BREADTH AND DEPTH OF ESL LEARNERS' LEXICAL KNOWLEDGE; EXPLORING ITS INTERPLAY WITH WRITTEN LANGUAGE PROFICIENCY.

Thesis submitted for the degree of Doctor of Philosophy

at the University of Leicester

by

Sihle Ndlovu

Department of Education

University of Leicester

2017

Abstract

Breadth and depth of ESL learners' lexical knowledge; exploring its interplay with written language proficiency.

Sihle Ndlovu

This thesis adopts a mixed methods approach to investigate the interplay between lexical knowledge and written language proficiency among learners of English as a Second Language (ESL). To achieve its objectives, the study examines how written language ability relates to a battery of size and depth lexical measures. The Word Associates Test (WAT), the Vocabulary Levels Test (VLT), the Vocabulary Profile (VocabProfile) tool and written compositions were used to produce quantitative data on the interplay between learners' breadth and depth of vocabulary knowledge on the one hand, and writing proficiency on the other hand. Purposive sampling was used to identify ESL participants in order to ensure that data generated would be capable of producing relevant insights to address research questions. Stratified random sampling was used to select 40 written language samples from the International Corpus of Learner English (ICLE) to ensure topic consistency with 18 essays from ESL students. This allowed for a comparative analysis between lower proficiency (ESL) and higher proficiency (ICLE) students. Following written assessments, the study employed the stimulated reconstruction procedure to obtain emic perspectives on the rationale behind the lexical choices that ESL learners made during the WAT. Quantitative findings obtained highlight aspects of both size and depth of lexical knowledge as important factors in the interplay between vocabulary knowledge and written language skills. Qualitative findings highlight the potential for multiple factors that could affect individual learners' trajectories. Taken together, the findings from quantitative and qualitative data deepen lexical insights by highlighting the complex interplay between lexis and writing. The study draws on the Dynamic Systems Theory (DST) as a lens for interpreting and reconciling these findings. To that effect, it offers methodological contributions by highlighting the relevance of DST to ESL developmental processes, which is a relatively new theory in the field of Applied Linguistics.

Acknowledgements

I feel very fortunate that during the arduous journey towards the award of my PhD, I had the support of numerous people whose generosity with their time and support was beyond imagination. I would like to sincerely thank my supervisor, Dr Agneta Svalberg and my second supervisor, Dr Pamela Rogerson-Revell. They both showed steadfast support, enduring patience and unquestionable belief in me even during times when I felt I had completely lost the cause. I cannot imagine how this project could have been a success without their insightful guidance and encouragement. I am also grateful to Professors Emma Smith and Glenn Fulcher for their insightful comments and encouragement particularly with statistical elements of my thesis. I must also extend a special word of appreciation to all tutor and student participants for their commitment and perseverance through the numerous assessment activities. Without their dedicated support, this thesis would have remained a dream.

I do not take for granted, the support of family, friends and colleagues who stood by me throughout the PhD journey. I owe you all a big thank you for your support and encouragement in numerous ways. To my beloved husband Tabani, I can only say, you are truly my rock; to my son Thando, you are a source of pride and together, the two of you are my unwavering pillars of inspiration. To you both, I warmly dedicate this lifetime achievement.

To my dad and late mum, I am indebted to you for planting the seed at an early age; and to my siblings and extended family, thank you for being such a big part of this journey all through the way. Last, but not least, I want to acknowledge this achievement as a true gift and honour from God, for which, I am sincerely grateful and eternally humbled.

Table of contents

List of Tables

List of Figures

Acronyms and abbreviations

AWL	Academic Word List
C&O	Cohesion and Organisation
CEFR	Common European Framework of Reference for Languages
CoP	Conventions of Presentation
DST	Dynamic Systems Theory
EAP	English for Academic Purposes
ELT	English Language Teaching
ESL	English as a Second Language
ESOL	English for Speakers of Other Languages
FonF	Focus on Forms
GC	Grammatical Control
ICLE	International Corpus of Learner English
IELTS	International English Language Testing System
K2	The first 2,000 most frequent words
K3	The first 3,000 most frequent words
K5	The first 5,000 most frequent words
L1	First language
L2	Second language

LFP	Lexical Frequency Profile
MMR	Mixed Methods Research
PET	Preliminary English Test
SPSS	Statistical Package for the Social Sciences
TESL	Teaching English as a Second Language
TOEFL	Test of English as a Foreign Language
VKS	Vocabulary Knowledge Scale
VLT	Vocabulary Levels Test
VLTAWL	The academic word level of the Vocabulary Levels Test
VLTK2	Vocabulary Levels Test at the 2000-word level band
VLTK3	Vocabulary Levels Test at the 3000-word level band
VLTK5	Vocabulary Levels Test at the 5000-word level band
VocabProfile(s)	Vocabulary Profile(s)
VocabprofileK1	The first 1,000 words on the Vocabprofile tool
Vocabprofile K2	The first 2,000 words on the Vocabprofile tool
WAT(s)	Word Association Test(s)

Chapter One:

Introduction and study background

1.0 Introduction and background to the study

The English language has been entrenched as the global language because of its pivotal role in the world of international business, science and technology as well as in research communities (Nunan, 2001). Within the United Kingdom, where English is the language of instruction, both native and non-native students need to be proficient users of the English language if they are to be able to effectively access relevant curricula related to their areas of study. This is particularly important because students need to read academic texts and appropriately respond in written form (essays, dissertations, theses or other assessment forms) in order to meet tertiary and higher education assessment standards. As such, English language proficiency is essential not only for academic achievement but also, for enhancing students' abilities to function effectively and get ahead in the global village where it is likely to enhance their competitive edge.

For the English as a Second Language (ESL) learner, this situation presents both opportunities and challenges. On the one hand, success in learning ESL opens up numerous doors as it can be a conduit for participation in the world of international business and commerce. On the other hand, the process of learning English can present a wide range of challenges, some of which reside at the lexical and grammatical levels both in written and spoken forms of the language.

As Meara (1996) observes, vocabulary skills are a major component of second language proficiency. To this end, various studies have explored the interaction between vocabulary knowledge and other aspects of linguistic competence such as reading and speaking (Hu

and Nation, 2000; Nation, 2006; Zeeland and Schmitt, 2013). The current study builds on previous studies that have explored the interaction between vocabulary knowledge and other aspects of linguistic competence such as reading, speaking and writing (see for example Hu and Nation, 2000; Nation, 2006; Zeeland and Schmitt, 2013; Crossley and McNamara, 2011).

1.1 Study rationale

The rationale for this study is based on factors operating at two different but related levels. Firstly, the research was inspired by my personal interest in lexis as a result of my professional experiences as an English teacher, particularly working with learners taking ESL. Secondly, my review of the research literature revealed the wider relevance of researching lexis, as Chapter 2 of this thesis will show.

1.1.1 Rationale based on professional experience

My interest in researching lexis was first ignited by the experiences I had teaching English Language and Literature at different levels of proficiency including Cambridge Ordinary and Advanced Levels, pre- and in-sessional English for Academic Purposes (EAP) as well as English Literacy for ESL learners. An important step at the beginning of most programmes was the completion of entrance/placement tests. In the majority of cases, the tendency was for vocabulary to be tested as a stand-alone component. In such cases, most tests tended to be cloze procedures where students had to supply a missing word in context. For grammar, the students were predominantly presented with ungrammatical structures and were required to correct the errors which involved key aspects of the language such as tense, punctuation, subject-verb agreement, clause and sentence structure.

What I noticed during my involvement with ESL programmes, both at an operational and course management level in the UK and abroad was that there were some learners whose linguistic profiles as reflected by their vocabulary scores did not predict written language ability. For the majority of the learners, high scores in entrance tests tended to be associated with relatively well-developed lexicogrammatical profiles and written language ability. However, this was not always the case. This caused concern among teachers who felt that such students were not performing as well as they could do. Speculations around individual differences such as motivation levels and overall learners' engagement levels became inevitable. There were also speculations among staff about whether the noted discrepancies might have been related to the extent to which the in-house entrance tests were able to generate accurate learners' vocabulary profiles. However, because for the majority of the students, higher initial test scores did predict higher scores in free writing and general performance on the course, I envisaged that the noted discrepancies could suggest that the relationship between vocabulary knowledge and written language ability may not be a straightforward one, and therefore required further investigation.

I was particularly interested in investigating writing as opposed to other productive skills such as speaking because in tertiary education institutions such as colleges of Further and Higher Education, learners are predominantly assessed on the basis of the written form. With this in mind, I singled out the investigation of lexis and writing as a worthwhile endeavour. This culminated in an interest in vocabulary studies at Masters Level and it was at this stage that my curiosity began to find substantiation in literature.

1.1.2 Relevance and value of researching lexis

Empirical evidence suggests that mistakes in lexical selection are likely to be less tolerated than mistakes in syntax (Sheorey, 1986; Carter, 1998). As an example, in a large-scale study involving 178 academic staff in an American university, Santos (1988) investigated the reactions of the participants to two 400-word compositions written by two non-native students. Academics were required to rate the compositions based on six 10-point scales, three of which were focused on content while the other three were focussed on language. They were also asked to identify the errors which they considered to be the most serious in terms of their impact on comprehensibility of text, their level of acceptability in an academic context, as well as the level of irritation caused by the error. Five language errors emerged as the most salient and these were subject verb agreement, use of the passive verb, lexical choice errors, article and pronoun errors. Out of these, lexical errors were considered to be the most serious and had the greatest impact on overall ratings of essays. This led Santos (1988) to conclude that this type of error 'impinges directly on content' because 'when the wrong word is used, the meaning is very likely to be obscured' (Santos, 1988 p.84). As such, scholars such as Sheorey (1986) and Carter (1998) have argued that mistakes in lexical selection are likely to be less tolerated than mistakes in syntax.

In light of the role of lexis in linguistic development, the current study seeks to shed light on the breadth and depth dimensions of vocabulary knowledge and how these interact with written language ability among ESL learners considering the key role of vocabulary knowledge in language contexts. In light of possible complexities

related to this interplay (see Section 1.1.1 above), the Dynamic Systems Theory (DST) was considered the most comprehensive theoretical framework for the study. According to De Bot et al. (2007 p.8):

> *DST …is originally about very simple systems such as the two coupled variables in a double pendulum. Even though such a system has only two interacting variables or degrees of freedom, the trajectory of the system is complex. When applied to a system that is by definition complex, such as a society or a human being, where innumerable variables may have degrees of freedom, DST becomes the science of complex systems.*

This suggests that as Dornyei et al. (2015) note, DST is able to take into account, the manifold issues and factors that interrelate in the process of second language development. Therefore because of the complex nature of second language processes in general, and vocabulary knowledge in particular, DST was considered to be an approach capable of offering valuable insights about the interplay between vocabulary knowledge and written language ability. Section 2.8.1 offers a more comprehensive discussion of DST in the context of the current study.

1.2 Significance of the study

Meara (1980) described vocabulary as an area of Applied Linguistics which was neglected in favour of grammar. Nevertheless, current literature suggests that vocabulary is increasingly being recognised as a key component of successful language learning (see for example Laufer et al., 2004; Daller et al., 2007). This shift of focus has culminated in extensive research on lexical issues, with researchers focusing on a wide range of themes including the role of vocabulary

knowledge in both first and second language learning (Grabe, 1991), issues around measurement of vocabulary knowledge (Chapelle, 1994; Laufer and Nation, 1995; Read, 2000) as well as the conceptualisation of vocabulary knowledge in terms of the size and depth dimensions (Schmitt, 2014). As will be further explored in the Literature Review chapter, a wide body of current research acknowledges that vocabulary knowledge can be understood in terms of the breadth (size) and depth (quality) dimensions (Laufer and Goldstein 2004; Melka 1997; Nassaji et al., 2010; Schmitt 2014), although alternative views about the value of distinguishing between the two exist (Vermeer, 2001). It would therefore seem reasonable that comprehensive approaches to vocabulary research should address both depth and breadth issues as Schmitt (2010) advises. Such research should seek to establish not only the interface between the vocabulary size and depth dimensions, but also, the refinement of conceptualisations of vocabulary knowledge in order to facilitate deeper understanding of this construct. To that effect, the current study utilised correlation analysis as the foundation to the exploration of the interplay between learners' vocabulary size, depth and written language skills. In line with DST, the study extended its findings by tracing individual ESL learners' performances across all the three written assessments which were administered to the ESL participants. This allowed for more in-depth analysis of individual profiles. Finally, through stimulated reconstructions, (Svalberg and Askham, 2016) the study elicited learners' thought processes which led to particular lexical choices during the Word Association Test (WAT).

A WAT is a test whereby participants are presented with a series of words. Depending on whether it is a receptive WAT or a productive

WAT (see Section 2.3.2 for a comprehensive discussion of WAT including the receptive and productive versions) the participant is required to either identify words that are related/associated with the key word, or provide the first word(s) that come(s) to mind when presented with the key word. The current study utilised the receptive WAT format, the rationale for which is provided in Section 3.4.4.

The current study also elicited students' attitudes towards the size and depth conceptualisations in order to gain a better understanding of vocabulary knowledge from the students' perspectives. While this study does not contest the value of judgements made by raters / assessors and tests such as the VLT, the triangulated methodology adopted adds a valuable dimension to vocabulary studies. It does this by drawing from both emic and etic perspectives in order to generate more comprehensive insights about the dynamics between vocabulary knowledge and learners' written language productions. As Morris, et al. (1999, p.782) note, "emic accounts describe thoughts and actions primarily in terms of the actors' self-understanding terms that are often culturally and historically bound". In contrast, "etic models describe phenomena in constructs that apply across cultures" (Morris, et al., 1999, p.782). As such, emic accounts give primacy to the participants' views about phenomena under investigation, while etic accounts typically focus on causal models, which can be generalised and applied to other contexts. In the context of the current study, emic perspectives derive from stimulated recalls / semi-structured interviews which were conducted on the basis of the WAT answers that participants provided. Etic perspectives derive from the quantitative analysis of scores from written assessments which provided correlational data between VLT, WAT, VocabProfiles and free

writing scores. The emic and etic perspectives are integrated in order to obtain deeper insights about the interplay between learners' vocabulary knowledge and their written language skills (Morris, et al., 1999).

It was important to interview the learners in order to understand, first of all, their perception of the vocabulary size and depth dimensions of vocabulary knowledge since this is an important yet contested area (See Section 2.2.8). Findings thereof are key to the understanding of learners' perceptions of what it means to know a word. Since beliefs shape behaviours, (Bracken, 2010) it is important that learners' conceptualisations of vocabulary knowledge are well aligned with the most up-to-date frameworks for understanding this concept. Findings from this study will therefore have implications for the teaching and learning of vocabulary.

1.3 Aims and objectives of the study

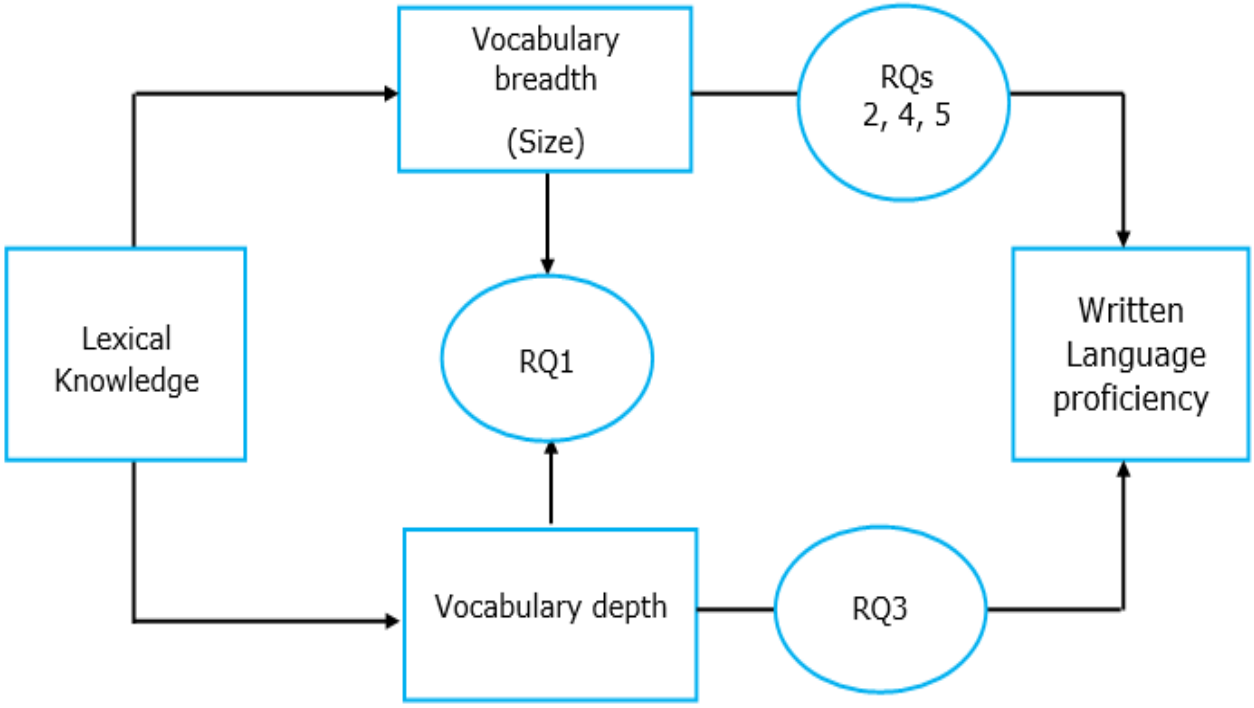
The aim of the study is to investigate the interplay between L2 learners' breadth and depth of vocabulary knowledge and their written language abilities. The study utilises data obtained from two groups of learners namely the ESL group which is the source of primary data and the ICLE group which is the source of secondary data. The ESL participants are all college students working at a lower proficiency level compared to the ICLE students who are university students and operating at a higher proficiency level (see Section 3.3.5 for a comprehensive description of the participants). Through the mixed methods approach adopted in the study, both quantitative and qualitative data collection and analysis methods are employed in order to triangulate correlational findings with interpretive insights, culminating in in-depth insights of lexis and writing for the learners

investigated in the study. The research questions that the study seeks to address are as follows:

1. RQ1 - What is the relationship between learners' vocabulary size (breadth) and the quality of their vocabulary knowledge (depth)?
2. RQ2 - What is the relationship between learners' vocabulary size and the quality of their written compositions?
3. RQ3 - What is the relationship between the quality of learners' vocabulary knowledge (depth) and the quality of their written compositions?
4. RQ4 - Is there a relationship between the learners' lexical profiles produced by VocabProfile, and the quality of their written compositions?
5. RQ5 - Is there a difference between the VocabProfiles of learners working at a lower proficiency level (ESL learners) and those at a higher proficiency level (ICLE students)?

Before addressing the above research questions, it is important to provide an overview of how the key concepts highlighted are inter-linked in the study. Therefore, Figure 1.1 provides an outline of the key research areas that underpin the current study and illustrates how these interlink and ultimately contribute towards the formulation of the research questions outlined above. The key research areas will be further discussed in the Literature Review chapter which explores how the concepts have been operationalised in the literature.

Figure 1.1 Overview of the areas that research questions relate to in the study



As Figure 1.1 suggests, the main agenda of this study is to generate empirical evidence which will provide insights on the nature of the interplay among vocabulary size, depth and written language ability.

1.4 Overview and structure of the thesis

This thesis comprises eight chapters. The first chapter provides an introduction to the study while Chapter 2 develops the theoretical background through reviewing literature relevant to the main concepts that constitute the study.

Chapter 3 presents methodological issues including the research paradigm, data collection/ data analysis procedures as well as ethical

and practical considerations. Ethical considerations are first presented from a general perspective before going on to provide a specific framework relevant to this particular study. Chapter 4 presents the findings from a pilot study carried out to ascertain the feasibility of this study.

This is followed by Chapter 5 which presents findings from written assessments. The chapter presents findings obtained through quantitative analyses while Chapter 6 concentrates on reporting findings from stimulated reconstructions / semi-structured interviews. Chapter 7 provides an interpretation of the findings by making linkages between the findings presented in Chapters 5 and 6 to relevant theoretical perspectives as well as previous literature in the field.

The thesis finishes by presenting a conclusion which brings to the fore the pedagogical and theoretical implications together with the study's contribution to knowledge. Therefore, the final chapter highlights the most significant insights obtained from the current research. To bring the thesis to a close, the final chapter also identifies the study's limitations which pave the way for the identification of possible avenues for further research in lexical studies.

Chapter Two:

Literature Review

2.0 Introduction to the literature review

The concept of vocabulary knowledge is at the core of this study. As such, this chapter starts with a discussion of what a *word* is and consequently, how word knowledge has been conceptualised. This is followed by an exploration of major concepts relevant to the study as dictated by the research questions. Based on the research questions, the concepts reviewed mainly relate to the following areas which are covered mainly within the sections identified below:

a) the nature of the vocabulary knowledge construct (Sections 2.1.1 – 2.2.3 and 2.2.7-2.2.9),

b) the assessment / measurement of vocabulary knowledge (Sections 2.2.4-2.2.6 and 2.3.1-2.3.2),

c) frameworks of vocabulary knowledge and measures of lexical richness (Sections 2.4-2.6.5),

d) pedagogical considerations for L2 lexis and writing (Sections 2.7-2.7.5), and finally,

e) theoretical underpinnings for the study (Sections 2.8 – 2.8.3)

2.1 Definition of terms

2.1.1 What is a word?

The overview presented in Figure 1.1 underscores the fact that word knowledge is at the core of this study, which makes the question of what a *word* means quite pertinent. At first glance, the concept of a *word* may appear to be a very simple and straight forward one. It brings to mind many examples of linguistic items that can fall into this category, for example, nouns and verbs such as 'child', 'come', 'sleep', 'boy' and 'girl'. Such conceptualisations consider a word as a unit of meaning in written form, separated by empty spaces in between (Cobb, 2013). However, as will be highlighted in the

following sections in this chapter, in an Applied Linguistics context, the concept of a word is not always easy to define (Milton and Fitzpatrick, 2014; Read, 2000). This owes to the fact that to a very large extent, the way that the concept is construed depends on the purposes for which the definition is intended (Gardner, 2007). Therefore, a word can be conceptualised as a type, a token, a single word or a multi-word. In addition, 'distinguishing word forms and word families is also important' (Read, 2000:19). This is because the way in which a researcher conceptualises and operationalises key constructs has implications for the validity of results obtained (Gardner, 2007). The following sections therefore provide an outline of the main conceptualisations of a word and identify how the concept of a *word* is applied to the current study.

2.1.2 Tokens and types

Tokens are defined as 'the number of running words in a text, while types are the number of different words' (Schmitt, 2010:188). Thus, a type-token analysis of the sentence 'The term *word* is difficult to define because it is used to describe different lexical items', would produce two tokens of each of the types 'is' and 'to' respectively and one token of each of the types 'the', 'term', 'word', 'difficult', 'define', 'because', 'it', 'used', 'describe', 'different', 'lexical' and 'items'. The sentence therefore has sixteen tokens but only fourteen types. This distinction is relevant to lexical studies because inappropriate decisions in terms of the treatment of these may lead to distorted inferences about a learner's vocabulary profile. For example, two learners may produce the same number of tokens in a text but one may produce many tokens of very few types while the other may produce many tokens of an extensive number of types, thus

suggesting differential vocabulary profiles. In the current study, Vocabprofile, a computer-based program is used to analyse the extent to which ESL students in a college and university context use vocabulary from lower and higher frequency bands. This provides an indication of how the learners' vocabulary knowledge is reflected in actual language use for this group of learners (Laufer and Nation, 1995). The programme provides output on the number of types and tokens that a learner utilises from each frequency band. It is the number of types that a learner uses which provides an estimate of the learner's ability to use vocabulary from a particular frequency band. The token figure would over-estimate the learner's vocabulary in each frequency band because repeated vocabulary would be counted as individual words in line with the definition of tokens as running words in a text (Schmitt, 2010).

2.1.3 Lemmas and word families

In vocabulary studies, the term *lemma* refers collectively to the base and inflected forms of a word (Read, 2000). Thus, the verb *visit* with its inflected forms *visits* and *visited* would be classified as one lemma because they belong to the same word class even though they are different forms of the same stem.

Word families are broader than lemmas because over and above the lemma relations described above, word families include derived forms which can change the class of the base word and alter its meaning (Schmitt, 2010). Thus, in addition to the inflected forms *visited* and *visits*, the verb *visit* also has the derivative noun *visitor* within its word family while the verb *dictate* would have the noun *dictator* and *dictation* included within its word family, together with the inflected form *dictates* and *dictated*.

In a study to examine native and non-native students' abilities to produce derivatives for given prompt words, Schmitt and Zimmerman (2002) found that students generally had partial knowledge of the prompt words. On average, participants were able to produce two out of the possible four derivatives for each word family provided. Cases where students did not know any derivatives at all for a particular word were very minimal, and so were cases of positive extremes where students knew all derivatives. This led the researchers to conclude that 'knowing one member of a word family does not imply knowledge of all (or even most) of the other word forms' (Schmitt and Zimmerman, 2002, p.158). This suggests that clearly differentiating words at lemma and word family level can be an important distinction in vocabulary studies depending on the research aims.

2.1.4 Multi-word units

Martinez and Schmitt (2012) note that multi-words or formulaic language constitutes an integral part of the lexicon. A multi-word or formulaic sequence can be defined as "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar." (Wray, 2000, p.9). As such, multi-word units enable language learners to write and speak in a native-like manner. This suggests that these linguistic structures have a major role to play in the development of fluency (Ellis et al., 2008). Notably, 'in normal interaction, the default setting is formulaicity, both for production and comprehension' (Wray and Perkins, 2000, p.19). This allows both the interlocutor and the listener to focus their attention

more on other aspects of the interactional process such as the evaluation of what has been said, predictions of what may come next and repairing any communication breakdowns.

The value of formulaic language in discourse is well-supported by research in corpus linguistics, for example, research carried out by Conklin and Schmitt (2008) highlights the psychological value of formulaic language. The researchers carried out an investigation to find out whether formulaic sequences are processed quicker than non-formulaic language amongst both native and non-native English speakers. Based on findings from 39 participants who took part in a reading comprehension exercise, the researchers found that native speakers processed formulaic sequences faster than creatively-constructed language sequences. This finding is consistent with findings from an earlier study carried out by Underwood, et al (2004).

Underwood et al (2004) tracked the eye movement of 20 native and 20 non-native participants during a reading activity involving a selection of formulaic sequences. Their results revealed less fixations on formulaic language, particularly for native participants. The researchers associated this finding with the lesser degree of familiarity with some of the sequences amongst non-native participants compared to native participants.

Notably though, both investigations above involved the processing of written text which raises the question of whether similar results would be obtained if the information was presented through a different channel such as audio format. Nonetheless, the finding that formulaic language is more efficiently processed than non-formulaic language has implications for language learning. Form a psycholinguistic

perspective, research suggests that formulaic language is processed quicker and as a result, these language forms aid communicative competence (Wray, 2000; Chen and Baker, 2010). Altenberg (1998) found that at least 80% of running words in the London-Lund Corpus were formulaic as they formed part of recurrent word combinations, thus highlighting the prevalence of these structures in language use.

However, as Wray (2002) observes, there may be discrepancies in such estimates depending on how formulaic language is conceptualised. In Altenburg's (1998) study, all strings of words which occurred more than once in identical form were counted on the basis that formulaic sequences defined as word sequences which are stored and retrieved whole from memory rather than decomposed at the point of use (Wray, 2010). On the basis of the above discussion highlighting the value of formulaic language, it can be argued that depending on the research aims, a study which conceptualises a word as a single unit may obtain different results from one that adopts a more comprehensive definition of a word which includes multi-words. A case in point is that of phrasal verbs such as *get along; give up; sit in; take off; move in; drill down*. When decomposed into individual words, other than *drill*, all of these words belong to the first 1000 most frequent words. However, as phrasal verbs, such words are likely to present a much higher cognitive load than decomposed individual words. As an example, Liu and Shaw (2001) investigated the quality of word-knowledge for the high frequency verb *make* amongst native and non-native speakers. The researchers found differences in the configuration of the verb between the native and non-native groups. The results of the study suggested that L2 learners were not able to

exploit the full potential of the word in their written productions particularly in terms of its versatility for formulaic use.

It would appear therefore that the most comprehensive conceptualisation of a word can be expected to include not only single items (types and tokens), but multi-words as well, as part of what is meant by a word. However, for purposes of the current study, the decision was made to adopt the dictionary definition of a word as a single item primarily for two main reasons. To date, no principled methodology exists for the inclusion of formulaic sequences in assessment tools such as vocabulary tests (Martinez and Schmitt, 2012). The VLT and Vocabprofile are no exception to this as they are based on wordlists such as the General Service List (West, 1953) and the Academic Word List (Coxhead, 2000).

Consequently, with the exception of closed form compound words (compound words that have merged into one word) such as *grandmother, footprint, fireworks* and *keyboard*, Vocabprofile processes and categorises multi-words, including compound words, as single lexical items. Therefore, such words are decomposed into individual constituent words during analysis. For example, open compound words (conventionally written with a space in between) such as *native speaker, post office* and *real estate* would be processed by Vocabprofile and classified as different *types* residing at different frequency levels. Hyphenated compound words (e.g *multi-word*) are also processed in the same way, that is, they are decomposed into their constituent parts.

Arguably, decomposition of multi-words is not ideal considering the varied cognitive load presented by such lexical items compared to individual words. However, as is evident from the research questions, the VLT and Vocabprofile tools are at the centre of the methodology for the current study so conceptualisation of words as multi-words would be incompatible with the parameters of word definition imposed by the main assessment tools in this study. In addition, one of the key variables in this study (the writing score) was derived from the evaluation of free writing. To date, assessment scales for free writing, including those for high stake examinations such as IELTS, do not address multi-words distinctly from the overall vocabulary assessment criteria. The result is that to a very large extent, major insights on the conceptualisation of lexis are based on the concept of a word as a single unit. Therefore, for comparability, this study adopts a similar conceptualisation of words as single units. Notwithstanding the above, this study acknowledges this as a weakness because such an approach may be misleading. As an example, in line with Liu and Shaw's (2001) observations, one may know the noun '*speaker*' but this does not mean they would also know the adjective 'native'.

2.2 The construct of lexical knowledge

A review of literature suggests that the term *lexical knowledge* has been used to refer to slightly different constructs by different researchers. The way in which researchers operationalise this construct seems to be highly dependent on the nature of their particular research interests and the objectives that they seek to achieve (Goulden et al., 1990; Henriksen, 1999; Qian, 2002). This

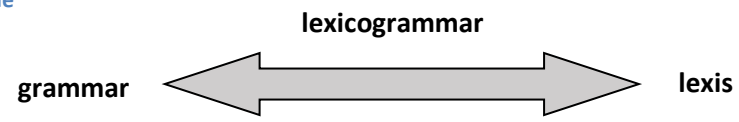
suggests that the concept needs unpacking in the context of this particular study.

Kamil and Hiebert (2005) posited that lexical knowledge refers to the kind of information that students must know about words in order to effectively engage with various language task demands. This definition provides a useful starting point towards an understanding of what word knowledge might entail. However, in the above study, the authors argued that students' perspectives of what they must know about words might differ from what other stakeholders such as teachers and applied linguists might deem as important. As Wesche and Paribakht (1996) observe, from a language learner's perspective, what is likely to be of paramount importance is the knowledge of words to facilitate understanding and communication within a particular context.

For the researcher, however, there are further pertinent issues such as "how much knowledge and which knowledge?" the learner needs in order to effectively carry out different communicative tasks (Wesche and Paribakht, 1996.p26). Therefore, a number of frameworks have been developed in order to facilitate a more comprehensive understanding of what lexical knowledge entails and these are discussed in the next section. Traditionally, the approach taken in second language / foreign language teaching and learning has been one that treats grammar and vocabulary as two disparate entities (Sinclair, 1991). However, as Tucker (1999) notes, such compartmentalisation of language is for the convenience of linguistics rather than a reflection of how languages works. In practice, lexis and grammar are closely interwoven. Indeed, in English language teaching and learning, it is not always easy to distinguish lexical

issues from grammatical issues. For example, if a learner produces the following output: 'The essay emphasises on the role of language in society…', it is evident that although the learner clearly understands the meaning of the lexical item, they have not fully grasped how the lexical item is used in context. This highlights that learning a list of academic vocabulary would not serve this learner well but what they need alongside such lexical knowledge is understanding of grammatical characteristics of the verb including how it is used in context, i.e. without a preposition. This highlights that, language form, meaning, and use should be approached as an integrated whole (Larsen-Freeman, 2003), a view which is aligned to the concept of lexicogrammar. Halliday (1961) first introduced the term l*exicogrammar* into the study of language and argued that "lexis can be defined as *most delicate grammar*" (Halliday,1961, p.259). This suggests that lexis and grammar constitute 'two poles of a single cline or continuum as shown in Figure 2.1 below:

Figure 2.1 The lexicogrammar cline



Adapted from Halliday and Matthiessen (2014, p.64)

The lexicogrammar cline highlights that since the use of a lexical item often has grammatical implications, lexical items are often grammatical in nature (Francis, 1993; Sinclair, 1991). In other words, grammar and lexis are the endpoints of an integrated structure (Tucker,2007), hence Halliday and Matthiessen (2014) describe conjunctions and prepositions as semi-grammatical structures suggesting the overlap between grammar and lexis in text. To that extent, in the current study, learner's written language is assessed in terms of lan-

guage use rather than vocabulary and grammar as separate entities (see Appendix 1). This is in line with Halliday's lexicogrammatical approach, "one in which there is no rigid compartmentalisation of grammar and lexis" (Tucker, 1999, p. vii). To that extent, while the aims of the current study fall within the lexical field, the study acknowledges that vocabulary and grammar are in fact interlinked. This view is aptly summarised by (Zhou, 2009, p.32) who contends that:

> *Linguistic resources include both syntactical knowledge and lexical knowledge, which are interrelated because some grammar issues are inseparable from lexical knowledge. For example, a proposition error appears to be a grammatical problem, yet it can be closely related to a learner's lack of lexical knowledge.*

Clearly then, both grammar and lexis are certainly important parts of second language learning. However, as Cobb (1999) notes, the role and importance of grammar in second language pedagogy has been overvalued for a very long time, often at the expense of other areas such as vocabulary growth.

2.2.1 Receptive vs. productive vocabulary knowledge

A key distinction in the study of vocabulary knowledge is the *receptive* versus *productive* notion, often referred to as *active* versus *passive* vocabulary knowledge (Qian, 1999; Read, 2000; Wesche and Paribakht, 1996). Receptive vocabulary knowledge refers to the ability to comprehend linguistic input presented in spoken and /or written form.

In contrast, productive vocabulary knowledge refers to the ability to produce linguistic output in written and/or spoken form (Nation, 2006). Thus, Laufer and Goldstein (2004, p. 205) observe that 'if two

people 'know' the meaning of a particular lexical item, their knowledge may not be identical' and an important aspect of this differential knowledge is whether a word is known receptively or productively.

