analyses were employed to quantitatively explore the strength of relations between these variables and the variance that could be explained by each word component in L2 writing ability and lexical proficiency.

The findings of this project demonstrated that vocabulary knowledge indeed can be broken down into various types of knowledge in a quasi-natural context. The study partially supports the popular theoretical line that form and meaning is the most fundamental lexical knowledge in context. Yet, depth components including productive form and meaning, particularly morphological forms, are mainly related to L2 writing ability and word use. Productive collocation and association predict less variance yet still correlate with productive skills and significantly improved the regression models in L2 writing and vocabulary component score. However, the receptive size measured by the VLT only has meagre relations with depth knowledge and L2 writing and word

use. Overall, this study provides empirical evidence for the theoretical word knowledge models and yields nuanced ideas regarding the smallest lexical predictors of L2 writing.

6.2 Contributions of the Study

6.2.1 Theoretical Contributions

Theoretically, the current study contributes to the vocabulary knowledge construct and provides empirical insights into the vocabulary knowledge component models formulated by Nation (2001, 2022) and Coxhead (2007). Nation (2001) developed the most comprehensive word knowledge framework, attempting to encapsulate a wide variety of the word knowledge aspects and types in the field of applied linguistics. Coxhead (2007) moved forward Nation's (2001) vocabulary component model towards lexical use in writing and tabulated anther model specifying word components needed in writing practice. Guided by the two word knowledge models, the current study is one of the attempts to inquire "explanatory theories out of the descriptive frameworks" (Schmitt & Meara, 1997, p. 33).

The current study devised a wide range of vocabulary tests to capture multiple aspects of word knowledge in a L2 writing context. This supplements the qualitative studies by Coxhead (2007, 2011, 2012) in an attempt to validate her model. This study also provides more empirical evidence to evaluate and refine the vocabulary knowledge models in L2 writing, as opposed to the previous studies either focusing on various word components only (Gonzalez-Fernandez & Schmitt, 2019), or merely receptive word aspects in context (Zhong, 2014, 2016), or a limited range of word types in writing (Lin, 2015; Dabbagh & Janebi Enayat, 2019; Sukying, 2023; Wu et al., 2019).

While the current study empirically confirms that word knowledge indeed structurally subsumes a variety of sub-knowledge, it also demonstrates that these knowledge components are strongly interrelated. In addition, the current study also suggests that receptive word size and word depth knowledge are separate and distinctive constructs. They may be related to each other to a certain degree, but should be conceptualized and operatinonalized as two different types of knowledge in assessment and research. The findings of the current study also reinforces the previous theories that all word knowledge aspects may be related to writing and lexical use (Coxhead, 2007; Nation,

2022). Yet, this study confirms that word knowledge components contribute to L2 writing and lexical proficiency to a varying extent. Productive form and meaning play the most critical role in L2 writing and word use, on a par with the receptive form and meaning aspects in sentence writing (Zhong, 2014, 2016).

Productive derivative, association and collocation also significantly increase the variance explained in the whole regression models in L2 writing. It is worth noting that derivative knowledge has most predictive power in L2 writing in the current study. This strongly supports the previous statements (Schmitt & Zimmerman, 2002; Sukying, 2022) relative to the importance of derivative knowledge. However, receptive word size has little relations with L2 writing ability and word use, which is a surprising finding in the current study inconsistent with the previous positive theories and findings. This warrants more replicate research to see whether it is generalizable to the whole Chinese university EFL context.

6.2.2 Pedagogical Contributions

In accordance with the results of the current study, mapping form and meaning links has been demonstrated as the most fundamental lexical ability in L2 writing and word use. Productive form and meaning accounted for approximately half of the variance explained in L2 writing and vocabulary component. It makes sense for Chinese university EFL learners and teachers to pay special attention to form and meaning knowledge. However, the current study revealed no relationship or contribution to L2 writing and word use made by their receptive word size. It may suggest that learners should not merely focus on the direct link of form and meaning. Most notably, the rote learning by memorizing L2 form to correspond to its L1 Chinese translation may not be an effective way to acquire word knowledge, particularly for improving productive skills. Apart form direct form-meaning links, they also need to deepen these links by learning and manipulating more productive from recall and L1-L2 word translations. In so doing, they may be more likely to fluently and readily use the form and meaning knowledge in speaking and writing performances.

Productive derivative or morphological knowledge in context should be given as much attention as form and meaning in word learning and teaching since productive derivative contributes most to L2 writing and word use in the current study. Learners in productive practices like L2 writing may not only have to recall the word form, they

must also recall and retrieve the corresponding derived forms according to the collocations they use and the specific syntactic structure and grammar. Thus, when learning and teaching a word, the other members of the word family should not be ignored. Association and collocation also significantly improve the whole regression models in L2 writing and vocabulary component score. This means that the two word components, though not directly contribute to L2 writing and word use, should not be overlooked. Chinese EFL learners and instructors need to dig deep into a word when learning and teaching it by exposing themselves to more types or components of the word. For example, English definitions of multiple shades of meaning should be learned and taught so that more synonyms of the word could be mastered. English dictionaries and materials should also be used to expand exposure to more typical and frequent collocations. These suggested pedagogical methods may be cumbersome and time-consuming, but are critical and indispensable for EFL communicating skills.

6.3 Limitations of the Study

6.3.1 Choice of target words

The current study selected 20 final target words from the AWL (Coxhead, 2001). Because participants had to integrate some of these words in their writing, the number of target words has been largely limited. Vocabulary knowledge components elicited from the 20 target words may under-represent the productive depth knowledge in EFL learners. It has been documented that non-native university EFL learners may generally have a vocabulary size ranging from 1500 to 7000 word families (Nation, 2006, 2016; Nurweni & Read, 1999). The limited number of word items may be general to all multi-component studies focusing on vocabulary depth knowledge in context. Studies of this kind need to strike a balance between the study aims and the practicality and feasibility of data collection. The current study considers that 20 target words may be suitable for the multiple separate word tests at sentence levels and L2 writing contexts.

6.3.2 Design of Instruments

The current study intended to measure one receptive size and five productive depth components of word knowledge. Yet, the synonymy depth test was a receptive format, requiring participants to choose answers from the box in which there was two to three keys. The receptive format of vocabulary tests are usually designed as a multiple choice

or matching question, in which test-takers have to recognize or comprehend a word in listening and reading (Read, 2000). This test format may contrasts with the definition of productive word knowledge in the current study, namely, the ability to retrieve and use multiple word knowledge components in a sentence and L2 writing tasks. Because the original WAF combines synonymy and collocation without context, the current study split the WAF and added a sentence context. In order to measure as many synonyms as possible, the current study maintained the multiple choice format. The added sentence to a certain extent puts the test in a controlled context and measures to what extent learners could recognize the synonyms at a sentence level.

Some of the depth test instruments may be overlapped to a certain degree in functions, which may be reflected from the correlation and regression coefficients in the discussion chapter. For example, the productive meaning test (L1-L2 word pair) captured not only meaning recall but also form recall; the productive derivative test also measured form recall ability; and the productive collocation test actually overlapped with the test format of form recall. This points to the close relationships between these multiple depth components. Actually, recent empirical studies (Gonzalez-Fernandez & Schmitt, 2020; González-Fernández, 2022) have cautioned that vocabulary knowledge may be more of a unidimensional construct instead of a multidimensional one due to the large relations among word knowledge components. As a result, more word tests need to be developed to measure different types of word knowledge. The large correlations between these depth components may not be generalizable to other studies or contexts.

In addition, the controlled L2 writing tasks may also be one of the major limitations of the current study. The writing tasks were originally designed to activate all measured word knowledge components in an actual context. Yet, the controlled writing tasks may not serve the purpose as it cannot activate form knowledge and the writing output may not directly reveal synonymy knowledge. As discussed in the previous chapter, the writing tasks can only partially retrieve productive L1-L2 meaning. This may lead to the fact that the contributions to L2 writing and word use made by these word components may be largely attributed to the prompted word form knowledge. As a result, the controlled L2 writing tasks may not fully manifest the importance of all

productive depth word components as intended.

6.3.3 Test Administration

The current study was merely a cross-sectional design, only measuring the multiple word knowledge components and L2 writing ability at a certain time point. Thus, it is unable to observe the dynamic development of this knowledge and the changes of the relations between these different constructs. Moreover, the VLT was administered online due to the restriction of the Covid-19, which may provide opportunity for test-takers to find the answers to the recognition-mapping questions from different sources. Likewise, only the researcher monitored the other depth tests, and test-takers may have chances to look up dictionaries in the vocabulary and writing tests. These factors may possibly increase their scores in all tests. In addition, the two-day arrangement of all vocabulary tests designed to avoid the test fatigue may also give test-takers chances to check the target words after the first-day tests. This may possibly affect their performance on the next day, although they were not informed that there was another set of tests with the same target words the next day. In order to make tests low-stake, they were also told that these tests were simply a mid-term examination, which would not affect their performance in this course.

6.4 Implications for Future Studies

6.4.1 Implications from the Limitations

The limitations on the design of instruments serve as a reminder that multiple word knowledge tests might capture constructs that are different from initial expectations. The extent of the construct that each word test could measure should be carefully examined not only based on theories but also on empirical evidence. Thus, future researchers and designers need to thoroughly evaluate the reliability and validity of the instruments. Also of particular concern is to seek empirical evidence from other studies or pilot studies to validate what construct the intended test would measure. This would ensure the accurate match between multiple word knowledge components and the respective test instrument.

In addition, given the limitations of the depth test instruments, only one battery of tests might be insufficient to conclude the correlation coefficients among the various word components. Future studies may develop a parallel battery of tests to capture the same

constructs and compare the results. Other analysis tools such as factor analysis could also be used to have a closer look whether there is difference between the results. The current study focused on the productive components of vocabulary depth knowledge, yet the receptive aspects of word knowledge were largely ignored. Only one overall receptive size was captured and the synonymy test was in receptive format in this study. Future studies may develop a corresponding battery of receptive tests to capture the same word knowledge components. This will not only provide a more rounded picture of the nature of word knowledge, as tabulated in Nation's (2022) model, but also elicit the development from receptive knowledge to productive use in a natural context. Schmitt (2010) and Milton and Fitzpatrick (2014) also encouraged the measurement of the various word components concurrently at both receptive and productive levels with a battery of tests.

Moreover, future studies may use different productive contexts to capture the actual word use ability in context, such as free writing tasks. The controlled sentence writing tasks (Read, 2000; Zhong, 2014) and the controlled L2 writing used in the current study have been proven unable to activate all the intended word components. The results were largely shaped by the controlled writing tasks. Thus, free writing tasks might be better to elicit a more comprehensive range of word knowledge components. In addition, since some word components, such as synonymy, cannot not be revealed in the writing output, qualitative methods are recommended for future studies. For example, a think-aloud protocol or follow-up interview might help to triangulate the quantitative results generated. In so doing, we may obtain more rich and reliable information about how test-takers use a word in their writing and what relations the various word components have with each other and writing proficiency.

In light of the cross-sectional design, future studies may replicate the current study in a longitudinal setting to capture the development of multiple word knowledge. The contributions made by different types of word knowledge might also change with the development of language proficiency. Finally, the tests could be administered in a more reasonable manner to minimize the cross-test effects. For example, all the vocabulary tests could be administered in classroom setting and more monitors could help to better organize the tests.