It is important to note that, while the receptive versus productive notion is widely used and applied to lexical research, the conceptual distinction between the terms quite often proves to be problematic (Read, 2000). As such, a notable caveat in the literature is that the two terms have not been conceptualised in the same way by different researchers. This has resulted in inconsistent results being obtained from research exploring the relationship between the two.

Of particular interest in the context of the current research is a study carried out by Waring (1997) who compared the receptive and productive vocabulary sizes of a group of second language learners using the VLT (Nation, 1990). The receptive VLT is a matching test in a multiple-choice format with words grouped by frequency so that learners' vocabulary knowledge can be assessed at the 2,000, 3,000, 5,000, and 10,000 academic vocabulary levels. Participants are provided with 6 words per item and three definitions. They are then required to match the definitions with corresponding words (see Section 2.2.5 for a comprehensive discussion of the receptive VLT).

The aim of Waring's (1997) study was to compare the receptive and productive vocabulary profiles of second language learners, with the aim of determining the quantitative difference between the two (i.e. receptive and productive vocabulary knowledge).

To that effect, Waring administered a receptive and a productive test to 76 Japanese ESL students who were each tested for receptive and productive knowledge for each of the words tested from the 1,000, 2,000, 3000 and 5,000 words frequency bands. The study adopted a 'within-subjects' design which means that each participant's own receptive and productive vocabulary profiles were compared.

The main finding was that all the 76 participants obtained higher scores on the receptive than the corresponding productive task, illustrating a larger receptive than productive repertoire for all subjects across all frequency bands. A commonly-cited explanation to this disparity is the view that productive vocabulary knowledge is a lot more sophisticated than receptive knowledge (Caspi and Lowie, 2014) because the learner needs to acquire multiple meaning dimensions of a particular word such as the spelling or pronunciation of a word in order to be able to use it productively (Melka, 1997; Webb, 2008).

Interestingly though, contrary to the findings from Waring's study where results unequivocally indicated a larger receptive than productive vocabulary size for all the learners tested, Webb's (2008) study found some variations. The study had similar objectives (a comparison of receptive and productive knowledge for a particular group of learners). Webb (2008) found that on some occasions, learners scored higher on productive than receptive vocabulary tests. Webb (2008) used a different methodology where translation tests were used to measure receptive and productive vocabulary ability amongst ESL learners. In the test for receptive knowledge, learners were given second language forms and they were required to provide the Japanese (first language) translations. On the productive

knowledge test, the participants were given first language forms (Japanese) and they were required to translate these into second language forms (English). One merit of Webb's methodology (translation tasks for both the receptive and productive assessment of vocabulary knowledge) is that it reduces opportunities for guessing, which is a potential caveat for the matching tests such as the receptive VLT. This is not possible with Webb's receptive and productive translation tasks. However, compared to matching tasks such as those comprising the receptive VLT, a translation task requires deeper word knowledge and is therefore likely to produce different outcomes from those of other tasks. Arguably, the translation task that Webb (2008) regarded as receptive is a production task in the sense that it requires learners to produce a particular form, albeit in the native language. These differences in the conceptualisation of receptive and productive might be responsible for the different results obtained.

However, dating as far back as pioneer studies such as that of Morgan and Oberdeck (1930), the majority of results in this area tended to suggest a gap between productive and receptive vocabulary knowledge with receptive vocabulary knowledge developmentally preceding productive vocabulary knowledge (Caspi and Lowie, 2014; Fan, 2000; Laufer and Paribakht, 1998). It would therefore appear that more studies investigating receptive and productive using consistent methodologies are necessary in order to ensure comparability of results.

As a way of introducing more clarity in relation to receptive and productive knowledge, Read (2000, p.155) used the terms 'recognition'

and 'recall' in relation to the two terms. The proposition here is that in recognition tests, 'learners are presented with the target word and are asked to show that they understand its meaning' whereas in 're-call' tests, learners are 'provided with some stimulus designed to elic-it the target word from their memory' (Read, 2000, p.155).

For the purposes of this study, Read's (2000) idea of *recall* and *recognition* is adopted. This is well aligned to the receptive VLT used to assess breadth of vocabulary knowledge among ESL learners. Thus, in the current study, the term *receptive* knowledge is used in relation to learners' abilities to *recognise* a lexical item in written form, and this includes the ability to match a given word with knowledge retrieved from their memory, hence the matching test.

Productive knowledge captures the ability to *recall* or retrieve an item from memory in order to achieve a linguistic task in written or spoken form. It is worth highlighting that a weakness of the productive test is that it is very difficult to control which words are tested and as such, a clue has to be provided. As an example, the Productive Levels Test provides first letters as a way of controlling lexical output from learn-ers (Laufer and Nation, 1999). An example of items from this test is provided in Figure 2.2 below.

Figure 2.2 The productive vocabulary levels test examples at the 2000 word frequency level

1. I'm glad we had this opp_____ to talk.
2. There are a doz_____ eggs in the basket.
3. Every working person must pay income t_____.
4. The pirates buried the trea_____ on a desert island.
5. Her beauty and ch_____ had a powerful effect on men.

Source: Laufer and Nation, (1999, p.46)

Since the first letters are provided, it remains to be determined what kind of output might be obtained from the same test items if learners were to produce free answers without being prompted by the first few letters.

2.2.2 The Mental Lexicon

The mental lexicon is a language user's mental store of words including the meanings and associations that an individual makes to particular words (Richards and Schmidt, 2013). Literature abounds with metaphors that different scholars have used to conceptualise the idea of the mental lexicon, for example, it has been likened to a dictionary, a thesaurus, an encyclopaedia, a library, a computer and the internet (McCarthy, 1990). While the dictionary / encyclopaedia metaphors hint at a more static phenomenon in relation to the mental lexicon (dictionary or thesaurus), later conceptualisations seem to capture the dynamism of the mental lexicon by alluding to the lexicon as a computer/ internet. This suggests that the mental lexicon is dynamic with new information being added and perhaps some older information being discarded through such processes as language attrition (Schmitt and Meara, 1997). Aitchison (2012) argues that while a lot has been written about the mental lexicon, relatively little is actually known about this subject. This is because, as the computer metaphor suggests, this is an area of immense complexity and dynamism, involving information sorting and organisation within the neural structure of the brain (Aitchison, 2012). This makes Word Association Tests (WAT) invaluable tools because of their ability to generate meaningful insights about words in the lexicon. Such data makes it possible to make inferences about the structure and development of such a complicated network. WAT have the ability to

measure depth of word knowledge amongst second language users and they can take the receptive or productive format (Agdam and Sadeghi, 2014). The main premise of such tests is that evidence of learners' knowledge of associations between words is used as a basis for making inferences about learners' knowledge of words.

2.2.3 Vocabulary breadth

The term vocabulary *breadth* is widely used in the research literature to refer to the number of words that an individual knows (Qian and Schedl, 2004; Haastrup and Henriksen, 2000; Wesche and Paribakht, 1996; Read 2004; David 2008) and is commonly used interchangeably with the term *vocabulary size.* Research in this area has produced useful insights which are relevant to pedagogy and research in lexical studies. As an example, Nation (2006) determined that knowledge of 8000–9000 word families is needed as it yields the ideal 98% text coverage when dealing with written text, while knowledge of 6000–7000 families is required for dealing with spoken text. Text coverage is defined as the percentage of running words in a text known by the reader (Hu and Nation, 2000; Nation, 2006).

Vocabulary knowledge has been found to correlate positively with other cognitive abilities such as academic success (Laufer and Nation, 1995; Morris and Cobb, 2004; Lumley 2002). Particularly relevant to the current study is Morris and Cobb's (2004) study which examined the extent to which vocabulary profiles could predict academic success among undergraduate Teaching English as a Second Language (TESL) trainees. The study found correlations between students' vocabulary scores obtained from the analysis of free written language with scores from the trainees' academic assessments. This

led the researchers to conclude that vocabulary profiles can potentially contribute towards predictions of academic performance.

In the same vein, Leki and Carson (1994) investigated perceptions of students about the contribution of English for Academic Purposes in the development of writing skills necessary to succeed in content studies at university level. The students highlighted vocabulary knowledge as the most important factor that they would like more tuition on during their English for Academic Purposes (EAP) courses. This suggests that vocabulary is considered key to language development by both researchers and students. All these insights are invaluable not only for diagnostic assessment purposes such as the identification of learners' language needs, but also for placing learners onto appropriate courses when they undertake further studies. Notwithstanding the above, as Wesche and Paribakht (1996) observe, a statistical measure of how many words a learner knows is not always adequate for understanding the learner's vocabulary knowledge. Such measures do not provide any insight as to how well words are known i.e. the depth of vocabulary knowledge. This aspect (vocabulary depth) is given further consideration in Section 2.2.7.

2.2.4 Measuring vocabulary breadth

On the basis of the value of the statistics related to the measurement of vocabulary size as outlined above, it is not surprising that a number of measurement tools have been devised for the assessment of vocabulary breadth among L2 learners, for example, Nation's (1990) VLT. The tool is widely used in lexical studies, for example, Nation and Waring (1997), Fan (2000), and Laufer and Paribakht (1998) all use the VLT in their investigations of receptive and

productive vocabulary knowledge. The following section provides a précis of the VLT particularly as this is the test adopted for generating vocabulary size profiles for learners in the current study.

2.2.5 The Vocabulary Levels Test (VLT)

First developed by Nation in 1983, the VLT provides an estimate of a learner's receptive vocabulary knowledge (Meara and Alcoy, 2010). The test assesses learners' vocabulary knowledge through generating vocabulary profiles based on a sample of 18 words drawn from each of the 2,000, 3,000, 5,000, 10,000-word frequency band. In addition to the four frequency bands, the VLT also offers the facility to assess knowledge of words from the vocabulary knowledge the Academic Word List (AWL) which was developed by Coxhead (2000). The inclusion of the AWL within the VLT is a very useful feature which is not offered by VST as a distinct category. This feature was particularly relevant to the current study where participants were tertiary education students whose vocabulary knowledge could be expected to show variability in terms of learners' use of words from the AWL category. The receptive VLT follows a multiple-choice format where test takers are required to choose synonyms or definitions which match given key words. Figure 2.3 below shows an example of a VLT item cluster taken from a receptive version of the test.

Figure 2.3 Example of an item from Nation's receptive VLT

1. original		
2. private	6	complete
3. royal	1	first
4. slow		
5. sorry	2	not public
6. total		

Source: Nation (1990, p.265)

The VLT was initially designed to meet a practical need to help English teachers plan for addressing the vocabulary needs of learners in a teaching and learning context. However, judging by its extensive use in research and language teaching contexts (see for example Pilar and Llach, 2009; Laufer, 1997) the test has gained popularity and become a valuable research and assessment tool in the field with Meara (1996, p.38) describing it as "the nearest thing we have to a standard test in vocabulary". To date, four versions of the test covering the assessment of receptive and productive vocabulary knowledge exist. The different versions offer flexibility and extend the utility of this test as it means that if required, the test can be used several times with the same learners at different intervals since different versions offer different items from the same frequency bands.

Validation evidence suggests that the VLT is a valid measure of vocabulary knowledge. For example, in their validation study of the 2000 Word Level and the AWL of the VLT, Beglar and Hunt (1999) found statistically-significant correlations between scores from the VLT and Teaching of English as a Foreign Language (TOEFL) scores on reading and grammar. The authors highlight this finding as part of the evidence that the VLT is a valid test which can be administered for the purposes of course planning and placement in language programs. Beglar and Hunt's (1999) views regarding the validity of VLT support the observation that numerous studies have utilised the VLT for research purposes (Pilar and Llach, 2009; Laufer, 1997). As Cameron (2002) notes, the tool has been assessed for its adequacy for different assessment purposes in the ESL context. This explains why new versions of the test have been developed in order to better

meet vocabulary assessment needs (Read, 1988; Schmitt, et al., 2001).

While the VLT is a well-established instrument for diagnostic and placement of students based on their vocabulary knowledge, (Kremmel and Schmitt, 2016), the test has its own drawbacks. As McLean and Kramer (2015, p.3) note, 'an assumption of test item analyses…is that the items demonstrate what is called *item independence'.* Item independence means that 'each item functions independently of others') in test (Beglar and Hunt, 1999, p.154). In other words, responses made to any items within a test should not impact consequent responses. Within the VLT, this assumption has not been shown to hold true given the format of the VLT test (Beglar and Hunt, 1999). This is because, as Figure 2.3 shows, the receptive VLT provides six items on the left which should be matched with only three definitions on the right. Once the first definition is identified, the number of available options decreases which can make further choices easier (McLean and Kramer, 2015). Therefore, the format of the VLT presents a potential weakness with regards to validity related to item dependency.

While the issue of item independence is one that Beglar and Hunter (1999) highlight as an area for future investigation, a body of research exists which suggests that while the test is not perfect, it is a valid measure of vocabulary knowledge (Meara, 1996; Beglar and Hunt, 1999; Schmitt, et al., 2001; Cameron, 2002; Stewart and White, 2011).

Another criticism against VLT is highlighted by McLean and Kramer (2015). The authors found that the test format where test clusters are presented in a ratio of six items on the left against three items on the right was problematic to their participants. McLean and Kramer (2015) found that the procedure was not immediately understood by the participants in their study. As a result, a lot of time was required to explain the procedure before the students were able to work through the VLT. For this reason, McLean and Kramer (2015) concluded that the format can be problematic compared to a traditional multiple choice format. In the context of the current study, the researcher ensured that a completed example was provided and a step-by-step verbal explanation provided. Overall, the VLT was considered to be a tool which is able to provide a quick (with only five possible levels) yet sufficiently detailed assessment of learners' vocabulary knowledge, including the assessment of AWL words. This was an important consideration since the participants were tertiary education students whose vocabulary knowledge of the AWL category was considered to be capable of generating useful insights.

2.2.6 The Vocabulary size test (VST)

The VST is a test of receptive vocabulary knowledge. It is capable of assessing learners' vocabulary knowledge from K1 up to the K20 word frequency level. For that reason, it is 'an appropriate instrument for separating students with a wide range of proficiencies' (McLean and Kramer, 2015, p.2). In Webb and Sasao's (2013, p.267) view, the VST is 'a far superior measure of second language vocabulary size as a whole (but not knowledge of particular levels)' because this tool assesses vocabulary knowledge up to the most frequent 14000 words or 20 000 words. In this test, test-takers are required to select the best definition or translation of each word provided out of the four

possible options given for each key word. The monolingual (English) format presents key words and their definitions in the English language as shown in Figure 2.4 below. The test can be completed in its paper-based version or its computer-based format as is the case with VLT.

Figure 2.4: VST item examples taken from Version A of the test

1. They <saw it> it.

a. closed it tightly

b. waited for it

c. looked at it

d. started it up

2. time: They have a lot of <time>

a. money

b. food

c. hours

d. friends

Source: Nation and Beglar (2007, p.75)

As can be seen from Figure 2.4, alongside with the key words, VST provides a brief context for the target words whereas the VLT is purely a matching test without the brief exemplar sentences demonstrating the key words in use. Relative to the VLT format, the brief context in which key words are presented offers a closer match to the authentic use of vocabulary in real life since vocabulary is generally encountered and used in context rather than as single entities. Arguably then, this is an advantage that VST offers over VLT. However, in the context of the current study, it was considered that the extra reading might present an additional cognitive load to be processed by the participants. This could extend the amount of time

required to complete the test, particularly considering the low proficiency level of the ESL students. Therefore, the single word, purely matching exercise offered by the VLT format was considered more appropriate for the learners in the current study. This provided part of the rationale for the choice of VLT over VST for the assessment of vocabulary size for the ESL students in the current study.

From the above discussion, it is evident that both the VLT and VST are valuable tools for the assessment of vocabulary knowledge. However, the VLT was chosen mainly because it offers 'separate slices of a learner's vocabulary (the 2nd 1000, the 3rd 1000, the 5th 1000, the Academic Word List and the 10th 1000' whereas the VST measures overall proficiency (Nation and Beglar, p.10). This was important for addressing the research questions in the current study.

While tests of vocabulary size produce invaluable insights about learners' lexical knowledge, such tests only provide a partial picture of the learner's vocabulary knowledge (Meara, 2009). The tests may therefore not be sufficient for contexts where a deeper understanding of learners' vocabulary knowledge is required. This should not be seen as a criticism of such tests because vocabulary size tests and investigations play an important role in the field of Applied Linguistics (see Section 2.2.3.) However, the argument that is being advanced here is that vocabulary size provides a crucial yet only partial picture of vocabulary knowledge, hence the need to take both vocabulary depth and breadth into account.

2.2.7 Vocabulary depth

Vocabulary *depth* relates to how well lexical items that constitute an individual's mental lexicon are known (Nation, 2001). It is a measure of the *quality* of the knowledge that an individual possesses about particular lexical items. As will be fully explored in Section 2.4.1, one of the most insightful frameworks for understanding the concept of lexical quality is provided by Richards (1976). Richards identifies seven aspects involved in knowing a word: knowledge of a word's semantic value; syntactic behaviour; network of associations; collocational behaviour; register; frequency of use in speech and writing in the target language as well as knowledge of underlying forms and derivatives for a particular lexical item. It is the development of associations between / among words in the lexicon that is at the core of depth of knowledge as investigated in this particular study. From Richards' (1976) word knowledge framework, it can be inferred that *depth* of knowledge relates to how well each of the seven aspects of vocabulary are developed. Bruton (2007) and Vermeer (2001) are amongst those scholars who have observed that no single measure of vocabulary knowledge can possibly capture all of the dimensions of vocabulary knowledge but separate measures are likely to capture the idiosyncrasies involved, thereby helping to build a more comprehensive picture of vocabulary knowledge. The WAT is one such measure and is therefore an invaluable tool in lexical studies (see section 2.3.2 for a more detailed discussion of WAT).

However, while the understanding of size and breadth of lexical knowledge is important in lexical studies, some researchers have pointed out that *breadth* and *depth* need not necessarily be perceived

as two distinct dimensions but rather, as part of a continuum. Schmitt (2010, p.216) for example, highlights that the notions of vocabulary size and depth are not completely isolated because for a lexical item to be counted as part of the learner's lexicon (size), 'some quality of knowledge, no matter how minimal, must be operationalised as the criterion of sufficient knowledge'. In other words, measuring the size of a learner's vocabulary cannot be completely isolated from some measure of quality. This overlap is also captured by Anderson and Freebody (1981) who highlight that an individual's vocabulary size includes all the words for which the individual knows at least some of the meaning aspects of relevant lexical items. Therefore, 'a person has sufficiently deep understanding of a word if it conveys to him or her all of the distinctions that would be understood by an ordinary adult under normal circumstances' (Anderson and Freebody, 1981, p.92-3). Arguably, the sufficiency of understanding may differ from individual to individual depending on their specific circumstances and reasons for developing word knowledge.

As an example, a budding fiction writer's perception of what sufficiently deep vocabulary knowledge should constitute will be different from what an average ESL student would perceive as sufficient vocabulary knowledge because of the particular uses to which the language (vocabulary) is intended for. However, the main point to be made is that Anderson and Freebody (1981) emphasise a certain expected standard i.e. understanding of knowledge aspects that would be understood by an 'ordinary adult under normal circumstances' (Anderson and Freebody, 1981, p.93). In some ways then, this definition acknowledges that acquiring depth of knowledge can be a highly subjective endeavour which is highly complex and

therefore difficult to fully quantify. From the discussions above, second language acquisition is complex and multifaceted, requiring theoretical lenses which are capable of capturing the myriad of idiosyncrasies involved. The Dynamic Systems Theory is one such framework, (see for example De Bot et al., 2007; Ellis and Larsen-Freeman, 2006; Larsen-Freeman and Cameron, 2008) hence the relevance of this theory to the current study as will be further explored in Section 2.8.1.

2.2.8 Vocabulary size and depth

While the size and depth conceptualisations of vocabulary knowledge are widely acknowledged in the literature (Beglar, 2000; Haastrup and Henriksen, 2000; Laufer et al , 2004; Qian, 1999; Read, 2004; Schmitt, 2012), there has been some contrasting perspectives suggesting that the distinction between size and depth is not helpful. In as far as the author of the current study is aware, these have been very limited but are considered relevant to the current study since size and depth of vocabulary knowledge are at the core of the investigation.

First among contrasting views is the study carried out by Vermeer, (2001). Vermeer (2001) set up two empirical studies involving Dutch monolingual and bilingual kindergarten children. The aim of the study was to explore the relationship between breadth and depth of vocabulary knowledge amongst the kindergarten children. In the first study, Vermeer's participants were required to complete a receptive vocabulary task (picture matching activity) and a description task where they were instructed to explain, describe, demonstrate or use

exemplification to demonstrate their knowledge of given lexical items. There was no requirement to provide definitions of lexical items provided. This constituted the vocabulary size measure for the participants. In the second study, Vermeer's participants were required to complete an association task as a way of obtaining a measure of the depth of knowledge for the vocabulary items investigated. The participants were examined on their abilities to supply formal definitions; features and components of given stimuli words; material and functional characteristics of these words as well as associations to do with how specific items on the stimuli list are used in real life. High correlations between size and depth measures were found in Vermeer's study, and this led the researcher to conclude that there is no conceptual difference between breadth and depth of vocabulary knowledge for both L1 and L2 learners since the participant group comprised monolingual and bilingual children.

As a reminder, Section 2.2.1 highlighted that differences in conceptualisations of major constructs can lead to different results in research findings. What is evident from the measures adopted in the study is the similarity between Vermeer's breadth and depth measures, particularly the description task which was intended to measure breadth of vocabulary knowledge and the association task which was intended to measure depth of knowledge. It is arguable that in describing words (description task) participants would inevitably use very similar lexis as they would use when required to supply descriptions related to the association task. In other words, Vermeer's size and depth measures bear some resemblance in terms of the skills that they assessed. This is problematic because arguably,

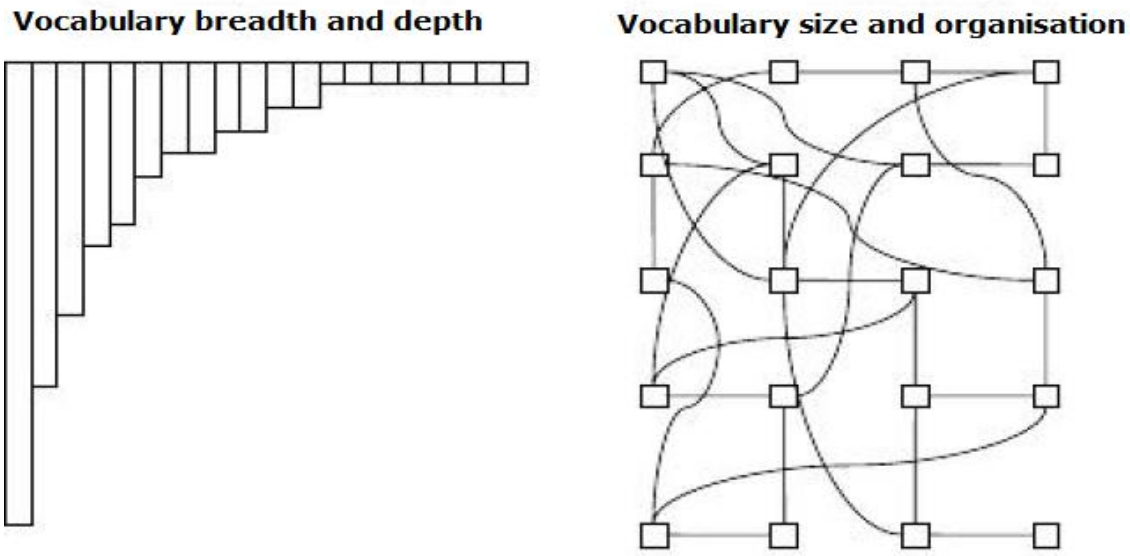
the very high correlations could be related to the similarities in the operationalization of the concepts under investigation.

Another point worth mentioning is that participants in Vermeer's study were kindergarten children and this could have also influenced the results obtained in the study. As pointed out by Feldman (2004), from a child development theory perspective such as Piaget's cognitive development theories, kindergarten children are still at the pre-operational stage of development. As such, they are more likely to view the world in concrete and contextual terms than in abstract and decontextualized terms (Orr, 1991). This could influence the types of associations and/or descriptions that children are likely to produce. In Vermeer's study, this would mean that the definitions and descriptions of lexical items for both the size and depth measures are likely to have been framed within the pre-operational stage of development framework. This provides another possible explanation to the high correlations between Vermeer's size and depth tests for the kindergarten children investigated in the study.

Another contrasting perspective to the size and depth conceptualisations is offered by Meara (2009). Meara argues that the concept of depth of knowledge is untenable, positing that assessing vocabulary depth on the basis of a vocabulary knowledge framework such as the one proposed by Richards' (1976) framework inevitably leads to more and more extensive and comprehensive testing of individual words. The author argues that such a situation is untenable particularly where numerous words have to be tested as is the case with most vocabulary tests. On this basis, Meara proposes that vocabulary knowledge should be perceived in terms of vocabulary

size and organisation rather than vocabulary size and depth. Figure 2.5 below is Meara's (2009) illustration of the vocabulary size and depth concepts juxtaposed with the proposed idea of vocabulary size and organisation.

Figure 2.5: Breadth and depth vs. size and organisation of vocabulary knowledge



Source: Meara, (2009, p.76)

On the vocabulary breadth and depth illustration (left side of Figure 2.5), each bar on the diagram represents a lexical item and the length of the bar represents the extent to which a word is known (depth). It is this depth dimension which Meara suggests could cause logistical problems. For instance, if fifty words needed investigation, then each of the seven dimensions of word knowledge specified by Richards (1976) would need to be considered. The diagram labelled 'Vocabulary size and organisation' (right hand side of Figure 2.5), represents what Meara refers to as 'organisation' of the lexicon where the number of networks is a key determining factor in terms of how well a word is known. Based on this view, the more networks

between / among lexical items, the more comprehensive the knowledge of the item can be expected to be.

Richards' (1976) model offers a framework for the various aspects of word knowledge that can be expected to develop as part of knowing a word (see Section 2.4.1). If such a model is applied indiscriminately to all lexical assessment contexts, this could indeed culminate in large numbers of test items per test which could result in fatigued and demotivated learners. This framework applied in its entirety might thus have both the logistical and pedagogical limitations that Meara warns of.

However, when applied with the understanding that individual studies can focus on particular aspects of vocabulary knowledge and produce insights which, when taken together, can contribute towards a theory of vocabulary knowledge, the model provides a comprehensive model for the cumulative generation of inferences about learners' lexical knowledge (Schmitt, 2010 ; Richards' 1976). This means that not every study needs to test words under investigation for each and every aspect of word knowledge. As an example, Schmitt and Zimmerman (2002) examined use of sixteen words from the Academic Word List (AWL) comprising verbs (e.g. assume; select); nouns (e.g. authority; philosophy); adjectives (e.g. traditional; liberal) and adverbs (e.g. inevitably).

Participants in Schmitt and Zimmerman's (2002) study comprised advanced non-native students who were either enrolled in an intensive pre-university English programme, an ESL writing course alongside other studies at university or a graduate ESL students

pursuing a Masters' degree in English Language Teaching. The main finding from the study was that participants often had problems producing all of the appropriate derivative forms within a given word family within the noun, verb, adjectival and adverbial word classes. The participants in this study rarely demonstrated either complete knowledge or no knowledge at all of the targeted words. Instead, the most common trend was for participants to demonstrate partial knowledge. Therefore, the study revealed the incremental nature of vocabulary knowledge based on one aspect of word knowledge namely knowledge of derivational forms.

Schmitt and Zimmerman's (2002) findings are contrary to Read's (1988) argument that 'if one knows the base word, little if any additional learning is required in order to understand its various inflectional and derived forms' (Read 1988, p.14). While evidence has been found to suggest that knowing some members of a word family may facilitate development of knowledge of unknown members within a word family particularly where stems are transparently linked to their derivatives (Nation, 2001), it has also been suggested that different word forms such as regular and irregular forms of a base word do not carry the same learning burden (Schmitt and Zimmerman, 2002). Therefore, Schmitt and Zimmerman's (2002) study provides further insight into vocabulary knowledge as it highlights that lexical development is not always as linear as implied by Read's (1988) comment that no additional learning is required for the acquisition of inflectional and derived forms once the base word is mastered. As was the case in Schmitt and Zimmerman's (2002) study, acquisition of one form of a word does not automatically imply acquisition of other forms of the same base. Similarly, Wolter's

(2001) study utilised depth of knowledge conceptualisations to explore the structure of the L2 lexicon compared to the L1 lexicon. The Vocabulary Knowledge Scale (see Section 2.3.1) and WAT were used to investigate how depth of knowledge might impact on the types of responses that participants provided. The researcher found that for both native speakers and L2 participants, words that were not well-known generally produced child-like/ non-native like response types. This led Wolter (2001) to conclude that the L2 mental lexicon is not structurally different from the L1 lexicon as had been previously assumed (Wolter, 2001). It may simply be at an earlier stage of development because fewer words are known and even these known words are likely to be known to a lesser degree compared to the L1's.

Nurweni and Read's (1999) study provides further insight on the interplay between vocabulary size and depth. The researchers carried out an investigation on the vocabulary abilities of first year university level students in Indonesia. The aim was to find out the students' understanding of the most frequent academic words. To that effect, the researchers measured the vocabulary size and depth for these learners and conducted correlation analysis. Overall, they found modest correlations between the two measures. However, when they split their participant group into three sub-groups according to proficiency level, (High, Middle and Low) they found that the relationship varied according to proficiency level. For the High-level students, the relationship between breadth and depth of knowledge was strong ($r = 0.81$) which suggests that not only did the students in this sub-group know the majority of the words tested, but they also knew the words well. When it came to the students in the Middle group, only a mod-

erate association between size and depth was found (r=0.43) and for the Low group, no meaningful relationship was found between their size and depth scores. This led the researchers to conclude that the size and depth interplay may be related to proficiency level.

From the above discussion, it is clear that not only is the size and depth relationship an important one in lexical studies, but it is also one that has caused some debate. The current research acknowledges the issues raised by Meara in relation to the logistical complications that may arise if, as Meara warns, a multiple-trait model such as the one offered by Richards' (1976) is applied in its entirety as a basis for assessing word knowledge for every lexical item tested. To that effect, the study does not attempt to investigate all of the dimensions of vocabulary knowledge but focuses on size (as measured by the VLT) and depth of knowledge (as measured through WAT). However, by virtue of focusing on particular aspects of vocabulary knowledge, the study acknowledges that vocabulary knowledge is a multiple-trait construct as suggested by Richards' (1976) framework which is fully explored in Section 2.4.1. Looking at the argument provided by Meara (1996) against the concept of depth of vocabulary knowledge on the one hand; and the empirical insights provided by studies adopting this approach on the other (see for example Wolter, 2001 and Schmitt and Zimmerman's, 2002). From this, it is arguable that the two perspectives are not mutually exclusive. Instead, they seem to contribute towards a hybrid view of the mental lexicon because if depth aids the development of connections in the mental lexicon as suggested by Wolter (2001) then such depth would contribute to the organisation of lexical networks which is emphasised by Meara's (1996) model. Therefore, Meara's

perspective is closely aligned to the network approaches to vocabulary knowledge which are further discussed in Section 2.4.2.

2.2.9 Vocabulary knowledge and writing

A considerable body of literature exists on how linguistic features such as lexical diversity, lexical sophistication and the ability to use cohesive devices can impact on the quality of written texts (see for example the works of Crossley and McNamara, 2012; McNamara et al., 2010; Jarvis et al., 2003, Jin, 2001; Engber, 1995). Overall, studies in this area have found that linguistic differences related to lexical use and text cohesion can be used to make inferences about the overall quality of texts as rated by human assessors (Crossley and McNamara, 2011). This is a valuable line of investigation, considering the role of written language proficiency in academic and professional contexts (Light, 2001).

In a study to investigate the extent to which lexical proficiency interacts with the quality of essays produced by ESL students, Engber (1995) analysed sixty-six essays which were produced by intermediate and advanced ESL students from multiple-linguistic backgrounds. The essays were holistically scored and the scores were compared to lexical richness scores for the essays. Engber (1995) found significant correlations between lexical variation and ratings on quality of essays so that essays that demonstrated fewer lexical errors and a wider range of vocabulary in use were awarded higher holistic scores for quality. In other words, the results suggest that not only does the diversity of lexical choice in writing significantly affect reader judgments of the quality of an essay, but also, the ability to choose appropriate lexical forms affect such judgements. This is an important finding particularly considering research which has highlighted that while the process approach to composing text (see Section 2.7.2 for process theories to writing) has been found to be highly successful

with advanced ESL learners, less advanced learners find lexical knowledge to be a stumbling block and that, even advanced learners do need a rich lexical background in order to compose high quality text (Engber, 1995). This suggests that understanding the processes involved in the production of written text is not sufficient for the production of high quality text. Lexis is an important resource for the effective execution of the writing process. Similarly, Crossley and McNamara (2011) investigated the relationship between holistic essay scores allocated by human raters and linguistic features utilised by both first language (L1) and second language (L2) writers of English. Crossley and McNamara (2011) analysed findings from a series of studies that compared computational indices of linguistic use such as lexical sophistication, syntactic complexity and text cohesion to human ratings of the same samples of writing (Crossley and McNamara, 2011). The main finding was that text quality scores allocated by human raters generally increased with increasing lexical sophistication, particularly with regards to lexical diversity and word frequency (Crossley and McNamara, 2011).

The use of low frequency words was also found to be significantly correlated to higher writing scores. In contrast, human allocated scores of text quality were not found to be strongly correlated to text quality in terms of linguistic cohesion or text comprehension. The studies analysed revealed that essays which were rated highly were not easy to read and comprehend because they contained more complex language (Crossley and McNamara, 2011). The authors found that although essay raters' evaluations of cohesion and coherence strongly correlated with overall scores allocated to essays, this was not necessarily based on the use of cohesion devices. Instead, there was an in-

verse relationship between high scores in overall quality of writing and the use of cohesion devices in writing. O'Reilly and McNamara (2007) suggest that this converse relationship may be because skilled readers such as expert assessors possess sufficient knowledge about the domains within which ESL students are typically assigned to write. As a result, their judgements of text quality are not affected by the lack of cohesive devises in a text as they can still read and understand the text even where such devices are not explicitly used. This has been referred to as the 'reverse cohesion effect' which describes the counter-intuitive finding that high-knowledge readers learn better from less cohesive texts (O'Reilly and McNamara, 2007, p.121; McNamara et al, 1996).

Crossley and McNamara's (2011) study also highlighted that although it was generally the case that the use of low frequency vocabulary was associated with more favourable judgements of essay quality, there were exceptions to this finding, for example in the case of polysemous words where the use of these words did not necessarily predict higher scores of overall essay quality. Since it is a widely-accepted view that polysemous words present challenges even to advanced English L2 learners, the point to be gleaned from Crossley and McNamara's (2011) findings is that the ability to effectively use diverse lexis as well as lower frequency vocabulary was associated with higher human judgements of text quality. This is an important finding which supports the view that vocabulary knowledge is implicated in the effective execution of key linguistic skills (Weigle, 2002; Olinghouse and Wilson, 2013).

Olinghouse and Wilson (2013) extended research on the relationship between vocabulary knowledge and writing by examining the role of

vocabulary in writing across three different genres (story, persuasive, and informative writing). A total of 105 compositions were produced across the three genres and all compositions were written under the same theme. The compositions were then rated using a holistic scale which determined the overall quality of writing. The compositions were also scored to determine vocabulary use, for example the diversity of vocabulary used and the extent to which students used academic words. The study found that genre had an impact on the type of vocabulary that students used. The story compositions exhibited highest lexical diversity, followed by persuasive text, with informative text exhibiting least lexical diversity. Furthermore, the study found that 'for story text, vocabulary diversity was a unique predictor, for persuasive text, content words and register were unique predictors while for informative text, content words were the strongest unique predictor explaining almost all of the total variance' (Olinghouse and Wilson 2013, p.45). Therefore, a key finding from the study is that not only was vocabulary knowledge found to correlate with written language ability but that this relationship differed according to writing genre. This finding is supported by Yu (2010) who found that different topics have an effect on lexical diversity with familiar topics bearing positively on lexical diversity.

Yu (2010) investigated the lexical diversity of a sample of written and spoken text for over 200 students. The study found significant correlations between lexical diversity and overall quality ratings for written compositions where lexical diversity accounted for 11% of the total variance in the overall quality ratings (Yu, 2010). The author considered this to be a high correlation in light of 'the myriad of other fac-

tors that may affect raters' judgement e.g. handwriting quality and syntactical complexity' (Yu, 2010, p.246).

More recently, Treffers-Daller et al. (2016) investigated the ability of lexical diversity measures to discriminate between essays produced by learners working at different proficiency levels as indicated by their Common European Framework of Reference for Languages (CEFR) levels. In the context of the current study, a particularly pertinent finding which emanated from Treffers-Daller et al's (2016) study is that lexical diversity measures were able to predict CEFR levels. The findings suggest that lexical indices play an important part in predicting overall writing ability. However, as Jarvis et al. (2003) note, most studies in the investigation of the interplay between lexis and writing have used correlation tests or tests that investigate differences between groups, thus assuming 'a linear relationship between writing quality and linguistic features' (Jarvis et al. 2003, p.378). To that extent, Jarvis et al. (2003) call for investigations which assume that there may be no single profile of highly rated texts in the same manner that there are multiple profiles of a 'good language learner' (Jarvis et al. 2003, p.378).

Therefore, the current study responds to this observation by not only utilising correlations to investigate the interplay between vocabulary size and depth on the one hand, and written language ability on the other hand; but also, exploring learners' own perceptions about the interplay between their vocabulary knowledge and their written language skills. This is achieved through stimulated reconstructions/semi-structured interviews. This approach opens avenues for multiple perspectives on the relationships under investigation (see Section 3.4.5 for a discussion of this approach). A number of instru-

ments have been developed in order to facilitate measurement of vocabulary depth. Amongst the most well-known measures are the Vocabulary Knowledge Scale (VKS) and the WAT which are discussed in the ensuing section.

2.3 Measuring vocabulary depth

2.3.1 Vocabulary Knowledge Scale (VKS)

The VKS is a depth of vocabulary knowledge test. It uses a five-point scale to integrate data from test-takers' performance with test-takers' self-reports (Wesche and Paribakht, 1996). It tracks the acquisition of lexical terms on a scale where the lowest level is absolute non-recognition of a lexical item and the highest level of knowledge is measured through the ability to use the target item in a sentence.

The first three items on the VKS prompt participants to provide self-reported word knowledge which is then validated through the fourth and fifth items which generate performance data by asking participants to provide synonyms and then write a sentence using the target items. Figure 2.6 below summarises the five prompts which constitute the VKS:

Figure 2.6: The Vocabulary Knowledge Scale

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: _____. (If you do this section, please also do Section IV)

Source: Wesche and Paribakht, (1996,p.30)

As can be seen from Figure 2.6, the scale measures participants' word knowledge in progressive degrees beginning with Level I through V.

Even though the scale captures both receptive and productive knowledge and elements of both size and depth, it is generally considered to be a test of depth of knowledge (Bruton, 2009). Participants are instructed to respond to all the prompts which describe their knowledge of a particular word. Levels I and II rely on participant responses. However, Levels III, IV and V capture verifiable knowledge of the item which is obtained via the production of synonyms and sentence writing. Output from the VKS is evaluated based on the marking scale shown in Figure 2.7 below.

Figure 2.7: The marking scale for the VKS

1 The word is not familiar at all.

2 The word is familiar but its meaning is not known.

3 A correct synonym or translation is given.

4 The word is used with semantic appropriateness in a sentence.

5 The word is used with semantic appropriateness and grammatical accuracy in a sentence.

Source: Wesche and Paribakht, (1996,p.30)

As Bruton (2009) noted, the VKS has been cited as one of the best known instruments for assessing the development of targeted words and as such, a number of studies have employed VKS or its modified versions for research on lexical development. One of the reasons why researchers may find this scale attractive is its ability to capture different degrees of developing word knowledge including no knowledge at all. This makes it possible to assess partial knowledge,

thus acknowledging that vocabulary knowledge is incremental in nature. Also, unlike other lexical development scales that have relied solely on self-reports, (see for example, Zimmerman, 1997), the VKS attempts to verify self-reports through eliciting productive knowledge.

However, as Schmitt (2010) observes, one of the criticisms that can be levied against the VKS is that not only does the scale require metalinguistic judgements from learners so that they can determine the degree of mastery of lexical items, but it also mixes receptive and productive measures of lexical knowledge. On the basis of the receptive vs. productive vocabulary knowledge conceptualizations provided in Section 2.2.1, it can be argued that only one item on the scale; the last item, which requires participants to use the word in a sentence, can be considered to measure depth of knowledge. Since this is a depth of knowledge scale, the expectation would be that if the scale combines vocabulary knowledge perhaps owing to the inevitable overlap between size and depth of vocabulary knowledge, then at least the scales could be balanced in favour of depth of vocabulary knowledge. For example, Questions 3 and 4 on the VKS see Figure 2.6) give test takers the option to provide a synonym or a translation. As was argued in Section 2.2.1, a translation task lends itself more to a receptive test than a productive test as it provides no evidence of a test taker's ability to use a given lexical item. To that effect, the test could be closer to a depth of knowledge test if learners are required to provide synonyms or definitions instead of synonyms or translations.

Another observation is that the scale seems to assume that if a participant has not seen a word before, then they do not know the

word. However, on the basis of Richards' (1976) multiple-trait model of vocabulary knowledge (Section 2.4.1), it is possible that someone who knows the pronunciation of a word might not know how to write it or read it. The individual might have had many verbal encounters with the word but may have never seen it in its written form. They might not even have the necessary reading skills to decipher the word when presented in written form. Such an individual would not be able to demonstrate their knowledge of the word within the parameters of the VKS.

2.3.2 Word Association Tests (WAT)

Since the 1950s, WAT have been used to draw inferences about the development and organisation of native speakers' mental lexicons, particularly for children (Fitzpatrick and Izura, 2011). To a very large extent, such studies suggested that as children grow older, their WAT responses become more paradigmatic than syntagmatic or clang (Cronin, 2002; Entwinstle, et al., 1964; Stolz and Tiffany, 1972). This led to the conclusion that development in the mental lexicon leads to a shift in response types from syntagmatic to paradigmatic response types (Söderman, 1993). Such perspectives consider the syntagmatic-paradigmatic to be a function of language exposure which leads to the development of word knowledge. Typically, this development is considered to start from clang to syntagmatic and then to a paradigmatic shift (Entwisle, 1966; Nissen and Henriksen, 2006).

Clang relations are words related in phonological terms only without semantic or syntactic relations (Wolter, 2001) as would be the case with *blog* and *clog*. Paradigmatic associations between/among words reside at the semantic or meaning level (Söderman, 1993) and an

example of such relations would be *right* and *left*; syntagmatic relations reside at the syntactic or collocational level (Qian, 1999) for example, *right* and *foot*. In WAT, the basic procedure is that participants are presented with a set of stimulus words and they have to demonstrate the associative networks in their lexical repertoire by identifying or producing relevant words associated with given stimulus words. Figure 2.8 is an example of a receptive WAT test item, with expected answers indicated with a tick (√):

Figure 2.8 Example item from the WAT

1. beautiful

☑ enjoyable	☐ expensive	☐ free	☐ loud	☐ education	☑ face	☑ music	☑ weather

Source: *Read, (1998, p.1)*

As can be seen from the example items provided in Figure 2.8, WAT provide opportunities for participants to demonstrate their knowledge of both paradigmatic and syntagmatic relations amongst given words. Therefore, although at first glance, the WAT assesses word knowledge at an individual word level, they in fact go beyond the individual word meanings to assessing the kind of links that words in their lexicon may have. As shown in Figure 2.8, in receptive association tests, participants have to choose what they consider to be the most appropriate associations to a particular stimulus word from given options. However, in productive association tests, participants are provided with stimulus words and instructed to write down the first word which comes to their mind when they see the stimulus word (Read 2004).

The main strength of WAT in either format is that they can reasonably be used to test a broad range of words while being fairly easy to administer. Notwithstanding the above, the advantage of the receptive test over the productive test is that it controls output from participants. This functionality makes objective scoring easier to achieve (Agdam and Sadeghi, 2014) compared to productive formats which inevitably lead to a wide range of learner outputs whose scoring quite often involves evaluative judgements regarding the degree of appropriateness of each response produced.

An illustration of the challenges associated with scoring productive WAT can be gleaned from Agdam and Sadeghi's (2014) study. The researchers report that in their productive WAT to measure depth of knowledge among elementary learners, learners were asked to provide words that came to their minds when they were given each key word. However, points could only be allocated if the answers produced met the researchers' criteria which did not necessarily cover all possible words that could be stimulated by the key words. The authors provide the following example:

Figure 2.9 Sample item from a productive WAT

Key word	**Student Responses**		
banana	yellow	fruit	delicious

Source : Adapted from Agdam and Sadeghi's (2014)

Agdam and Sadeghi (2014) note that on the basis of the output captured in Figure 2.9 above, the learner would earn points for 'fruit' because this response denotes a subordinate association at a paradigmatic level with the key word. The learner would also earn

points for 'yellow' because researchers considered this to capture a conceptual association with the key word. However, the word 'delicious' did not have any points attached to it so although this word is clearly relevant to 'banana', it would not earn any points, yet with it being a free task, the key words are likely to generate an extensive range of answers (Read, 1998). This illustrates the lack of objectivity associated with the test. For receptive WAT, objectivity is enhanced because a finite set of responses is provided and it is on the basis of these controlled responses that learners' performance is assessed.

As an example, in the sample item provided in Figure 2.8, the correct answers were identified as 'enjoyable' at the paradigmatic association level, together with 'face', 'music' and 'weather' at the syntagmatic level. This would apply to all learners assessed, therefore making the test easier to mark. This is one of the reasons why the receptive WAT was the chosen method for purposes of the current study. The fact that it is easier to score objectively was considered to be an important factor that would contribute to the reliability of the results obtained. Another reason why the receptive WAT was the preferred option is that the test is well aligned to the VLT test which was used in the current study, making the two variables (WAT and VLT) comparable.

In addition, WAT are an extensively utilised test format for assessing depth of knowledge because they are considered to have the potential to economically measure language users' knowledge of target words (Higginbotham, 2010). Their validity has been tested through empirical studies, for example, Qian and Schedl (2004) carried out a study whose main aim was to evaluate a depth of vocabulary knowledge test in the format of WAT. They found strong

correlations between the Depth of Vocabulary Knowledge test and items on the vocabulary part of the TOEFL test. This led the researchers to the conclude that the WAT format has great potential as a framework for developing items to assess reading comprehension for ESL learners in high stake tests such as TOEFL examinations. Despite the above, it is worth highlighting that a different perspective on WAT format tests is offered by Schmitt, Ng and Garras, (2011). The authors carried out a validation study on WAT and on the basis of their findings, they concluded that while the WAT format is economical and has the potential to assess depth of word knowledge, the selective (receptive) WAT test is unlikely to be robust enough to become part of high-stakes standardized tests because of its susceptibility to guessing and the consequent challenges with the interpretation of scores.

Guessing is a strategy that many learners rely on to compensate for lack of vocabulary knowledge and overcome related obstacles in assessment activities (Kaivanpanah and Moghaddam, 2012; Read, 2000). Guessing can be preceded by a well-informed reflection which involves an individual making a logical conclusion based on other information already known. In lexical studies, this is commonly referred to as lexical inferencing (De Bot, et al, 1997; Nassaji, et al., 2010; Nazmia and Sima, 2004). Describing this strategy, Haastrup (1991, p.40) notes that lexical inferencing 'involves making informed guesses as to the meaning of a word in the light of all available linguistic cues in combination with the learner's general knowledge of the world, her awareness of the co-text and her relevant linguistic knowledge'. However, notwithstanding this criticism, overall, Schmitt, Ng, and Garras, (2011) concede that WAT format tests have the

ability to generate valuable insights about the quality of vocabulary knowledge amongst second language learners.

For a long time, WAT items were derived from the Kent-Rasanoff (1910) list which comprises high frequency nouns, adjectives and verbs originally selected for psychological studies (Kent and Rosanoff, 1910). However, one problem associated with the test is that the use of high frequency words to elicit word association responses is likely to affect the types of responses produced. The mental lexicon is made up of words from all frequency bands so examining the mental lexicon based on high frequency words only suggests that the results obtained may not be applicable to the organisation of the rest of the lexicon (Higginbotham, 2010). To that effect, in more recent WAT studies, researchers select prompt words in a more methodologically-considered way in line with their research aims instead of using high frequency words only. For example, Higginbotham (2010) used nouns from different frequency bands to investigate the impact of word frequency on the types of word associations produced by students. The prompt items used included nouns such as *body, book, business, car, case, child, church, class* and *door.* In the current study, the frequency variable is an important consideration. This is because the interplay between vocabulary knowledge and written language can be expected to be affected by frequency since words at different frequency levels are likely to be acquired to different levels. As such, the WAT used covers K1, K2, K3 and AWL words (see Appendix 3).

In recent years, word association oriented tasks have permeated L2 research aimed at addressing issues related to the structure, development and organisation of the L2 mental lexicon (see for

example Fitzpatrick, 2007; Schmitt, Ng and Garras, 2011). Consequently, word association tasks have been used to investigate a wide range of issues related to the mental lexicon such as responses produced by L2 learners' to WAT tasks compared to those produced by native speakers (Wolter, 2001). Although researchers have been attempting to develop an understanding of the mental lexicon for decades, there has not been a consensus on a number of areas including how the second language mental lexicon differs from that of native speakers (Higginbotham, 2010; Fitzpatrick, 2007). As an example, Higginbotham (2010) argues that as proficiency develops, responses produced by second language learners would shift from syntagmatic to paradigmatic, thus paralleling the development noted in children's L1 development Another view is offered by those who argue that word associations produced by adult native speakers are homogenous and stable enough to reflect what could be considered to be native-speaker norms (Meara, 1983). This assumption is challenged by Fitzpatrick (2007). To this end, Fitzpatrick carried out an investigation of word association output from a group of L1 English speakers who completed two-word association tasks. The study revealed significant within-group variations in the response patterns of English L1 participants in the study where high standard deviations from the mean for each word association category were reported. This highlighted individual differences within native speakers' responses. It was on this basis that Fitzpatrick challenged the assumption that native speakers' response types were homogenous and stable. What was found to be more consistent was the way in which individual profiles matched the same individuals' profiles between the two-word association tasks administered to the group.

In other words, individuals were consistent in their behaviour on the two tasks but there was significant variation between individuals within the group. This led Fitzpatrick to the conclusion that 'adult native speakers were not homogenous or predictable in their response behaviour as a group' (Fitzpatrick, 2007). On this basis, Fitzpatrick proposed that a more rewarding line of enquiry could be the analysis of word association data from an individual perspective rather than from a group perspective which attempts to establish native versus non-native norms.

The views proposed by Fitzpatrick find support in studies that have investigated WAT responses for native and non-natives based on words unfamiliar to both native and non-native populations. Findings from this line of enquiry reveal that when native and non-native speakers are tested on rarer vocabulary not familiar to either, they tend to be similar in terms of the proportions of response types generated i.e. paradigmatic, syntagmatic, and clang responses produced (Wolter, 2001; Fitzpatrick, 2006; Zareva, 2007).

In the same vein, Namei (2004) carried out a comparative study to determine the organisation of native vs. non-native lexicon. The study found that words that are barely known elicited phonologically-based associations for both natives and non-natives. Those that were partially known had a strong syntactic organization, while well-known words were connected to other words mainly on a semantic basis. This suggested that different types of word associations amongst both native and non-natives occur as a function of the degree of word knowledge rather than L1 background per se, thus suggesting that proficiency is a main factor in word association tasks. These findings

support earlier investigations such as those of Lambert (1956); Politzer (1978) and Meara (1983) which suggested that there is a relationship between word association knowledge and language proficiency in general. This suggests that differences found in second language learners' WAT responses could be a function of their levels of proficiency, and therefore that higher proficiency levels among L2 learners would correspond with higher proportions of paradigmatic associations while lower proficiency would correspond with syntagmatic and clang associations.

Notwithstanding the proficiency arguments highlighted above, Wolter 2002) offers a different perspective. Wolter investigated the possibility of using WAT activities to assess proficiency. His study found only moderate correlations between word association task results and scores from a C-test. The C-test is a measure of overall language proficiency structured very much like a cloze test. In this test, learners demonstrate their language skills by providing missing parts to words in context. As only entirely correct responses can be accepted, the test is considered to provide a robust assessment of the learner's general proficiency skills (Katona and Dornyei, 1993). The fact that Wolter's (2002) study did not reveal a strong correlation between word associations and the C-test was therefore interpreted as suggesting that there was no clear link between proficiency in a foreign language and word association tasks.

Therefore, what can be gleaned from the literature in word association tasks is that although the study of word association tasks and the illuminations that can be obtained through them about the L2 mental lexicon continue to grow (see for example, Fitzpatrick, 2007;

Zareva, 2010; Schmitt, Ng and Garras, 2011; Clenton, 2015) there are still a number of contested areas which need further research. Two of these are i) the issue of whether performance on WAT can be predicted along the native speaker vs. non-native speaker dichotomy (see for example, Fitzpatrick, 2007; Zareva, 2010) and ii) the interplay between WAT output and second language proficiency (see Wolter, 2002; Clenton, 2015). Considering that the main aim of this thesis is to investigate how lexical knowledge relates to written language ability, it is the latter that resonates with the aims of this study.

2.4 Frameworks of lexical knowledge

Section 2.1.1 highlighted the complexities associated with defining word knowledge owing to the multiple aspects involved in knowing a word. This section presents some of the frameworks that have been proposed to encapsulate what word knowledge might comprise. I consider two of such approaches, the components approach and the network approach to vocabulary knowledge. These two frameworks relate to the vocabulary breadth and depth issues which are at the core of the current study.

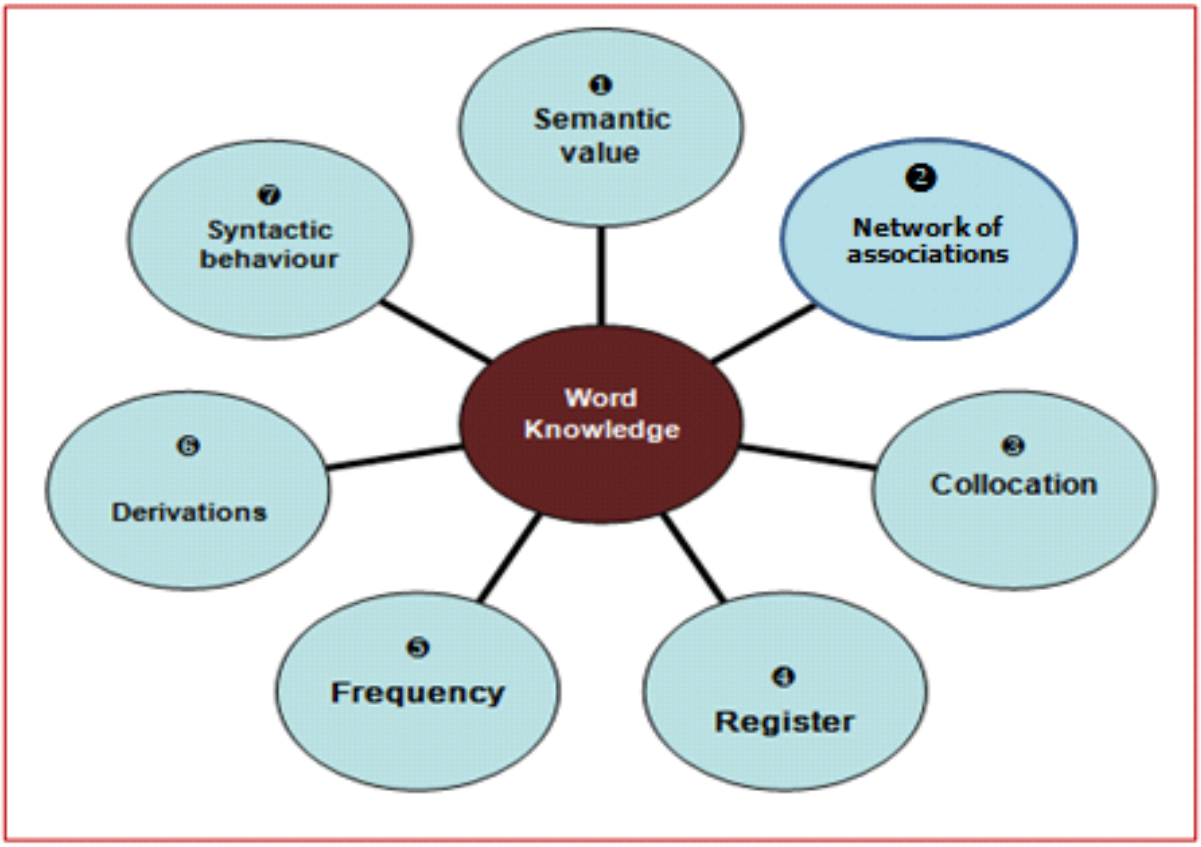
2.4.1 The components approach to vocabulary knowledge

Stahl (2005) states that knowing a word encompasses not only knowledge of the definition of the word, but also, knowledge of how that word fits into the world. This conceptualisation fits in with what has been termed the *dimensions or components* approach to lexical knowledge. It involves specifying the various types of word knowledge one can have about lexical items. This approach dates back to Richards' (1976) seminal paper on vocabulary teaching where

the author identified seven dimensions that are considered to be at the core of the concept of what it means to know a word. These parameters are summarised in Figure 2.10 below.

As Figure 2.10 shows, from the framework provided by Richards (1976), word knowledge is not only about the form-meaning link for a particular lexical item. The principal idea that can be gleaned from Richards' model is the notion of vocabulary knowledge as a multifaceted phenomenon, hence the multiple-dimensions that constitute the construct. If vocabulary knowledge is indeed a multifaceted phenomenon, then this implies that its research, teaching and assessment ought to acknowledge this in order to get more comprehensive insights on learners' vocabulary knowledge.

Figure 2.10 What it means to know a word



Source: *Adapted from Richards, (1976, p.81-82)*

Therefore, since its publication, Richards' components approach to vocabulary knowledge has continued to influence and inspire research in the field. It has seen many researchers building on this foundation and refining it in line with new developments in the field.

As an example, Nation (1990) distinguished eight main dimensions of knowing a word which he believed were part of mastery of lexis. These word knowledge aspects are summarised below:

1. The spoken form of a word,
2. The written form of a word,
3. The grammatical behaviour of the word,
4. The collocational behaviour of the word,
5. How frequent the word is,
6. The stylistic register constraints of a word,
7. The conceptual meaning of a word and
8. The associations a word has with other related words.

The eight dimensions are not dissimilar to Richards' (1976) conceptualisation and, like its predecessor, Nation's framework identifies vocabulary knowledge as a multifaceted construct whose development depends on several aspects that all work together. However, Nation (2001) subsequently developed a revised and extended version of this framework, shown in Table 2.1 below. It is this updated version of the framework which is now widely used and extensively cited in the research literature (Read, 2004).

Table 2.1 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?
Meaning	Word Parts	**R** What parts are recognisable in this word? **P** What word parts are needed to express this meaning
	Form and meaning	**R** What meaning does this word 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 one?
	Constraints on use (register, frequency)	**R** Where, when and how often would we expect to meet this word? **P** Where, when and how often can we use this word?

R = Receptive P=Productive

Source: Nation (2001, p. 27)

Schmitt (2010) identifies the framework as the best specification of the range of word knowledge aspects in the field. Perhaps the greatest strength of the revised framework, is its ability to differentiate between productive and receptive vocabulary knowledge within each of the dimensions of word knowledge (i.e. form, meaning and use). As Schmitt (2010) observes, it is very common for language learners to understand lexical items when listening or reading, but not be able to produce those items in written or spoken form. From this perspective, Nation's framework has the potential to help teachers in the selection and evaluation of classroom activities to facilitate receptive and/or productive knowledge.

2.4.2 Network approaches

Proponents of network approaches such as Aitchison (2012) and Meara (2009) argue that rather than looking at properties of lexical

items, a more plausible approach towards understanding the nature of vocabulary knowledge is to analyse properties of the lexicon as a whole (see Section 2.2.7). This view posits that words become enmeshed in the network of related words through the development of a variety of links. Such links could be through synonymy, subordination or super-ordination; syntagmatic relations (links through a sequential or collocational relationship); or phonological relations (where words can be related because they include similar sounds / clang relations, even though they may not have a semantic relation), and most commonly, the links could be paradigmatic (meaning related (Aitchson, 2012; Wolter, 2001)). From the network approaches perspective, it is through the process of network building that lexical items find their place within the learner's lexicon (Haastrup and Henriksen, 2000). Therefore, the number of links between and among lexical items provides an indication of how well an item is known. In Meara's view, this building of networks is what characterises *depth* of knowledge so that well known words would be characterised by an increasing number of links with other words and vice-versa (Meara 2009). This would have an impact on lexical availability for both receptive and productive use so that words with more links will be more available for productive use while those with less developed links will be more available for receptive use.

Henriksen (1999) takes what seems to be an eclectic approach in his conceptualisation of vocabulary knowledge. The term eclectic seems appropriate because Henriksen's framework combines elements of the dimensions and network approaches to lexical knowledge. The researcher proposes a three-pronged framework where vocabulary knowledge is considered to comprise three interrelated dimensions

namely the partial-to-precise dimension; the depth of knowledge dimension and the receptive-to-productive dimension.

On Henriksen's partial-to-precise dimension of vocabulary knowledge, learners are envisaged to start off with vague meanings. Lexical development therefore involves the gradual movement or development of knowledge from 'rough categorisation' to 'more precision and mastery of finer shades of meaning' (Henriksen, 1999, p.311). This acknowledges a key notion of lexical knowledge namely that it is not an all or nothing phenomenon, instead, it is incremental in nature, with learners' knowledge of vocabulary items starting from zero and gradually growing towards better precision. However, describing this dimension as the 'partial-to-precise' dimension may be misleading because 'precise' suggests a complete state yet, as the author acknowledges, 'no native speaker will ever develop an exhaustive knowledge of a word's meaning potential'. This is because 'understanding is gradually changed and increased as experience both of the world and of the language is expanded' (Henriksen, 1999. p.311). With this in mind, if the term 'precise' is taken to mean 'exactness' (Stevenson, 2010, p.1397), as defined in the Oxford Dictionary of English it would be an unrealistic goal to always aim for complete mastery in each aspect of vocabulary knowledge.

The third dimension in Henriksen's model entails the development of vocabulary knowledge from the receptive to the productive dimension. As discussed in section 2.2.2, while conceptualisations of the two dimensions bear some inconsistencies in the literature, this is a widely recognised distinction with a wide consensus on the view

that receptive mastery generally precedes productive ability (Laufer and Paribakht, 1998).

In discussing the three dimensions, Henriksen emphasises the interrelations among them suggesting that development in one has an impact on the other dimensions. So not only is knowledge of individual words growing along the different continua, but this also leads to the development of links/networks amongst lexical items. Thus, acquisition of more knowledge about a lexical item helps the learner to develop better networks between the new lexical item and other items already stored in the mental lexicon. Notwithstanding the use of the debatable term 'precision' as highlighted earlier, Henriksen's framework offers a very useful lens for understanding lexical knowledge. Not only is the framework multiple-pronged and therefore reflective of the multiple-trait nature of vocabulary knowledge but it also highlights the incremental nature of vocabulary knowledge. From this perspective, the framework taps into both component and network approaches to vocabulary knowledge.

2.5 Word frequency

Schmitt (2010) suggested that an educated adult native speaker is likely to know around 16,000-20,000 words These figures are based on the view that native speakers add roughly 1,000 word families a year to their vocabulary size (Nation and Waring, 1997). From this perspective, a young native speaker will have a vocabulary of around 4,000 to 5,000 word families when they start school and by the time they graduate from university, their vocabulary size is anticipated to have grown to approximately 20,000 word families (Goulden, et al., 1990). However, findings from a recent study by Milton and Treffers-

Daller (2013) suggest that these figures might be an overestimation. The study investigated the relationship between vocabulary size and the academic achievement of monolingual university level students. It used the frequency based vocabulary size test to measure learners' vocabulary sizes and then compared these with students' academic performances. Findings from the study suggested that monolingual university level students knew about 10,000 word families and that on average, a figure which suggests that uptake of about 500 words per year is required before an individual reaches university level (Milton and Treffers-Daller, 2013).

Milton and Treffers-Daller (2013) used the same frequency based vocabulary size test as Goulden, Nation and Read (1990) who give an estimate of 20,000 word families. However, for each of the words that learners indicated they knew, Milton and Treffers-Daller (2013) asked their participants to provide a synonym, an explanation or to give an illustration of how the word is used in context (Milton and Traffers-Daller, 2013). This was an important step in verifying self-reported knowledge and may explain some of the variation in the estimates of learners' vocabulary sizes. While the lower estimates suggest a more achievable task than previous larger estimates, the figures still provide a very clear indication that learning vocabulary represents a considerable burden to ESL learners. It is therefore not surprising that research has identified vocabulary as an area that presents substantial difficulty in L2 learning (Leki and Carson, 1994; Muncie, 2002). To this end, word frequency has long been identified as one of the most important criteria of vocabulary selection for pedagogic purposes because it provides a framework for systematic vocabulary selection and teaching (Honeyfield, 1977). This helps ensure that

vocabulary is taught and learned in the most efficient ways (Nation, 2001). An unstructured approach which exposes learners to words in a random manner is likely to culminate in unrealistic goals that can frustrate both ESL students and teachers alike.

Word frequency counts traditionally identify the first 2,000 word families as high frequency words, a figure derived from the General Service List (GSL) first developed by West in 1953 (Schmitt and Schmitt, 2012). The GSL has, for many decades, helped teachers and material writers by providing a framework for the identification of words that are likely to be of greatest utility to learners rather than attempting to teach all words in a haphazard manner. Put another way, not all words in a language are equally useful to learners and a commonly used guide for determining usefulness is word frequency, that is, how often the word occurs in normal use of the language (Nation and Waring, 1997). It is the most frequent words in English that most ESL learners need to learn as a priority as these will facilitate their ability to use the target language. As such, it makes sense for such words to be part of the vocabulary core which is taught systematically. The next section provides a more comprehensive outline of how frequency bands have been conceptualised in the literature. In the current study, frequency is an important variable because it is one of the main features that are used to determine the quality of vocabulary used by the lower and higher proficiency groups investigated in the study as Section 2.5 highlights.

2.5.1 High Frequency Vocabulary

As already noted, 'there is a relationship between word frequency and the likelihood that words will be encountered or used by second lan-

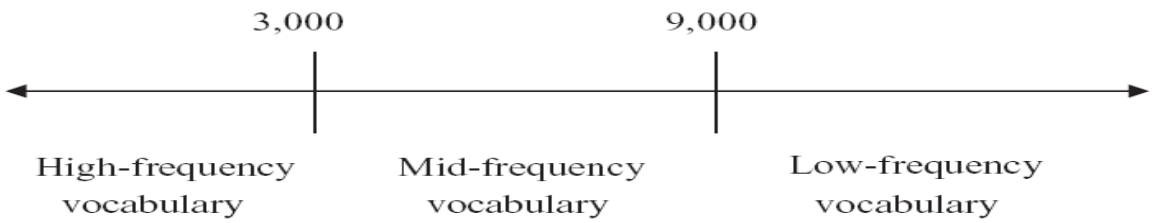
guage learners' (Davidson et al., 2008, p.133). However, within the high frequency band, the first 1,000 words have been found to be used substantially more than their second 1,000 word counterparts. For example, on the basis of data from the British National Corpus, Leech and Rayson (2014) note that the first 1,000 words constitute 78-81% of written text while the second 1000 words cover only 8-9% of text. Nation (2006) found a similar trend for spoken language where 81-84 % of spoken text came from the first 1000 words and only 5-6% came from the second 1000 words. Another point to highlight in relation to high frequency words relates to the word families included. Schmitt and Schmitt (2012) make a very poignant observation that there is great pedagogic value in extending the low frequency band to the first 3,000 words. Looking at Nation's (2006) findings, Schmitt and Schmitt's recommendation seems justifiable because the rate of occurrence for words beyond the 2,000-word band differs only slightly from the second 1,000-word band. Indeed, Nation's study cited earlier found that the third 1,000-word level provided 3-5% of written text coverage and 2-3% of spoken text coverage.

Considering that Nation's (2006) investigation led him to conclude that 98% coverage or more is required for learners to effectively decode and comprehend written text, the third 1,000-word level represents a substantial lexical resource for learners. This suggests that excluding the third 1,000 words from the essential part of the learners' core vocabulary development could cause challenges with regards to their ability to use language in authentic situations. On this basis, Schmitt and Schmitt's (2012) proposition for an additional 1,000 words to be considered as part of essential vocabulary seems well justified.

2.5.2 Mid Frequency Vocabulary

From the high frequency band discussed in Section 2.5.1, the classification in the research literature then switches to the Academic Word List (AWL) and then to low frequency words which Nation (2013) puts at beyond the 9000-word level. This creates a gap of vocabulary which is unaccounted for. This is the gap that the authors propose needs to be closed through what they refer to as 'mid-frequency vocabulary' (Schmitt and Schmitt, 2012, p.12) as Figure 2.11 below shows:

Figure 2.11 The nature of mid-frequency level vocabulary



Source: Schmitt and Schmitt, (2012, p.12)

As Figure 2.11 shows, the mid-frequency range accounts for vocabulary outside the 3,000-word level and below the 9,000-word level. Mapping of this range could, for example, allow material developers to focus on this vocabulary range to ensure the availability of teaching / learning materials that are specifically targeted at this frequency band. While it may not be feasible for teachers to target all of this vocabulary for explicit teaching, raising awareness of the value of this mid-range vocabulary band may encourage learners to engage with it independently and help them expand their overall lexical resources in the long run.