6.4.2 Implications from the Results

The current study also has some implications for future studies based on the results it produced. The results indicated that the VLT, though referred to as one of best measures of vocabulary knowledge, cannot significantly correlate or predict L2 writing scores. It is recommended that future research replicate this study in the same context and confirm whether the same result could be generalizable. It is noted that the VLT size was also weakly correlated with other depth knowledge, which is a surprising result. Future studies may sample the target words from different sections of the VLT and look into the internal various depth components of these words. This likely helps to inquire the relationship between recognition size and recall depth of the same target words.

The current study also revealed that derivative knowledge is of particular importance in word use and L2 writing. This knowledge is the key contributor to the depth construct and correlates most with the productive proficiency. Thus, morphological awareness should be one of the major considerations in future vocabulary research. Future studies may also elicit the receptive derivative knowledge, as opposed to the productive counterparts captured in the current study and investigate the incremental process of this knowledge. The development from receptive to productive derivative knowledge may also impact word use in productive skills like sentence writing or L2 writing. Since derivative has been demonstrated to correlate with lexical ability and writing proficiency (Leontjev et al., 2016; McCutchen & Stull, 2015), the development of this depth knowledge needs to be examined associated with writing ability, be it sentence writing or controlled L2 writing tasks.

Research has hitherto examined the role played by size and depth word components in reading (Li & Kirby, 2015; McLean et al., 2020; Qian, 1999, 2002; Warnby, 2023) and listening (Stæhr, 2009; Wang & Treffers-Daller, 2017). It seems that the multi-task approach to vocabulary knowledge has largely been employed in receptive skills, yet similar studies in productive skills like writing and speaking are not many, particularly speaking ability. The current study recommends that more vocabulary studies related to speaking proficiency should be done. Future studies may devise written word tests as well as spoken forms of word tests and identify the correlations and predictive power of these types of word knowledge to speaking proficiency. This may scaffold a deeper

and more profound understanding of vocabulary knowledge as an indicator of productive language skills. In so doing, it may ascertain what types of word knowledge would be recalled and retrieved in the oral production of a word in speaking performance.

Finally, since the current study focuses on Chinese intermediate-level university EFL learners, future studies are recommended to replicate the study by extending to more groups of participants. For example, future studies could sample participants from lower or more advanced proficiency levels and compare the difference of the results. ESL learners or different L1 learners and native speakers should also be part of the sample. The replication study could also be done in a different context, as suggested by Lin (2015) who examined the relations between vocabulary knowledge and L2 writing in the contexts of Mainland China and Hong Kong. This replication could help to see to what extent the results are generalizable or specific to a certain group of learners or English learning context.

6.5 Chapter Summary

This chapter summarizes the thesis project and reiterates the contributions, limitations and implications of this study from different perspectives. This study is a part of a broader L2 vocabulary acquisition and applied linguistics, attempting to take a multi-task approach to vocabulary knowledge in the actual use of a word in L2 writing. The current study, though not without some limitations, generated results that contribute to the existing insights into the nature of vocabulary knowledge and extend the understanding of multiple word components, particularly productive word knowledge in L2 writing proficiency. It not only adds to the previous receptive tests in sentence writing (Zhong, 2014) but also quantitatively supplements the previous qualitative studies in this area (Coxhead, 2007; Huang, 2010; Schmitt, 1998). As a stepping stone, the current study hopes to inspire more future studies taking the challenge to include more receptive and productive word components and apply this knowledge to a wider range of language use.

REFERENCES

REFERENCES

Aberson, C. L. (2010). *Applied Power Analysis for the Behavioral Sciences.* New York, NY: Taylor and Francis Group, LLC.

Akbari, N. (2017). Lexical diversity and the use of academic and lower frequency words in the academic writing of EFL students. *Australian Review of Applied Linguistics, 40*(1), 3-18.

Albrechtsen, D., Haastrup, K. & Henriksen, B. (2008). *Vocabulary and Writing in a First and Second Language: Processes and Development.* PALGRAVE MACMILLAN.

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: the interface between learning and assessment*. London: Continuum.

Alireza, B. E., (2017). *Measuring Productive Depth of Vocabulary Knowledge of the Most Frequent Words.* [Doctoral dissertation, The University of Western Ontario].

Alsahafi, M. (2023). The Relationship Between Depth of Academic English Vocabulary Knowledge and Academic Success of Second Language University Students. *SAGE Open, 13*(1), 1-9.

Angelova, M., & Riazantseva, A. (1999). If you don't tell me, how can I know?: A case study of four international students learning to write the US way. *Writing Communication, 16*(4), 491-525.

Anderson, R.C., & Freebody, P. (1981). Vocabulary knowledge. In Guthrie, J.T. (ed.), *Comprehension and Teaching: Research Reviews.* Newark, DE: International Reading Association.

Astika, G.G. (1993). Analytical Assessments of Foreign Students' Writing. *RELC Journal, 24*(1), 61-70.

Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing, 18*(3), 191-208.

Batty, A. (2012). Identifying Dimensions of Vocabulary Knowledgein the Word Associates Test. *Vocabulary Learning and Instruction, 1*(1), 70-77.

Bates, E & Goodman, J.C. (1997). On the insaparability of grammar and the lexicon: evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes, 12*(5), 507-584.

Barkaoui, K. (2019). What can writer's pausing behavior tell us about their L2 writing performances? *Studies in Second Language Acquisition, 41*(2019), 529-554.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice.* Oxford: Oxford University Press.

Bachman, L.F., & Palmer, A.S. (2010). *Language Assessment in Practice: Developing Language Assessments and Justifying their Use in the Real World.* Oxford: Oxford University Press.

Bauer, L. & Nation, P. (1993). Word Families. *International Journal of Lexicography, 6*(4), 253-279.

Bestgen, Y. (2017). Beyond single-word measures: L2 writing assessment, lexical richness and formulaic competence. *System, 69*(2017), 65-78.

Beglar, D., & Hunt, A. (1999). "Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing, 16*(2), 131-162.

Berninger, V. W., Nagy, W., & Beers, S. (2011). Child writers' construction and reconstruction of single sentences and construction of multi-sentence texts: Contributions of syntax and transcription to translation. *Reading and Writing: An Interdisciplinary Journal, 24*(2011), 151-182.

Berker, A. (2018). Not to scale? An argument-based inquiry into the validity of an L2 writing rating scale. *Assessing Writing, 37*(2018), 1-12.

Bolton, K., & Graddol, D. (2012). English in China Today. *English Today, 28*(3), 3-9.

Biria, Reza., & Karimi, Zahra. (2015). The Effects of Pre-task Planning on the Writing Fluency of Iranian EFL Learners. *Journal of Language Teaching and Research, 6*(2), 357-365.

Brun-Mercer, N., & Zimmerman, C. B.(2015). Fostering Academic Vocabulary Use in Writing. *The CATESOL Journal, 27*(1), 131-148.

Brown, D. (2010). What aspects of vocabulary knowledge do textbooks give attention to? *Language Teaching Research, 15*(1), 83-97.

Canale, M., & Swain, M. (1980). Theoretical bases of communication approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47.

Cameron, L. (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research, 6*(2), 145-173.

Cheng, L.Y. (2008). The key to success: English language testing in China. *Language Testing, 25*(1), 15-37.

Cho, Y. (2003) Assessing writing: Are we bound by only one method? *Assessing Writing, 8*(3), 165-191.

Cho, E. (2016). "Making Reliability Reliable". Organizational Research Methods. *SAGE Publications, 19*(4), 651–682.

Chui, A. (2006). A study of the English vocabulary knowledge of university students in Hong Kong. *Asian Journal of English Language Teaching, 16*(2006), 1-23.

Chen, C. & Truscott, J. (2010). The Effects of Repetition and L1 Lexicalization on Incidental Vocabulary Acquisition. *Applied Linguistics, 31*(5), 693-713.

Chapelle, C.A. (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research, 10*(2), 157-187.

Cheng, J.Y. & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing, 35*(1), 1-23.

Choi, Y. (2017). Roles of receptive and productive vocabulary knowledge in L2 writing through the mediation of L2 reading ability. *English Teaching, 72*(1), 3-23.

Corson, D.J. (1995). *Using English Words.* Dordrecht: Kluwer Academic Publishers.

Coxhead, A. (2000). A new Academic Word List. *TESOL Quarterly, 34*(2), 213-38.

Coxhead, A. (2007). Factors and aspects of knowledge affecting L2 word use in writing. In P. Davidson, C. Coombe, D. Lloyd, & D. Palfreyman (Eds.), *Teaching and Learning Vocabulary in another Language* (pp. 331–342). Dubai, UAE: TESOL Arabia.

Coxhead, A. (2011). The Academic Word List 10 years on: research and teaching implications. *TESOL Quarterly, 45*(2), 355-620.

Coxhead, A. (2012). Academic Vocabulary, Writing and English for Academic Purposes: Perspectives from Second Language Learners. *RELC Journal, 43*(1) 137-145.

Coxhead, A. (2021). Vocabulary in English in Tertiary Contexts: Connecting Research and Learning. *LEARN Journal: Language Education and Acquisition Research Network, 14*(1), 1-14.

Coxhead, A., & Hirsh, D. (2007). A pilot science-specific word list. *Rev. franç. de linguistique appliquée, 12*(2), 65-78.

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing, 16*(2007), 129-147.

Coxhead, A., Nation, P., & Sim, D. (2015). Measuring the Vocabulary Size of Native Speakers of English in New Zealand Secondary Schools. *NZ J Educ Stud, 50*(2015), 121-135.

Coxhead, A. & Boutorwick, T. J. (2018). Longitudinal Vocabulary Development in an EMI International School Context: Learners and Texts in EAL, Maths, and Science. *TESOL Quarterly, 52*(3), 588-610.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

Crossley, S.A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research, 11*(3), 415-443.

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). What is lexical proficiency? Some answers from computational models of speech data, *TESOL Quarterly, 45*(1), 182-93.

Crossley, S. A., Salsbury, T., McNamara, D. S. (2015). Assessing Lexical Proficiency Using Analytic Ratings: A Case for Collocation Accuracy. *Applied Linguistics, 36*(5), 570-590.

Crossley, S. A., & McNamara, D. S. (2011). Text coherence and judgments of essay quality: Models of quality and coherence. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society*. (pp. 1236-1241). Austin, TX: Cognitive Science Society.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Csomay, E. & Prades, A. (2018). Academic vocabulary in ESL student papers: A corpus-based study. *Journal of English for Academic Purposes*, *33*(2018), 100-118.

Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written Communication, 7*(4), 482-511.

Cumming, A. (2001) Learning to write in a second language: Two decades of research. *International Journal of English Studies, 1*(2), 1-23.

Dale, E. (1965). Vocabulary measurement: techniques and major findings. *Elementary English, 42*(8), 895-901

Daller, H.M. & Xue, H.J. (2009). Vocabulary Knowledge and Academic Success: a Study of Chinese Students in UK Higher Education. In: Richards, B., Daller, M.H., Malvern, D.D., Meara, P., Milton, J., Treffers-Daller, J. (eds) *Vocabulary Studies in First and Second Language Acquisition* (pp.179-193). Palgrave Macmillan, London.

Daller, H.M., & Treffers-Daller, J. (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.