2.5.3 The Academic Word List (AWL)

It is a widely accepted view in the literature that academic vocabulary knowledge is an indispensable component of academic success (Albretchtsen et al., 2008; Alderson, 2005; Laufer and Goldstein, 2004). This presents teachers and material writers with the challenge of how to choose vocabulary that would best prepare learners for the academic world. Coxhead and Nation (2001) therefore devised the AWL as an extension of the GSL and intended it to offer a guide to words that are frequently used in academic discourse but are rare in other text genres. Such a list is essential for extending learners' vocabulary knowledge once the high frequency words have been mastered and comprises approximately 570 words these words which have been brought together as the AWL (Nation, 1990). Cobb (2007) notes that when combined with the 2,000 most frequent word list, the list of 2,570 words constitutes approximately 90% coverage of an academic text.

On the basis of the above figures, a learner who has developed knowledge of the first 2,000 words as well as the 570 words from the AWL can be expected to know approximately 90% of the words encountered in reading academic texts. As such, academic vocabulary is essential for effective execution of most tasks in the Higher Education context. According to (Nation, 2001), numerous exposures are required for word knowledge to be acquired. However, a word of caution is in place here because while the impact of frequency is widely acknowledged in the literature, research has also found that some very frequent words may cause difficulty because of their polysemous nature. This means that learners may learn some but not all of the

meanings of a word, thus making a seemingly basic word quite difficult.

As an example, Altenberg and Granger (2001) investigated high frequency verbs, in particular the verb 'make' which is one of the verbs topping any corpus-based list of high frequency verbs in English (Altenberg and Granger, 2001). As the authors note, such polysemous verbs will typically carry abstract meaning extensions resulting in specialised meanings, collocations and idioms. In other words, they tend to be semantically anti-intuitive and quite often very context-specific. Such attributes combine to make learning words in a foreign language more difficult than would otherwise be expected. This view is supported by (Sinclair, 1991) who argues that language learners sometimes find basic verbs, particularly when used as phrasal verbs, so challenging that they avoid them and rely on rarer and sometimes clumsier words which make their writing sound stilted and awkward. It is therefore not surprising that such verbs are widely reported as being prone to errors in spoken and written learner language. They may not necessarily be easy to learn and as Altenberg and Granger (2001) found, even advanced proficiency learners face great challenges with high frequency verbs such as 'make'.

2.5.4 Offlist

Low frequency vocabulary has traditionally been thought of as beyond the 10,000 frequency level (Nation, 2006; Schmitt and Schmitt, 2014). Traditionally, frequency tools such as Laufer and Nation's (1995) Lexical Frequency Profile tool classified words into the first 1000 and 2000; the AWL and then the Offlist category, suggesting

that any words that do not fall within the first three categories inevitably fell into the Offlist category, thus representing words which could not be categorised. However, since the 2000s, the tool has seen a number of adaptations resulting in more refined tools such as the Vocabulary Profile, commonly referred to as VocaProfile (Cobb, 2013).

Since Vocabprofile has the ability to categorise words into as many as 25 frequency bands, the tool offers clearer distinctions between texts (Cobb, 2013). This reduces the number of words that cannot be categorised using current Vocabprofile tools. The next section considers some of the most commonly utilised measures of lexical richness.

2.6 Measures of Lexical Richness

An extensive body of research exists regarding various textual measures that can be utilised for the measurement of lexical features in text (Cohen, 1988; Laufer and Nation, 1995; Tweedie and Baayen, 1998). However, the most widely used include measures of lexical originality, lexical density, lexical sophistication and lexical variation (Laufer and Nation, 1995). Sections 2.6.1 – 2.6.5 provide a brief overview of common procedures that have been utilised for the measurement of lexical features in written language.

2.6.1 Lexical Originality

Lexical Originality measures the percentage of words which one writer uses and which are used by no other writers in the group (Laufer and Nation, 1995; Muncie, 2002). Since the Lexical Originality Index measures the learner's unique performance in comparison to other members of a particular group, this means that the index only remains stable if the group does not change, but if the group

changes, then the index changes too. This raises questions about the reliability of this index.

2.6.2 Lexical Density

The term Lexical density was first proposed by Ure (1971) in an investigation of language patterning in text. Ure (1971) describes the procedure which she adopted as follows:

> *First, the number of orthographical words was counted for each text. Then the number of words with lexical properties was counted, and a percentage arrived at for each text. The term proposed for this proportional occurrence of lexis is lexical density (Ure, 1971, p.445)*

In her study, Ure (1971) found that overall, written text produced higher lexical density indices (40% and above) while spoken discourse generally produced lower indices below 40%). Therefore, one of the conclusions drawn was that lexical density varied as a function of a number of factors including the text medium (spoken or written) as well as whether text was planned or not, where planned texts generally produced higher lexical indices than unplanned texts. Therefore, as Ure (1971) highlights, lexical diversity provides a measure of the percentage of content words used by a writer in a text relative to the total words used in the text including function words. Lexical words are the types of words that are considered to be absolutely necessary in a text because they embody lexical meaning so they include nouns, verbs, adjectives and adverbs.

Function words on the other hand are words that function at a grammatical rather than lexical level so they are used to express how words, phrases or sentences relate to each other in a text. Examples of such words would be articles, prepositions, conjunctions, auxiliary verbs, particles and pronouns. It follows then that in order for a text

to exhibit high density, the text has to contain a wide range of content words in relation to the total number of tokens in the composition (content words and functional words put together).

2.6.3 Lexical sophistication

Also referred to as lexical rareness in the literature (see for example, Lu, 2012) lexical sophistication provides an index of "the proportion of relatively unusual or advanced words in the learner's text" (Read, 2000,p.203). Lexical sophistication measures the percentage of 'advanced' words compared to frequent words in a text.

As Laufer and Nation (1995) highlight, what is considered to be advanced lexis would depend on the researcher's definition, yet at the same time, such a definition would also depend on the researcher's particular context and research aims. As an example, proficiency level would be a factor in making decisions about whether a vocabulary item can be considered advanced. A word can represent advanced lexis for low proficiency learners while at the same time be considered as unsophisticated with higher proficiency students. This suggests that producing consistent results based on the lexical sophistication index can be difficult. From this perspective, the index can be considered prone to instability and difficult to apply with objectivity and robustness (Laufer and Nation, 1995).

The impact of this subjectivity is notable in the literature, for example, in the investigation of the quality of compositions written by forty-two Swedish learners of English, compared to those written by twenty-one native English speakers. Linnarud (1986) only considered English words introduced at grade nine or later as sophisticated English vocabulary. This was on the basis of the grading system in

Sweden where grading of vocabulary was, at the time, largely based on considerations of frequency (Faerch,1987) The study revealed significant differences between native English speakers and the Swedish English learner group in terms of lexical sophistication.

Notwithstanding the above, in a comparative study of near-native second-language learners of Swedish vs. native Swedish speakers, Hyltenstam (1988) investigated lexical usage and defined sophisticated lexical words as those words beyond the 7,000 most frequent Swedish words. The results obtained from this study did not reveal any significant differences between the two groups. While the inconsistent results may be an artefact of the different participant groups investigated, the fact that lexical sophistication was conceptualised in different ways makes it difficult to compare the results, hence the cited limitations of the lexical sophistication index as far as objectivity is concerned (Laufer and Nation, 1995).

2.6.4 Lexical variation

The lexical variation index records the ratio between types and tokens in a text and is also referred to as lexical diversity in the literature (Malvern et al., 2004). It captures the range of a learner's vocabulary as displayed in his or her language use (Lu, 2012). In other words, lexical variation can be viewed as simply the number of different words used in a written text or the type/ token ratio, that is, the ratio in per cent between the different words in the text and the total number of running words (Laufer and Nation, 1995). A major criticism of this measure is that it can be affected by differences in texts because, the longer a text is the more likely that certain words will be repeated. In addition, repeated words are likely to be high

frequency words (McCarthy and Jarvis, 2007) which will impact on the resultant type/ token ratio. In addition, as was highlighted in section 2.1.1, a number of conceptualisations relating to what a word is exist. Consequently, the lexical variation ratio will be affected by the definition which a particular researcher opts for.

As Laufer and Nation (1995) note, if derivatives are considered to be different words, then this will increase the type-token ratio but if a word family is considered to be one word, the opposite effect is likely to be achieved. This means that researchers have to be careful firstly, about how they define a word in such contexts, and secondly about what conclusions can be drawn from the findings. As an example, in cases where a word and its derivatives are considered as different types, lexical variation may not necessarily be as indicative of linguistic achievement as may otherwise be expected.

A more recent lexical richness tool is the Vocabprofile tool which is an online vocabulary profiling tool. A more comprehensive discussion of this online tool is offered in Section 2.6.5 below since this is the tool, which is utilised for the analysis of learners' lexical use in the current study.

2.6.5 The VocabProfile tool

Vocabprofile is a freely available online tool, which uses frequency lists to compute lexical use in a written text. The tool breaks down vocabulary in a text to show the percentage of words that a writer uses from each one-thousand-word frequency band from the first to the twenty-five-thousand-word frequency band. The classic (original version) of the Vocabprofile tool categorises words into four bands comprising K1 (1-1,000); K2 (1,001-2,000); AWL (Academic Word

List) and Offlist. However, the tool has been developed to account for a much wider range of frequency bands so that in its current state, VocabProfile can measure learners' vocabulary sizes as extensive as the 25,000 word level (Nation and Beglar, 2007). This makes it a very comprehensive automated lexical profiling tool, hence its extensive use for research purposes (Nation and Heatley, 1996; Laufer and Nation,1995; Akbarian, 2010).

However, a criticism that can be raised against the VocabProfile tool pertains to its treatment of the concept of a word as a single unit. Since current conceptualisations of word knowledge include multi-word units, a welcome development on the VocabProfile tool for both research and pedagogic purposes would be the inclusion of multi-word units.

From a DST perspective, such a development would be helpful for tracing learners' individual trajectories in their vocabulary learning so that gains (and losses) in vocabulary learning can be picked up both in terms of word knowledge development at the level of words as single units and word development knowledge at the level of words as multi-word units. This is particularly so because multi-words have been found to cause a lot of challenges not only to beginner learners but also to advanced learners (Kuiper, et al., 2006). Therefore, comprehensive methodologies for tracing the development of lexical knowledge can be expected to have the ability to trace not only knowledge of single words but also, knowledge of words as multi-units.

As part of the preparation of texts before the automated analysis can be completed, it is recommended that basic spelling errors which do

not compromise recognition of the particular word are corrected as this enables the computer programme to recognise the words (Laufer and Nation, 1995). It is only the incorrectly used words that should be discarded together with proper nouns. This is important in order to avoid inflating vocabulary profiles for the texts under analysis.

At first glance, the fact that basic spelling errors are corrected as part of text preparation for the automated analysis can be a basis for criticism. It could be argued that since Vocabprofile computes words that have been partially corrected for form, to a certain extent, it rewards learners not only for known words but also, for word knowledge which is only partially developed. However, I would argue that it makes sense for imperfect knowledge to be acknowledged and rewarded because of the multidimensional nature of word knowledge (see Section 2.4.1) which inevitably means that even for native speakers, lexical knowledge is always in a state of development. Arguably then, failing to acknowledge partial knowledge is, in essence, failing to acknowledge a key attribute of the process of vocabulary learning; that is, that development is gradual along different dimensions. The question perhaps should be the degree to which partial knowledge should be acknowledged – in other words, a question of how to acknowledge partial knowledge as opposed to whether to acknowledge it. To this end, one would expect the Vocabprofile program to differentiate between partial knowledge (as represented by words that have been corrected before the program carries out the analysis) and words that have been mastered to a level that is adequate for Vocabprofile analysis without correction. This would avoid potentially over-estimating learners' vocabulary knowledge.

2.7 Writing in a foreign language

Weigle (2002, p.1) notes that 'the ability to write effectively is becoming increasingly important in our global community'. As such, instruction in writing, including its assessment constitute an integral part of any language English Language teaching programme. Writing assessment is key because outcomes from such assessments are used to make far–reaching decisions such as enrolling students onto courses. On the surface, writing assessment can be seen as a straightforward process whereby language learners complete a writing activity and a teacher or examiner evaluates the piece of writing and allocates a score which reflects the quality of writing. However, both writing and its assessment have long been recognised as complex activities (Weigle, 2002). Writing is affected by a number of factors operating at both the macro and micro levels, and such factors include (but are not limited to) the writer's purpose for writing, understanding of audience, understanding of text characteristics, linguistic resources available to the writer and/ or cultural expectations (Friginal, et al., 2014; Jarvis et al., 2003). The current study acknowledges that writing is a multi-componential skill and so its development and mastery is likely to be influenced by numerous factors including but not limited to lexis (Uccelli, eta l., 2013).

Cumming (2001) suggests that perspectives and conceptualisations of learning how to write in a second language have expanded and have been refined over the years. As Cumming notes, researchers in this area have identified at least three main areas of focus; firstly, research which focuses on the quality of texts that learners produce; secondly, research which focuses on the processes of students' composing and finally, research which focuses on the specific socio-

cultural contexts in which this learning occurs. However, in practice, these domains integrate during the writing process rather than operating as distinct facets. This highlights the complexities involved in the process of composing, and indeed, evaluating texts. For second language learners, the process of composing text is likely to pose even more challenges because not only do ESL learners have to deal with the coordination of multiple cognitive challenges associated with writing in general, but they also face further complexities associated with writing in a non-native language such as limited grammatical, syntactic and vocabulary choices. This explains why the current study found the investigation of vocabulary knowledge and its role in written language proficiency a worthwhile research agenda.

2.7.1 Assessing Writing

Considering its contribution to high stakes English Language assessments such as the International English language testing system (IELTS), it is indisputable that the skill of composing text (writing) is considered an essential part of language proficiency and communicative competence. At surface value, assessing writing is a straightforward process whereby learners produce written text and the assessor evaluates the writing and allocates it a score. However, as the following discussion will show, the assessment of writing ability poses a number of challenges both to the assessors and those being assessed.

As Stomp (2012, p.81) notes, 'the construct of writing ability is broad, multifaceted, situated, contextual, and resistant to a monolithic, stable definition'. For that reason, the concept is 'difficult to pin down'. Therefore, as Weigle (2002) notes, a key consideration

that should be made when assessing writing ability is the purpose for which writing is carried out, in other words, the reason why it is important to evaluate writing ability for a particular group of learners. Bachman and Palmer, (1996) identify two main purposes for the assessment of writing. These are to make inferences about language ability and to make decisions on the basis of those inferences (Bachman and Palmer, 1996) Inferences can be made at an individual, classroom, or even program level (Weigle, 2002). At an individual level, inferences can be made about a particular learner's performance in accordance with a particular grading system. Since an individual's language ability does not lend itself to direct observation, it is their response to test items which forms the basis for making inferences about the individual's writing ability (Weigle, 2002).

An important distinction about writing tests is whether they are high stakes or low stakes test (Amrein and Berliner, 2002). This determines the importance of inferences that are made on the basis of these tests. As Weigle (2002, p.40) notes, "high stakes decisions have a significant impact on the lives of individuals or on programs, and are not easily reversed, so that errors in these decisions can be difficult to correct". Such tests are more likely to lead to the occurrence of positive washback (see Section 3.3.7) because of the implications of any inferences made from such tests such as admission to a program of study. For this reason, a number of considerations have to be taken into account when designing and assessing language tests in general, of which written language is a key part.

Bachman and Palmer (1996) use the term 'performance assessment' to refer to assessment procedures which are based on behaviours ob-

served in authentic, real life situations or simulations of such situations. To that extent, "any writing test that involves actual writing, as opposed to completing multiple-choice items, for example, can be considered to be a performance test since the written product represents a performance of writing" (Weigle, 2002, p.46). This is an important consideration of test design as it contributes to the usefulness of the test to the test-takers. In the current study, students completed free writing tests which are well aligned with Bachman and Palmer's (1996) view of performance tests.

2.7.1.1 Rating scales

Rating scales can be mainly classified into multiple-trait, holistic, and primary-trait scoring methods (Barkaoui, 2007). Multiple-trait scales involve assigning multiple sub-scores to individual traits or dimensions such as language and organization (Weigle, 2002). These sub-scores are then summed up in order to obtain an integrated score for the whole text. In holistic scoring procedures, the assessor evaluates performance indicators as stipulated by the rating scale and only one integrated score is assigned as a judgement of the overall performance in a piece of writing (Goulden, 1992). The primary-trait scoring approach entails the assessor making an overall judgement based on whether the specific aim of the task has been fully addressed (Goulden, 1992; Weigle, 2002).

Hamp-Lyons and Henning (1991) argued that highly structured scales such as multiple-trait scales are more likely to increase inter and intra-rater reliability because of the more focused guidelines on each key trait on the scale. This is likely to reduce the likelihood of a rater assigning disproportionate values to some traits compared to others.

On the other hand, holistic scoring is likely to increase opportunities for assessors to include additional traits not listed within the scale criterion and use personal judgment to determine the importance of a specific trait to the overall score (Goulden, 1994).

In recognition of the complexity of setting standards by which to objectively evaluate student writing, Broad (2003) notes that assessing writing quality can be notoriously obscure in nature. As such, it is untenable to set one single written standard that can be universally accepted as representative of the ideal written product in English. In the same vein, Bacha (2001) argues that perhaps the best advice for educators faced with the challenge of choosing which scale to use is to consider the purpose of essay writing as well as the content and construct validity of the essay task. Content validity refers to a scale's ability to evaluate "writers' performances on the kind of writing tasks they are normally required to do in the classroom" (Jacobs, et al., 1981,p.74). Construct validity on the other hand refers to how effectively an instrument is able to distinguish between, or among, abilities within the parameters of what it sets out to measure (Bacha, 2001). As can be expected, it is important to have such measures in place if reliable results are to be obtained. Rating scales can impact on the marking of free compositions and ultimately, the scores that assessors allocate to a piece of writing (Barkaoui, 2007).

In a mixed-methods study seeking to investigate the effects of two different rating scales on EFL essay scores, (Barkaoui, 2007) investigated the use of a holistic scale and then a multiple-trait rating scale among Tunisian EFL teachers. A set of 24 EFL essays were rated and scores from the two scales were analysed and compared. The

finding was that the holistic scale resulted in higher inter-rater agreement. Section 3.4.2 discusses the rationale behind the scale that was utilised in the current study.

2.7.2 Writing Process Theories

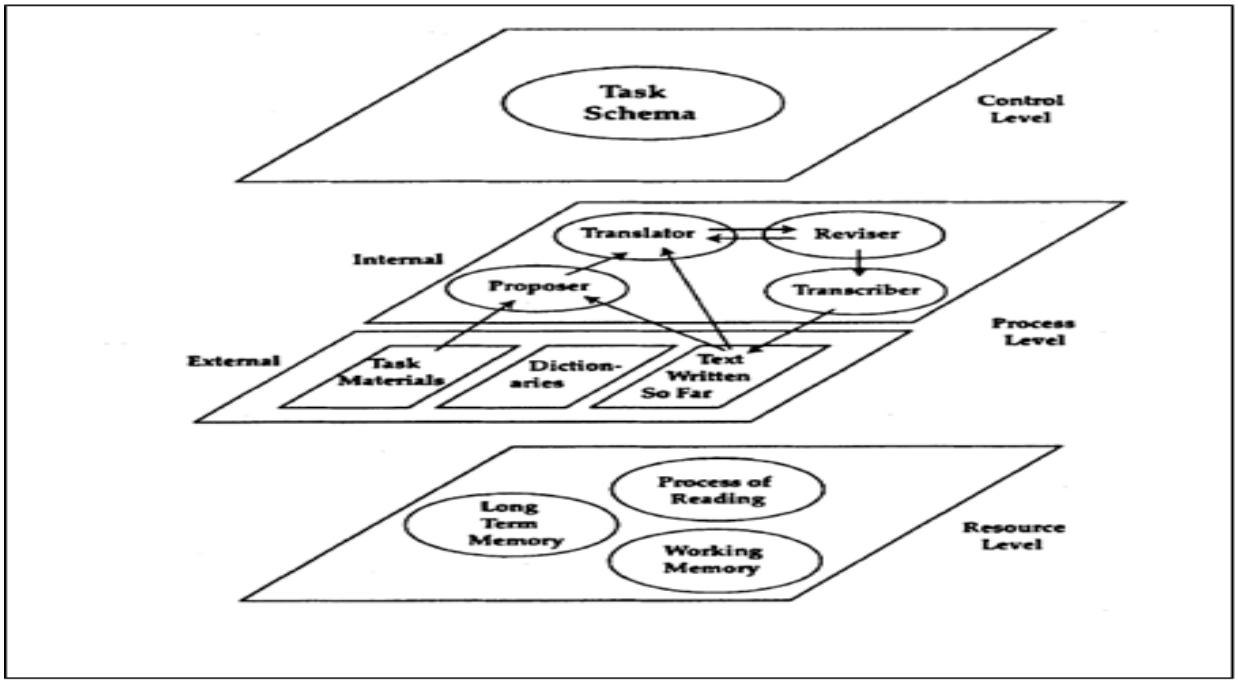
Cognitive writing process theories have been used as a framework for understanding the cognitive processes that writers go through when composing text (Flower and Hayes, 2008). One of the central views related to cognitive process writing theories is that 'writers are constantly, instant by instant, orchestrating a battery of cognitive processes as they integrate planning, remembering, writing and re-reading' (Flower and Hayes 2008, p.387). Thus, from this perspective, writing is not just about the end product (written piece of writing) but also the process which writers go through. As such, the process of writing is considered to be an artefact of the degree of writing expertise as evidenced by the different strategies that writers adopt, the different skills that they display and the different processes that they go through.

It is largely because of the simultaneous orchestration of different cognitive skills during the writing process that writing presents considerable challenges to both native and non-native novice writers (Hayes, 1996; Kormos, 2012). Cognitive writing process theories shed light on the complex combination of skills and processes that skilled writers adopt in order to produce text.

By so doing, such models provide opportunities for researchers and practitioners to better understand the main processes and variables that may be at play during the writing process. This has practical

implications because such models can provide a lens through which a complex process such as writing can be understood. Such models also highlight possible challenges that those learning to write may face. Cognitive Process models were first developed on the basis of results obtained from the analysis of data from think -aloud protocols, which suggested that the writing process encompasses multiple domains as shown in Figure 2.12

Figure 2.12 Model of written language production from a cognitive writing process theory perspective



Source: *Chenoweth and Hayes (2001, p.84)*

As Figure 2.12 shows, according to Chenoweth and Hayes' (2001) model of written language production, the writing process is made up of three main components; the resource level, the process level and the control level. The process level is at the centre of the writing process because this is where text production fundamentally takes place (Chenoweth and Hayes, 2001). This level is facilitated by resources from the resource level and operating within the framework

provided by the control level. Within the process level, Chenoweth and Hayes (2001) distinguish between external and internal elements and processes.

External elements include the task material, dictionaries and text written so far while internal processes are made up of four elements - the proposer, the translator, the reviser, and the transcriber. The proposer is responsible for the generation of pre-linguistic ideas to be expressed to satisfy goals of the writing task; the translator then converts these pre-linguistic ideas into language constructions such as phrases, clauses and sentences with appropriate word order and grammar. The reviser evaluates this information to determine its suitability and appropriacy, while the transcriber turns all this content into written language.

The reviser may accept or reject ideas and lexicogrammatical output or potential output (before or after it has been transcribed) depending on whether this output is judged / evaluated as being appropriate or not within the relevant framework for the task. The framework for judging appropriacy of output operates at the Control Level. If the content is rejected, then the process might have to start again resulting in revision or multi-revision of ideas and/or text. The resource level, as the name suggests, provides a supply of resources required during the text production process. These may include information for the generation of content in a free writing task which is stored in the writer's memory.

As such, the Long Term Memory (LTM) plays a major role during text production because without resources (information) from the LTM, it would be difficult for writers to generate appropriate content for free

writing activities. Another key component of the model is text that the reader has already produced; hence the 'process of reading' is highlighted as a key constituent of the resource level. This is because through reading what has already been written, the writer is able to generate further relevant content to enable completion of the writing task.

From a slightly different but related theoretical perspective, the cognitive load theory suggests purports that working memory is limited in capacity but that these limitations are drastically reduced or eliminated when dealing with familiar information that has been learned and stored in cognitive schemas (Paas and Ayres, 2014). When a particular skill requires a lot of effort for it to be executed, cognitive overload occurs as a result of the individual's cognitive capacity for that particular activity being superseded. This can lead to the individual's cognitive system consuming disproportionately more cognitive resources which would otherwise be used for executing other important skills within a particular cognitive process.

In the context of the current study, this implies that if any of the skills required for effective writing is not sufficiently developed, for example, lexical knowledge and/or content knowledge related to the topic that the writer is tasked to write about, then cognitive capacity is likely to be superseded. In the context of scores obtained from a free writing task, this would mean that an insufficiently developed lexical repertoire is likely to be related to low scores not just on the language-related assessment criteria (the Language category in this study), but also the other key categories (the Task and Organisation categories in this study). This supports what has been termed the inhibition theory (van Gelderen et al., 2010). According to the

inhibition theory "inexperienced writers' inefficient use of grammatical and lexical knowledge impedes their monitoring of text production on the level of content and their use of higher order strategies for optimising text quality on a global level" (van Gelderen et al. 2010, p283). This means that as a result of the cognitive load presented by dysfluency at the lexico-grammatical levels, less proficient language users may fail to be oriented to text-level skills in writing such as the generation of relevant and appropriate content and ensuring the production of coherent text.

In relation to the process writing theories, the implication of the inhibition theory is that the better each of the resource level skills such as lexis and grammar is learnt and automatised, the more efficiently the whole writing process can be executed therefore leading to higher scores in a free writing task such as the one that was administered to participants in this study. This is because efficient retrieval of lexis, for example, is likely to free cognitive resources that can be used for generating relevant content and organising writing in a more coherent manner. This suggests that a skill or linguistic attribute that improves the rate at which resources such as lexis can be retrieved from the mind during a productive task may have a positive impact on the overall efficiency of the productive task such as the compositions produced by the ESL students in the current study.

van Gelderen et al., (2010) highlight an opposing view offered by the compensation theory. According to the compensation theory, underdeveloped linguistic skills such as an underdeveloped mental lexicon do not necessarily inhibit other cognitive processes. Instead,

the compensation theory argues that global tasks such as organisation and coherence in text can in fact be fulfilled in a sequential rather than simultaneous manner as suggested by the inhibition theory and related cognitive load hypothesis. This would mean that even with limited availability and retrieval of certain linguistic skills such as lexis, it would be possible for novice writers to efficiently deploy cognitive abilities in a way that would allow them to achieve high scores in other aspects of writing such as organisation, coherence and task fulfilment. Van Gelderen et al. (2010) investigated the effect of a writing course with a lexical training element on free writing scores.

The researchers subjected two participant groups (Dutch speakers learning English) to different treatments where one group received tuition in writing skills including lexical fluency training while the other group received tuition in writing skills without the lexical fluency element. Written language scores of both experimental groups were then compared to a control group which did not receive any training at all. The researchers found improvements in global text quality of both groups that received either of the training types but no improvements were found for the control group. This led the researchers to conclude that the experimental treatments were effective. However, the fact that the group that received lexical fluency training did not perform any better than the group that received more generalised writing skills training led the researchers to conclude that the inhibition theory could not be empirically justified in the context of their study. This will be discussed further in Chapter 7, in light of the findings from the current study.

2.7.3 Scores from written compositions

Since the main aim of the current study is to explore the relationship between vocabulary knowledge and written language ability among ESL learners, a key part of the current study pertains to the ratings allocated to students' writing on the basis of linguistic competence evident in these written samples. While such written language samples provide a very useful measure of learners' linguistic resources, it is important to note that such measures may not provide an accurate measure of learners' linguistic resources particularly at the level of lexis.

This is because learners may choose to use certain words not because such words are truly representative of their lexical resources but because their knowledge of certain lexis is not yet secure. As such, to avoid making lexical mistakes and consequently being penalised for this, learners may simply avoid using any vocabulary that they feel they do not understand very well. As an example, Schmitt and Zimmerman (2002) note that even though derivational knowledge presented challenges to the L2 learners that they investigated, derivational errors are not as widespread as results of their research would suggest. The authors hypothesised that one reason why certain errors that could otherwise be expected from a particular L2 group may not be evident in learners' productions is that learners may choose to avoid structures that they find difficult to deal with such as derivative forms. While such a strategy might help learners effectively navigate the terrain, it unfortunately means that learners may, as a consequence to the avoidance strategy, deprive themselves of the opportunity to utilise some of their lexical resources with the result that they may present as learners with limited mental lexicons.

Therefore, the vocabulary knowledge assumed for the learners in the current study may not fully capture the learners' actual vocabulary knowledge as the free writing scores captured evidence of lexical knowledge based on what the learners chose to produce within the limitations of a free writing task. It would be beneficial for future studies investigating productive language to tap into partial knowledge which learners avoid using.

2.7.4 Formative and Summative Assessments

As detailed in Section 1.1.1, part of the rationale out of which the current study was conceived is linked to observations related to the assessment of vocabulary writing in the L2 context. It was the observed inconsistencies between students' initial assessment scores on language performance overall, and lexis in particular, and their ultimate performance on the course as reflected by their written language skills, that prompted questions regarding the interplay between vocabulary knowledge and written language skills. As such, it is important to provide a brief overview of assessment practices relevant to L2 learning and teaching.

Traditionally, a distinction is generally made between summative and formative assessment activities (Johnson and Jenkins, 2009). To that extent, Colby-Kelly and Turner, (2007, p.11) state that summative assessment is 'the process of seeking and interpreting evidence for making substantively grounded decisions or judgements about the product of a learning task'. By contrast, formative assessment is viewed as:

> 'the process of seeking and interpreting evidence for making substantively grounded decisions or judgements about the product of a learning task in order to decide where the learners are in their learning, where they need to go, and how best to get there' (Colby-Kelly and Turner, 2007, p.11).

Therefore, a major difference between the two forms of assessment relates to the intended purpose for the evidence obtained from an assessment activity. If the fundamental purpose of an assessment is to elicit evidence to be used for further action-planning to improve future performance, then it serves more of a formative than summative purpose (Black and William, 2006).

Formative assessment constitutes an integral part of learner-empowering pedagogical methods and can lead to enhanced classroom learning (Cheng, et al., 2004; van de Watering and van der Rijt, 2006). It is therefore not surprising that in recent years, educators have encouraged and supported learners to take more responsibility for their own learning. With appropriate training, learners can utilise a wide range of formative assessment methodologies both in the classroom and independently, for example, in the form of independent self-assessments, peer–to-peer assessments and technology-based practices (Colby-Kelly and Turner, 2007). It is for this reason that formative assessment is considered to be aligned with a more recent dynamic assessment approach referred to as Assessment for Learning (AfL) (Leung, 2007). Owing to its interactive nature and focus on learner involvement, AfL gives learners the opportunity to actively participate in their own learning and assessment through teacher-student collaborative setting and independent, student-led monitoring of learning goals supported by reflection on feedback obtained either from tutors or technology-based assessment tools such as the VocabProfile tool utilised in the current study (Black

and Wiliam, 2006). To that extent, AfL empowers learners and encourages autonomy which is a contributing factor towards achievement as it also helps teachers to better identify what learners need to learn (Colby-Kelly and Turner, 2007). Table 2.2 summarises the main features of AfL as opposed to AoL practices:

Table 2.2: A comparison of AoL and AfL assessment practices

	Assessment of learning (AoL)	**Assessment for Learning (AfL)**
Time	Carried out at the end of an instructional unit	Carried out during a unit of instruction
Purpose	Measures and summarises students' knowledge and skills; designed to certify learning and to make judgments about students' progress; serves summative purposes	Monitors and supports learning; designed to improve learning; serves formative purposes
Form of assessment	Takes the form of tests or examinations	Relies on a wide range of assessment data – e.g. questioning, observation and conferencing
Feedback	Feedback is expressed in the form of marks or grades that distinguish high-perfuming students without giving much information about students' mastery of skills and concepts; serves reporting purposes.	Feedback highlights students' strengths and weaknesses and provides descriptions that inform individual students' learning, helping teachers scaffold the next steps of instruction
Teacher and student role	Teachers dominate the assessment process	Teachers play a central role, but they share responsibility with students by involving them in the assessment process.

Source: Lee (2011, p.20)

As can be seen from Table 2.2, AoL represents 'a paradigm shift from a focus on product to one on process in language assessment' (Lee, 2011, p.19). As will be further explored in Section 2.8, an important epistemological framework in this study is that L2 development is a dynamic and complex process (Larsen-Freeman, 2012). To that effect, this theoretical framework resonates with AoL assessment practices.

2.7.5 Course books in ESL teaching

It is not within the scope of the current study to provide a comprehensive evaluation of the role of the course book in the ESL learning and teaching context. However, considering the role of the course book in ESL programmes where many courses (including the ESL programme from which the participants for this study were drawn) tend to have at least one identified core text and/or course book package, it is important to provide an overview of the role of the course book. This is because if the course book plays such a pivotal role in the teaching and learning of L2 skills in general and lexis in particular, then the learning practices that learners adopt are likely to be impacted by the framework provided the course book. To that extent, in a study which explores the interplay between lexis and written language production not only on the basis of performance data (from assessments completed by the participants) but also on the basis of learners' conceptualisations of what it means to know a word, it is important to consider the course book as a tool that may shape learner cognitions.

On the basis of the view that teachers' and students' beliefs about course books can impact on the ultimate utilisation of such resources, McGrath (2006) investigated the metaphors and similes that students and their teachers used to describe course books. McGrath (2006) collected and compared metaphors used by each of the two groups within the same teaching/ learning context in Hong Kong. The analysis of findings from the 75 teachers who were involved in the study revealed both positive and negative attributions to course books. Positively, course books came out as a resource; a source of guidance and a support mechanism. For example, metaphors such as 'a text-

book is like oil in cooking—a useful base ingredient'; 'textbooks are like ladies' handbags because we can take what we need from them and ladies tend to take handbags wherever they go' and ' a textbook is the stone from which a sculpture will be made (needing bits chopped off, added on and occasionally a little crushing)' were quite revealing as they all highlighted very positive beliefs about course-books (McGrath, 2006, p.174). On the negative, course books were viewed as a constraint where metaphors of course books as 'road-blocks' 'straitjacket' and 'millstones' signified less favourable images of course books by teachers (McGrath, 2006, p.174).

Similarly, data from several hundreds of students highlighted both positive and negative attributions. Over and above the four categories obtained from teacher data, student data also produced metaphors which were categorised as indicative of course books as sources of boredom, instilling fear and anxiety as well as worthless tools.

It is worth noting that although both the teacher and learner groups in Mcgrath's (2006) study produced a combination of positive and negative attributions about the value of course books, the learner group produced more extensive negative attributions than the teachers. McGrath (2006) concluded that this discrepancy suggests that 'there is value in teachers researching their learners' beliefs and attitudes—in relation to course books and other aspects of the teaching-learning environment—and reflecting on and comparing these with their own' (McGrath, 2006,p.171). Since the course book plays such an important role in ESL teaching and learning, McGrath's (2006) observation that the learner perspective on the selection of course materials they use has been relatively neglected in the field is a matter of concern because as the foregoing discussion has highlighted,

learners can be quite discerning and sensitive as captured by the metaphors used to describe course books in McGrath's (2006) study.

In the context of the current study, this lends support to the epistemological stance taken to give voice to learners' views, beliefs and attitudes through stimulated reconstructions/ semi-structured interviews. McGrath's (2006) highlights the debate around the value of the course book in L2 teaching. Those in favour of course books argue that a course book can be a valuable resource for both students and tutors alike. Within this line of argument, the use of course books gives students the opportunity to look ahead into work planned for the future as well as refresh and therefore reinforce previous learning (O'Neill, 1982; Gutiérrez-Bermúdez, 2014). This is an important factor because textbooks can encourage and provide opportunities for self-directed learning (Cunnings (1995). This is key to the development of autonomy and independence; a skill which is highly valued in tertiary education (Tomlinson and Masuhara, 2013).

Haycrot (1998) argues that the textbook provides assurance and psychological security to students who may feel that their achievements and progress can be measured against the standards set by the textbook. Such standards tend to be perceived as more credible than those offered through materials designed in house by the teacher (Haycroft, 1998). This view finds support in literature which cites one of the main advantages for the use of course books as being 'the element of quality assurance, as a well-developed course book exposes students to tried and tested materials based on sound learning principles (Richards, 2001, p.13). In today's tertiary education terrain where education in general and many ESL programs in particular are

associated with high tuition costs, such views can be expected to constitute an important consideration for educational institutions.

Another reason why course books can be the preferred resource for many teachers is its ability to alleviate time pressure as it is a readily made resource. In addition, since course books are compiled by experts in the field, teachers expect course books to provide systematically structured content which facilitates appropriate instruction and development at each level, based on current theoretical approaches and teaching methodologies (Tomlinson and Masuhara, 2011). Since commercially produced course books come in a range of levels designed to guide teaching and learning from the most basic stages all the way through to advanced levels of proficiency, the course book provides teachers with consistent support, direction, and a variety of activities to help maximise returns from classroom time (Mares, 2003).

For inexperienced teachers in particular, course books may be an invaluable tool as such teachers may lack the confidence and / or requisite skills for compiling their own teaching resources (Ruby, 2003). As such, it is not surprising that the textbook is widely accepted as a universal element in ESL learning and teaching and in many contexts, it seems that a learning and teaching programme is incomplete without a relevant course book attached to it (Hutchinson and Torres, 1994). Not only do most EFL courses have an identified course book, but also, most modern course books are accompanied by a student workbook with a corresponding answer book as well as the teacher's resource book, sometimes including associated websites and digital resources which together constitute the course book package (Allen, 2015). As Allen notes, even with the digital era making inroads into

the ESL teaching and learning context, the course book package continues to hold a central position even amongst the new generation of digital native teachers and learners.

Arguments against course books include the view that course books can be superficial, reductionist and stifling to teachers' creativity and innovation (Hutchinson and Torres, 1994), and this is one of the reasons as to why their ability to facilitate effective teaching and fully cater for the diverse learning for the diverse needs of learners has been questioned by some. As Gray (2010) argues, ESL materials need to meet the communicative needs of learners; the need for learners to fulfil their long-term goals; the need for authenticity while being student centred. Those who argue against course books suggest that course books are not always able to meet such core needs for learners (Soleimani and Dabbaghi, 2012).

Arguably, not every course book can meet all the learning needs of all ESL learners considering that ESL is delivered to students from across the globe which means that the materials are targeted to an extraordinarily diverse population. This makes it difficult, perhaps impossible, for material writers to develop materials that can be a perfect fit for any teaching/learning situation. For this reason, it seems reasonable to argue that even the best textbook in the market should be used discerningly with teachers adapting course books, in order to provide the best fit between the course book and their learners' needs (Cunningsworth, 1995). Such adaptations may include considerations of the specific environment in which the learners are situated, demographics within the classroom; the learners' objectives and aspirations related to learning English as well as the range of abili-

ties/proficiency levels represented within a group of learners. This gives some flexibility to teachers together with opportunities for creativity and innovation, which, coupled with the direction and support provided by the course book, can result in a highly effective curriculum. It would therefore appear that if used discerningly, a good course book has a major role to play in the ESL teaching and learning context. Since the course book can be such an important part of any ESL programme, a pertinent question that arises is that of the extent to which course books can be considered to meet the needs of teachers and learners as major stakeholders in the process of learning and teaching.

As highlighted in Section 1.1.2 one particular area that has gained primacy in the field of Applied Linguistics in recent years is the importance of lexical knowledge as a part of language development for English L2 learners. Since research suggests that lexical knowledge is an important part of linguistic development, then the expectation is that this research should be evident in the manner in which lexical knowledge is reflected in course books. As (Cobb, 1995) notes, many commercial ESL books now claim to adopt a lexical approach or to contain a strong lexical component (Cobb, 1995; Eldridge and Neufield, 2009). For example, the introductions to The New Cambridge English Course emphasizes that the books offer opportunities for systematic vocabulary learning. Similarly, another widely used course book, New Headway, has been described by its publishers as one that provides a strong lexical syllabus (O'Loughlin, 2012) while Cutting Edge is considered to be one that places particular attention on high-frequency vocabulary (O'Loughlin, 2012). All of these examples are positive developments because they suggest that vocabulary

learning has started getting more attention not only from researchers but from material writers as well. However, as course books are in essence commercial products in the market, a major concern from a pedagogical perspective is that in some cases, the publishers might prioritise their need to make profit instead of the pedagogical aims and needs of the material users (Gutiérrez-Bermúdez, 2014; Sheldon, 1988). As a result, the need for profitability can get in the way of the design of effective materials (McGrath, 2013).

Following from the above, it is therefore imperative for key stakeholders such as researchers, teachers and programme managers to evaluate the extent to which course books do actually meet the manifestos that they provide in terms of their course book offers. Since commercially published course books can be a key resource for teaching and learning in the ESL context, it goes without saying that ESL students will be looking up to the course book as an important part of their L2 learning in general and their vocabulary development in particular. In the context of the current study which investigates the interplay between vocabulary knowledge and written language ability, the question of how vocabulary knowledge is treated in ESL course books becomes relevant.

2.8 Theoretical background of the study

Larsen-Freeman and Long (1991, p.227) posited that 'at least forty theories of SLA have been proposed' to date. The multiple theories seem inevitable in light of the complexity of the second language development processes. However, none of these theories present a complete explanation for the phenomenon (Menezes, 2013). From a linear model perspective, this can be problematic because each model

/ theory explains the process from its particular epistemological position and is therefore unlikely to present a complete picture. For example, behaviorism, acculturation, universal grammar hypothesis, sociocultural theory and connectionism are some of the most well-cited broad theoretical perspectives in the literature (see Menezes, 2013, for a comprehensive discussion of these). While each theory is capable of explaining L2 phenomena from its particular perspective, this raises the question of how different theoretical perspectives can be reconciled (Menezes, 2013) in order to take advantage of insights offered by different theories. It is for this reason that DST has been proposed as a paradigm which reconciles different theoretical perspectives that are necessary to understand several puzzling L2 phenomena (Dornyei et al., 2015).

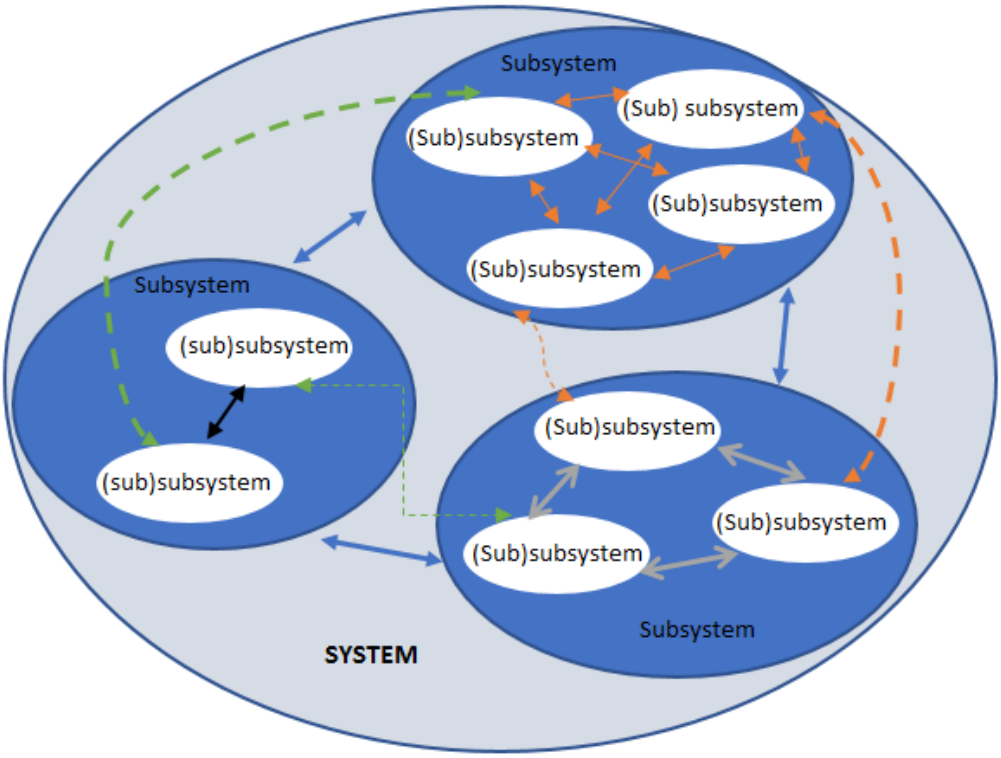
2.8.1 Dynamic Systems Theory (DST)

Traditionally, variability is considered to be a result of measurement error in research, hence it has been referred to as 'noise' in data and therefore not worthy of investigation (Dijk and Geert, 2007; Thelen and Smith, 1994). As Schermer (2012) observes, such views give primacy to paradigms of linear cause and effect with the vision of "an orderly universe where paradox, probabilities, turbulence and disorder are viewed as incidental 'noise' in the system that need no further explanation (Schermer, 2012, p.277). This would justify the total exclusion of non-linear occurrences in research insights, in favour of linear cause and effect findings. However, DST provides a lens for legitimately viewing and understanding complex phenomena and their development (Hasko, 2013; Larsen-Freeman, 2012; van Geert, 2007). Judging from the recent increase in studies that adopt a DST approach to second language research (see for example

Svalberg and Askham, 2015; Dornyei, 2014; Caspi and Lowie, 2014; Verspoor and Schmitt, 2012; Spoelman and Verspoor, 2010; van Geert and Steenbeek, 2005; van Geert, 2007), DST offers a promising avenue for researching and understanding of L2 processes.

Process and change are at the centre of complex dynamical systems (Rosmawati, 2013). Therefore, DST is a theory of development which focuses not only on the result of change but also, on the process of change. Dynamic Systems are complex systems because they comprise multiple components / subsystems which do not operate in isolation but are intricately interlinked and interdependent for the survival of the system. De Bot et al. (2007) argue that while growth / development is at the centre of a dynamically complex system, an essential condition for such growth to take place is the availability of resources as these ensure the sustenance of the system. The authors distinguish between internal and external resources, where internal resources are considered to be resources from within the learning individual such as their capacity to learn, the time that they are willing to invest into learning, as well as their motivation to learn. External resources are considered to be those resources that exist outside the second language learner (De Bot et al., 2007). To that effect, the authors cite the learning environment as well as resources offered by the environment such as books and the media as examples of external resources. Figure 2.13 provides a simple illustration of an interconnected dynamic system.

Figure 2.13 Illustration of multiple factors that interact within dynamic systems



Source: *Adapted from Rosmawati, (2013, p.68)*

Figure 2.13 illustrates the multiple interconnected components / sub-systems of a system. As the coloured lines show, the components / subsystems are nested in the system but there can also be further (sub)subsystems nested in the subsystem, giving rise to a complex system which contains many nested, interconnected, interdependent systems (Rosmawati, 2013).

In complex systems, interactions between components give rise to an important feature, namely dynamic change. Both the components of the system and their interactions are subject to change over time. Consequently, the system and the interactions within it continuously

shape themselves based on the output from the interactions at the preceding stage. In the literature, this feature is typically described as sensitive dependence on initial conditions (Larsen-Freeman, 1997) and leads to another central concept of complex dynamic systems namely emergence. Emergence refers to the process whereby behavioural patterns or degrees of regularity emerge from the interaction of components within a system even though such patterns are not exhibited at individual component / subsystem level. Thus, 'the behaviour of complex systems emerges from the interactions of its components, it is not built into any one component' (Larsen-Freeman, 1997) but emerges from the complex interactions.

Another important feature of development from a DST perspective is that it is typically described as being chaotic (Larsen-Freeman, 1997). Traditional notions of chaos bear connotations of disorganisation and disorderliness but within the DST framework, the term carries wider implications of a system which is capable of producing disproportionately large changes as a result of small interventions. In other words, small interventions on the system may produce unexpected results and vice-versa.

A commonly-cited analogy in this regard is that of the butterfly effect where a single flap of a butterfly can produce future weather changes as significant as a tornado (Larsen-Freeman and Cameron, 2008). It is this sensitivity to initial conditions which renders complex systems unpredictable. In the context of second language learning, unpredictability might mean that students in a learning environment may be exposed to the same input from the same teacher and have the same access to learning materials. However, the results of the pedagogical

intervention may be far from similar and the reasons for this variation may not be obvious. If learners in the same class receive similar levels of input from the teacher and similar levels of access to resources, all things being equal, it could be expected that the same intervention could produce relatively consistent results. However, this is not always the case and the variability can be attributed to the role of the environment / context which constitutes an integral part of the system (Larsen-Freeman, 2012).

Traditional views on the impact of the environment / context on the investigation of development (as is typical in second language studies) commonly presented context as a backdrop feature. However, DST views context as a focal point of interest and therefore challenge the idea that context is simply background to the main action (Beckner et al., 2009; Dornyei, 2000; Verspoor and Schmitt, 2012). This is because from a DST perspective, 'everything is always interacting and interfacing with others and the environment organically' (Cilliers, 2001,p.142).

In the literature, one of the studies that clearly illustrates the centrality of the environment/ context in complex dynamic systems is that of Churchill (2007). Churchill investigated processes involved in the development of word knowledge for a single word and found that his ability to learn a new Japanese lexical item was affected by affordances offered by his environment, mainly medical written and oral output from nurses and doctors within the hospital context where he was situated. This led to revisions and adaptations of his knowledge as he continued to build his understanding of the lexical item. The investigation led him to conclude that there is a case for

applying a DST perspective to language learning, arguing that when an individual is situated in a particular context, he is oriented to his surroundings (Churchill, 2007). Such surroundings influence linguistic development and cognitions about particular concepts within the L2 learning context.

2.8.2 Second language development as a complex system

On the basis of the above discussion of some of the key characteristic features of complex dynamic systems, it is not surprising that second language development is considered to be a complex dynamic / adaptive system (Churchill, 2007; Dornyei, 2014; Larsen-freeman, 1997; Macintyre and Legatto, 2010; Menezes, 2013; van Geert and Steenbeek, 2005; Verspoor and Schmitt, 2012; Verspoor, 2012). Indeed, second language development comprises multiple components operating at different levels including the phonological, morphological, syntactical, semantical and pragmatic levels. It can therefore be expected that the interaction of these components can lead to the emergence of different levels of proficiency during L2 development.

Since DST offers the opportunity to reconcile different theoretical perspectives in L2 developmental processes, it is not surprising that this theory does not exist in isolation from perspectives offered by traditional L2 perspectives. For example, insights from connectionism offer perspectives which are amenable to DST approaches (Menezes, 2013). This theory highlights that language learning is understood as the process whereby experience and the repetition of experiences leads to the development, strengthening and/or weakening of the connections between / among linguistic features such as lexis (Verspoor and Schmitt, 2012).

This suggests that language learning is not linear but emerges from the multiple interactions associated with "simple learning mechanisms, operating in and across the human systems for perception, motor-action and cognition as they are exposed to language data as part of a social environment' (Ellis, 1998, p.631). Through DST approaches, multiple theoretical perspectives can be reconciled as subsystems of the same complex system. Different L2 theories can be 'treated as explanations of parts of a whole, since each captures a different aspect of SLA' (Menezes, 2013,p.404). Such interaction between / among such theories is likely to produce broader and deeper conceptions of the second language development processes (Menezes, 2013). In the context of the current study, the implication is that if second language development is a complex, dynamic system, then each ESL learner can be taken to represent a dynamic system where lexical knowledge of each of the learners investigated, as well as their written language skills, are subsystems nested within the bigger second language development system. As a subsystem, it can be expected that other environmental / contextual constraints will constitute part of the interacting (sub) subsystems leading to development where variability is typical.

Larsen-Freeman's (2006) study is another example of L2 studies that have adopted a DST approach to the investigation of second language learning issues. Larsen-Freeman (2006) investigated the oral and written productions of five ESL learners from a Chinese L1 background. The aim was to find out how the learners' written language skills were developing in terms of complexity, fluency, and accuracy.

The study involved five adult participants who were classified as functioning at the high intermediate proficiency level for English (Larsen-

Freeman, 2006). The author utilised a time-series design whereby a series of observations were carried out at different stages of the learners' written language development. Learners were required to perform the same task several times. The use of the same task was an important part of the study design because from a DST perspective, small changes can have profound impacts (Larsen-Freeman and Cameron, 2008). The requirement was for the participants to identify a previous experience that they were comfortable writing about and sharing verbally with the researcher. Once the texts were composed, the participants were asked to put them away for three days and then they were required to provide a verbal narrative of the story to the researcher. The verbal narratives were recorded and transcribed, before the participants were asked to re-write the same story without receiving any feedback from the researcher or their tutor.

The narratives obtained from the five participants were then analysed for fluency, grammatical complexity and lexical accuracy. In addition, the written and oral samples were qualitatively analysed to find out how different versions of the same story differed from one episode to another. As can be expected, findings from the study revealed that on average, learners' writing skills improved over time. However, Larsen-Freeman also found that average group data had no particular validity related to any of the five participants and identified this as a limitation of aggregated data (Larsen-Freeman, 2007). The study also identified that individual learners had distinctive orientations and paths that they exhibited over time. As an example, for some learners, when grammatical complexity was compared to vocabulary complexity the resultant profiles suggested that the participant had seen development in one area more than the other suggesting varying de-

grees of success in their development despite exposure to the same treatment conditions. Larsen-Freeman (2007) therefore concluded that development does not unfold in a predetermined way but that development occurred 'as a system adapting to a changing context, in which the language resources of each individual are uniquely transformed through use' (Larsen-Freeman, 2006, p. 590).

More recently, Svalberg and Askham (2015) investigated the development of grammar knowledge by two trainee teachers, one with English as an L1 and the other with English as an L2. The students were enrolled on a course designed to develop their grammar awareness and were working towards a postgraduate qualification. Svalberg and Askham (2015) analysed each participant's pathway as a separate case, an approach inspired by a DST perspective (Larsen-Freeman, 2007). To that extent, the researchers sought to obtain 'an in-depth understanding of the individual pathways they (students) had chosen and factors which may have impacted on those choices' (Svalberg and Askham, 2015, p.172). On the basis of data obtained from the students' diaries of grammar workshops, their interactional behaviour during grammar workshops as well as interviews, the authors derived two contrasting profiles of postgraduate grammar students. Their analysis of the learners' individual profiles revealed that there was no single way of learning and being an effective student so that for example, both the more interactional and less interactional approaches worked in each case as they tapped into each participant's personality, strengths and cognitions. To that extent, the study highlighted that L2 students are complex systems where each individual participant's strategies, personality traits and attitudes play an important part in their L2 development. DST studies such as these

acknowledge the complexity of L2 processes by focussing on in-depth analysis of L2 developmental issues at an individual level in order to obtain insights that may otherwise be missed out.

2.8.3 Critique of DST

Notwithstanding the merits and relevance of DST to the current study as highlighted in Section 2.8.2 above, DST is not without critique (Jiang and Dewaele, 2015). A particularly critical perspective to DST is provided by Gregg (2010) who describes Larsen-Freeman and Cameron's (2008) application of DST to Applied Linguistics as "shallow draughts" (Gregg 2010, p. 549). One of the reasons that Gregg (2010) puts forward for this view is what the scholar describes as a "failure to engage with the facts" (Gregg, 2010, p.557). To this end, Gregg (2010) considers Larsen-Freeman and Cameroon's (2008) examples of the applications of DST in second language learning to be 'idiosyncratic' (Gregg, 2010, p.557).

Dornyei et al. (2015) note that proposals for a paradigm shift towards DST approaches in the field of Applied Linguistics have seen some developments over the years. However, the authors note that during the pioneering years of DST in Applied Linguistics, it was evident that while the body of literature in this area was growing, 'very little of this work was empirical in nature'. Owing to this, for many researchers, there was 'the sense of being at a loss as to how exactly to go about researching dynamic systems' (Dornyei et al., 2015, pp. 1-2). Arguably then, this may explain Gregg's (2010) criticism of DST applications in L2 contexts.

Evidently, empirical studies are now infiltrating the field (see Svalberg and Akham, 2015; Churchill, 2007; Dornyei, 2014). However, there

is still an imbalance between conceptual and empirical studies in DST research, hence the current study's contribution towards addressing this gap.

2.9 Chapter summary

In recognition of the aims of this thesis, this chapter offered a critical review of literature on vocabulary research and writing studies. It highlighted evidence suggesting that vocabulary has moved from being a Cinderella subject in L2 research (Meara, 1980) to its current state of being recognised as a key player in second language acquisition and in facilitating learners' abilities to function in the L2. To that effect, the Literature Review highlighted the relevance and justification of the aims of the current study. The chapter also offered an overview of the key concepts on word knowledge and theoretical perspectives on what it means to know a word. These ranged from individual word knowledge frameworks such as the one offered by Nation (2001), to network approaches postulated by scholars such as Meara and Wolter (2004).

Since the main aim of the study is to explore the interplay between vocabulary knowledge and written language ability, the chapter also brought to the fore, literature from writing research and included perspectives on the processes considered to be key when composing text. Some of the reasons why writing in a foreign language can present challenges to L2 writers were also considered since the current research straddles the line between lexical and writing research.

Chapter Three:

Methodology

3.0 Chapter introduction

The main purpose of this chapter is to make explicit, the philosophical principles behind the current study and how these were translated into systematic steps that were taken to complete the study as revealed by the methods adopted. The chapter starts by exploring what epistemology and ontology mean and how these relate to methodological considerations in the current study. This is in line with the view that good research, like a building, is attributed to its underlying architecture (Lewis et al., 2009). This highlights the importance of addressing philosophical issues and underscores the need to be aware of the philosophical commitment inherent in a researcher's choice of research strategy.

3.1 Aims and objectives of the study

Since the main aim of this study is to explore the relationship between lexical knowledge and the written language abilities of English L2 learners, the following research questions were identified as being at the centre of the investigation:

RQ1 - What is the relationship between learners' vocabulary size (breadth) and the quality of their vocabulary knowledge (depth)?

RQ2 - What is the relationship between learners' vocabulary size and the quality of their written compositions?

RQ3 - What is the relationship between the quality of learners' vocabulary knowledge (depth) and the quality of their written compositions?

RQ4 - Is there a relationship between the learners' lexical profiles produced by VocabProfile, and the quality of their written compositions?

RQ5 - Is there a difference between the VocabProfiles of learners working at a lower proficiency level (ESL learners) and those at a higher proficiency level (ICLE students)?

The research questions are fully expanded upon in Section 3.2.

3.2 Research Paradigm and methods adopted

The term *paradigm* has been subject to wide and slightly varied interpretations, with some scholars using the term to refer to philosophical orientations whereas others have used it in a narrower sense to refer to models of research within a specific discipline (Morgan, 2007). For purposes of this study, the former view, which is the more commonly adopted in social inquiry, is adopted in line with Guba and Lincoln's definition of a research paradigm (Guba and Lincoln, 1994). Therefore, in the current study, the author views a paradigm as "the basic belief system or worldview that guides the investigation, not only in choices of method but in ontologically and epistemologically-fundamental ways" (Guba and Lincoln, 1994, p.104). To that effect, the observations and interpretations of findings in the current study are framed within the pragmatic philosophical stance as will be fully explored later in this section. Paradigms are therefore considered to be an important part of this research project because they influenced how the researcher designed the study and the theoretical interpretations that the researcher attached to the findings. This is in line with Guba's (1990) view that paradigms help to frame and guide action in any disciplined inquiry.

Research paradigms are understood through their ontology, epistemology and methodology (Johnson et al., 2007; Johnson and Onwuegbuzie, 2004). Therefore, these three concepts (ontology, epistemology and methodology) are intimately related because

ontological assumptions give rise to epistemological assumptions, which lead to methodological considerations, which in turn give rise to issues of instrumentation and data collection (Cohen et al., 2009). Ontology refers to beliefs about the nature of reality and what constitutes acceptable knowledge in a field (Hennink, Hutter, and Bailey, 2011). Such beliefs determine what can be known. Epistemology focuses on beliefs about the nature of the relationship between the researcher and knowledge during an inquiry (Bracken, 2010). It is the ontological beliefs that determine how subjective or objective the nature of the relationship between the researcher and the knowledge is (Lincoln and Guba, 1985). Methodology refers to the researcher's strategy for discovering knowledge; in other words, the procedures that they adopt in order to acquire knowledge during an investigation (Punch, 2005; Barbie, 2007).

In the current study the pragmatic philosophical stance adopted means that the study is framed on the premise that external reality exists but values also play a role in the interpretation of findings (Harrits, 2011). On the basis of this belief, the study adopts a mixed methods approach for both data collection and data analysis as highlighted in Table 3.1. As such, the study does not adopt purist philosophical positions which view positivism and interpretivism as irreconcilable as has been reflected in the 'the paradigm wars' between qualitative and quantitative researchers (Maxwell, 2010; Mertens and Hesse-Biber, 2012). Instead, in line with pragmatic traditions, the study is designed on the premise that 'it is a valid and acceptable proposition to assert that there is a single real world or reality but that individuals may have their own unique interpretations of that world' (Morgan, 2007, p.72), hence the MMR approach adopted.

The pragmatic stance adopted in the current investigation is evident in the design of the study. Thus, the first insights to be obtained about the interplay between L2 learners' vocabulary knowledge and their written language skills derive from a positivist ontology. To that effect, students' test scores from a vocabulary size test are correlated with scores from a depth of knowledge test (WAT) as well as to written language scores allocated by tutors. Such an approach provides a positivist perspective because positivism is founded on a belief in the existence of an objective reality which implies that a researcher needs to maintain distance from the research (Tashakkori and Teddlie, 2009). In the current study, this was realised through the statistical analyses of learners' test scores. For this part of the study, an experimental design was adopted and the statistical analyses produced quantitative results which are objective and not context-dependent (Howe, 1988). However, the study further interrogates test scores qualitatively through tracing individual trajectories in order to understand learners' abilities at an individual level. In addition, learners are interviewed in order to explore emic views to the interplay between their vocabulary knowledge and their written language skills (see Section 1.2 for an overview of epic and emic perspectives). The interviews which were carried out in the study sit within an interpretivist philosophical ontology characterised by a belief that participants and the researcher co-construct knowledge through interaction with the context / environment (Guba, 1990; Morgan, 2007).

Therefore, the combination of elements from positivist and interpretivist dimensions in the current study imply a belief in the ability of both subjective and objective knowledge dimensions to work together, supplementing insights from each knowledge dimension in order

to offer valuable, in-depth contributions to research. As Table 3.1highlights, the data collection and the data analysis procedures adopted in the current study are a reflection of the underlying pragmatic worldview that informs the current study.

Table 3.1: Analysis of Research Questions

Research Question	Quantitative data	Qualitative data	Research Instrument	Data analysis
RQ1. What is the relationship between learners' vocabulary size (breadth) and the quality of their vocabulary knowledge (depth)?	VLT scores WAT scores	Semi-structured interviews/stimulated recall	VLT WAT test	Quantitative and qualitative
RQ2. What is the relationship between learners' vocabulary size and the quality of their written compositions?	Vocabulary size Teacher allocated scores	Semi-structured interviews/stimulated recall	Vocabulary Levels Test Written language compositions	Quantitative and Qualitative
RQ3. What is the relationship between the quality of learners' vocabulary knowledge (depth) and the quality of their written compositions?	Word association scores Teacher/researcher allocated scores	Semi-structured interviews/stimulated recall	Word Association Test Written language compositions	Quantitative and qualitative
RQ4. Is there a relationship between the learners' lexical profiles produced by VocabProfile and the quality of their written compositions?	VocabProfile scores Teacher/researcher allocated scores	Semi-structured interviews/stimulated recall	VocabProfile online tool Written language compositions	Quantitative and qualitative
RQ5. Is there a difference between the VocabProfiles of learners working at a lower proficiency level (ESL learners) and those at a higher proficiency level (ICLE students)?	VocabProfile scores Teacher/researcher allocated scores	Semi-structured interviews/stimulated recall	VocabProfile online tool Written language compositions	Quantitative

As can be seen from Table 3.1, in order to address all the research questions relevant to this study, it was necessary to collect quantitative and qualitative data. Similarly, once relevant data was collected, the data analysis procedures also comprised qualitative and quantitative approaches. The semi-structured interviews / stimulated reconstructions provided qualitative data which was invaluable for understanding the complexity of the interplay between vocabulary knowledge and aspects of learner performance. Section 3.2.1 below sheds more light on the role of research questions in MMR designs in general and the current study in particular.

3.2.1 The role of research questions in MMR

Research questions play an important role in any research context but in MMR, metaphors such as the 'hub of the research process' (Plano Clark and Badiee, 2010, p.276) and the 'dictatorship of the research question' (Tashakkori and Teddlie, 2003, p.679) have been used in relation to the role of research questions. Such descriptions highlight the pivotal role of research questions in MMR in terms of determining and guiding the key decisions made throughout the research process. In the current study, an analysis of the five research questions revealed that a major focus is on 'relationships', either between different aspects of lexical knowledge or between lexical knowledge and other linguistic skills such as writing. To that extent, part of the exploration of these relationships involved the use of quantitative procedures where performance data from written assessments provided key variables for exploration of correlations through the Statistical Package for the Social Sciences (SPSS) procedures. However, qualitative data from stimulated recalls / semi-structured interviews helped not only to provide possible explanations

to the correlations (and or lack of) between aspects of learners' vocabulary knowledge and their scores obtained from the battery of written assessments at a group level but also, qualitative data helped provide insights about individual learners' lexical trajectories which would have been impossible to get through an approach exclusive to quantitative data and analysis.

Therefore, what concerns MMR researchers in general, and the current study in particular, are not the philosophical debates on whether qualitative or quantitative research approaches should be adopted in any particular context. Instead, consideration is given to the aims of the research and ultimately, to the research questions. A blend of qualitative and quantitative approaches was deemed to be most capable of addressing the aims of the current research, and this resulted in the MMR approach which runs across the thesis. This approach resonates Tashakkori and Teddlie's (2010) view that it is the research question(s) that should determine the methods employed within any given study. It is for this reason that the current study embraces both positivist and interpretive approaches.

3.3 Data collection

3.3.1 Setting

The setting for the study is a Further Education college located in the Midlands of the United Kingdom. The institution provides a wide range of education and training opportunities, from pre-entry through to university level courses and attracts students from all over the world. It typically runs courses between September and June but also offers short summer courses between June and September. Students mostly study General English but may choose to work towards IELTs

examinations. The class sizes typically range from 14 to 25 students and comprise adult learners (18 years and above).

3.3.2 Sampling Technique

Probability and non-probability sampling have been identified as the two main sampling strategies that can be adopted in a study (Schillewaert, Langerak and Duhamel, 1998); Schreuder, et al., 2001). However, Teddlie and Yu (2007) contend that there are infact four broad categories with convenience and mixed methods sampling techniques as the two additional sampling procedures. Figure 3.1 highlights Teddlie and Yu's taxonomy of sampling techniques (Teddlie and Yu, 2007).

Figure 3.1 Taxonomy of sampling techniques

I. Probability Sampling

A. Random Sampling
B. Stratified Sampling
C. Cluster Sampling
D. Sampling Using Multiple Probability Techniques

II. Purposive Sampling

A. Sampling to Achieve Representativeness or Comparability
B. Sampling Special or Unique Cases
C. Sequential Sampling
D. Sampling Using Multiple Purposive Techniques

III. Convenience Sampling

A. Captive Sample
B. Volunteer Sample

IV. Mixed Methods Sampling

A. Basic Mixed Methods Sampling
B. Sequential Mixed Methods Sampling
C. Concurrent Mixed Methods Sampling
D. Multilevel Mixed Methods Sampling
E. Combination of Mixed Methods Sampling Strategies

Source: Teddlie and Yu (2007, p.78)

A comprehensive discussion of all the sampling procedures as identified by Teddlie and Yu (2007) is outside the scope of the current project. However, an overview of the main sampling categories is necessary in order to set the scene and provide grounding and rationale for the discussion of the sampling technique adopted in the current study. In probability sampling, the sample is selected in a random manner which means that members of the population have equal chances of being selected as part of the sample (Tashakkori and Teddlie, 2003). Purposive sampling involves the subjective selection of study participants by the researcher based on the population's characteristics (Sieber, 1998). The researcher's objective is to identify samples with characteristics that will best facilitate the answering of the research questions in a study (Teddlie and Yu, 2007). Convenience sampling entails the use of samples that are easily accessible to the researcher either as captive or volunteer participants (Teddlie and Yu, 2007). The last category (mixed methods sampling) involves the selection of study participants using both purposive and probability strategies.

Combining multiple sampling techniques in a single study is one of the strategies that can be employed to generate complementary data (Collins, 2010). In the current study, a combination of convenience and purposive sampling techniques were adopted at different stages of the research process. The triangulated approach adopted in this study was a conscious effort to enhance the study's ability to provide in-depth insights needed to comprehensively answer the research questions to allow for naturalistic rather than scientific generalisability

(see Section 3.6 for a discussion of generalisability in this study). From this perspective, small sample sizes, even single cases have the ability to inform and enlighten (Melrose, 2009).

3.3.2.1 Convenience sampling in the current study

As I was a full-time employee at a college of Further education, ease of participant recruitment and logistical convenience were a contributing factor in the choice of the adopted sampling technique. The college provided practical and convenient accessibility and proximity to participants. However, convenience had to be considered in light of the sample's ability to produce data appropriate for answering the research questions as discussed in the sections above. Therefore, although convenience was an important factor, the suitability of the participant pool to produce relevant data was an even more important factor.