Daller, H., & Phelan, D. (2007). What is in a teacher mind? Teacher ratings of EFL essays and different aspects of lexical richness. In J. J. Daller, J. Milton & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 133-149). Cambridge: Cambridge University Press.

Daskalovska, N. (2015). Acquisition of three word knowledge components through reading. *The Journal of Educational Research, 109*(1), 68-80.

Dabbagh, A., & Janebi Enayat, M. (2019). The role of vocabulary breadth and depth in predicting second language descriptive writing performance. *The Language Learning Journal, 47*(5), 575-590.

Dabbagh, A., & Janebi Enayat, M. (2022). Comparing two measures of L2 depth of vocabulary knowledge using the association with vocabulary size. *Journal of Language and Education, 8*(3), 52-64.

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition, 34*(1), 5-34.

Dodigovic, M., Ma, C.C., & Jing, S. (2017). Lexical Transfer in the Writing of Chinese Learners of English. *TESOL International Journal, 12*(1), 75-90.

Douglas, S.R. (2015). The Relationship Between Lexical Frequency Profiling Measures and Rater Judgments of Spoken and Written General English Language Proficiency on the CELPIP-General Test. *TESL Canada Journal, 32*(9), 43-64.

Durrant, P. (2014). Discipline- and level-specificity in university students' written vocabulary. *Applied Linguistics, 35*(3), 328-356.

Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes, 43*(2016) 49-61.

Engber, C.A. (1995). The Relationship of Lexical Proficiency to the Quality of ESL Compositions. *Journal of Second Language Writing, 4*(2), 139-155.

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367-413.

Ellis, N.C., & Beaton, A. (1993). Factors affecting foreign language vocabulary: imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology, 46*(3), 533-558.

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

Ellis, R. (1995). *The Study of Second Language Acquisition.* Oxford: Oxford University Press.

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford: Oxford University Press.

Evans, S., & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes, 6*(1), 3-17.

Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal, 31*(2), 105-119.

Ferris, D. (1994). Lexical and syntactic features in ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly, 28*(2), 414-420.

Ferris, D.R., & Hedgcock, J.S. (2005). *Teaching ESL composition: Purpose, Process and Practice.* Lawrence Erlbaum Associates, Inc.

Field, A. (2009). *Discovering Statistics Using SPSS (3rd ed).* London, England: Sage.

Fitzpatrick, T. (2006). Habits and rabbits: word associations and the L2 lexicon. *EUROSLA Yearbook, 6*(1), 121-145.

Fitzpatrick, T. (2007a). Word association patterns: Unpacking the assumptions. *International Journal of Applied Linguistics, 17*(3), 319-331.

Fitzpatrick, T. (2007b). Productive vocabulary tests and the search for concurrent validity. In Daller, Helmut, Milton, James and Treffers-Daller, Jeanine, (Eds). *Modelling and Assessing Vocabulary Knowledge* (pp. 116-133). Cambridge Applied Linguistics, Cambridge: Cambridge University Press.

Fitzpatrick, T., & Clenton, J. (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing, 27*(4) 537-554.

Fitzpatrick, T., & Clenton, J. (2017). Making Sense of Learner Performance on Tests of Productive Vocabulary Knowledge. *TESOL Quarterly, 51*(4), 844-867.

Fitzpatrick, T., & Thwaites, P. (2020). Word Association Research and the L2 Lexicon. *Language Teaching, 53*(3):237-274.

Gardner, D., & Davies, M. (2014). A New Academic Vocabulary List. *Applied Linguistics, 35*(3), 305-327.

Gass,S., & Selinker, L. (2001). *Second language acquisition: An introductory course.* Mahwah, NJ:Erlbaum.

Gass, S., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd Ed.). New York: Taylor & Francis.

Gan, Z., Humphreys, G. & Hamp-Lyons, L. (2004), Understanding Successful and Unsuccessful EFL Students in Chinese Universities. *The Modern Language Journal, 88*(2), 229-244.

González, E.F., Trejo, N.P., & Roux, R. (2017). Assessing EFL University Students' Writing: A Study of Score Reliability. *REDIE, 19*(2), 1607-4041.

González-Fernández, B. (2022). Conceptualizing L2 vocabulary knowledge: An empirical examination of the dimensionality of word knowledge. *Studies in Second Language Acquisition, 44*(4), 1124-1154.

González-Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *International Journal of Applied Linguistics, 166*(1), 94-126.

González-Fernández, B., & Schmitt, N. (2020). Word Knowledge: Exploring the Relationships and Order of Acquisition of Vocabulary Knowledge components. *Applied Linguistics, 41*(4), 48-505.

Grabe, W. (2001) Notes towards a theory of second language writing. In T. Silva and P.K. Matsuda (eds) *On Second Language Writing* (pp. 39-58). Hillsdale, NJ: Erlbaum.

Griffin, G.F. (1992). *components of the psychology of second language vocabulary list learning.* [Doctoral dissertation, University of Warwick].

Griffin, G.F., & Harley, T. (1996). List Learning of Second Language Vocabulary. *Applied Psycholinguistics, 17*(4), 443-60.

Greidanus, T., & Nienhuis, L. (2001). Testing the quality of word knowledge in a second language by means of word associations: Types of distractors and types of associations. *Modern Language Journal, 85*(4), 567-577.

Ghalib, T.K., & Al-Hattami, A.A. (2015). Holistic versus Analytic Evaluation of EFL Writing: A Case Study. *English Language Teaching, 8*(7), 225-236.

Harmer, J. (2001). *The Practice of English Language Teaching* (3rd Ed.), Edinburgh: Pearson Education.

Ha, H.S. (2019). Lexical Richness in EFL Undergraduate Students' Academic Writing. *English Teaching, 74*(3), 3-28.

Halliday, M.A.K., & Hasan, R. (1989). *Language, context, and text: components of language in a social-semiotic perspective.* Oxford: Oxford University Press.

Hamp-Lyons, L. (2011). Writing assessment: Shifting Issues, new tools, enduring questions. *Assessing Writing, 16*(1), 3-5.

Hadley, E.B., Dickinson, D.K., & Kathy, H.P. (2015). Examining the Acquisition of Vocabulary Knowledge Depth among Preschool Students. *Reading Research Quarterly, 51*(2), 181-198.

Hadley, E.B., Dickinson, D.K., & Kathy, H.P. (2018). Building Semantic Networks: The Impact of a Vocabulary Intervention on Preschoolers' Depth of Word Knowledge. *Reading Research Quarterly, 54*(1), 41-61.

Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae, 30*(2), 87-93.

Haastrup, K., & Henriksen, B. (2000). Vocabulary acquisition: Acquiring depth of knowledge through network building. *International Journal of Applied Linguistics, 10*(2), 221-240.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*(2), 303-317.

Hills, T.T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Categorical structure among shared features in networks of early-learned nouns. *Cognition, 112*(3), 381-396.

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal 36*(2), 153-166.

Hinkel, E. (2002). *Second Language Writers' Text: Linguistic and Rhetorical Features.* London: Lawrence Erlbaum Associates.

Hinkel, E. (2004). *Teaching ESL Academic Writing: Practical Techniques in Vocabulary and Grammar.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hinkel, E. (2006). Current perspectives on teaching the four skills. *TESOL Quarterly, 40*(1), 109-131.

Huang, H.-T. (2010). *How does Second Language Vocabulary Grow over Time ? A Multi-Methodological Study of Incremental Vocabulary Knowledge Development* [Doctoral dissertation]. Retrieved from ProQuest Dissertations & Theses database. (UMI No. 3415908)

Hu, G. W. (2003). English language teaching in China: Regional differences and contributing factors. *Journal of Multilingual and Multicultural Development, 24*(4), 290-318.

Huang, L.P. & Shu, Y. (2020). An overview of the English vocabulary teaching in China: status quo, problems and prospects. *Foreign Language Research, 15*(2020), 21-45.

Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*(3), 539-558.

Hyland, K. (2011). Learning to write: Issues in theory, research and pedagogy. In R. M. Manchon (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 17-35). Amsterdam: John Benjamins.

Hyland, K. (2003) *Second Language Writing.* Cambridge: Cambridge University Press.

Hyland, K. (1997). Is EAP necessary? A survey of Hong Kong undergraduates. *Asian Journal of English Language Teaching, 7*(1997), 77-99.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly, 41*(2), 235-253.

Ishii, T. & Schmitt. N.(2009). Developing an integrated diagnostic test of vocabulary size and depth. *RELC Journal, 40*(1), 5-22.

Ishii, T. (2005). *Diagnostic Test of Vocabulary Knowledge for Japanese Learners of English.* [Doctoral dissertation, University of Nottingham]. Nottingham, UK.

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning, 63*(1), 87-106.

Janebi Enayat, M., & Amirian, S.M.R. (2016). Vocabulary Levels Test and Word Associates Test: Can they Measure Language Proficiency? *International Journal of Assessment and Evaluation in Education, 6*(2016), 17-26.

Janebi Enayat, M., & Amirian, S.M.R. (2020). The Relationship between Vocabulary Size and Depth for Iranian EFL Learners at Different Language Proficiency Levels. *Iranian Journal of Language Teaching Research, 8*(2), 97-114.

Janebi Enayat, M., & Derakhshan, A. (2021). Vocabulary size and depth as predictors of second language speaking ability. *System, 99* (2021), 102-121.

Jacobs, H., Zingraf, A., Warmuth, D., Hartfiel, V. & Hughey, J. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.

Janssen, Gerriet., Meier, Valerie., & Trace, Jonathan. (2015). Building a better rubric: Mixed methods rubric revision. *Assessing Writing, 26*(2015), 51-66.

Jeffery, J. V. (2009). Constructs of writing proficiency in US state and national writing assessments: Exploring variability. *Assessing Writing, 14*(1), 3-24.

Jin, Y., & Fan, J.S. (2011). Test for English Majors (TEM) in China. *Language Testing, 28*(4), 589 -596.

Jordan, R.R. (1997). *English for Academic Purposes. A Guide and Resource Book for Teachers.* Cambridge: Cambridge University Press.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21*(1), 47-77.

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 24*(4), 617-637.

Jiang, N. (2004). Semantic transfer and development in adult L2 vocabulary acquisition. In P. Bogaards & B. Laufer (Eds). *Vocabulary in a Second Language: Selection, Acquisition and Testing* (pp. 101-126). Philadelphia, PA: John Benjamins.

Joe, A. (1998). What effects do text-based tasks prompting generation have on incidental vocabulary acquisition? *Applied Linguistics, 19*(3), 357-377.

Johnson, M., Acevedo, A., & Mercado, L. (2016). Vocabulary Knowledge and Vocabulary Use in Second Language Writing. *TESOL Journal, 7*(3), 700-715.

Karakoç, D., & Köse, G.D. (2017). The impact of vocabulary knowledge on reading, writing and proficiency scores of EFL learners. *Journal of Language and Linguistic Studies, 13*(1), 352-378.

Keating, G. (2008). Task effectiveness and word learning in a second language. *Language Teaching Research, 12*(3), 365-386.

Keith, T. Z. (2015). *Multiple Regression and Beyond (2nd edition)*. Boston, MA: Pearson Education Inc.

Kilic, M. (2019). Vocabulary Knowledge as a Predictor of Performance in Writing and Speaking: A Case of Turkish EFL Learners. *PASAA, 57*(2019), 133-164.

Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 58*(2), 285–325

Kim, Y.S., & Schatschneider, C. (2017). Expanding the developmental models of writing: A direct and indirect effects model of developmental writing (DIEW). *J Educ Psychol, 109*(1): 35-50.

Kim, M., Crossley, S.A., & Kim, B.K. (2022). Second language reading and writing in relation to first language, vocabulary knowledge, and learning backgrounds. *International Journal of Bilingual Education and Bilingualism, 25*(6), 1992-2005.

Kieffer, M.J. & Lesaux, N.K. (2012). Knowledge of Words, Knowledge about Words: Dimensions of Vocabulary in First and Second Language Learners in Sixth Grade. *Reading and Writing , 25*(2012), 347-373.

Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing, 26*(2), 275-304.

Knoch, U., & Chapelle, C.A. (2017). Validation of rating processes within an argument-based framework. *Language Testing, 35*(4), 477-499.

Koizumi, R., & In'nami, Y. (2020). Structural Equation Modeling of Vocabulary Size and Depth Using Conventional and Bayesian Methods. *Frontiers in Psychology, 11*(2020), 618-629.

Koizumi, R., & In'nami, Y. (2013). Vocabulary Knowledge and Speaking Proficiency among Second Language Learners from Novice to Intermediate Levels. *Journal of Language Teaching and Research, 4*(5), 900-913.

Kojima, M., & Yamashita, J. (2014). Reliability of lexical richness measures based on word lists in short second language productions. *System, 42*(2014), 23-33.

Kremmel, B. & Schmitt, N. (2016). Interpreting Vocabulary Test Scores: What Do Various Item Formats Tell Us about Learners' Ability to Employ Words? *Language Assessment Quarterly, 13*(4), 377-392.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings and application. *TESOL Quarterly, 49*(4), 757-786.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In Arnaud, P.J.L. and Béjoint, H. (eds), *Vocabulary and Applied Linguistics* (pp. 126-132). London: Macmillan.

Laufer, B. (1994). The Lexical Profile of Second Language Writing: Does it Change Over Time? *RELC Journal, 25*(2), 21-33.

Laufer, B. (1998). The development of passive and active vocabulary: same or different? *Applied Linguistics, 19*(2), 255-271.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *The Canadian Modern Language Review, 59*(4), 567–587.

Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook, 5*(1), 223-250.

Laufer, B. (2013). Vocabulary and writing. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*(pp. 117-127). Boston: Wiley-Blackwell.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive vocabulary. *Language Testing, 16*(1), 33-51.

Laufer. B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics, 16*(3), 307-322.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54*(3), 399-436.

Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning, 48*(3), 365-391.

Laufer, B. C., Elder, K., Hill,P., & Congdon. (2004). Size and strength: do we need both to measure vocabulary knowledge? *Language Testing, 21*(2), 202-226.

Laufer, B. & Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learner' English. *Language Learning, 61*(2), 647-672.

Laufer, B.C., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1-26.

Laufer, B.C., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it ? *RELC Journal, 28*(1), 89-108.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15-30.

Lawrence, J.F., Hagen, A.M., Hwang, J.K., Lin, G., & Lerva, A. (2018). Academic vocabulary and reading comprehension: exploring the relationships across measures of vocabulary knowledge. *Reading and Writing, 32*(2019), 285-306.

Levitzky-Aviad, T., & Laufer, B. (2013). Lexical properties in the writing of foreign language learners over eight years of study: single words and collocations. In C. Bardel, C. Lindquist & B. Laufer (Eds). L2 Vocabulary Acquisition, Knowledge and Use, new perspectives on assessment and corpus analysis. (pp. 127-148). European Second Language Association.

Levitzky-Aviad, T., & Laufer, B. & Goldstein, Z. (2019). The New Computer Adaptive Test of Size and Strength (CATSS): Development and Validation. *Language Assessment Quarterly, 16*(3), 345-368.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation.* Cambridge, MA: MIT Press.

Lee, S.H. (2003). ESL learners' vocabulary use in writing and the effects of explicit vocabulary instruction. *System, 31*(2003), 537-561.

Lee, S. H., & Muncie, J. (2006). From receptive to productive: Improving ESL learners' use of vocabulary in a post-reading composition task. *TESOL Quarterly, 40*(2), 295-320.

Leontjev, D., Huhta, A., & Mantyla, K. (2016). Word derivational knowledge and writing proficiency: How do they link? *System, 59*(2016), 73-89.

Leki, I., & Carson, J. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly, 28*(1), 81-101.

Lemmouh, Z. (2010). *The Relationship among Vocabulary Knowledge, Academic Achievement and the Lexical Richness in Writing in Swedish University Students of English*. [Doctoral dissertation, Stockholm University]. Stockholm.

Lin, H.F. (2015). *Roles of Depth of Vocabulary Knowledge in EFL Learners' Writing Proficiency*. [Doctoral dissertation, The Hong Kong Polytechnic University]. Hong Kong.

Lin, C., & Hirsh, D. (2012). Manipulating instructional method: The effect on productive vocabulary use. In Hirsh, D. (Ed.), *Current perspectives in second language vocabulary research* (pp. 117–148). Bern, Switzerland: Peter Lang.

Li, M., & Kirby, J. R. (2015). The effects of vocabulary breadth and depth on English reading. *Applied Linguistics, 36*(5), 611-34.

Li, H. & He, L.Z. (2015). A Comparison of EFL Raters' Essay-Rating Processes Across Two Types of Rating Scales. *Language Assessment Quarterly, 12*(2), 178-212.

Li, Y. (2022). The Role of Vocabulary Knowledge in Second Language Writing. In Hirsh, D. (Ed.), *Research Perspectives in Language and Education* (pp. 24-36). Peter Lang Group.

Liu, M., & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System, 33*(4), 623-636.

Llach, M.P.A. (2005). The relationship of lexical error and their types to the quality of ESL compositions: an empirical study. *Porta Linguarum, 3*(2005), 45-57.

Llach, M.P.A. (2011). *Lexical errors and accuracy in foreign language writing.* Second Language Acquisition. Multilingual Matters.

Lowry, M. (1999). Lexical Issues in University ESL Writing Class. *The CATESOL Journal, 12*(1999), 7-37.

Mancho´n, R. M. (2011). Writing to learn the language: Issues in theory and research. In R. M. Mancho´n (Ed.), *Learning-to-write and writing-to learn in an additional language* (pp. 61-82). Amsterdam: John Benjamins.

Mancho´n, R. M., Murphy L., & Roca J. (2007). Lexical Retrieval Processes and Strategies in Second Language Writing: A Synthesis of Empirical Research. *IJES, 7*(2), 149-174.

Malvern, D.D., & Richards, B.J. (1997) Anew measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language* (pp. 58-71). Clevedon, UK: Multilingual Matters.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical richness and language development: Quantification and assessment.* Houndmills, Basingstoke: Palgrave Macmillan.

McCarthy, P.M. (2005). An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD). *Dissertation Abstracts International,* 66(12), 117-136.

McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing, 24*(4), 459-488.

McLean, S., & Kramer, B. (2015). The Creation of a New Vocabulary Levels Test. *Shiken, 19*(2), 1-32.

McCutchen, D., & Stull, S. (2015). Morphological awareness and children's writing: accuracy, error, and invention. *Read Writ, 28*(2015), 271-289.

McLean, S., Stewart, J., & Olaf Batty, A. (2020). Predicting L2 reading proficiency with modalities of vocabulary knowledge: A bootstrapping approach. *Language Testing, 37*(3), 1-23.

McNamara, D.S., Crossley, S.A., & McCarthy, P.M. (2010). Linguistic Features of Writing Quality. *Written Communication, 27*(1), 57 -86.

Meara, P. M. (1984). *The study of lexis in interlanguage.* In A. Davies, C. Criper, & A. Howatt (Eds.), *Interlanguage* (p. 225-240). Edinburgh, Scotland: University of Edinburgh Press.

Meara, P. M. (1990). A note on passive vocabulary. *Second Language Research, 6*(2),150-154.

Meara, P. M. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Competence and performance in language learning* (pp. 35-53). New York: Cambridge University Press.

Meara, P. M. (1997). Towards a new approach to modeling vocabulary acquisition. In Schmitt, N. and McCarthy, M. (eds), *Vocabulary: Description, Acquisition, and Pedagogy* (pp. 109-121*)*. Cambridge: Cambridge University Press.

Meara, P. M. (2002). The rediscovery of vocabulary. *Second Language Research, 18*(4), 393-407.

Meara, P. M. (2005). Lexical Frequency Profiles: A Monte Carlo Analysis. *Applied Linguistics, 26*(1), 32-47.

Meara, P. M. (2009). *Connected Words: Word Associations And Second Language Vocabulary Acquisition*. Amsterdam, the Netherlands: John Benjamins.

Meara, P.M., & Jones, G. (1990). *Eurocentres Vocabulary Size Test, Version E1.1/K10*. Zurich: Eurocentres Learning Service.

Meara, P.M., & Bell, H. (2001). P_Lex: a simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect, 16*(3), 5-19.

Meara, P. M., & Fitzpatrick, T. (2000). "Lex30: an improved method of assessing productive vocabulary in an L2". *System, 28(1),* 19-30.

Meara, P. M., & Wolter, B. (2004). Beyond vocabulary depth. In D. Albrechtsen, K. Haastrup, & B. Henriksen (Eds.), *Angles on the English speaking world 4* (pp. 85-96). Copenhagen, Denmark: Museum Tusculanum Press.

Meara, P., Carlos, J., & Alcoy, O. (2010). Words as species: An alternative approach to estimating productive vocabulary size. *Reading in a Foreign Language, 22*(1), 222-236.

Melka, F. (1997). Receptive vs. productive components of vocabulary. In Schmitt, N. and McCarthy, M. (eds), *Vocabulary: Description, Acquisition, and Pedagogy* (pp. 84-102). Cambridge: Cambridge University Press.

Milton, J. (2009). *Measuring second language vocabulary acquisition,* Vol. 54, Bristol: Multilingual Matters.

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In Bardel, C. Lindqvist, C. and Laufer, B. (Eds.), *L2 vocabulary acquisition, knowledge and use* (pp. 57-78). Eurosla.

Milton, J., & Fitzpatrick, T. (Eds.). (2014). *Dimensions of vocabulary knowledge.* Basingstroke: Palgrave Macmillan.

Miller, G. A. (1999). On knowing a word. *Annual Review of Psychology, 50*(1), 1-19.

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacon-Beltran, C. Abello-Contesse, M. Torreblanca-Lopez, and M. Lopez-Jimenez. *Further Insights into Non-Native Vocabulary Teaching and Learning* (pp. 115-130). Cambridge University Press.

Morris, L., & Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System, 32*(2004), 75-87.

Mondria, J., &Wiersma, B. (2004). Receptive, Productive, and Receptive and Productive L2 Vocabulary Learning: What Difference Does it Make?', In P. Bogaards and B. Laufer (Eds), *Vocabulary in a Second Language: Selection, Acquisition and Testing* (pp. 79-100). Amsterdam: Benjamins.

Nation, I.S.P. (1983). Testing and teaching vocabulary. *Guidelines, 5*(1), 12-25.

Nation, I.S.P. (1990). *Teaching and learning vocabulary.* Boston: Heinle & Heinle Publishers.

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

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59-82.

Nation, I.S.P. (2007). The Four Strands. *Innovation in Language Learning and Teaching, 1*(1), 2-13.

Nation, I. S. P. (2012). The BNC/COCA word family lists. http://www.victoria.ac.nz

Nation, I.S.P. (2013). *Learning vocabulary in another language (2nd ed.)*. Cambridge, England: Cambridge University Press.

Nation, I. S. P. (2020). The different aspects of vocabulary knowledge. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 15–29). Routledge.

Nation, I.S.P. (2022). *Learning vocabulary in another language (3rd ed.)*. Cambridge, England: Cambridge University Press.

Nation, I.S.P., & Waring, R. (1997). 'Vocabulary size, text coverage and word lists' in Schmitt N. and Nation, P (eds), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 6-19). Cambridge university press.

Nation, I.S.P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Nation, I. S. P., & Gu, Y. (2007). *Focus on vocabulary.* Sydney: National Centre for English Language Teaching and Research, Macquarie University.

Nation, I.S.P., & Webb, S. (2011). *Researching and analyzing vocabulary.* Boston, MA: Heinle.

Nagy, W. E., & Scott, J. (2000). Vocabulary processes. In M. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (pp. 269-284). Mahwah: Lawrence Erlbaum Associates.

Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *Modern Language Journal, 90*(3), 387-40.

Namaziandost, E., Hosseini E., & Utomo, D.W. (2020). A comparative effect of high involvement load versus lack of involvement load on vocabulary learning among Iranian sophomore EFL learners, *Cogent Arts & Humanities, 7*(1), 2-14.

Neumann, A. (2012). Advantages and Disadvantages of Different Text Coding Procedures for Research and Practice in a School Context. In A. Neumann (Eds). *Measuring Writing: Recent Insights into theory, methodology and practices* (pp. 33-54). Brill.

Nguyen, T.M.H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research, 21*(3), 298-320.

Nizonkiza, D.& Van den Berg, K. (2014). The dimensional approach to vocabulary testing: What can we learn from past and present practices? *Stellenbosch Papers in Linguistics Plus, 43*(2014), 45-61.

Nizonkiza, D. (2016). First-year university students' receptive and productive use of academic vocabulary. *Stellenbosch Papers in Linguistics, 45*(2016), 169-187.

Nontasee,W., & Sukying, A. (2021). The learnability of word knowledge components in Thai EFL high school learners. *Journal of Language and Linguistic Studies, 17*(1), 34-55.

Olinghouse, N.G., & Leaird, J.T. (2009). The relationship between measures of vocabulary and narrative writing quality in second and fourth-grade students. *Springer*, *22*(2), 545-565.

Olinghouse, N.G., & Santangelo, T. (2010). Assessing the Writing of Struggling Learners. *Focus on Exceptional Children*, *43*(4), 1-27.

Olinghouse, N.G., Santangelo, T., & Wilson, J. (2012). Examining the Validity of Single-Occasion, Single-Genre, Holistically Scored Writing Assessments. Measuring Writing: Recent Insights into Theory, Methodology and Practices *Studies in Writing*, *27*(2), 55-82

Palmberg, R. (1987). Patterns of vocabulary development in foreign-language learners. *Studies in Second Language Acquisition*, *9*(2), 201-220.

Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis.* London, England: Continuum International.

Phakiti, A. (2015). Quantitative research and analysis. In B. Paltridge, & A. Phakiti (Eds). *Research methods in applied linguistics: A practical resource* (pp. 27-48). Bloomsbury.

Paribakht, T.S., & Wesche, M. (1996). Assessing second language vocabulary knowledge: Breadth vs Depth. *Canadian Modern Language Review*, *53*(1), 13-39.

Paribakht, T.S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In Coady, J. and Huckin, T. (eds), *Second Language Vocabulary Acquisition (*pp. 174-200). Cambridge University Press.

Paribakht, T.S. (2005). The influence of L1 lexicalization on L2 lexical inferencing: AStudy of Farsi-speaking EFL learners. *Language Learning*, *55*(4), 701-748.

Pellicer-Sanchez, A. (2019). Examining second language vocabulary growth: Replications of Schmitt (1998) and Webb & Chang (2012). *Language Teaching-Replication Paper.*

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42*(2), 282-296.

Peng, Y.X., Sun, J., Quan, J.Q., Wang, Y.Q., Lv, C.Y., & Zhang, H.M. (2023). Predicting Chinese EFL Learners' Human-rated Writing Quality in Argumentative Writing Through Multidimensional Computational Indices of Lexical Complexity. *Assessing Writing, 56*(2023), 1-15.

Pichette, F., de Serres, L., & LaFontaine, M. (2013). Sentence reading and writing for second language vocabulary acquisition. *Applied Linguistics, 33*(1), 66-82.

Polio, C. (1997) Measures of linguistic accuracy in second language writing research. *Language Learning, 47*(1), 101-143.

Polio, C. (2001) Research methodology in second language writing research: The case of text-based studies. In T. Silva and P.K. Matsuda (Eds). *On Second Language Writing* (pp. 181-201). Hillsdale, NJ: Erlbaum.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review, 56*(2), 282-308.

Qian, D.D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513-536.

Qian,D.D. (2008). From Single Words to Passages: Contextual Effects on Predictive Power of Vocabulary Measures for Assessing Reading Performance. *Language Assessment Quarterly, 5*(1), 1-19

Qian, D.D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing, 21*(1), 28-52.

Qian, D. D., & Lin, L. H. F. (2020). The relationship between vocabulary knowledge and language proficiency. In S. Webb (Ed.), The Routledge Handbook of Vocabulary Studies, (pp. 66-80). Routledge.

Qi, L. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education: Principles, Policy & Practice, 14*(1), 51-74.

Qi, G.Y. (2016). The importance of English in primary school education in China: perceptions of students. *Multilingual Education, 6*(1), 1-27.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing. 10*(3): 355-371.

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. J. Kunnan (Ed.), *Validation in language assessment: Selected papers from 17th Language Testing Research Colloquium, Long Beach* (pp. 41–60). Lawrence Erlbaum Associates Publishers.

Read, J. (2000). *Accessing Vocabulary.* Cambridge University Press 2000.

Read, J. (2004). Plumbing the Depths: How Should the Construct of Vocabulary Knowledge Be Defined. In P. Bogaards, & B. Laufer (Eds.), Vocabulary in a Second Language (pp. 209-227). Amsterdam: John Benjamin.

Read, J. (2007). Second Language Vocabulary Assessment: Current Practices and New Directions. *IJES, 7*(2), 105-125.

Read, J. (2020). Key issues in measuring vocabulary knowledge. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 545–560). Routledge.

Read, J., & Chapelle, C.A. (2001). A framework for second language vocabulary assessment. *Language Testing, 18*(1), 1-32.

Read, J., & Dang, T.N.Y. (2022). Measuring depth of academic vocabulary knowledge. Language Teaching Research, 1-27.

Ren, H.L., & Liu, Y.L. (2020). The Study of Effects of Tasks with the Same Involvement Load on the Incidental Vocabulary Acquisition. *China Academic Journal Electronic Publishing House,17*(2), 23-30.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly, 10*(1), 77-89.

Richard, J. (2011). Does size matter? The relationship between vocabulary breadth and depth. *Sophia International Review, 16*(2011), 107-120.

Robinson, P. J. (1988). *Procedural and Declarative Knowledge in Vocabulary Learning: Communication and the Language Learner's Lexicon.* Annual Meeting of the Teachers of English to Speakers of Other Languages, Chicago.

Ruhl, C. (1989). *On monosemy: A study in linguistic semantics.* Albany: StateUniversity of New York Press.

Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly, 22*(1), 69-90.

Saville-Troike, M. (1984). What really matters in second language learning for academic achievement? *TESOL Quarterly, 18*(2), 199-219.

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016). Lexical profiles of comprehensible second language speech. *Studies in Second Language Acquisition, 38*(4), 677-701.

Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning, 48*(2), 81-317.

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

Schmitt, N. (2008). Review article instructed second language vocabulary learning. *Language Teaching Research,12*(3), 329-363.

Schmitt, N. (2010). *Researching vocabulary*: A vocabulary research manual. Basingstoke, England: Palgrave Macmillan.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64*(4), 913-951.

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word Associations and Verbal Suffixes. *Studies in Second Language Acquisition*, *19*(1), 17-36.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55-88.

Schmitt, N., Ng, J.W.C., & Garras, J. (2011). The word associates format: Validation evidence. *Language Testing, 28*(1), 105-126.

Schmitt, N., Dornyei, Z., Adolphs, S., & Durrow, V. (2004). Knowledge and acquisition of formulaic sequences: A longitudinal study. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp.55-86). Philadelphia: John Benjamins.

Schipolowski, S. & Böhme, K. (2016). Assessment of writing ability in secondary education: comparison of analytic and holistic scoring systems for use in large-scale assessments. *L1-Educational Studies in Language and Literature*, *16*(2016), 1-22.

Schoonen, R. (2003). First language and second language writing: the role of linguistic knowledge, speed of processing and metacognitive knowledge. *Language Learning 53*(1), 165-202.

Schoonen, R. (2012). The Validity and Generalizability of Writing Scores: The Effect of Rater, Task and Language. Measuring Writing: Recent Insights into Theory, Methodology and Practices. *Studies in Writing*, *27*(2012), 1-22.

Schoonen, R., & Verhallen, M. (2008). The assessment of deep word knowledge in young first and second language learners. *Language Testing*, *25*(2), 211-236.

Schoonen, R., Snellings, P., Stevenson, M., & van Gelderen, A. (2009). Toward a blueprint of the foreign language writer: The linguistic and cognitive demands of foreign language writing. In R. Mancho´n (Ed.), *Writing in foreign language contexts* (pp. 77-101). Bristol: Multilingual Matters.

Schoonen, R., Van Gelderen, A., Stoel, R., Hulstijn, J., & De Glopper, K. (2011). Modeling the development of L1 and EFL writing proficiency of secondary-school students. *Language Learning, 6*(1), 31-79.

Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f 2 , a measure of local effect size, from PROC MIXED. *Frontiers in Psychology, 3*(111), 1-6.

Shaw, S. (2004). *IELTS Writing: revising assessment criteria and scales (phase 3).* Research and validation group.

Shaw, S., & Falvey, P. (2008). *The IELTS Writing Assessment Revision Project: Towards a revised rating scale*. Cambridge: University of Cambridge ESOL Examinations.

Shi, L.M. & Qian, D. (2012). How Does Vocabulary Knowledge Affect Chinese EFL Learners' Writing Quality in Web-based Settings?—Evaluating the Relationships Among Three Dimensions of Vocabulary Knowledge and Writing Quality. *Chinese Journal of Applied Linguistics (Quarterly),35*(1), 117-127.

Singleton, D. (1999). *Exploring the second language mental lexicon.* Cambridge: Cambridge University Press.

Siyanova–Chantura, A., & Martinez, R. (2015). The idiom principle revisited. Applied Linguistics, 36(5), 549-569.

Skjelde, K., & Coxhead, A. (2020). Mind the gap: Academic vocabulary knowledge as a predictor of English grades. *Acta Didactica Norden,14*(3), 1-20.

Slomp, D.H. (2012). Challenges in assessing the development of writing ability: Theories, constructs and methods. *Assessing Writing, 17*(2012), 81-91.