3.3.2.2 Purposive sampling in the current study

The two most important considerations for sampling were that firstly, two groups of learners working at two different proficiency levels were required and secondly, students from both groups had to be able to generate the required quantity of words as explained in Section 3.3.2.1.

3.3.2.1 Proficiency levels

To answer the research questions, it was important to identify students working at different proficiency levels. At the institution where the current study was carried out, all ESL students were required to complete a language assessment as part of their entrance examination. The entrance examination was completed online and except for the free writing part of the assessment, the assessment

was marked through an automated system which allocated marks and grades to each learner. The levels resulting from this assessment were Entry 1, Entry 2, Entry 3, Level 1 and Level 2, with Entry 1 representing the lowest proficiency and Level 2 representing the highest. Entry 1 and 2 learners were considered to be beginners; Entry 3 were lower intermediate; Level 1 higher intermediate and Level 2 were the college's advanced group. In addition, in order to be placed within the Entry level groups, students had to score 4.5 and below on an IELTS type writing assessment. The assessment was delivered as part of the initial assessment programme because the college in-house assessment did not contain a free writing exercise. The free writing assessment comprised one writing task, a free writing task of about 300 words on an argumentative type of essay. Therefore, the current study utilised this streaming system in order to identify suitable ESL students as the target population for the study. Entry 3 students were considered to be as a suitable population to study as this level represents a lower level of proficiency compared to the ICLE data sample (see Section 3.3.5 for more details about the higher proficiency group).

Laufer and Nation (1995) utilised the Lexical Frequency Profile (LFP) to investigate vocabulary use in written compositions of 22 university students in New Zealand. LFP, the original version of VocabProfile (Cobb, (2017), personal communication), is an online tool developed by Laufer and Nation (1995). During text analysis using LFP, Laufer and Nation (1995) found that texts of over 200 words produced stable results in the sense that where the same students produced multiple compositions within the same genre, there were no significant differences in the LFPs produced. Since VocabProfile, the

lexical analysis tool adopted in the current study, utilises a similar methodology for analysing lexical use in text i.e. frequency profiling, this study used Laufer and Nation's (1995) guidelines for the length of essays that were deemed appropriate for analysis. Therefore, when collecting samples of written texts both from the primary and secondary data sources, the researcher ensured that a minimum of 200 words in continuous text was obtained from each participant.

The minimum requirement of 200 words for the free writing texts meant that absolute beginners would have found it difficult to meet the study requirements without any scaffolding or support during the writing activity. Entry 3 students were deemed able to produce the required quantity of data and at the same time considered to be sufficiently different from Level 2 in terms of their proficiency level. Collecting data at two different levels was essential for addressing the requirements of Research Question 4 (Is there a relationship between the learners' VocabProfiles and the quality of their written compositions?) and Research Question 5 (Is there a difference between the VocabProfiles of learners at different language proficiencies?). Therefore, for purposes of the current study, Entry 3 was identified for the lower proficiency group while Level 2 was identified for the advanced group.

At the beginning of the study, the aim was to collect data from two low proficiency classes and two high proficiency classes so that put together, the sample would comprise two groups of learners working at two different proficiency levels. Across the intended four classes, this would provide a total of 72 students. However, at the data collection stage, two main challenges were encountered which

threatened the successful completion of the study. The first challenge was when the departmental head retracted on previously granted permission for the collection of data from the Level 2 groups as higher proficiency groups were said to be under a lot of pressure and could not be allowed to take time to complete the necessary activities for the research. This reduced the sample size by almost fifty percent as it left 39 lower proficiency students only.

Unfortunately, this data would not suffice for addressing Research Question 5 since it required a comparison of VocabProfiles from low and high proficiency students. To overcome this obstacle, I decided that I would obtain a similar number of written language essays from the ICLE so that these would provide the required higher proficiency data. Since the ICLE essays were all produced by university level students (see Section 3.3.4 for more details about ICLE) the essays offered sufficient scope for addressing Research Questions 4 and 5. The second set-back occurred soon after the commencement of data collection, when one of the teachers withdrew their class from participating in the study, highlighting time pressure as the reason. This left me with one group of low proficiency students comprising 18 students from mixed nationalities (see Table 3.3). As a result, although the initial plan was to investigate the interplay between vocabulary knowledge and written language skills among lower and higher proficiency groups in a college setting based on primary data, the study eventually utilised both primary and secondary data due to the unexpected challenges faced with access to participants. 40 essays were eventually obtained from the ICLE in order to compensate for the challenges faced with getting access to primary

data. However, as Table 3.2 shows, convenience and purposive sampling remained the key sampling techniques for the current study.

Table 3.2: Sampling techniques in the current study

Primary Data (ESL students)	Secondary Data (Corpus data)
Purposive sampling used to obtain data from second language students working at a low proficiency level but capable of producing between 200 and 250 words of continuous text.	**Purposive Sampling** used to obtain corpus data from students operating at a higher proficiency level than the primary data set and representing multiple L1 backgrounds as was the case with the profile of the primary data source.
Convenience sampling as the sample was drawn from researcher's institution for practicality reasons.	

As can be seen from Table 3.2, the study utilised two data sets through the adoption of non-probability sampling techniques. Data from both primary and secondary sources was integrated in order to address the research questions.

3.3.3 Sample size

A small sample size may generate insights which fail to reach acceptable significance levels (Lowie and Seton, 2013). In other words, a small sample size increases chances of getting a result which is more likely to be due to chance (Burns and Burns, 2008). In the current study, for reasons discussed in Section 3.3.3 above, there were less than thirty students per group in the primary data set. This is a small sample size in statistical traditions (Lowie and Seton, 2013). Considering that in statistical analysis, a small sample size may skew the results in terms of its reflection of the characteristics of the population it was drawn from, the MMR approach adopted was considered to be an important factor in complementing findings from the statisti-

cal analysis (Fielding, 2012). It provided the opportunity to qualitatively delve deeper into the findings from the statistical analysis in order to enhance the validity of the study, hence the integral role of data from stimulated reconstructions/interviews. In addition, since corpus data is readily available, the study utilised a larger sample size (40 essays) for the higher proficiency group as another step towards mitigating the small sample size obtained from the primary data source. Nonetheless, the study acknowledges issues of lack of comparability of data between the two samples as discussed in Section 3.4.6.

3.3.4 Participants: The primary data

The primary source was a group comprising 18 ESL students enrolled in an English course designed to prepare the students for further study within the college and ultimately help students with their progression into undergraduate studies. Both in the pilot and main study, participants were from the researcher's institute but not in any of the researcher's classes. Students who participated in the pilot study were excluded from participating in the actual study to avoid possible washback effect from the pilot study. Bailey (1996) views washback as the potential impact that a test or assessment may have on those involved in the teaching and learning process (see Section 3.3.7 for a more comprehensive discussion of washback). In the context of the current study, this suggests that if the same participants had been used for both the pilot study reported in Chapter Four, as well as the main study, then there is a chance that this experience could have had an impact on the students' performance particularly if they were to re-take a similar test. For example, students might have

altered their language development practices in preparation for consequent vocabulary tests and this could have inflated the test scores.

The students were all adult learners aged between 18 and 30 (see Table 3.3). They were all studying in an English institution for the first time. All the students had had very limited exposure to the English language in their countries of origin; they all declared that they had not been taught content subjects in the English medium before they arrived in the UK. The students were on the course for personal reasons including their desire to study in the UK, their desire to develop skills to enable them to help their children with homework, and a desire to learn the language because of its status in their country of origin. As such, the learning programme was not imposed on any of them and so their motivational levels could be expected to be high. However, the participants were drawn from a short summer course and were therefore under time pressure, as is typical on intensive summer courses. For some learners, this was combined with pressure from family and social commitments. This meant that some participants who completed written assessments (VLT, WAT and Writing) could not complete interviews. Out of all the participants, only one student had previously studied a degree-level course although the tuition was delivered in French.

For purposes of the current study, and compared to the learners providing the secondary data, the ESL group was the lower proficiency group while the ICLE was the higher proficiency group (see Section 3.3.5 below for a description of how the advanced level group was identified).

As highlighted in Section 3.3.2, Entry 3 students were those who came out at 4.5 or lower in an IELTs type writing assessment at the beginning of the course. The 18 students who made up this group were drawn from a class of 25. Out of the 25, seven were removed from the study. This was either because the students were not able to complete some of the written elements of the study or they were absent at the time when any one of the written assessments were completed. The procedure followed for absent learners was a key ethical consideration and is therefore discussed in Section 3.5.4. Table 3.3 below provides a summary of the ESL student profiles.

Table 3.3: ESL participants' profiles

Student	Age	Gender	First Language
S1	27	F	French
S2	30	M	Hindi
S3	22	M	Russian
S4	22	F	Polish
S5	24	F	Chichewa
S6	18	M	French
S7	23	F	Urdu
S8	22	M	Chinese
S9	19	F	Kurdish
S10	24	F	Arabic
S11	21	M	Urdu
S12	25	F	Bengali
S13	19	F	Sinhala
S14	23	F	Hindi
S15	25	M	Polish
S16	21	F	Polish
S17	22	F	Afghan
S18	23	F	Turkish

N = 18 **Mean age** =22.3 years

As Table 3.3 shows, the students represented a very wide range of L1 backgrounds / nationalities, as is typical in an ESL teaching environment. The backgrounds represented include Europe, Eastern Europe, Russia, Asia, the Middle East and Africa. Their ages ranged between 18 and 30, with an average age of 22.8 years. The students were required to choose one topic out of the 9 essays which were provided in line with the institutional guidance where the study took place. The range of essay topics provided was in line with the requirement of the institution where the study took place. Indeed, tutors are required to give students a wide range of choices for free writing. This is considered to increase chances of each learner getting a topic that they feel comfortable writing about within a particular genre. This study acknowledges that in a research context, asking participants to produce texts in a wide range of topics within a particular study can pose a threat to the validity of the findings because different topics are likely to elicit different kinds of vocabulary knowledge. This is because, as with variability in genre, variability in topics is likely to have an impact on the type of vocabulary that students produce (Olinghouse and Wilson, 2013).

Therefore, to control for the potential impact of variability in genre on learners' vocabulary use, in the current study, the argumentative writing genre was chosen as it was aligned with the learners' curriculum**.** Out of the 9 topics that were provided to the students, only 2 topics were chosen. The topics chosen were as follows:

Topic 1: '*Money is the root of all evil'. Do you agree?*'

Topic 2: '*The prison system is old fashioned. No society should punish its criminals by sending them to prison. It should help them become good citizens. What are your views about this statement?*'.

Within the ESL group, 10 students chose topic 1 while 8 chose Topic 2 (44% vs 56% respectively for Topics 1 and 2). No student chose the remaining 7 topics at all. As part of their ESL programme, participants were required to write a minimum of five assessed pieces of free writing. Out of these, two (mid-term and end of course assessments) can be considered to be fully aligned with AoL practices the assessments were delivered summatively (see Section 2.7.4 for an overview of different types of assessments). However, the other three assessments satisfy some of the criteria for AfL practices because although the tutor marked and gave feedback which was intended to help the students further develop their writing skills, the assessments were carried out at the end of particular units of instruction. Unlike the mid-term and end of term assessments, the other three assessments were completed at home, under non-examination conditions. However, students worked independently with neither the teacher's nor peers' support until they submitted their work for marking. In addition, although the students were encouraged to submit multiple drafts, this was not a stipulated requirement so students made individual choices. However, the students were made aware that they would gain marks for submitting a coherent plan with their composition when they sit the final assessment at the end of the course.

3.3.5 Participants: The secondary data

In terms of secondary data, 40 essays were obtained from the ICLE which comprises essays produced by undergraduate students (see Section 3.4.6 for a comprehensive discussion of corpora). As was highlighted in Table 3.3, for ESL students, multiple L1 backgrounds

were represented (Europe, Eastern Europe, Russia, Asia, the Middle East and Africa). ICLE essays were derived from students drawn from Asia/Middle East (Turkey), Europe, Eastern Europe as well as Africa. This was the closest match that could be obtained where students had produced essays under the two essay topics that were used by the ESL students.

Indeed, the current study considered that it was more important to obtain essays written under the same topic as the primary data set because of the impact that different topics may have on the type of lexis that students produce (see Section 2.2.9). Therefore, all the essays selected had to be essays written under the same topics as the ones chosen by ESL students (see Section 3.3.4 above). Just over 50% of the essays chosen were based on Topic 1 while the remaining essays were based on Topic 2. This was considered an appropriate proportion based on the split that was obtained from ESL students.

The age range for the ICLE students was between 20 and 27years, with an average age of 22.3 years (Granger, et al., 2009). This is fairly comparable to the ESL students who ranged between 18 and 30 years, with an average age of 22.8 years (see Table 3.3). Therefore, as far as possible, for purposes of the current study, corpus data was closely matched to the ESL sample, both in terms of L1 backgrounds represented, the topics selected as well as the ratio of essays obtained under each topic. However, the learners represented in the corpus were all undergraduates as opposed to the college students comprising the primary data source. To that extent, the study acknowledges issues of lack of representativeness of the corpus data as will be further explored in Section 3.4.6.

In terms of proficiency levels, as Granger, et al. (2009, p.3) note, one of the main requirements which were set right from the beginning of the ICLE project was that essays would be collected from 'young adults (university undergraduates) operating at 'advanced proficiency level'. As Granger, et al. (2009) further advise, the classification of ICLE students as advanced learners was defined from two perspectives. Firstly, the notion referred to the level of study where the students were university undergraduates in English 'usually in their third or fourth year' (Granger, et al, 2009, p.11). Secondly, in compiling the corpus, the scholars took into consideration, 'the highly heterogenous nature of learner language' (Granger, et al., 2009, p.3).

This means that amongst any group of learners, there is likely to be differential profiles so that categorising students as advanced learners based on a single criterion (i.e. the fact that they were university students) may not be a robust enough measure. For that reason, the scholars took steps to gain further insights into the proficiency levels of the students represented in the ICLE corpus. To that effect, they randomly selected 20 essays per subcorpus which gave a total of 320 essays, and asked an independent rater to rate these according to the CEFR descriptors for the assessment of writing. The finding was that the essays mostly fell within the higher intermediate and the advanced range. On both the above premises, the scholars argue that 'it seems appropriate to say that the proficiency level ranges from higher intermediate to advanced' (Granger, et al., 2009, p.11). While it may be argued that the sample which was checked represents a rather small percentage of the total essays (320 out of 6085 essays), the results provide a useful premise for categorising the students as

the higher proficiency group. In the context of the current study, this provides sufficient differentiation in terms of proficiency levels considering that the ESL students were all operating at the pre-intermediate level (see Section 3.3.4.).

3.3.6 Validity

Researchers have generally used the terms validity, trustworthiness and credibility to refer to aspects of quality in research (Tashakkori and Teddlie, 2008). However, in describing a typology for assessing the quality of research in MMR designs, Tashakkori and Teddlie (2008, p.106) introduce the term inference quality as an umbrella term to refer to the quality of findings that a researcher draws from a particular study. The authors argue that 'inferences are the most important aspects or outcomes of any study' where inferences are defined as 'a researcher's construction of the relationships among people, events and variables as well as his or her construction of respondents' perceptions, behaviours, and feelings and how these relate to each other in a coherent and systematic manner' (Tashakkori and Teddlie, 2003, p.692). Therefore, validity should be consistently given priority throughout the research process so as to ensure high quality research outputs which are consistent with the epistemological and ontological bases of a study. Thus, the ensuing discussion considers validity issues in the current study against the backdrop of MMR designs.

Mixing methods requires the researcher to 'appreciate the threats inherent in the methods being combined' (Fielding, 2012, p.126). Each method brings with it its own potential threats to validity. Consequently, the researcher needs to carefully synthesize the methods used in order to avoid compounding threats. In order to

uphold validity in this study, the following steps, inspired by Fielding (2012) were taken:

i. The framing of the research questions: both qualitative and quantitative data were utilised in order to address the research questions. Neither the qualitative nor quantitative dimension were given priority in this research so a pragmatic stance was taken (see Table 3.1).

ii. Appropriateness of paradigm which underpins the study: Pragmatism as a philosophical stance upholds multiplicity of views / perspectives and this links very well with the MMR approach which seeks to combine insights from different ontological and epistemological assumptions. This is evident from the blended qualitative and quantitative elements in data collection and data analysis captured in Table 3.1. This allowed for qualitative insights from the learners' vocabulary profiles to supplement findings from quantitative analysis.

iii. Choice of setting for the study: Both secondary and primary data were obtained from ESL settings. This enhanced the internal validity of the study. However, the study acknowledges that the two settings are not equivalent as the ESL data was derived from a Further Education context in the UK whereas the ICLE data was drawn from university context around the world.

iv. Sampling techniques: Purposive and convenience sampling techniques were chosen as they offered the flexibility to adopt multiple samples that produced appropriate data for addressing the research questions.

v. The way in which data is recorded and stored: Data obtained during the data collection process was recorded and stored systematically. With regards to interviews, all interviews were recorded and transcribed before any analysis took place.

vi. The extent to which data is seen as trustworthy: Data was collected from learners in an authentic ESL learning context within a college and was complemented by secondary data from an authentic learner corpus (ICLE).

The above-mentioned steps were considered key to the quality of inferences that could be made from the current study.

3.5.5 Integration of findings

Tashakkori and Teddlie (2008) emphasise the concept of meta-inference as part of ascertaining validity in MMR designs. This means that inferences should be drawn from both strands (qualitative and quantitative) of a research project. This is very important because a major challenge faced by MMR researchers is demonstrating clearly how the adoption of MMR as opposed to a mono-method approach has benefited the study (Bergman, 2012). The current study draws from a multi-method approach to ensure that all research questions are fully addressed as was highlighted in Table 3.1.

3.3.7 Washback effect

In the current study, washback was considered to be one of the factors that could pose a threat to the validity of the findings, hence required appropriate control. Washback is the effect of a testing procedure which encourages the adoption of practices that are in line with what is considered to be the current best thinking in the filed in relation to teaching and learning (Weigle, 2002; Alderson et al.,

2004). It can be positive or negative depending on the behaviours that a test stimulates among test takers. As Taylor (2005) notes, where washback impacts negatively on teaching and learning, it can constrain the teaching/learning experience. An illustration of negative washback would be a situation whereby writing is tested only by multiple choice items. This would put great pressure for students to practise such items rather and encourage teaching and learning activities that are focused on the actual skill of writing itself (Davies et al., 1999). In other words, testing is working against the overall against intended curriculum outcome of developing the writing skills that students need in order to be able to compose texts.

Therefore, negative washback is any effect of testing that leads to or encourages the adaptation of practices/activities that are 'counterproductive' to learning (Weigle, 2002, p.54). In contrast, positive washback occurs when a testing procedure encourages positive teaching and learning practices (Taylor, 2005). In the current study, the potential impact of washback effect was identified in relation to the participants who were involved in the study. If the participants who participated in the pilot were also recruited for the next stages of the study, from a research ethics perspective, students would have needed to be told that they would be participating in the next stage of the project. This could have then instigated practices related to trying to prepare for the tests in order to obtain positive outcomes from the assessments. This is particularly so in light of the researcher's position as a tutor at the institution (see discussion of social desirability bias in Section 3.5.4). It was for this reason that

participants who participated in the pilot study were not recruited to participate in the main study.

3.3.8 Procedure

In order to fully address the research questions, it was important for all participants to complete all three written assessments. At the beginning of the written assessments, this was explained as part of the introduction to the study as a follow up to information already provided to students in written form. After the assessments, all students were sent emails with a reminder about the request for interviews. Students signed up for interviews as they came to get feedback for their assessments.

3.4 Research Instruments

3.4.1 The VLT

The VLT was administered to all participants in order to establish their lexical knowledge at the beginning of the study. This involved learners completing a vocabulary matching exercise based on a sample of 18 words drawn from each of the 2000, 3000, 5000 and the AWL category of the test. The learners were required to choose synonyms or definitions that matched given key words. The test was completed in class under the supervision of the researcher and learners did not have access to dictionaries or any other reference tools since the purpose of the test was to assess their vocabulary knowledge.

It was important for the students to complete the VLT because although the students were already streamed according to ability through the college initial assessment tool, the screening only allocated a level (e.g. Entry 3) as an overall assessment score rather

than an actual score specific to different linguistic skills. Therefore, in order to obtain data that would identify the variation between or among learners within the streamed group, VLT was administered. Schmitt (2010) observes that one of the issues that can potentially pose a threat to validity in vocabulary studies is failure to control for prior knowledge as this makes it difficult to judge whether the findings obtained from a study are related to the phenomenon under investigation or are simply attributable to learners' different levels of proficiency. Assessing all participants using the VLT helped control for this potential flaw. The VLT, WAT and free writing were all administered within the same week at the end of scheduled classes. Ideally, these tests would have been administered in one sitting to prevent any potential vocabulary gains between the time that VLT was administered and the time that WAT and free writing compositions were completed. However, this was not possible because the students' timetables would not allow much more than an hour per day.

3.4.2 Free Writing

Written compositions provided data to address Research Question 5. The scale used for the assessment of the written compositions was the in-house scale used in the college for summative and formative assessments. As Appendix 1 shows, the scale is a multiple-trait scale and the rationale behind the choice of the in-house scale as the preferred assessment tool for the writing task was derived from a number of factors.

Firstly, as highlighted in Section 2.7.1., multiple trait scales offer higher levels of objectivity relative to holistic scales. Therefore, the

use of the scale contributed towards enhanced objectivity of scores that tutors allocated to learners' essays and most likely contributed towards inter-rater reliability in comparison to what may have been achieved using a holistic scale (Mitchell and Anderson,1986).

Secondly, the assessment categories (Task, Organisation and Language) are well-aligned to the IELTs writing assessment criteria where compositions are assessed under the four categories of Task Achievement; Coherence and Cohesion; Lexical Resource and Grammatical Range and Accuracy. However, the in-house scale subsumes lexis and grammar under Language whereas the IELTs scale separates these two into Grammatical Range and Accuracy. Considering that the IELTs are high stake examinations which are internationally recognised, the preference would have been for a scale which more closely resembles the IELTs scale. However, insights from the pilot study (see Section 4.3) suggested that inter-rater reliability amongst raters was likely to pose validity threats if assessors used a rating scale that they were unfamiliar with, as was with the case with the piloting of Lumley's scale (see Appendix 7). Further to this, the researcher was also mindful of research which suggests that lexis and grammar are interrelated (see Halliday, 1961; Halliday and Matthiessen 2014; Tucker, 2007) and as such, a scale that subsumes these two categories does not detract from L2 research.

In addition, as with both the IELTs writing assessment scale and Lumley's scale, the in-house scale is a multiple-trait scale and has the advantage of the potential to offer higher inter and intra rater-reliability (Weigle, 2002). For the current study, inter-rater variability was between one and four marks. The study adopted the inter-

reliability framework used in the college whereby five marks is the maximum difference that can be averaged between two assessors during double marking. If the difference between the two scores allocated by the two assessors is higher than five marks, then a third assessor is brought in to harmonise the scores. In the current study, the services of the third marker were never required because of the high inter-rater reliability scores. The high inter-rater reliability scores could be attributed to the multiple-trait scale used as highlighted above. However, the assessors' background (all experienced in the use of the in-house scale) might have also played a role.

Therefore, owing to the content validity of the in-house scale (with it being the scale that the institution of research used for their formative and summative assessment of free writing), its potential to offer high inter-rater reliability because of assessor familiarity with the scale, as well as its alignment with current research in lexicogrammar, I concluded that the in-house scale would sufficiently meet the needs of the current study.

3.4.3 The writing assessment (in-house) scale

The scale is divided into three main categories, each focusing on a specified set of skills. The first category (*Task Fulfilment*) assesses the learners' ability to produce text which addresses the topic (relevance); the balance of their argument; their writing style as well as the extent to which the text presented is comprehensible. Writing style in this context alludes to the learner's ability to produce text which is appropriate to a specific context. For example, a learner who adopts a formal register in an email to a friend would be marked down on this regardless of whether the writing is accurate at the lexicogrammatical level or not. In other words, such a learner would

be considered to have failed to fulfil the requirements of the task in terms of recognising the audience and adopting an appropriate style for their audience.

Task Fulfilment carries 20% of the scale weighting and is referred to as *Task* in the current study. In the in-house scale, *Task* differs from Lumley's scale because although both specifications focus broadly on text content/subject matter in terms of its relevance to the task specifications, the in-house scale considers 'appropriacy of style' mainly in terms of the register used whereas Lumley's scale includes an assessment of whether 'vocabulary choices are appropriate and effective' (see Appendix 7). In the in-house scale, vocabulary choices are subsumed within the Language Control category. Therefore, this category looks at the learner's lexico-grammatical control as appropriate to the context. This includes both the range of grammatical structures and vocabulary used. It also includes sentence structure, spellings and punctuation. The section carries 60% of the total weighting and is referred to as *Language* in this study.

The *Organisation of text and content* category focuses on the way in which the content is organised. Key to organisation in this context is the learners' ability to organise information logically within paragraphs and across the whole text; their control of cohesive devices, as well as the appropriacy of the script layout. This section also carries 20% of the overall weighting and is referred to as *Organisation* in the current study. A key assumption of the notion of cohesion is that the linking of ideas plays an important role in the creation of coherent discourse and is therefore central to discourse effectiveness (Halliday and Hasan, 1976; Crossley and McNamara,

2011). The inclusion of this category in the in-house writing assessment scale, which is also the case with Lumley's scale, seems aligned to this view.

3.4.4 Word Association Tests (WAT)

WAT have been identified as the main method currently used for assessing depth of lexical knowledge among ESL learners (Fitzpatrick and Clenton, 2010; Meara and Wolter, 2004; Zareva, 2005). The test was designed by Read (1993) to provide information about links among lexical items in the learner's mental lexicon. Since it goes beyond the form-meaning link, it was considered to be highly suited for exploring depth of knowledge in the current study. The test requires participants to produce single word responses to a given stimulus word so it is relatively easy to administer, is very economical in terms of the time required for learners to complete it, and allows for objectivity in marking since learners were required to choose the correct lexical item rather produce any of their own free output. This makes it highly suited to the design of this study where a battery of other data collection tools was used to assess ESL learners' vocabulary knowledge.

3.4.5 Stimulated reconstructions/ semi-structured interviews

Following the WAT, learners were interviewed to find out the rationale and thought processes which led to incorrect answers. This procedure was inspired by the think aloud protocol commonly used in second language research (Bowles, 2010) which involves learners analysing their own thought processes during the completion of a particular cognitive task. This offers insights to cognitive perspectives from emic (insider) perspectives which may otherwise be difficult to obtain

(Morris, et al., 1999). As Matsuda (1997), notes, generalisations made by researchers and tutors about students' experiences are immensely enriched by consulting the students themselves about their intentions. In line with this view, the semi-structured interviews/stimulated reconstructions were carried out to give voice to the student participants whose vocabulary knowledge and written language skills were under investigation. This allowed for the investigation to be carried out not only on the basis of correlations; but also, by exploring learners' own perceptions about the interplay between their vocabulary knowledge and their written language skills. This approach opened avenues for multiple perspectives on the phenomenon under investigation.

However, since the interviews on the lexical choices that participants provided were carried out in retrospect rather than during the WAT, the interviews were considered to be stimulated reconstructions, a term proposed by Svalberg and Askham (2015) to describe reflective interview procedures that are not immediate but which aid memory of a previous experience. Other than the methodological attraction of stimulated recalls whereby the stimulated recalls were considered to have the ability to generate requisite insights, another attraction offered by the approach was the potential benefit that it could offer to participants.

In a study to investigate the impact of stimulated recall on noticing and overall language awareness among ESL students in the L2 writing classroom, Lindgren and Sullivan (2003) examined how stimulated recall impacted upon the revisions made by students on their writing, and therefore, whether stimulated recall triggered noticing and enhanced language awareness among the students. The students were

required to engage in a free writing exercise, after which they were invited to reflect and comment on the writing exercise leading to revisions on the texts. The main finding from the study was that the stimulated recall sessions had a positive impact on awareness raising among the participants, and that they triggered more revisions which were evidenced by the number of additional words that the students added to their edited versions. Noticing is an important part of language development as it can be linked to hypothesis formulation and consequent testing of understanding leading to language development. To that effect, in the current study, stimulated reconstructions were found to be an attractive tool not only at a methodological level because of its ability to generate relevant data for study, but also, at an ethical level as it promised to offer an opportunity to obtain pedagogic gains in terms of their language awareness and noticing of linguistic (vocabulary) items during the study.

Another reason why this study opted for stimulated reconstruction is that compared to other methodologies such as 'think aloud' protocols stimulated reconstruction allowed the students to focus exclusively on their WAT output thereby obtaining scores that are as accurate as possible in terms of their reflection of the students' vocabulary abilities. The stimulated reconstructions/semi-structured interviews took place within a period of between two and three weeks after WAT assessments were completed for all the participants.

Gass and Mackey (2000) note that while the stimulated recall methodology is capable of producing some of the most insightful data in second language research, the methodology has its own potential shortfalls which the researcher needs to be aware of, one of which re-

lates to timing. The advice that the authors offer is that it is important for the researcher to do everything they can to avoid under or overestimating the time required to complete the procedure for a particular study (Gass and Mackey, 2000). Other than allowing the smooth running of the schedule, accurately estimating time can avoid over-fatiguing participants due to over-estimation or rushing through parts of the procedure owing to underestimation. Another important consideration is that in order to facilitate access to uncorrupted memory structures, it is recommended that stimulated recall procedures are carried out as soon as possible after the event (Gass and Mackey, 2000). To that effect, the two-to-three weeks window which was allowed for the completion of stimulated recalls within the current study seems relatively long. However, it was in fact the earliest possible time because not only were the participants required to complete multiple written tasks but also, the stimulated reconstructions had to be completed after the written assessments had been completed and marked to ensure that students could receive feedback on their free writing alongside the stimulated recalls completed for purposes of the research. Nonetheless, the study acknowledges that the two—three weeks window that students had to wait before the stimulated recalls could be conducted paused a threat to the reliability of findings from the procedure. This is because stimulated recall methods offer highest levels of reliability for obtaining data about the thoughts which participants had while performing that task if the stimulated recall interview is conducted within a short period of time, with some authors suggesting a forty-eight-hour time frame (Henderson and Tallman, 2006). This view has a psychological basis because 'if there is no gap (or only a very brief one) between the event and

the recollection, the information is still available in short-term memory for access' (Meier and Vogt, 2015, p.48). Unfortunately, the time lapse in the current study could not be avoided because of the intensive programme which led the gatekeepers to restrict access to the participants. In instances where stimulated reconstructions cannot be carried out within recommended timeframes such as the forty-eight-hour time frame suggested by Henderson and Tallman (2006), Morgan (2007) advises that visual aids such as videos can be used as a way of helping participants remember the thoughts that participants had when completing a task. In the context of the current study, to reduce the impact of the long wait, the researcher ensured that WAT scripts were available as support systems for the stimulated reconstructions. These were given to each participant at the beginning of the stimulated recall session. The participants were allowed at least 15 minutes at the beginning of the interview to go through their answer script to stimulate their memories. Participants were also allowed to refer to the answer script during the course of the interview if they so wished.

In addition, where an interview question was considered to be centred on specialist linguistic knowledge, the researcher started by explaining the terms/concepts involved before posing the interview question. The second question in Section 2 of the Interview Guide (See Appendix 2) is an example of such a question. The question was focused on metalinguistic knowledge and particularly sought the learners' views on the importance of vocabulary size and depth on the development of reading, writing, speaking and listening skills. The researcher started by carefully reminding students of what the terms 'size' and 'depth' mean in lexical studies. This was important to en-

sure construct validity as misunderstanding / misinterpretation of the terms could lead to unintended inaccurate data from the learners.

3.4.6 Corpus data

Tognini-Bonelli (2001, p. 2) defines a corpus as:

> *A collection of texts assumed to be representative of a given language, put together so that it can be used for linguistic analysis. Usually, the assumption is that the language stored in a corpus is naturally occurring, that it is gathered according to explicit design criteria, with a specific purpose in mind, and with a claim to represent larger chunks of language selected according to a specific typology'*

To that extent, corpora offer great utility for teaching and learning as well as research activities because they provide access to samples of authentic language which may otherwise not be accessible to an individual researcher, teacher or language learner (Granger, 2003). Indeed, one of the challenges currently facing second language learning researchers is the difficulty associated with collecting samples of written language productions (Baker, 2006; Schmitt, 2010) as this can be a time-consuming and daunting process. In the current study, challenges with accessing requisite data for addressing the research questions highlighted the value of the ICLE (see Section 3.3.3). As such, the corpus was an invaluable source of authentic written essays. However, as Tognini-Bonelli (2001) notes, corpora should be representative of the variety of language captured. To that effect, the ICLE captures university level interlanguage for intermediate and advanced learners. In addition, the ICLE captures augmentative language as used by university level students within a teaching and learning environment.

In the context of the current study, since the primary source which provides data for the comparison of lexical use between lower and higher proficiency students is in fact based on college rather than university level students, the ICLE corpus does not offer full representativeness. This may have affected the findings from the study. For example, lexis that students produced in each environment/context might have differed because of the different affordances which may impact on the development of vocabulary knowledge. This makes it difficult to differentiate whether lexical use for a particular group of learners may be explained by their proficiency levels or other contextual factors such as the language used within their context (Churchill, 2007). Therefore, the ICLE corpus presents comparability issues which study acknowledges as a limitation.

In acknowledgement of the limitations of the ICLE corpus data in the context of the current study, steps were taken to limit the impact of the lack of representativeness of the corpus, and these were related to the topics addressed and the nationalities investigated in the study. In terms of the topics investigated, the current study ensured that all the topics addressed within the study were of the argumentative type. In addition, the essays selected from the corpus for analysis matched those that were used by ESL students who constituted the primary data set. The topic match was not only in terms of genre but also, in terms of the actual topics addressed so that only two topics were addressed by both ESL and ICLE students (see Section 3.3.5). This is in line with the view that topic familiarity can influence the type of lexis that learners produce (de Larios et al., 2008)

In terms of nationalities, with 16 different mother tongue backgrounds represented in the ICLE compared to 12 represented within the primary data set, both data sets cover a wide range of linguistic backgrounds including European, Eastern European, Russian, Asian, the Middle East and Africa. However, this does not offer exact representativeness and may influence the type of data produced by learners. As Swan (1997) notes, 'the mother tongue may 'help, hinder, or simply stand aside' during second language vocabulary development (Swan, 1997, p.162). This highlights the possible impact of interlanguage influence which may be positive, as in the case of cognates in some European languages, but may equally be negative, in the sense of false cognates. This relates to situations where lexical items appear to be similar between two languages, for example, at the phonological or semantic level, but are infact derived from different etymologies (Swan, 1997). To that effect, the fact that the secondary and primary data sets could not be matched exactly highlights another potential threat to the study with regards to the representativeness of data which was obtained from the ICLE corpus.

In terms of size, the ICLE corpus contains 3.7million words (Granger et al., 2009). Compared to corpora such as the British National Corpus which contains over 100million words (Leech, 1992), the ICLE could be considered to be very small. However, as Cobb (2003) notes, SLA research has traditionally been based on very limited linguistic evidence. For that reason, in the context of SLA studies, the ICLE 'is sufficiently large to provide insights into a large number of EFL features' (Granger et al., 2009, p.39) and was therefore deemed an

appropriate source of learner language in the context of the current study.

3.5 Data analysis

Data from interviews was qualitatively analysed to identify relevant themes that helped to clarify the thought processes that informed lexical choices made during the WAT task. The written essays were analysed using VocabProfile. Data obtained from VLT, the free writing activity and WAT were analysed using for correlations using SPSS in the first instance. For free writing, the data was scored and subsequently analysed numerically alongside the VLT and WAT scores.

3.5.1 Integrating idiographic and nomothetic perspectives

Emanating from psychological research, the distinction between idiographic and nomothetic perspectives focuses on the research approach adopted in a study (Nesselroade et al., 2007). Nomothetic approaches focus on finding universal rules while idiographic perspectives seek to understand individuals at an in-depth and personal level (Grice, 2004) Therefore, idiographic perspectives seek a unique understanding of the individual rather than general laws applicable to all in a group. In the context of Applied Linguistics, the importance of idiographic perspectives is highlighted by researchers such as Meara (2004). In a study carried out to explore the process of vocabulary loss as part of the second language learning trajectory, Meara (2004) found that even though there was evidence of substantial similarities within learners' vocabulary trajectories, considerable individual variability was also found.

The patterns of vocabulary loss that Meara found varied between individuals and this led him to conclude that averaging data may not produce robust insights particularly when dealing with a complex phenomenon such as second language learning. This suggests that the adoption of a combination of idiographic and nomothetic approaches in linguistic enquiry can reveal new and interesting insights. The above view resonates with findings from Larsen-Freeman (2006) who investigated the oral and written productions of Chinese ESL learners. Larsen-Freeman found that linguistic competence did not develop homogeneously for the investigated learners but that each individual's development exhibited some unique features (see section 2.8.2). Other researchers also concede on the importance of variability in learner competence and performance (see for example (Fitzpatrick , 2007; Higginbotham, 2010; Young, 1991). In that regard, the analysis in this study is extended to explore individual lexical profiles and interviews in order to generate deeper insights about the vocabulary knowledge of individual learners. This approach sits within the domain of DST (see Section 2.8.2).

3.5.2 Tracing individual profiles

As part of the analysis of individual trajectories, a classification system for each of the assessments was devised with the aim of categorising learner scores into two colour-coded categories according to the level of achievement as per Figure 3.2

Figure 3.2 Categories for attainment on VLT, WAT and free writing

	LLA = Lower Level Attainment
	HLA = Higher Level Attainment

The categorisation provided in Figure 3.2 is intended to be a rudimentary nomenclature designed to trace and highlight emerging patterns in learners' profiles. The analysis of learners' performance and conclusions were intended to highlight nuances of learners' profiles at an individual level across the three assessments in order to reveal further insights beyond the group averages that were gleaned from the statistical analyses. For the VLT test, the classification was modelled on Nation's guideline for the scoring of the VLT test which stipulates that 12 out of the 18 possible is the minimum score at which a learner can be considered to have mastered lexis at a particular frequency band (Laufer and Nation, 1995). To that effect, a score of 11 and below is classified as Low Level Attainment (LLA) for purposes of the analysis of the learners' trajectories. Since 12 is the beginning of mastery, this score was deemed to mark the beginning of Higher Level Attainment (HLA).

The WAT and free writing scores do not lend themselves to a similar scoring procedure as VLT. However, they were both marked out of a total of 100 marks. Therefore, a score of 49 and below was classified as LLA; 50 and above was the HLA category.

3.5.3 The Interview Guide

A framework for the semi-structured interviews was devised prior to commencing interviews and used as a guide for the researcher. It was divided into four sections. The first section provided a framework for capturing learners' background information including their motivation for learning English and their educational backgrounds.

The second section comprised questions related to the learners' metacognitive knowledge of key lexical issues with a particular focus

on learners' perspectives on breadth vs. depth of vocabulary issues. The third section sought to identify the kind of activities that learners engaged in as part of the development of their language skills. The fourth and final section related to the rationale for the WAT answers provided, and this is the section that inspires learners to reconstruct the cognitive processes that led to the answers provided on the WAT.

Dornyei (2007) advises against asking learners personal background information questions right at the beginning of a data collection procedure (e.g. a questionnaire or interview). This is because such questions can potentially be construed as threatening by the participants. However, for purposes of this study, the researcher had already met the learners during the administration of written assessments. This allowed for the development of good rapport between the learners and the researcher so it was very unlikely that the learners could feel threatened by the questions at this stage.

In addition, looking at the Speaking and Listening activities that the learners had been previously engaged in for their ESL lessons, talking about self and immediate environment was one of the most common themes they had been working on. It was therefore reasonable to assume that the structure of the interview would suit this particular group. In fact, it turned out that students eased into the interviews really well as they probably had better linguistic resources to talk about a familiar topic (themselves) than discussing their experiences of vocabulary learning or reflecting on the answers produced during a vocabulary test (WAT).

Therefore, the structure of the interviews where questions around the rationale of their lexical choices in the WAT constituted the fourth and last section of the interview guide proved appropriate.

3.5.4 Ethical considerations

Recent years have seen an increase in both awareness and concern over ethical issues and moral dilemmas in educational research and the research community at large (Neuman, 2006). All researchers therefore have an ethical obligation both to their colleagues in terms of the quality of their research outputs but also to their study population, for example, in terms of upholding their rights, privacy and welfare. What follows is an outline of the ethical responsibilities that were upheld in the current study. The issues are classified according to different research phases in line with Creswell et al.'s (2003) classification, which groups ethical issues into five categories pertaining to:

1. The research problem statement
2. The purpose statement and research questions
3. Data collection
4. Data analysis and interpretation
5. Writing and dissemination

The first two categories constitute what can be referred to as ethical issues relating directly to the research agenda. Creswell et al. (2003) concedes with Hennink et al. (2011) that the researcher needs to ensure that his / her research benefits rather than marginalises or disempowers participants. In the current study, the tests carried out gave learners the opportunity to get educational benefits from the

experience through positive washback. Positive washback refers to the learning opportunities that learners benefit from as a result of going through an assessment. Therefore, content validity was envisaged to benefit students as it is widely recognised that a test with high content validity can facilitate educational gains (Munoz and Alvarez, 2010).

In order to be a learning tool, an assessment needs to be relevant to the learners' particular needs. To that effect, in the current study, efforts were made to ensure that the assessments were relevant to the learners. As an example, the writing activity used was of the argumentative type. It was deemed to be an opportunity for students to practise their writing skills in a context that matched the needs of their course as argumentative writing was one of the essay types that learners were tested on for both formative and summative assessments. The essays were scored using the in-house scale which provided an opportunity for students to get feedback relevant to their overall learning goals. This was an important ethical consideration because primary participants in the study were adult students on a short intensive course and therefore operating under time pressure.
An assessment which detracts from learning goals would disadvantage learners by taking away time that they could have otherwise used to develop their English Language skills, thus arguably causing psychological harm in the form of undue pressure and / or stress. Therefore, the adoption of a relevant assessment task on the basis of assessment criteria relevant to the students' needs contributed towards making the writing assessment a worthwhile

learning activity for the learners rather than an activity completed purely for research purposes.

In addition, during the Pilot Study reported in Chapter 4, virtually every student expressed an interest in getting verbal feedback from the researcher. This presented another opportunity to reward the students for participating in my study. Therefore, as well as marking, rating and providing written feedback for each essay, students were also offered the opportunity to get individual verbal feedback on their writing. Although this was a very time-consuming activity, the students found the interaction highly beneficial.

In terms of ethical issues relating to data collection, the principle of guarding against potential harm (Creswell et al., 2003; Sieber, 1998) was adhered to in a number of ways. Firstly, authority to access classes was sought from the Head of Department who facilitated meetings with staff so that the researcher could liaise with teachers of relevant classes. Written communication was sent to both staff and students to seek consent for participating in the study. This communication detailed the nature of research, the amount of time that the study was anticipated to take as well as the expected research outcomes (please see Appendix 6 for a sample of the communication). This was intended to help all concerned individuals to make informed decisions.

Secondly, from the beginning of the data collection process and at every key data collection encounter with the participants, the right to participate voluntarily, as well as the right to withdraw at any stage of the process was reiterated to participants and their teachers. This

was an important ethical consideration because as Clandinin and Connelly (2000) observe, human beings are highly dynamic and can change their perspectives depending on time and context. What an individual may find acceptable today may be deemed unacceptable in future because people's understanding and interpretations of the world around them change over time. Maintaining consistent dialogue and reiterating the participants' right at every key stage of the process was considered an essential part of the process of respecting this dynamism.

Finally, another important consideration during the data collection phase related to learners who were absent during any of the data collection sessions. If a learner was absent on the day of any of the tests, they were given an opportunity to complete the missed assessment (and thereby be included as part of the project) only if they requested this opportunity. This was an ethical consideration which was put in place following the teacher's recommendation that most of the students came from cultural backgrounds where they would find it difficult to opt out of a study. Based on his experience working with this group of learners and similar cohorts in the past, the teacher strongly felt that his students were more likely to try and find a way of avoiding the requested task than tell the tutor or the researcher that they would prefer not to take part in the study even if this was the case. The tutor therefore envisaged that some of the learners might simply not turn up on the day of the assessments as a way of getting out of it. From this perspective, the researcher had to find a way of respecting these learners' needs without withholding the opportunity to participate in the study from those who might have been absent for authentic reasons. It was for this reason that a

decision was made to offer make-up sessions only when learners requested this. This arrangement was fully communicated to the learners and the tutor.

In terms of data analysis and interpretation, anonymity of participants was a key ethical issue (Sieber, 1998). Once data is collected, Creswell et al. (2003) advise that it should be kept only for a reasonable time so that it does not end up being inadvertently used for purposes other than those it was intended for. In this study, participants' identities were kept anonymous by ensuring that individual names were always kept separate from the data through a code system only used for helping the researcher match data and participants in order to build individual profiles.

Writing about ethical issues in relation to writing and disseminating information, Neuman (2000) identifies falsifying, suppressing or inventing findings for the benefit of the researcher as a fraudulent practice which constitutes scientific misconduct. In the current study, marking free compositions was identified as one of the areas where the researcher's subjective judgements might influence students' scores. As such, double marking was used to ensure accurate judgements of the quality of learners' compositions (see section 4.1.1).

In qualitative inquiry, an interview can be defined as a situation where a researcher meets or interacts with an individual or individuals in order to obtain specific information (Kvale, 2006). In other words, interviews are intended to achieve a specific goal, which makes them a 'conversation with a purpose' (Kvale, 2006, p.483). What can be deduced from this definition of an interview is the poten-

tial for interviews to be perceived as non-egalitarian encounters or 'exercises of power in education' rather than the open dialogues that researchers would generally intend them to be. Therefore, it is important for the researcher to recognise the potential of power dynamics during the social construction of knowledge as this is necessary 'to ascertain objectivity and ethicality of interview research' (Kvale, 2006, p.480).

Elwood and Martin (2000) argue that power dynamics can take many forms, for example, selecting appropriate sites for conducting interviews. This may seem to be a relatively mundane part of the research process, but might in fact be a complicated decision with wide-reaching implications (Elwood and Martin, 2000). This is because the interview site might confirm certain power dynamics, for example, a student(s) being interviewed by a professor, in the professor's office might reinforce the view of the professor as the expert and therefore influence the objectivity of the data that is collected. To that extent, Elwood and Martin (2000) argue that subtleties such as interview sites need to be given careful consideration, as the interview site itself may impact on the interactions during an interview. Indeed, when Elwood and Martin (2000) gave their participants choices regarding where they wanted to meet the researchers for interviews, the participants made choices related to a wide range of factors including community solidarity and cohesion, for example, where participants chose to be interviewed in neighbourhood association centres which they argued, made them feel that their voices were listened to.

In the current study, pragmatic considerations meant that participants could not be offered choices related to the site for interviews. This is because the college was a place where all participants could easily travel to. In addition, since participants were fairly new to the UK, it is likely that they would have found getting to a different location burdensome and perhaps unnerving. A more relevant consideration in the current study related to the teacher-student relationship. As a teacher at the college, it was possible for me to investigate students from my own classes. However, as Bamber, et al. (2000) note, one of the challenges facing social science research relates to social desirability bias.

Social desirability arises because people generally prefer to provide socially acceptable responses and as a result, individuals may deny or under-report socially undesirable actions while admitting and/or over-reporting behaviours that they consider to be socially desirable (Chung and Monroe, 2003). Therefore, in the current study, I realised that interviewing my own students about their vocabulary knowledge could potentially make the students feel that their knowledge was being 'assessed' which could lead to apprehension and perhaps exaggerated positive responses about the students' learning experiences. Therefore, to mitigate against this potential bias in the responses, I decided not to investigate my own classes even in the face of the challenges that I faced with getting access to participants for the study (see Section 3.3.3). Therefore, all of the groups selected for the study were not my classes. They were students in my institution but they were taught by other colleagues within the department. This

was a decision taken to reduce the impact of the social desirability factor in the current study.

In addition, the study recognises that power relations may influence the quality of data obtained (De Laine, 2000; Seidman, 2015). Other than its influence on the quality of data, it can also carry ethical implications. Therefore, as a researcher within my own institution, I had to be conscious of potential pitfalls that might distort research output particularly from interviews

3.6. Generalisability

The current study does not aim for scientific generalisability but instead, aims to generate insights that allow for understanding of the interplay between lexis and written language proficiency in the context of this particular study (Stake, 2000). This allows for naturalistic generalisability, a process whereby readers gain insights by reflecting on the details and descriptions presented in a study and then identifying similarities that resonate with their own experiences (Stake, 2000). This makes it possible for readers to make their own judgements in terms of the extent to which findings presented in a particular study may be applicable or transferable to their own or other situations.

Therefore, naturalistic generalisation involves readers making inferences based on insights derived from in-depth analysis. As readers consider the in-depth minutiae in a particular study, they begin to view similar circumstances in their lives from different perspectives and/or extend their knowledge and interpretations of certain experiences in their lives (Hellstrom, 2008; Lincoln and Guba,

1985; Melrose, 2009). This has the potential to instigate new ways of thinking and viewing existing knowledge and is therefore a valuable contribution to research.

3.7 Chapter summary

This chapter started by offering an overview of the relevance of philosophical underpinnings in research before linking these to the current study. The pragmatic stance, and related MMR approach adopted in this study, together with associated views were explored and their relevance to the current study outlined. An analysis of the research questions highlighted the need for both quantitative and qualitative data collection and analysis procedures and this justified the adoption of the MMR approach and related pragmatic philosophical stance in the current study.

The chapter described how data in this study was collected through a battery of qualitative and quantitative tools including the VLT, the WAT task and the free writing activity as the sources from which quantitative data were derived while semi-structured interviews/stimulated reconstructions provided a source of qualitative data. This allowed the research to capture both emic and etic perspectives on lexical knowledge. The emic or insider approach was considered important for the current study because it made it possible to capture the rationale behind lexical choices made by the learners from the learners themselves as opposed to hypothesised explanations. Since the idea is to capture not only group averages in relation to the learners' lexical trajectories but also, the individual trajectories highlighting variability at an individual level, the MMR

design is at the core of the success of the study. To that extent, the chapter highlights that in line with the pragmatic philosophy underpinning the current study, none of the approaches (quantitative or qualitative) were privileged over the other but both were fully utilised to derive in-depth insights about learners' lexical knowledge and its interplay with free writing skills.

Chapter Four:

The Pilot Study

4.0 Chapter introduction

The term 'pilot study' can be used to refer to a feasibility study (where a researcher carries out a mini version of a full-scale study) or the testing of a particular research instrument before the actual study is carried out (van Teijlingen and Hundley, 2002). The choice regarding the focus of a pilot is likely to be guided by both the researcher's aims for carrying out the pilot study and the practicality of actually completing the requisite study including time and resources. However, the literature concurs that pilot studies are an integral part of good study design (Jacobson, and Wood, 2006; Nunes et al., 2010). This is because pilot studies have the potential to provide valuable insights about the intended project. This means that pilot studies can contribute significantly towards the overall rigour and consequent success of a project. For example, they can help identify whether a particular study is feasible at all, based on actual empirical work rather than conjecture. Piloting therefore offers the researcher an opportunity to revise and refine a study design or aspects of it based on illuminations obtained henceforth.

van Teijlingen and Hundley (1998) identify a number of reasons for the value of pilot studies. The authors suggest that pilot studies can be useful for the development of testing instruments, for example, testing research instruments in terms of the extent to which they are likely to generate intended output or even assessing the feasibility of a full-scale study before it is undertaken. Further to this, pilot studies can be used for trialling and assessing the proposed data analysis techniques to find any potential problems; developing and/or refining research questions and a realistic research plan as well as training a

researcher on key elements of the research process as much as possible (Kezar, 2000; Prescott and Soeken, 1989). Therefore, pilot studies play a major role in preventing or reducing chances of poor study-design. In the context of the current study, the rationale behind conducting a pilot was to increase the likelihood of successful implementation of the main study. Considering that participants were all enrolled on a summer programme, it was absolutely crucial to ensure that the study was feasible within the time constraints imposed by the programme. In addition to checking feasibility, another key aim of the pilot study was to identify possible opportunities for improving the study design as well as any other possible pitfalls throughout the different research stages. Section 4.4 provides a summary of the adaptations that were implemented based on insights obtained from the pilot study.

Considering that there were four different assessments to be completed within a group setting (the VLT, the WAT, the free writing activity) as well as interviews to be completed on a one-to-one basis, the pilot was an important part of the study as it helped me to set realistic goals regarding the amount of work that could be done within any given timeframe. As an example, I wanted to determine whether the time that I had been allowed for the study was sufficient for me to carry out all the assessments planned.

Due to time constraints related to pilot study participants being on a short intensive English Language course, it was not possible to pilot all the assessments. However, taking inspiration from Baker's (1994) observation that a pilot study can involve pre-testing a specific

research instrument or strategy, it was envisaged that even though the practicalities of the current study were such that it could not be trialled in its entirety, there was still a lot of value to be obtained from piloting specific parts of the project. On that basis, specific elements of the study were identified and targeted for piloting. Section 4.1 details the areas that were selected and explains the rationale behind the choices made.

4.1 The areas selected for Piloting the Study and Rationale

4.1.1 The Free Writing Activity

The Literature Review chapter identified multiple-trait and holistic scales as the main types of scales for the evaluation of free writing. While such scales are helpful towards achieving objectivity, rigour and consistency in the marking and scoring of free language output, accuracy in the evaluation and scoring of free writing remains one of the more serious challenges in the design of tests, including national tests (East, 2009; Hawkey and Barker, 2004; Huang, 2008). According to the rater-type hypothesis, one of the challenges within the assessment of free writing is that assessors of writing, whether highly experienced or not, do not always agree on the aspects of writing that they focus on the most when they assess writing (Eckes, 2008). This suggests that individual differences among raters are a force to contend with even though there are well-developed scales and most raters do in fact receive training prior to carrying out ratings. Such differences suggest an element of subjectivity in the evaluation and scoring of free writing assessments.

With this in mind, it was envisaged that piloting the writing activity would improve the validity of my study. Therefore, a free writing activity was administered to ten students. All ten completed an argumentative essay based on a topic selected from the options provided by the researcher. The essays were marked using Lumley's (2002) analytical scale. A score was allocated for each of the four criteria set out in Lumley's scale i.e. Task Fulfilment and Appropriacy (TFA); Conventions of Presentation (CoP); Cohesion and Organisation (C and O) as well as Grammatical Control (GC). Each criterion carried five marks with 0 representing absolute lack of fulfilment of specific criteria and 5 representing full marks on fulfilment of that criterion.

The scores were recorded separately from the answer sheets in order to get an independent score from a second marker. After I marked all the essays, the second marker was asked to randomly double mark and score three out of the ten essays that were collected from the pilot. For the three essays that were randomly selected for double-marking, the difference in allocated scores was 4, 4 and 6 marks respectively. I therefore sought a third marker for the essay where the margin was 6 in order to harmonise the marks allocated. For the first two essays where the difference was four marks, an average score based on the two scores allocated by the two markers was allocated. Since the difference between the first two markers was over the five-mark limit for one of the essays which was double-marked, I sought feedback from the two tutors involved in terms of how the marking and scoring procedure could possibly be improved. The feedback from both tutors was that they felt the scale did not

have an intuitive feel to it, and that they consequently had to keep referring to the scale in a way that they found to be unhelpful.

On the basis of the feedback and all the input from the three tutors involved in marking, Lumley's scale was adapted in terms of its scoring rubrics so that instead of rating each category on a scale of 0 to 5, each category was allocated a weighting of 25 marks in order to obtain a total of 100%. This adapted version was then piloted on a further 5 essays which were obtained from one of the tutors' archives of essays from previous Entry 3 students. The essays were again double-marked. Unfortunately, there were no improvements in inter-rater reliability scores at 10, 7, 7, 8 and 6 marks difference for each of the double-marked essays. On the basis of the inter-reliability challenges faced, I decided that the in-house scale would be a more pragmatic choice as all the tutors involved had experience of using the scale. In addition, the policy at the college where all three tutors work is that all ESL tutors go through the standardization process for marking argumentative essays at least once a year. Considering that all tutors involved had at least three years of experience working as ESL tutors in the college, and were all qualified ESL tutors, using the in-house scale seemed to offer the best option. This factor, in addition to the advantages of content validity of the in-house assessment scale became a key consideration for the adoption of the in-house scale in the main study.

Hale (1992) observes that time limits during assessments can be a major source of anxiety to test takers who may be worried about not being able to complete the task effectively within the time limits. This

view is supported by Powers and Fowles (1997, p.2) who argue that 'some individuals, or sub-groups, may be disadvantaged because of their characteristic test-taking styles (e.g., slow and deliberate) or their particular cultural values (e.g., a low premium on quickness or a high priority on reflectivity'. This suggests that learners who may be able to attain high performance scores when allowed sufficient time may obtain low scores when working under pressure within strict time limits.

For purposes of the current study, the focus was on the vocabulary knowledge that the students possessed – in other words what the students knew -rather than how quickly they could produce the vocabulary. Therefore, no time limits were imposed per se. Instead, an indicative timeframe of 45 minutes was considered to be a reasonable starting point. It was important to set an indicative timeframe so that this could be shared with relevant authorities for purposes of logistics with regards to identification of suitable timeslots that could be utilised within the learners' timetables. Through the pilot study, it was possible to test the appropriacy of this timeframe and appropriate adjustments were made as reported in Section 4.4.

4.1.2 The Word Association Test (WAT)

Section 2.3.2 not only highlighted WAT as one of the most commonly used tests for assessing depth of vocabulary knowledge, but also identified some of the challenges associated with the test. One of these is the scoring of the test, particularly the productive version. This was not anticipated to be a problem in the current study as the selective rather than the productive version was used.

The selective WAT was deemed the most appropriate for this study as it is well aligned to the receptive VLT used to get an indication of learners' vocabulary size. However, the challenge likely to be presented by a selective test such as WAT is the likelihood of guessing. This is because the answers are already provided so guessing and getting the answers right is a possibility. With these potential issues in mind, I decided to pilot the WAT test to see what potential challenges could emanate from my specific research context and how these could be curtailed.

4.2 Participants for the study

As with the participant group for the main study and those from the corpus, participants for the pilot study came from a wide range of nationalities. Their ages ranged from 18-34 years. They had been in the UK for a relatively short period of time ranging between three and eleven months. Their English learning background was variable but they were all assessed through the institution's in-house initial testing system which is used for placing students on courses (see Section 3.3.2.1). Based on this assessment, the students were deemed to be working at Entry Level 3. This is the same level as the student group selected for the main study.

Table 4.1: Profiles of participants for the pilot study

Student	Age	Gender	First Language
S1	18	F	Arabic
S2	23	M	Korean
S3	22	F	Arabic
S4	32	F	Arabic
S5	34	F	Portuguese
S6	25	M	French
S7	21	M	French
S8	28	F	French
S9	19	F	Chinese
S10	27	M	Chinese

4.3 Insights from the pilot study

4.3.1 Pilot data analysis and findings from the LFP

The LFP is an online tool developed by Laufer and Nation (1995). It uses frequency lists to compute profiles of the learners' vocabulary knowledge. This is achieved by breaking down vocabulary in a text to show the percentage of K1, K2, and AWL words. The analysis also identifies the percentage of words that do not belong to any of the three identified categories and classifies these as 'Offlist'. The analysis of text using LFP involves a straight forward process of either typing up or uploading relevant text into the programme. The analysis results in an automatically produced lexical profile of lexical use in a text.

Since the LFP utilises free writing text to generate lexical profiles, I considered that once free writing samples were collected, it would be beneficial to pilot the LFP in order to familiarise myself with the tool in preparation for the main study. In addition, since the analysis of text using the LFP requires manual preparation of data (checking spellings and discarding unidentifiable words) before the text can be subjected to the automated analysis process, the pilot was considered to be an ideal opportunity for trialling this process before attempting it on the main study. To this end, the LFP program was utilised to derive lexical profiles for the ten students who took part in the pilot study.

4.3.2 Insights from data analysis

This section provides details of the overall insights that were obtained from the Pilot Study:

Lexical categorisation of text: As highlighted in Section 4.3.1, the LFP categorises lexis in a text into K1, K2, AWL and Offlist words. The inclusion of AWL lexis as a distinct category was a major consideration in the choice of LFP as a tool for analysis lexical use since compositions for the main study were derived from students working within an academic context. However, an observation that was made from the LFPs that were obtained from the compositions of the 10 students who took part in the pilot was that each of the profiles had contained a considerable amount of words which were classified as Offlist. This made the researcher hypothesise that the limited frequency levels that are represented within this tool (four frequency levels) could be a contributing factor to the number of words classified as Offlist. For this reason, on the main study, the LFP tool was replaced by the VocabProfile tool which offers a more

comprehensive categorisation of lexical use from K1 up to K25 frequency levels, hence enhancing chances of words being classified into specific frequency bands than the Offlist category. This was considered an appropriate move because while the participants in the main study are based in academic institutions, the focus of the research question is on the investigation of lexical use across different categories rather than within the AWL category per se.

More realistic expectations in terms of time commitment: Marking and annotating learners' work proved to be much more time-consuming than was initially envisaged. Secondly, learners enthusiastically took up the opportunity for one-to-one feedback as they considered it an additional learning opportunity. This meant that the time required for completing the data collection stage had to be extended significantly beyond the initially planned timeline to accommodate the time required for feedback.

The complexities involved in making sense of lexical errors from an etic position: Analysing and making sense of errors proved to be more complicated than initially expected. This implies that it is not enough to understand lexical issues from the researcher's perspective only. An emic perspective as offered by the learners themselves is likely to be illuminating. This confirmed the value of the semi-structured interviews / stimulated reconstructions which were carried out in the main study.

Time limits for the writing activity: During the pilot study, no time limit was imposed and so learners insisted on re-drafting. Muncie (2002) carried out a study to find out whether lexical use would improve across drafts produced by second language University level

students. The study revealed that lexical use became more sophisticated in later drafts. From this perspective, the idea of drafting and re-drafting raised questions about the implications of comparing first-attempts against redrafted productions. This could have an impact on the results obtained. On this basis, I decided that a consistent approach would be important so all the learners were encouraged to produce a plan and a draft of their work before they produced a final draft. This procedure is typical of the ESL assessment procedure that the students were required to follow during their formative and summative assessments. The plans and drafts were neither marked nor considered part of the assessment. However, in their English assessments on the course they were studying, the students were awarded marks for producing a plan and / or a draft. Therefore, encouraging students to plan and/or draft their work was beneficial to the students as it gave them the opportunity to practice free writing under conditions that were a very close match with the context of their final assessments on the course. The last student to complete the writing assessment did so within about 55 minutes while the first student to finish did so within just over 30minutes. On this basis, the timeframe for the actual study was set at one hour. This was a guide that the researcher was prepared to work flexibly around and this was communicated to both the learners and their tutor to ensure that students did not feel under extensive time pressure.

Controlling for essay length: While it is not possible to draw any conclusive insights based on the mini pilot study, the pilot study findings indicated that the three students who made the highest

number of lexical errors also produced more than 350 words each where the average number of words produced by the group was 200 words. This suggests that the quality of these students' work might have been compromised by quantity. It is possible that the students did not devote much time/effort to proofreading and careful thought about their vocabulary. In order to control for this potential effect, during the main study, learners were required to stick to the stipulated word limit of between 200 and 250 words.

Judging from the requirements set for high stakes English Language proficiency examinations such as IELTS, the ability to control for the number of words seems to be part of being a skilled writer. It is also part of college and university level assessments where word limits are a key part of assessment criteria. Therefore, this was another opportunity for learners in the main study to practice a skill that they needed in their current programme of study and beyond in their academic careers.

4.4 Adaptations of tests as a result of the pilot study

The results of the pilot study highlighted that all the students who participated in the pilot had particular difficulty with words from the 10,000-word frequency band. As such, during feedback sessions, the majority of students from the pilot study indicated that they did not attempt words from this frequency band at all. It was also evident that those who did attempt such words found them very difficult and obtained extremely low marks in this section compared to other areas.

In particular, the following insights pertaining to the 10,000-word frequency band were obtained during one-to-one feedback sessions:

i. Three students reported that they did not know the words so they found it frustrating to try and work out answers for words that did not mean anything at all to them.

ii. The majority of the students (seven in total) reported that they left the questions blank because they did not see the point in trying at all as they knew they would not get the answers correctly.

iii. Two students said that they guessed because they had no idea of what the words meant.

iv. One student said that she spent most of her time trying to work out the difficult questions. Then, in the end, she did not finish the test because she felt under pressure to finish when other students left.

v. All the students reported negative feelings associated with not being able to answer some of the questions. These included anxiety and feelings of inadequacy as some learners felt that their performance in the test was an indication of their general language performance.

vi. The LFP tool was replaced by the VocaProfile tool as discussed in Section 4.3.2 above.

The main point that I derived from the feedback provided by students (i – v above) is that it is important to be cognisant of students' abilities and ensure that the assessments administered are not completely outside the students' Zone of Proximal Development (ZPD). Devel-

oped by Vygotsky (1896-1934), the ZPD refers to the range of skills that an individual can develop with guidance or support, in other words, it is an individual's developmental potential (Ohta, 2005).

In a language learning context, 'the ZPD is the distance between the actual developmental level as determined by individual linguistic production, and the level of potential development as determined through language produced collaboratively with a teacher or peer' (Fani and Ghaemi, 2011, p.1553). Subjecting learners to a highly complex assessment which is well beyond their current level of potential development would be unethical in the sense that it would expose learners to unnecessary levels of anxiety and deprive them of the potential pedagogic benefits associated with test-taking (Leung and Mohan, 2004). Therefore, for purposes of the main study, the words from the 10000-word level were excluded as they were deemed inappropriate for Entry 3 learners.

4.5 Chapter summary

This chapter presented details of the pilot study which was carried out to test the feasibility of undertaking the actual study. An overview of the relevance of piloting in research was presented first at a general level before a rationale specific to the current study was identified and presented. On the basis of literature highlighting the potential value of carrying out pilot studies, the expectation was that some useful insights would be obtained from the pilot study. Indeed, the pilot successfully helped refine the research design and helped with the identification of potential pitfalls that could compromise the validity of the study. The insights obtained also highlighted potential ethical dilemmas and were therefore used to guide further

development of the research plan and research instruments. Thus, piloting the study proved to be a vital part of the current study design and therefore enhanced the overall success and validity of the study.

Chapter Five:

Findings and data analysis Part 1

5.0 Chapter Introduction

This chapter presents the study findings from the quantitative and qualitative analyses of data collected and analysed in order to address the research questions. The chapter draws findings from data collected through the VLT, the WAT and the free writing activity. These are integrated with findings from the tracing of individual trajectories where learners' performances are traced across all the three assessments, resulting in a triangulated approach. Throughout the chapter, the results from each part of the analysis are linked to the research question(s) that they are intended to address. This helps illustrate how specific research objectives and the overarching research aims are met.

5.1 Findings from quantitative analysis

5.1.2 Research Question 1: What is the relationship between learners' vocabulary size (breadth) and the quality of their vocabulary knowledge (depth)?

In order to address RQ 1, scores from the VLT were correlated with scores from the WAT as this was used as a measure of the learners' depth of vocabulary knowledge. Descriptive statistics as shown in Table 5.1 below were computed for all the three tests that the learners completed i.e. VLT, WAT and free writing. The descriptive statistics are reported before the correlation data in order to provide an overview of the learners' performance across the three tests.

Table 5.1: Descriptive Statistics for ESL students

	Mean	Standard Deviation	N
VLTK2	85	15	18
VLTK3	86	12	18
VLTK5	72	24	18
VLTAWL	59	23	18
WAT	61	16	18
Total VLT	54	11	18

As can be seen from Table 5.1, the highest means were recorded within the higher frequency bands (VLTK2 and VLK3) compared to the lower frequency bands (VLTK5 and VLTAWL). This highlights the higher scores, which ESL students obtained from higher frequency bands compared to lower frequency bands. However, the mean score of 72 for the VLTK5 category is somehow unexpected for a low proficiency group such as the ESL group investigated in this study. According to Laufer and Nation (1995), in as far as VLT scoring is concerned, mastery of a particular frequency band begins at 67% (based on a minimum score of 12 out of the 18 items tested per frequency band). Therefore, at first sight, a mean of 72 suggests that on average, the ESL learners investigated in this group have mastered the VLTK5 words. However, viewed alongside the standard deviation of 24 (which is the highest recorded standard deviation across all categories) it is clear that there may have been outliers who obtained high scores within this category. Therefore, overall, the descriptive statistics reflect the higher levels of lexical knowledge within the high frequency categories compared to low frequency bands. Considering that this is a lower proficiency group, this result is to be expected.

In terms of the correlations obtained for the ESL group, as shown in Table 5.2 below, for individual VLT categories, the results did not reveal any statistically significant correlations between vocabulary size and depth. However, when VLT bands were integrated and computed against the WAT scores, a statistically significant relationship was obtained at a confidence interval of 95% as shown in Table 5.2.

Table 5.2: Spearman's correlation coefficients for VLT and WAT scores for the ESL students

	WAT
VLTK1	.39
VLTK2	.28
VLTK3	.38
VLTK4	.40
VLTK5	.40
TotalVLT	.49*

VLTK2: the 2000-word level band; VLTK3= the 3000-word level band; VLTK5=the 5000-word level band; VLTAWL=the academic word level. TotalVLT=total VLT score. N=18

* Correlation is significant at the 0.05 level (2 tailed)
**Correlation is significant at the 0.01 level (2 tailed)

According to Cohen (1988), a correlation index under the value of 3 (rho=.10-.29) is considered weak; an index below 5 (rho=.30-.49) is classified as a medium relation and an index of 5 and above (r=.50-1.0) would be strong. Therefore, the correlation obtained between total VLT scores and the WAT scores in the current study reveals a medium relation. Possible interpretations of these results are proffered in the Discussion chapter.