Soleymanzadeha, L., & Gholami, J. (2014). Scoring Argumentative Essays Based on Thematic Progression Patterns and IELTS Analytic Scoring Criteria. *Procedia Social and Behavioral Sciences, 98*(2014), 1811-1819.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal, 36*(2), 139-152.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition 31*(4), 577-607.

Sukying, A. (2017). *The relationship between receptive and productive affix knowledge andvocabulary size in an EFL Context*. [Doctoral Dissertation, University of Sydney].

Sukying, A. (2018). Investigating receptive and productive affix knowledge in EFL learners. In David Hirsh (Ed.), *Explorations in Second Language Vocabulary Research* (Vol. 254, pp. 183-218). Bern: Peter Lang

Sukying, A. (2022). A Taxonomy of English Affix Acquisition in EFL Learners. In David Hirsh (Ed.), *Research Perspectives in language and education* (Vol. 289, pp. 49-84). Bern: Peter Lang

Sukying, A. (2023). The Role of Vocabulary Size and Depth in Predicting Postgraduate Students' Second Language Writing Performance. *LEARN Journal: Language Education and Acquisition Research Network, 16*(1), 575-603.

Sukying, A. & Nontasee, W. (2022). The acquisition order of vocabulary knowledge aspects in Thai EFL learners. *World Journal of English Language, 12*(5), 306-319.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp.235-253). Rowley, MA: Newbury House.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step toward second language learning. *Applied Linguistics, 16*(3), 371-391.

Teng, F. (2014). Assessing the Depth and Breadth of Vocabulary Knowledge with Listening Comprehension. *PASAA, 48*(2014), 29-56.

Teng, F. (2016). An In-depth Investigation into the Relationship between Vocabulary Knowledge and Academic Listening Comprehension. *The Electronic Journal for English as a Second Language, 20*(2), 1072-4303.

Teng, F. (2015). Assessing the Relationship between Vocabulary Learning Strategy Use and Vocabulary Knowledge. *PASAA, 49*(2015), 39-65.

Treffers-Daller, J. and Huang, J. (2020) Measuring reading and vocabulary with the Test for English Majors Band 4: a concurrent validity study. In: Clenton, J. and Booth, P. (eds.) *Vocabulary and the four skills-current issues future concerns* (pp. 91-113). Taylor & Francis, Abingdon.

Treffers-Daller, J., & Milton, J. (2013) Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review, 4*(1), 151-172.

Troitschanskaia, O.Z., Shavelson, R.J., Schmidt, S., & Beck, K. (2019). On the complementarity of holistic and analytic approaches to performance assessment scoring. *British Journal of Educational Psychology, 89*(3), 468-484.

Tseng, W-T. (2011). *Modeling vocabulary knowledge: A mixed modeling approach.* Paper presented at 2011 Language Testing Research Colloquium, June 23–25. Ann Arbor, MI: University of Michigan.

Tseng, W.-T., & Schmitt, N. (2008). Toward a model of motivated vocabulary learning: A structural equation modeling approach. *Language Learning, 58*(2), 357-400.

Uchihara, T., & Saito, K. (2016). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *The Language Learning Journal, 47*(1), 64-75.

Uludag, Pakize., & McDonough, Kim. (2022). Validating a rubric for assessing integrated writing in an EAP context. *Assessing Writing, 52*(2022), 100-609.

Uysal, H.H. (2009). A critical review of the IELTS writing test. *ELT Journal, 64*(3), 314-320.

Varnaseri, M., & Farvardin, M. (2016). The relationship between depth and breadth of vocabulary knowledge and writing performance of Iranian MA students of TEFL. *Modern Journal of Language Teaching Methods, 6*(2), 544-554.

Van Zeeland, H. (2013). L2 vocabulary knowledge in and out of context: Is it the same for reading and listening? *Australian Review of Applied Linguistics, 36*(1), 52-70.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics, 22*(1), 217-234.

Vermeer, A. (2004). The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. In P. Bogaards, & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 173-189). Amsterdam and Philadelphia, PA: John Benjamins.

Vögelin, C., Jansen, T., Keller, S.D.Machts, N., & Möller, J. (2020). Corrigendum to "The influence of lexical features on teacher judgments of ESL argumentative essays". *Assessing Writing, 43*(2020), 100-448.

Vögelin, C., Jansen, T., Keller, S.D.Machts, N., & Möller, J. (2021). The impact of vocabulary and spelling on judgments of ESL essays: an analysis of teacher comments, *The Language Learning Journal, 49*(6), 631-647.

Wang, W. Q. (2014). Students' perceptions of rubric-referenced peer feedback on EFL writing: A longitudinal inquiry. *Assessing Writing, 19*(2014), 80-96.

Warnby, M. (2023). Academic vocabulary knowledge among adolescents in university preparatory programme. *Journal of English for Academic Purposes, 61*(2023), 101-203.

Waring, R. (1997). A Study of Receptive and Productive Learning from Word Cards. *Studies in Foreign Languages and Literature, 21*(1997), 94-114.

Waring, R. (1998). Receptive and productive vocabulary: Do we know what we are talking about? *PACSLRF Conference,* March 26. Tokyo, Japan.

Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: the contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System, 65*(2017), 139-150.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 27*(1), 33-52.

Webb, S. (2007a). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46-65.

Webb, S. (2007b). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research, 11*(1), 63-81.

Webb, S. (2008a). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition, 30*(1), 79-95.

Webb, S. (2008b). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language, 20*(2), 232-245.

Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal, 40*(3), 360-76.

Webb, S. (2013). Depth of Vocabulary Knowledge. In A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics,* (pp. 1-8). Blackwell Publishing Ltd.

Webb, S., & Chang, A.C.S. (2012). Second language vocabulary growth. *RELC Journal, 43*(1), 113-26.

Webb, S., & Chang, A.B.S. (2015). Second Language Vocabulary Learning Through Extensive Reading with Audio Support: How Do Frequency and Distribution of Occurrence Affect Learning? *Language Teaching Research, 19*(6), 667-686.

Webb, S., & Sasao, Y. (2013). New Directions in Vocabulary Testing. *RELC Journal, 44*(3), 263-277.

Webb, S., Sasao, Y., & Oliver, B. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *International Journal of Applied Linguistics 168*(2017), 33-69.

Weigle, S.C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.

Weigle, S.C. (2007). Teaching writing teachers about assessment. *Journal of Second Language Writing, 16*(2007), 194-209.

Wesche, M., & Paribakht, T.S. (2009). *Lexical Inferencing in a First and Second Language:Cross-linguistic Dimensions.* Clevedon: Multilingual Matters.

Wesche, M. & Paribakht, S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review, 53*(1), 13-40.

Whitmore, J. M., Shore, W. J., & Smith, P. H. (2004). Partial knowledge of word meanings: Thematic and taxonomic representations. *Journal of psycholinguistic research, 33*(2), 137-164.

Wilkins, D.A. (1972). *Linguistics in Language Teaching.* London: Arnold.

William, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing, 21*(2012), 321-331.

Wiseman, C.S. (2012). A Comparison of the Performance of Analytic vs. Holistic Scoring Rubrics to Assess L2 Writing. *Iranian Journal of Language Testing, 2*(1), 225-252.

Winke, Paula. & Lim, Hyo Jung. (2015). ESL essay raters' cognitive processes in applying the Jacobs et al. rubric: An eye-movement study. *Assessing Writing, 25*(2015), 38-54.

Wilks, C., & Meara, P. (2002). Understanding word webs: Graph theory and the notion of density in second language word association networks. *Second Language Research, 18*(4), 303-324.

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition, 23*(1), 41-69.

Wolter, B. W. (2005). *V-links : a new approach to assessing depth of word knowledge.* [Doctoral dissertation, University of Wales Swansea].

Wolff, D. (2000). Some reflections on the importance of writing in a foreign language. In I. Plag & K. Schneider (Eds.), *Language use, language acquisition and language history* (pp. 213-226). Trier, Germany: WVT Wissenschaftlicher Verlag Trier.

Wu, S., Dixon, Q., Sun, H.T., & Zhang, P. (2019). Breadth or depth: the role of vocabulary in Chinese English Language beginning writers' development. *International Journal of Bilingual Education and Bilingualism, 24*(9), 1356-1372.

Xue, G., & Nation, P. (1984). A university word list. *Language learning and communication, 3*(2), 215-229.

Yi, Jyi-yeon. (2009). Defining writing ability for classroom writing assessment in high schools. *Journal of Pan-Pacific Association of Applied Linguistics, 13*(1), 53-69.

You, X. (2010). *Writing in the devil's tongue: A history of English composition in China*. Carbondale: Southern Illinois University Press.

Yoon, Hyung-Jo. (2018). The Development of ESL Writing Quality and Lexical Proficiency: Suggestions for Assessing Writing Achievement. *Language Assessment Quarterly, 15*(4), 387-405 .

Yoon, Hyung-Jo. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System, 66*(2017), 130-141.

Yu, G.X. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics, 31*(2), 236-259.

Yanagisawa, A., & Webb, S. (2020). Measuring depth of vocabulary knowledge. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 371–386). Routledge.

Zareva, A. (2005). Models of lexical knowledge assessment of second language learners of English at higher levels of language proficiency. *System, 33*(4), 547-562.

Zareva, A., Schwanenflugel, P., & Nikolova,Y. (2005). Relationship between lexical competence and language proficiency. *Studies in Second Language Acquisition., 27*(4), 567-595.

Zhang, D., & Yang, X. (2017). Chinese L2 Learners' Depth of Vocabulary Knowledge and Its Role in Reading Comprehension. *Foreign Language Annals, 49*(4), 699-715.

Zhang, X. & Lu, X. (2015). The Relationship Between Vocabulary Learning Strategies and Breadth and Depth of Vocabulary Knowledge. *The Modern Language Journal, 99*(4), 740-753.

Zhang, D.B. & Koda, K. (2017). Assessing L2 vocabulary depth with word associates format tests: issues, findings, and suggestions. *Asian-Pacific Journal of Second and Foreign Language Education, 2*(1), 1-30.

Zhang, X.P., Lu, X.F., & Li, W.W. (2021). Beyond Differences: Assessing Effects of Shared Linguistic Features on L2 Writing Quality of Two Genres. *Applied Linguistics, 43*(1), 168-195.

Zhong, H., & Hirsh, D. (2009). Vocabulary growth in English as a foreign language context. *University of Sydney Papers in TESOL, 4*(2009), 85-113.

Zhong, H. (2014). The Interface between Receptive and Productive Vocabulary Knowledge: Vocabulary Knowledge as a Multi-component Construct. [Doctoral dissertation. Sydney University]. Sydney.

Zhong, H. (2016). The relationship between receptive and productive vocabulary knowledge: a perspective from vocabulary use in sentence writing, *The Language Learning Journal, 46*(4), 357-370.

APPENDICES

Appendix I Instruments for Pilot Study

Appendix I-a The Form-recall Test

Instructions

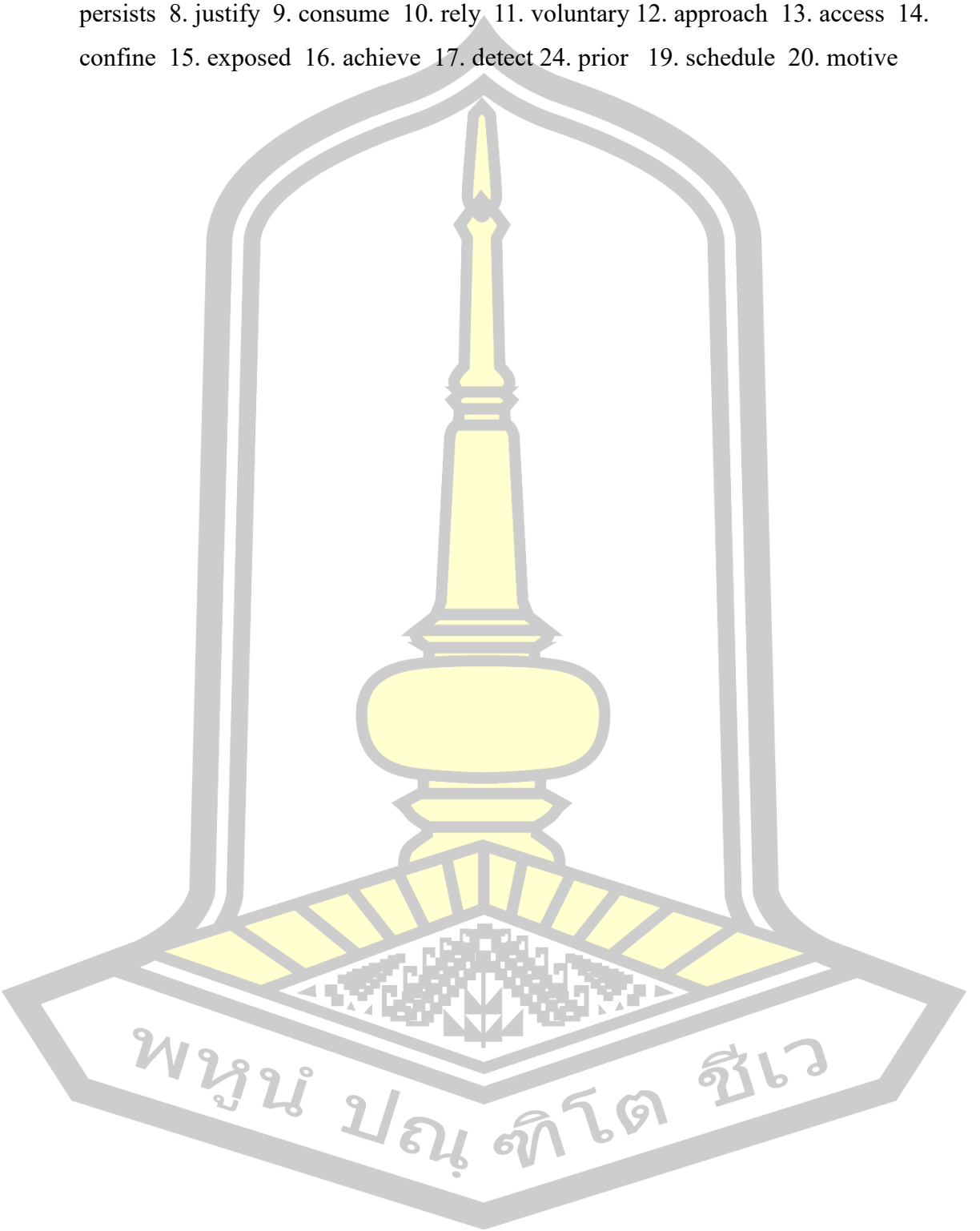
Complete the underlined words in the sentence. An example has been done for you.

请将下列句子中的单词填写完整，注意语境和语法，请参照示例完成。

Example The police inspected the factory for safety after the earthquake.

1. The hotel wants to exp_____ its business by adding two more stores.
2. She was inc_____ to become an English teacher after I graduate from the Normal University.
3. To be good at the game, you need a reasonable level of inte_____.
4. An inspiring teacher can sti_____ students to work harder.
5. Research ind_____ that over 81% of teachers are dissatisfied with their salary.
6. He wanted to de_____ all his time and energies to writing films.
7. The snow on the top of the mountain per_____throughout the year.
8. The Prime Minister has been asked to jus_____ the decision to the public.
9. The electricity factory will con_____a large amount of fossil fuels such as coal.
10. These days, some students re_____ too much on teachers to tell them all the things they have to learn.
11. She didn't want to make more money; she did these things only to provide some vol_____ services.
12. He decided to adopt a new app____ and teach the course through story-telling.
13. You need a password to get acc_____ to the computer system.
14. It is cruel to con____ the bird in a cage; she needs freedom in the sky.
15. The report revealed that workers had been exp_____ to high levels of danger.
16. They could not ach_____ their target of less than 3% inflation.
17. The disease is difficult to det_____ at its early stage.
18. You do not need any pr_____knowledge of the subject. Anyone can attend the course anytime.
19. I have a busy sch_____ for the next few days.
20. There seemed to be no mot_____ for the murder.

Keys: 1. expand 2. inclined 3. intelligence 4. stimulate 5. indicates 6. devote 7. persists 8. justify 9. consume 10. rely 11. voluntary 12. approach 13. access 14. confine 15. exposed 16. achieve 17. detect 24. prior 19. schedule 20. motive



Appendix I-b The Productive Word-Pair Test

Instructions

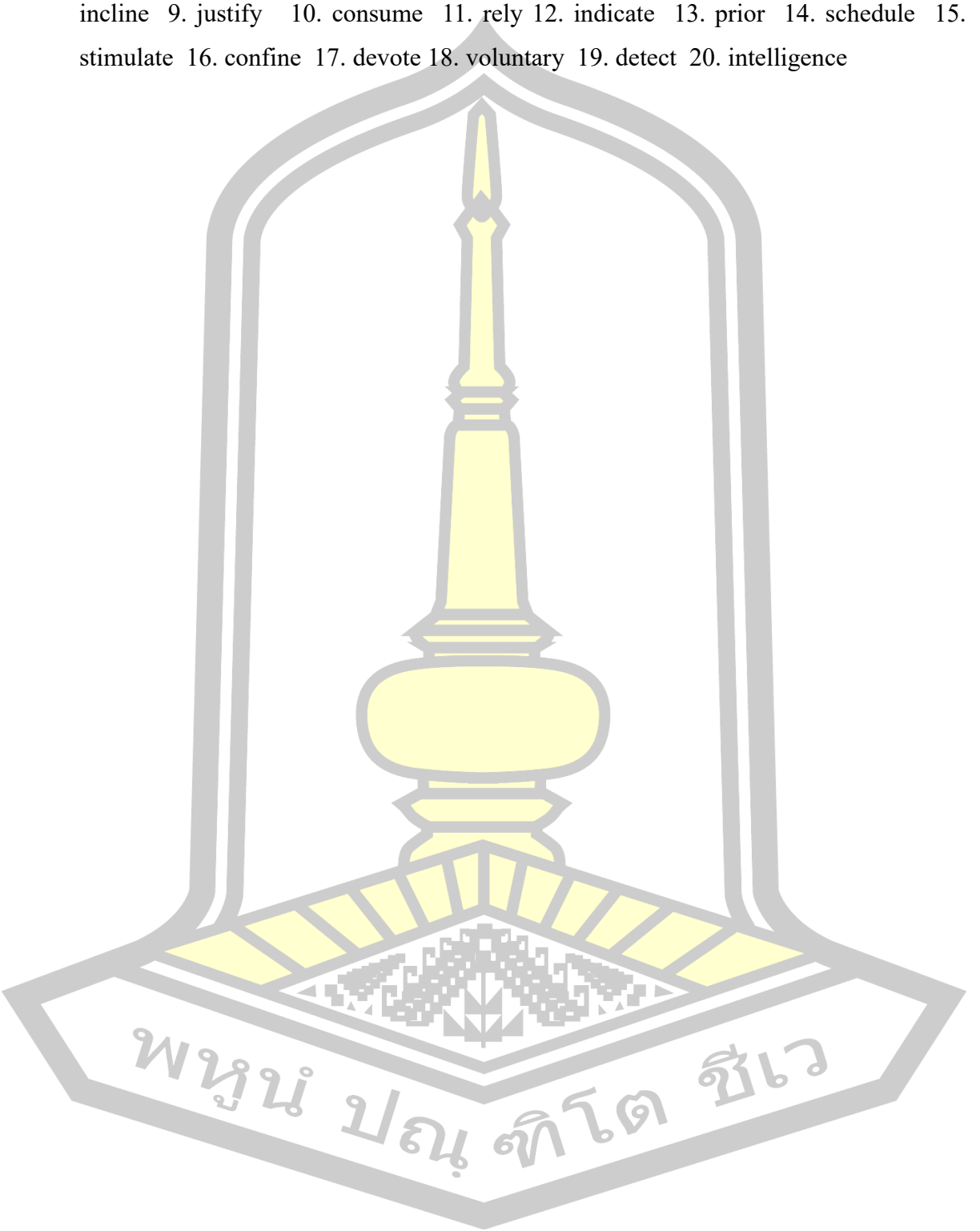
Complete the word according to the Chinese translation. An example has been done for you.

请将以下中文词汇翻译成对应的英文单词，英文单词的第一个字母已给出，请按照示例完成。

Example 检查 --- inspect

1. 扩张/增长 --- e_____
2. 方法/途径 --- a_____
3. 持续存在 --- p_____
4. 动机 --- m_____
5. 获得/完成 --- a_____
6. (进入某地的) 方法/权利 --- a_____
7. 暴露/揭穿 --- e_____
8. 倾向于 --- i_____
9. 证明…合理 --- j_____
10. 消耗 --- c_____
11. 依赖 --- r_____
12. 显示 --- i_____
13. 在前面的 --- p_____
14. 进度表/安排 --- s_____
15. 刺激/激励 --- s_____
16. 限制/禁闭 --- c_____
17. 献身于 --- d_____
18. 自愿的 --- v_____
19. 察觉到 --- d_____
20. 智力/聪明 --- i_____

Keys: 1. expand 2. approach 3. persist 4. motive 5. Achieve 6. access 7. expose 8. incline 9. justify 10. consume 11. rely 12. indicate 13. prior 14. schedule 15. stimulate 16. confine 17. devote 18. voluntary 19. detect 20. intelligence



Appendix I-c The Productive Derivative Test

Instructions

Fill in the blanks with the appropriate forms of the prompt words. Please note the context of the sentence and the grammar of the word.

请按示例将目标单词转变成恰当的形式填入下列句子中，请注意单词的语境和语法。

Example

determine
Noun: He fought the illness with courage and determination.
Adjective: He was determined that the same mistakes would not be repeated.

1. achieve

Noun: It was a remarkable _____ for such a young player.

Adjective: It is often a good idea to start with smaller, easily_____ goals.

2. approach

Adjective: Despite being a big star, she is very _____.

Verb: With winter _____, many animals are storing food.

3. expand

Noun: The rapid _____ of cities can cause social and economic problems.

Adjective: She opened her arms wide in an_____ gesture of welcome.

4. consume

Noun: The meat was declared unfit for human _____.

Adjective: _____ goods are products that are intended to be used and replaced.

5. incline

Adjective: It was Sunday morning, and she was not _____ to get up yet.

Noun: Teachers simply do not have the time or the _____ to investigate these matters.

6. motive

Verb: A good teacher has to be able to _____ her students.

Adjective: The key to a successful modern economy is a well-educated and _____ workforce.

7. prior

Noun: After several accidents in the area, security is now a high _____.

Verb: You need to _____ your tasks; that is time management.

8. justify

Noun: These children have a strong sense of _____.

Adjective: The experiment was stopped because it was not morally _____.

9. rely

Noun: This learning method encourages too much _____ on the teacher.

Adjective: Most companies are_____ on computer technology to work everyday.

10. detect

Noun: She hired a _____ to find out if her husband was having an affair.

Adjective: The noise is barely _____ by the human ear.

11. schedule

Adjective: Meetings are _____ to take place all over the country.

Noun: We are working to a tight _____.

12. confine

Noun: They were held in _____for three years.

Adjective: It wasn't easy to sleep in such a _____ place.

13. indicate

Noun: He gave no _____at all of his own feelings.

Adjective: Their failure to act is _____ of their lack of interest.

14. stimulate

Noun: Teaching and learning styles should vary as widely as possible in order to provide maximum _____ for both teachers and pupils.

Adjective: The _____ effects of coffee and tea can make it hard to fall asleep.

15. persist

Adjective: If she hadn't been so _____, she might not have gotten the job.

Noun: His _____ paid off when he was offered the job of manager.

16. access

Verb: With this new software, I finally _____ my computer remotely.

Adjective: Computers should be made readily _____ to teachers and pupils.

17. expose

Noun: Long-time _____ to the sun can cause skin cancer.

Adjective: She was left feeling _____ and helpless after her father died.

18. intelligence

Adjective: His lecture was readily_____ to all the students.

Adjective: He was once described as the most _____ woman in America.

19. voluntary

Verb: Helen _____ to have Thanksgiving at her house this year.

Noun: I need some _____ to help with the washing-up.

20. devote

Noun: Alan has always shown intense _____ to her children.

Adjective: After the car accident, Isabella was _____ to her brother for the rest of her life.

Keys: 1. achievement, achievable 2. approachable, approaching 3. expansion, expansive 4. consumption, consumable 5. inclined, inclination 6. motivate, motivation 7. priority, prioritize 8. justice, justified/justifiable 9. reliance, reliant 10. detective, detectable 11. scheduled, schedule 12. confinement, confined 13. indication, indicative 14. stimulation, stimulating 15. persistent, persistence 16. accessed, accessible 17. exposure, exposed 18. intelligible, intelligent 19. volunteered, volunteers 20. devotion, devoted

Appendix I-d The Association Test

Instructions

Tick off words in the below box with the similar meanings as the underlined word in the sentence. There may be 2-3 keys in the box. **DO NOT** choose more than 3 words.

请在下面方框中勾选出与句子中划线单词意思相近的词，每个方框中最少有两个最多有三个同义词，所以请不要勾选超过三个词，否则该题为零分。

Example

Some stains are difficult to <u>remove</u> with ordinary washing powder.

delete	resolve	revise	erase	retain
√		√		

We <u>outlined</u> our proposal to the committee.

generalize	elaborate	describe	depict	discuss
√	√	√		

1. The water froze inside the pipe, causing it to <u>expand</u> and burst.

enlarge	swell	contract	dilate	explode

2. As a writer she takes a completely new <u>approach</u> to the classic horror story.

way	means	method	insight	request

3. How can they <u>justify</u> paying such huge salaries?

reason	warrant	confirm	fend for	accept

4. The report <u>exposes</u> the weakness of modern medical practice.

disclose	account	reveal	outline	uncover

5. All the teachers did the work on a voluntary basis.

compulsory	conscious	optional	reluctant	self-imposed

6. The police believe the motive for this murder was jealousy.

reason	cause	mode	motion	excuse

7. Changes may not be made without the prior approval of the council.

extra	preceding	official	positive	advanced

8. The majority of holiday flights depart and arrive on schedule.

plot	timetable	scheme	agenda	device

9. I wish the speaker would confine himself to the subject.

confide	limit	shrink	restrict	prevent

10. Few people are able to devote all of his time fully to their career.

employ	dedicate	devour	deplete	spend

11. He achieved career success after years of hard work.

arrive at	accomplish	attain	promote	win

12. Her interest in art was stimulated by her father.

inspire	stifle	encourage	arose	irritate

13. The electricity industry consumes large amounts of fossil fuel.

reserve	waste	use	expend	eat up

14. They have to rely on the river for their water.

look for	depend	aspire	hinge	surround

15. Research indicates that over 81% of teachers are dissatisfied with their salary.

maintain	suggest	insist	signal	reveal

16. Access to the papers is restricted to senior management.

right	source	inclusion	privilege	permission

17. The accident inclined him to reconsider his career.

influence	convince	affect	tend	sway

18. When the water pipe burst, she had the intelligence to turn off the water at the main.

ingenuity	intellect	innovation	wisdom	courage

19. I thought I detected a hint of irony in her words..

realize	notice	detest	discern	observe

20. If the pain persists, you must see a doctor.

continue	last	cling	adhere	go on

Keys: 1. enlarge swell dilate 2. way means method 3. warrant fend for 4. disclose reveal uncover 5. optional self-imposed 6. reason cause excuse 7. preceding advanced 8. timetable agenda 9. limit restrict 10. employ dedicate spend 11. accomplish attain win 12. inspire encourage arouse 13. use expend eat up 14. depend hinge 15. suggest signal reveal 16. right privilege permission 17. influence affect sway 18. intellect wisdom 19. notice discern observe 20. continue last go on

Appendix I-e The Collocation Production Test

Instructions

Complete the sentences with an appropriate collocation (习惯搭配). Collocation means phrases in which the word given always appear with other words in sentences. The first letters of the words (including prepositions) you have to use to make the collocation has been provided. These target collocations you have to complete include different types: noun+preposition, adjective+noun, verb+noun, among other combinations. The Chinese sentence has prompted the target collocation.

请用习惯搭配填入下列句子。习惯搭配是指所给单词经常和其它词以词组的形式出现在句子中。词组中的单词（包括介词）的首字母已给出。需要填出的词汇搭配包括不同类型如名词+介词搭配，形容词+名词搭配，动词+名词搭配等等。中文语境信息已对目标词组做了提示，请注意单词语法。

Examples

上班通勤对很多人来说都很头疼，尤其是在早晚高峰期。

Many commuters have to bear the congestion during peak hours.

她最近需要参加朋友的婚礼，可是去年买的裙子小了。

She has determined to go on a diet to lose some weight.

1. 读写结合模式可以说是一种新的实践性很强的语言教学途径。

Many teachers believe that this can be a new, practical a_____ t_____ language teaching .

2. 大学生做一些义工有助于丰富他们的人生。

She did different kinds of v_____ w_____ during her university life.

3. 有些孩子虽然天资聪颖，但仍需后天的努力才能成功。

John showed h_____ i_____, but ended up a loser in his career.

4. 她将个人的时间精力全部倾注在自己的事业中，投身于医学的发展。

She d_____ herself t_____ her growing cause of medical treatment.

5. 对大学生来说，书本上的东西远远不够，他们需要更多实践来开阔眼界。

College students have to experience more to e_____ their h_____.

6. 学校可以组织学生参加更多课外活动来激发学生的学习兴趣。

The exhibition has s_____ i_____ in her study of History.

7. 不要局限自己，因为很多时候我们并不了解自己的潜力。

Owen did not c_____ himself t_____ writing only one type of poem.

8. 他经常有些不着边际的想法，但是在这个问题上我更倾向于他的观点。

I i_____ t_____ the view that we should take no action at this stage.

9. 日常生活中对计算机的过度依赖使我们无法应对突发状况。

Nowadays, we should not r_____ h_____ o_____ computers to do all the work.

10. 有的人比较敏感，能够轻易察觉出别人语气中的情绪。

Do I d_____a n_____ of irony in your voice?

11. 在这样的法治社会，所有的罪犯都应该被绳之以法。

The police promised to b_____ j_____ t_____ all the criminals.

12. 农村的孩子很少有机会接触到古典音乐。

Some children in rural areas are rarely e_____ t_____ classical music.

13. 对于这个工作岗位，学过相关课程的会给予优先考虑。

People who have finished the course will be g_____ p_____ o_____those who have not.

14. 他新接手的这项工作十分复杂，需要耗费大量时间。

The new project was complex and t_____-c_____.

15. 学生成绩的上升不仅仅需要自身努力，还需要老师家长鼓励，以使他们感到动力十足。

Teachers and parents should always encourage students to make them f_____ m_____.

16. 教师是一份崇高的职业，因为它可以给人带来一种成就感。

Many students wish to become a teacher because they believe being a teacher can give them a s_____of a_____.

The study i_____the p_____ of a link between poverty and crime.

17. 尽管她遇到巨大困难，但是她始终坚持努力读书。

She p_____ w_____ her studies in spite of financial problems.

18. 如今，很多人认为获得更好的教育资源是成功的第一步。

Many parents believe that their child must h_____ a_____ t_____ good resources to be admitted to prestigious universities.

19. 由于工期太紧，工人们加班加点终于提前完成任务。

The new bridge has been finished two years a_____ of s_____.

Keys: 1. approach to 2. voluntary work 3. high intelligence 4. devoted to 5. expand horizon 6. stimulate interest 7. confine to 8. incline to/ towards 9. rely heavily on 10. detect note 11. bring justice to 12. exposed to 13. given priority over 14. time-consuming 15. feel motivated 16. sense achievement 17. indicates possibility 18. persisted with 19. have access to 20. ahead schedule

Appendix I-f L2 Writing Test

Instructions: Please write an argumentative essay on the following questions. You are required to use at least **5** words (the more the better) randomly selected from the below box. Please integrate the words you choose in your essay in the most natural manner cohesively and grammatically. You can use any derived form of the word (i.e., verbs to adjectives) you choose and mark them in your essay with a circle or underline. You should write 250-300 words within one hour (60 minutes).

(请按照下列要求写作议论文：下面方框中有十个单词，请随机挑选至少五个单词（越多越好）用在你的作文中。请将所选的单词自然融入写作中，语法正确，语篇连贯，单词的任何变形（动词变为形容词）都可以使用，请用圆圈或下划线标注你所选用的单词。作文长度需控制在250-300单词。)

Writing task one: Competition for places at university is increasing. Why do more and more people want to study at university? Is this a positive or negative development?

indicate	achieve	justify	prior	exposed
intelligence	stimulate	detect	devoted	persist

Writing task two: Today, millions of university students have to enroll in online learning for higher education. Colleges and universities offer e-learning programs and courses. Do you agree or disagree with the popularity of online learning?

approach	consume	rely	access	expand
motive	voluntary	schedule	confine	incline

BIOGRAPHY

NAME	Mr. Min Changyong
DATE OF BIRTH	February 7, 1989
PLACE OF BIRTH	China
ADDRESS	No. 339, Taiyu Road, Xiaodian District, Taiyuan city, China
POSITION	An English teacher
PLACE OF WORK	Shanxi Normal university, China
EDUCATION	2013 Bachelor of Arts in Southwest Forestry University in China 2016 Master of Liberal Arts in Business English Studies, Guangdong University of Foreign Studies 2024 Doctor of Philosophy in English Language Teaching (Ph.D.), Mahasarakham University