5.1.3 Research Question 2: What is the relationship between learners' vocabulary size and the quality of their written compositions

In order to address Research Question 2, correlation analysis was carried out between VLT variables and written language ability scores. Findings from descriptive statistics computed as preliminary analyses are presented in Table 5.3 below:

Table 5.3: Descriptive Statistics for Writing and VLT for the ESL Group

	Mean	Standard Deviation	N
WritingTotSc	48	7	18
Total VLT	54	11	18

As Table 5.3 shows, the descriptive statistics provide a useful overview of the learners' written language skills as well as their vocabulary knowledge. As was highlighted in the Methodology chapter (Section 3.5.2) a 50% score and above) on the free writing test represents the starting point for what is considered to be an HLA score in this study. Therefore, a mean of 48% is lower than this cut off point. This supports the view taken in this study that the ESL learners are a lower proficiency group. For the VLT scores, a mean of 54% also falls below the level of mastery of a particular frequency band as this is set at 67% (Laufer and Nation, 1999).

Table 5.4 provides a summary of the results obtained from the Spearman correlation analysis. The total score that assessors awarded for the free writing activity comprised three components of written language ability, each of which was awarded an individual score (Task, Organisation and Language categories; see Appendix 1 for further details). The total score was therefore an aggregate of

these individual scores. The correlation analysis followed the same format which was adopted by teachers in the marking of compositions i.e. computations were obtained for each sub-category first (task, organisation and language) as well as the total writing score against each of the lexical indices.

Table 5.4: Spearman's correlation coefficients between writing and VLT scores for ESL learners

	Task	**Organisation**	**Language**	**Writing**
VLTK1	.30	.43	.32	.48
VLTK2	.26	.46	.29	.38
VLK3	.32	.41	.24	.39
VLTK4	.33	.20	.11	.20
VLTK5	.39	.29	.09	.25
TotalVLT	.52*	.39	.18	.51*

VLTK2=Vocabulary Levels Test at the 2000-word level band; VLTK3=Vocabulary Levels Test at the 3000-word level band; VLTK5=Vocabulary Levels Test at the 5000-word level band; VLTAWL= the academic word level of the Vocabulary Levels Test. VLTTotSc=total VLT score. N=18
*Correlation is significant at the 0.05 level (2 tailed)
**Correlation is significant at the 0.01 level (2 tailed)

As can be seen from Table 5.4, statistically significant associations were found between the Task sub-category of the writing score and the total VLT score, as well as the total VLT score computed against the total writing score. This suggests that lexical development has an impact on the quality of students' written language productions. An interpretation of these findings is provided in the Discussion chapter.

5.1.4 Research Question 3: What is the relationship between the quality of learners' vocabulary knowledge (depth) and the quality of their written compositions?

The expectation here was that higher scores on the vocabulary depth measure (WAT) would be associated with higher scores for overall written language proficiency. The finding was that out of the three variables constituting the total writing score (i.e. Task, Organisation and Language), the Task and Organisation scores reached statistical significance and explained 34% and 25% variance respectively. The Language score and the total writing score each computed against the WAT score did not reveal any associations, a result which was rather unexpected on the basis of the central role of lexical knowledge in the development of linguistic proficiency' (see Section 2.2.9). Table 5.5 summarises findings for the WAT vs. Writing correlation analysis.

Table 5.5: Spearman's correlation coefficients for free writing and WAT scores

	Task	Organisation	Language	Total Writing Score
WAT	.58*	.50*	.17	.45

VLTK2=Vocabulary Levels Test at the 2000-word level band; VLTK3=Vocabulary Levels Test at the 3000-word level band; VLTK5=Vocabulary Levels Test at the 5000-word level band; VLTAWL= the academic word level of the Vocabulary Levels Test. N=18

*Correlation is significant at the 0.05 level (2 tailed)
**Correlation is significant at the 0.01 level (2 tailed)

5.1.5 Research Question 4: Is there a relationship between the learners' lexical profiles produced by VocabProfile, and the quality of their written compositions?

In order to address Research Question 4, correlation analysis was carried out to investigate the relationship between learners' scores on

the different frequency bands as produced by VocabProfile on the one hand and their written language proficiency scores on the other hand. Table 5.6 shows the results obtained from the Spearman correlation analysis for VocabProfile variables and writing scores.

Table 5.6: Spearman's correlation coefficients for ESL learners' writing and VocabProfiles

	Task	Organisation	Language	Writing
VocabProfK2	.55*	.21	.26	.31

VocabProfK2 = The first 2000 most common words category on the Vocabprofile tool. N=18
*Correlation is significant at the 0.05 level (2 tailed) **Correlation is significant at the 0.01 level (2 tailed)

As can be seen from Table 5.6, students who demonstrated an ability to use the first 2,000 most frequent words also ranked highly in terms of their ability to fulfil expectations of the task (Task category) during the free writing activity. Therefore, the association between Task and VocabProfK2 explained 30% of the variability in learner profiles at the .05 significance level. Associations amongst all the other Vocabprofile variables computed against the overall writing score were remarkably low and did not reach statistical significance. In contrast, for the 40 ICLE students, correlation analyses revealed statistically significant between the Task, Language and total Writing scores on one hand, and the Offlist lexical category of the VocabProfile tool on the other hand. These results are summarised in Table 5.7 below.

Table 5.7: Spearman's correlation coefficients for ICLE learners' writing and VocabProfiles

	Task	Organisation	Language	Writing
VocabProfOfflist	.24	.32*	.53**	.47**

VocabProfOfflist = The Offlist category of the VocabProfile tool.

*Correlation is significant at the 0.05 level (2 tailed) **Correlation is significant at the 0.01 level (2 tailed)
N=40

These findings are given further consideration in the discussion section where possible interpretations and theories that may help explain these results are explored.

5.1.6 Research Question 5: Is there a difference between the VocabProfiles of learners working at a lower proficiency level (ESL learners) and those at a higher proficiency level (ICLE students)?

Based on literature highlighting the importance of lexis in the development of linguistic skills (see for example Albretchtsen, et al., 2008; Alderson, 2005) the expectation in terms of the relationship between VocabProfiles and learners' proficiency was that there would be a significant difference between the profiles of the two groups. In other words, the expectation was that learners' lexical profiles would reflect their linguistic abilities as demonstrated by their use of vocabulary in free language compositions. This would mean that learners operating at lower levels of proficiency would rely more on high frequency words (as reflected by their VocabProfiles) compared to counterpart learners operating at higher proficiency who should show evidence of better access to low frequency and Academic Word List words.

As highlighted in the Methodology chapter, learners from the ESL group are generally operating at lower levels of language proficiency compared to ICLE students. Based on computations of VocabProfile scores for each group, comparisons between the two groups provide some empirical evidence which suggests that for the learners investigated in the current study, VocabProfiles can be a

distinguishing factor between/ among learners operating at different proficiency levels in terms of their written language output. Since it was hypothesised that higher proficiency learners will produce lexical profiles characterised by use of higher level vocabulary, the Mann Whitney test was used to compare the two groups in terms of their use of words from different frequency bands. To that effect, data from ESL students was analysed against data from the ICLE students, results Mann-Whitney U which are presented in Table 5.8 below.

Table 5.8: Mann-Whitney U results for ICLE and ICLE students' VocabProfiles

	VocabProfK1	**VocabProfK2**	**VocabOfflist**
Mann-Whitney U	193	335	161
Wilcoxon W	1013	506	332
Z	-2.81	-.428	-3.40
Significance	.005	.669	.001

Significance (Bonferroni adjusted) = $p \leq 0.03$

Once the Mann Whitney analysis was completed, a rule of Bonferroni adjustment was applied in order to lower the significance level (p-value), and thereby reduce the impact of Type 1 error (inflated significance) which is likely to be associated with multiple comparisons (Lowie and Seton, 2013). As Table 5.8 highlights, the results obtained from the Mann Whitney analysis suggest statistically significant differences in the way that the two learner groups accessed and used vocabulary from the K1 and Offlist frequency bands as assessed through the VocabProfile tool. The results for the VocabProfK2 category did not reach statistical significance.

5.2 Analysis of individual trajectories for the ESL students

The insights obtained through the tracing of individual learners' trajectories provide further empirical evidence to help answer the following research questions in terms of the interplay between lexis and written language proficiency at an individual rather than group level:

1. RQ1 - What is the relationship between learners' vocabulary size (breadth) and the quality of their vocabulary knowledge (depth)?
2. RQ2 - What is the relationship between learners' vocabulary size and the quality of their written compositions?
3. RQ3 - What is the relationship between the quality of learners' vocabulary knowledge (depth) and the quality of their written compositions?

Therefore, the tracing of individual trajectories extends findings related to the first three research questions. Sections 5.1.2 and 5.1.3 above highlighted some significant correlations, which were obtained between vocabulary size, vocabulary depth and aspects of written language ability. For RQ1, the main finding was that although there was no correlation found between vocabulary size and writing when lexical knowledge was broken down according to frequency bands as derived from the VLT, when total VLT scores were computed against written language skills, a statistically significant correlation of .49 was found. With regards to RQ2, the total VLT score was found to be significantly correlated to the total writing score as well as the Task category of the Writing sub-scale at .51 and .52 respectively. For RQ3, depth of vocabulary knowledge was found to be significantly correlated to the Task and Organisation subscales of the Writing

scale. All of these findings from the study provide insights at group level.

Therefore, the tracing of each learner's performance across all assessments is intended to add a new dimension to the findings in terms of the interplay between vocabulary knowledge and written language proficiency at an individual level. To that extent, findings from the qualitative analysis will complement statistical analysis and/or provide alternative insights into the relationship already revealed by statistical analyses. Therefore, the aim here was to triangulate these qualitative insights with quantitative insights obtained from the correlation analysis. Table 5.9 presents each learner's profile captured through their scores across the three categories (vocabulary size, vocabulary depth, and written language skills). These are colour-coded to reflect performance levels within each skill (see section 3.5.2 of the Methodology chapter for the method adopted for categorising this data).

Table 5.9 shows that each leaners' performance was classified into the lower and higher attainment categories (see section 3.5.2 for a more detailed discussion of the methodology and rationale behind it). This allowed for the tracing of individual learner performance across the three assessments. From Table 5.9, it can be seen that the highest number of HLA scores was recorded within the VLTK2 category (16 students) followed by the WAT category where 14 students achieved an HLA score. These results support findings from correlation analysis where WAT and VLT produced a statistically significant result of medium strength (.49).

Table 5.9: Findings from the analysis of learners' individual trajectories

	2K	WAT	Writing
S1	18	80	55
S2	15	68	50
S3	18	70	49
S4	17	59	46
S5	18	40	45
S6	16	38	56
S7	17	55	46
S8	17	82	46
S9	9	73	41
S10	10	33	39
S11	15	63	43
S12	14	50	43
S13	14	61	55
S14	17	76	53
S15	17	43	41
S16	18	91	70
S17	12	65	49
S18	15	57	46

Key

Higher Level Attainment (LLA)

Lower Level Attainment (LLA)

Free Writing recorded the lowest number of HLA scores with only 6 students achieving scores within the HLA band. Chapter 2 highlighted writing as a highly complex skill which requires the simultaneous orchestration of multiple skills including efficient retrieval of appropriate lexical knowledge (see Section 2.7.2). To that effect, the expectation would be for a trajectory where HLA VLT is associated with an HLA WAT score and ultimately, an HLA Writing score. This trajectory is evident in 5 out of the 18 students (S1, S2, S13, S14 and S16). On the basis that S3, S4, S7, S8, S11, S12, S17 and S18

achieved HLA VLT and HLA WAT, it can be inferred that these 7 students' profiles may also be developing along a similar trajectory as the first 5. Therefore, a total of 12 out of 18 students' trajectories can be said to be potentially following a developmental trajectory where HLA VLT is associated with HLA WAT which in turn is associated with HLA Writing. However, S6 and S9 were exceptions to this trend. A further 3 students (S5, S10 and S15) could not be classified as either conforming or detracting from the expected trajectories. This is because S5 and S10 only obtained one HLA (within the VLT category) which means that the students' developmental trajectories could potentially conform or diverge from the observed trends. Similarly, S15 achieved one HLA score (VLT), so it is not possible to predict this student's trajectory based on their current performance. The profiles of S6 and S9 are also noteworthy because S6 is the only student who obtained an HLA for writing after obtaining an LLA for WAT, while S9 is the only student who obtained an HLA for WAT after obtaining an LLA for VLT. Therefore, the primary finding from the tracing of the individual profiles of the ESL students presented in Table 5.9 is that the majority of the learners' trajectories (67%) followed a developmental trajectory whereby high VLT scores were associated with high WAT scores, which in turn were associated with high written language proficiency scores. The findings support the statistically significant correlations that were found between VLT, WAT and written language proficiency. However, there were also exceptions where students' profiles diverged from the typical trajectories. Such trajectories could not be gleaned from statistical analysis alone as statistical analyses generate insights based on averages and

therefore do not offer insights into profiles of individual learners as was offered by the analysis of individual trajectories.

Since the analysis of individual trajectories highlights the nuances of learners' profiles that could be concealed by an approach that focuses exclusively on group-aggregated data, the trajectories are an important part of the study design. They provide qualitative insights to the quantitative insights generated through the SPSS analysis procedures. I therefore link the main observations from the learners' trajectories to each of the first three research questions as a way comprehensively generating corroborative qualitative findings to the investigation of the interplay between vocabulary knowledge and written language proficiency which is at the core of the current study.

5.2.1 Corroborating qualitative findings from the tracing of individual trajectories with quantitative findings

Recall that a significant correlation was found between the total VLT and the WAT scores for the ESL learners investigated in the current study. Similarly, for Research Question 2 which sought to investigate the relationship between vocabulary size and written language ability, statistically significant associations were found not only between the Task sub-category of the writing assessment scale and the total VLT score but also, between the total writing score and the total VLT score. As was reported in Section 5.2, when individual trajectories were traced for each learner, a total of 12 out of 18 (67%) students' trajectories were considered to have followed a developmental trajectory which suggested that HLA VLT scores were associated with HLA WAT scores, which in turn were associated with HLA writing scores. This suggests that for the majority of the students, a high vocabulary

size was largely associated with better-developed depth of vocabulary knowledge compared to their peers, and was also related to higher written language proficiency.

However, in light of the cases of S6 and S9 who diverged from this trend, an important finding which comes out is that of variability. Since the sample size for the current study was small, the qualitative results are particulary important as they provide opportunities, as illustrated above, to generate additional insights that could have been averaged out by an analysis exclusive to the group level. These findings are not surprising because as the Literature Review chapter highlighted, language learning is a complex system and so its development is likely to show dynamic variation rather than straighfoward linearity. This view of L2 development as a complex system is further explored in the Discussion chapter where findings from the study are given theoretical interpretations.

5.3 Chapter summary

A number of findings emanated from the quantitative analysis of data which utilised learners' profiles derived from the three written assessments completed for this study. The tracing of individual trajectories provided qualitative evidence which revealed additional insights to those obtained from statistical analyses. The findings from this chapter are further utilised in Chapter 6 to provide a starting point for further, in-depth, qualitative analysis of learners' lexical knowledge and how this relates to the skill of writing. This allows for further triangulation of insights from quantitative and qualitative analysis which ultimately enhances the validity of the study. This is an

important consideration in the context of the MMR approach adopted in the study.

Chapter Six:

Findings and data analysis Part 2

6.0 Chapter introduction

The previous chapter analysed qualitative data (obtained from the tracing of individual learner profiles) alongside quantitative analysis (SPSS procedures) for the ESL students in the study. The chapter utilised the MMR approach adopted in this study in order to provide deeper insights as was argued in the Methodology chapter. The MMR approach adopted in this study continues to be visible in this chapter where qualitative data is used to provide further insights on the research questions which were posed in the study relating to the interplay between vocabulary size and depth, vocabulary size and free writing skills as well as vocabulary depth and free writing skills. The quantitative analysis highlighted the correlations that were found between these variables, results of which were corroborated with findings from the tracing of individual trajectories.

6.1 Analysis of data from stimulated reconstructions and semi-structured interviews

In this section, the analysis moves forward to data obtained from stimulated reconstructions / semi-structured interviews which were based on incorrect lexical choices made by ESL students on the WAT. To that effect, the stimulated reconstructions / semi-structured interviews provide data which is based on the learners' own perspectives about their vocabulary knowledge. Throughout the chapter, links will be made between findings from the stimulated reconstructions and the way in which they extend insights on the research questions posed at the beginning of the study. The stimulated reconstructions help to answer the research questions raised in the study by providing learner perspectives on the interplay

between vocabulary size and depth, vocabulary size and writing as well as vocabulary depth and writing.

6.1.1 The relationship between vocabulary knowledge and writing

As shown in Appendix 4, the first question that was posed to interviewees sought to find out which of the three tests (VLT, WAT and Writing) presented the least challenges and which presented the most challenges to the participants. The expectation here was that higher achieving learners such as Student 16 who outperformed peers across all three tests would find the writing task more manageable on the basis of both the significant correlations found between lexical variables (VLT and WAT) on the one hand, and free writing on the other hand. Thus, if higher performing students such as Student 16 had reported that they had found writing easy, then this would have suggested that perhaps the learner's strong VLT and WAT scores might have been contributing factors to the ease of the task. This would support both the correlational findings where significant findings were obtained between lexical variables and writing. This would also shed further light on the lexical trajectories discussed in Section 5.2 above where writing seemed to be preceded by high VLT and WAT scores. However, this was not the case, i.e. high scores in written tests did not necessarily predict particular response in terms of which of the tests the students found difficult or easy.

All interviewees unequivocally highlighted free writing as the test that they found to be the most challenging while VLT was reported to be

the test that posed the least challenges across the board regardless of students' performances on the three tests.

6.1.2 Learners' perspectives on size vs. depth of vocabulary knowledge

In this part of the interview, the questions posed to the students were intended to probe into their views on the concept of size and depth of vocabulary knowledge. The overarching aim was to tease out whether the students felt one had greater utility than the other for their language development (reading, writing, speaking and listening) and their reasons for any views held. Students were prepared for this part of the interview to ensure that they understood the concepts of size and depth as applied to vocabulary knowledge.

The distinction between vocabulary size and depth could not be assumed as prior knowledge for the ESL students as they were a low proficiency group so this was explained to the learners. In addition, during interviews, the discussion of size and depth always ended with a direct question from the interviewer asking the interviewees which of the two (size or depth) they considered more important. This was important because it gave the interviewees another opportunity to clarify what size and depth mean, thus helping to maintain construct validity. Asking for direct confirmation from participants as a way of a conclusion to the discussion about size and depth was also an important way of ensuring that the learners provided their own interpretations of the narratives that they provided. As a result, the decision as to whether each participant considered size or depth to be more important was firmly grounded on the data provided by the participants themselves.

The interviews revealed that for the majority of study participants, vocabulary size was considered to be more important than depth. Only two learners argued the case for developing word knowledge further than the form-meaning link, with one of these learners (Interviewee Seven) suggesting that size and depth are equally important. The Interviewee highlights her frustration with inappropriate use of words in written discourse when she says:

> *"I do not like it when my teacher she find wrong words all the time … I think now I know this word and my teacher she says it's the wrong word*…it's hard. That's why I think now I want to study more informations about my words, in dictionary, from teacher, my friends, everywhere about some important words "

The commentary above suggests that the learner wants to develop the quality of lexical knowledge to avoid penalties associated with inappropriate use of words in writing. The learner highlights an interest in learning more about the words already known, and identifies use of independent resources (dictionary), the teaching staff and friends as possible sources of help in terms of lexical development. This suggests a that the learner is keen to develop depth of knowledge. However, it is also possible that the learner may have been affected by the social desirability bias (see Section 3.5.4.). Indeed, it is possible that this bias may have affected any of the other learners who participated in the study. As such, participants were also asked to indicate the type of activities that they actually engaged in as part of their language development and responses provided some verification data. These findings are presented in Section 6.1.5.

Interviewee Three is the other student who identified the importance of depth of knowledge and for this learner, *'quality is better than just many words in my notebook'.* The interviewee highlights that *'even notebook is not useful for exam because how I can use the word in a sentence?'* The learner therefore argues that to avoid challenges with word usage in the exam, learning more about words should be a priority, hence the interviewee's assertion that *'learning how to use the words really is big priority for me'.*

For the remaining six interviewees who confirmed that size is more important to them than depth, the overarching finding was that the form-meaning link was identified as the most important part of learning vocabulary. Once the basic form-meaning link was established, the learners considered a word 'learned' and did not appreciate the value of further work on such lexical items. This view was noted by all except one interviewee (Interviewee Two). To this end, Interviewee One argued that '*you must do the meaning of word and put in vocabulary book all of the new words…because if you don't know the words, then it's useless'.* Similarly, Interviewee Four argued that *'If I know meaning of that word I use it in writing and it will make my writing good'.*

For Interviewee Four therefore, knowing the meaning of a word guarantees an ability to use it in writing which in turn results in written language ability. For Interview Five, any further activities related to a word whose form-meaning link has already been established is a waste of time:

> *My dream just to learn many word not wasting any time to do the thing I know already. If I know already that word, what it*

mean, that is very good…Why I can practise again the same simple word because I must learn more word for my writing so every week I have new word to learn.

Similar sentiments are echoed by Interviewee Six who argues that:

I try to learn what means the new words. I write it in my book of new word. I get many word like this now so I have no time to do more because already I know this meaning

Interviewee Eight suggests that as a non-native speaker, it is perhaps not necessary to go beyond the form-meaning link as revealed by the following narrative:

English not my first language so may be no need to learn all meaning about the words because if I know many words I will not be in difficult position.

Interviewee Two seemed to be content with limiting learning to the classroom environment and did not seem to have much of a rationale behind the learning activities that he engaged in:

For me it don't matter, I learn many word at school because just I need to do all of my works.

Therefore, overall, there were more students who considered size to be more important than depth and the form-meaning link came out as being synonymous to vocabulary learning. The implications of these findings are given further consideration in the Discussion chapter.

6.1.3 Learners' perspectives on the role of lexical knowledge in the development of the skills of reading, writing, speaking and listening

When asked about their views on how important they consider vocabulary knowledge to be for the development of the four skills of

reading, writing, speaking and listening, narratives from the students in the current study suggest that the learners generally felt that vocabulary knowledge is key to all aspects of linguistic development. A selection of narratives to this effect is presented in Table 6.1 below:

Table 6.1: Vocabulary knowledge and the skills of reading, writing, speaking and listening

	Excerpts from student narratives
Int1	*For me I think all of them, yes **all of them vocabulary very very important**. Because I say before, if you don't know the words, then what will you do? No writing, no speaking or anything like that, so yes, all of them very important*
Int2	*I think first, ask any student, just any from my class. I think all of them they will just tell you that **vocabulary is important to them.***
Int3	*All of us, **we need vocabulary to speak, to write, to read as well so** I think no one is better than the other but yeah, all of them very important for vocabulary.*
Int4	*I think **vocabulary is important but to me, writing is my biggest problem**. Sometimes I have vocabularies for speaking but writing, **many times I can't find the word even in my head I know my idea but saying it real is a problem.***
Int5	*As you will evidence, speaking, listening and reading, all of them, **vocabulary is very important to me**.*
Int6	*For me, vocabulary is more important.... **very important because I need this for all of the works. For the writing, speaking, listening all of these.***
Int7	*You talk about all of these things. **You cannot do these things if you don't know the words.***
Int8	***There is no way of doing the speaking or the writing test if you no have many vocabularies***

Int1 – Int8 = Interviewee One – Interview Eight

The positive beliefs about the role of vocabulary knowledge in the development of key linguistic skills are particularly embodied in the emboldened text. Such beliefs have practical implications in ESL learning contexts as they are likely to contribute towards developing and sustaining motivation levels during the process of vocabulary learning.

6.1.4 Learners' perspectives on the role of low frequency vs. high frequency words in writing

Based on the literature review which suggests that high frequency words are likely to be easier to learn than less frequent, more sophisticated vocabulary (see Section 2.5), learners' views on the use of high frequency compared to low frequency words were also sought. This part of the interview focused on the type of vocabulary that learners thought was important for them to learn so that they could attain requisite scores on their assessments, particularly the writing assessment which is central to the aims of the current research. The researcher therefore sought to find out interviewees' views about the types of words (short, easy words that most people are likely to know compared to more difficult words that most people are not likely to know).

The main point raised by Interviewee One in this regard was that '*you cannot just use small word in writing and you think maybe you will get good luck'.* In as far as this learner is concerned then, a major contributing factor in passing writing examination (which was the focus of this particular interview question), is the learners' ability to use sophisticated vocabulary. As such, the learner argues that:

> *If you just use small word all of your writing, that will fail the writing examination because everybody see this word all the time in book, TV or somethings like that, so nothing is special with that kind of writing*

In general, high frequency words would fall into the kinds of words that Interviewee One perceives as 'not special' because of their

frequency of occurrence. The implications of such perspectives are considered in the Discussion chapter in light of the results obtained from both VocabProfile and assessment scores for learners' written compositions.

Similarly, Interviewee Two feels that what makes writing difficult is in fact the need to use sophisticated words. The learner expresses confidence in his vocabulary size but feels that the types of words that he knows are limited to 'easy' words. He therefore comments that:

> *I know many many words, but I think they just not really good words…. so I fail writing but always by few marks….it frustrate me a lot because obvious to me that my English still no good*.

To this learner, failing writing is attributed to the use of 'small' words.

For Interviewee Four, the '*number one reason I want to learn on this course is to learn many different words, many of big words to use then I will pass writing test'*. Interviewee Four expresses feelings of frustration related to the realisation that he quite often has good ideas which he cannot express effectively, hence his argument that '*I know all the idea but I can't find the right word or the good word'.* The learner ascribes the lack of desired vocabulary knowledge to poor memory as he says that *'In the end I just write easy word because I cannot remember this word I want; I have bad memory'.*

Interviewee Seven's attitude towards low frequency vocabulary comes out through the student's reliance on the thesaurus for

identifying low frequency vocabulary to use during written language assessments:

> *I write down the word…then I will look at my phone to see what is another word it mean this word, then I will choose best words; not the easy word everybody will know.*

Interview Eight highlights disagreement with the ESL teacher's advice that sophisticated vocabulary does not necessarily guarantee good writing. This is evident from the interviewee's narration which states that:

> *My teacher he say all the time it's not long word that make do writing good, but I don't agree to him on this thing'.*

As the selected extracts highlight, it is evident that the dominant perspective among students interviewed in this study was that knowledge of less frequent vocabulary is an important feature of linguistic development including the development of written language skills. These findings and their implications are considered in the Discussion chapter.

6.1.5 Learners' involvement in language development activities outside class

Learners' involvement in language development activities outside class was considered an important part of the current investigation. Data from this part of the interview provided an opportunity to verify self-reported data about beliefs on the importance of size and depth dimensions of vocabulary knowledge. The findings here were that out of the eight participants, six students confirmed that they view inde-

pendent learning activities as a key part of their language development. Examples of activities that learners said they find useful and therefore engage in included reading English books with children' (Interviewee One); doing grammar exercises then using answers at the back of the book for self-assessment and reading English books (Interviewee Three) doing vocabulary activities then using answers at the back of the book for self-assessment (Interviewee Four); doing internet based research on areas of interest (Interviewee Five); taking part in Language study groups organised by local libraries (Interviewee Seven); and reading magazines in areas of interest, watching English channels on television and holding conversations with family and neighbours (Interviewee Eight).

Table 6.2 provides a synthesised view of the learners' profiles based on their perspectives on vocabulary size and depth issues; the role of vocabulary knowledge on the development of writing, reading, speaking and listening skills, their perspectives on low frequency vocabulary as well as their perspectives on independent learning as part of their learning. A tick (√) under the heading *'Size'* means that evidence from semi-structured interviews suggests that the learner is aware of the importance of vocabulary size in the development of linguistic skills. A tick (√) under the *'Depth'* column means evidence from semi-structured interviews suggests that the learner is aware of the importance of vocabulary size in the development of linguistic skills. Similarly, a tick (√) under *'Rare vocabulary'* means evidence from semi-structured interviews suggests that the learner believes that rare/sophisticated vocabulary is key to the development of linguistic skills and finally, a tick (√) under *'Independent work'* means

evidence from interviews suggests that the learner engages in independent work outside class to improve reading, writing, speaking and/or listening skills.

Table 6.2: Synthesis of quantitative and qualitative results relating to ESL learners

	Size	Depth	Rare Vocabulary	Independent work	VLTK2	WATS	WRITING
S1 Int 1	√		√	√	18	80	55
S3 Int 2	√		√		18	70	49
S6 Int 3	√	√	√	√	16	38	56
S8 Int 4	√		√	√	17	82	46
S10 Int 5	√		√	√	10	33	49
S13 Int 6	√				15	61	55
S16 Int 7	√	√	√	√	18	91	70
S9 Int 8	√		√	√	9	73	41

The synthesised data presented in Table 6.2 reveals two main notable findings. Firstly, it provides further evidence to suggest that even though all ESL students interviewed in this study are evidently aware of the role of lexical knowledge in their linguistic development, the majority of interviewees believe that development of their vocabulary size is a more important goal for them compared to the development of depth of vocabulary knowledge. To this end, 75% of the learners said that size is more important to them than depth (6 out of the 8 interviewed) while narratives from the remaining 25% (two students only) identify the importance of depth. Secondly, having identified in the current part of the study that only two students highlighted depth of vocabulary knowledge as an important part of their overall linguistic trajectories, it became imperative to further probe into the two learners' profiles in order to establish whether there were any peculiar features of their profiles which could be linked to this finding. The finding in this regard was that one of the two learners (Interviewee Eight) obtained the highest-ranking scores across all the three assessments administered to the students. The other learner (Interviewee Three) scored the second highest ranking score for Writing, a relatively high ranking score for VLT but one of the lowest ranking scores for WAT.

In addition, while it was not the aim of this thesis to investigate tutor vs. learner beliefs on vocabulary knowledge issues, as captured in the excerpt from Interviewee Five, it is evident that the tutor had sensitised learners to the importance of not just the accumulation of new words but also, the value of developing depth of this knowledge. However, based on this learner's narrative, some learners did not necessarily agree with the tutor's views and therefore found it

difficult to appreciate the value of depth of knowledge, hence the comment from Interviewee Five that:

> *My teacher he thinks good thing to practise the old words also but I know I don't learn enough vocabularies yet so I must learn new words all the time. Me, my friends, all of us do this same thing to learn many many words. We want to learn more vocabularies.*

The same was found to be applicable to issues around the use of low frequency vocabulary where evidence was found to suggest that the tutor encouraged students to learn and use words appropriately rather than excessively focusing on using less frequent words, highlighting the potential impact of inappropriate use of lexical items. In this regard, Interviewee Eight highlights that the tutor does not always encourage the use of big words:

... *but my teacher he will not agree, just one time or two time.*

A similar sentiment is expressed by Interviewee Eight in the following narrative:

> *I remember that my teacher he say all the time it's not the long word that make do writing good, but I don't agree to him on this thing because good word is what all of the people here want to make their writings good.*

Clearly then, there is evidence of some strong beliefs held by learners which may have implications on their vocabulary learning.

6.2 Findings from the exploration of incorrect lexical choices made on the WAT

Other insights on the size and depth dimension obtained from interview data and presented in Sections 6.1.1 to 6.1.5, the study sought to understand the quality of knowledge possessed by each learner on the basis of the rationale provided for incorrect lexical choices that they made. This gave insights about individual differences in terms of how the quality of word knowledge (depth) differed from learner to learner and whether this had an impact on written language skills amongst learners compared to their performance across the three tests (VLT, WAT and free writing). As the account in this section will show, a key finding from the analysis of the learners' commentaries during the stimulated reconstructions was that while there were some common factors that seemed to impact on the quality of lexical knowledge that learners possessed, there were also some individual factors at play. An important aim for the stimulated reconstructions was to explore whether the rationale behind lexical choices would give insights to learners' vocabulary use in writing so that inferences could be drawn between lexical choices, in particular the rationale for these, and learners' written language skills and or performance in other assessment activities completed in this study. To that effect, the stimulated reconstructions provide further insights on the exploration of the interplay between vocabulary knowledge and written language skills.

Four main factors were identified as impacting on learners' lexical choices. However, none of these were clearly attributable to the learners' written language proficiency levels, or indeed, the students' performances across the other tests administered in the study. The four factors identified were as follows:

i) Cultural association

ii) Educational background

iii) Guessing

iv) Partially-developed word knowledge

There were some unclear cases which could not be rationalised and categorised. These were included within the Guessing category (see Section 2.3.2).

6.2.1 Cultural association

Responses which were categorised as 'cultural associations' were those that seemed to derive from experiences within learners' own cultural backgrounds. This categorisation was inspired by the widely-recognised view that that collaborative activities with other members of a particular culture is key to the process of constructing meaning within a particular culture (Mitchell and Myles, 2004). In the context of the current study, this was evidenced by the narratives that the learners brought to the fore when explaining the reasons behind the lexical choices that they made during the WAT. To that effect, analysis of interviewee data revealed that some of the learners' logic for making certain lexical choices had its origins in experiences unique to their own cultural environments, revealing the intricate relationship between culture and language (Moore-Hart, 1995). For example, Interviewee Seven produced the following output for Item sixteen:

Tight

close rough uncomfortable wet	bend pants surface wood

Correct lexical choice Incorrect lexical choice

For each of the WAT items capturing students' output as in Interviewee Seven's Item sixteen, the boxes are laid out as in the WAT test. Thus, the lexical items in the left box relate paradigmatically to the key word in bold (provided above the lexical options), while the lexical items in the right-hand box relate syntagmatically to the key word. The syntagmatic and paradigmatic relations are not spelt out to the learners but the learners are instructed to choose a total of four answers with a requirement to choose answers from both the left and the right-hand boxes (see Appendix 3). As the above output for the key word '*tight*' shows, Interviewee Seven was able to successfully identify three of the four correct lexical relations of the key word '*tight'*.

However, the third choice seemed so far-fetched that it could be assumed to be a result of random guessing, particularly considering that the WAT used in the study was the receptive version where learners were only required to choose options rather than produce their own lexical output. However, evidence from the learner's narrative suggests differently. When probed about this test item, it became evident that the learner relied quite heavily on his own cultural experiences to arrive at the decision to match the key word *'tight'* with '*wood'.* The learner provides the following commentary:

> Ok, *wood and tight I think that these two, they go together. 'Cause I think, like when you put woods together, then you must make it tight .....like in my country, when you put woods together, when you're going to make a fire, when you tie them.... when you tie them together, they have to be tight, because you can't carry them if like...if you don't make them like, tight. This is like big big thing for a new bride...if they can't do it then maybe husband will complain or something'*

From the student's commentary, it would appear that in his cultural context, the act of tying wood (firewood) carries quite significant cultural connotations and is a key part of a woman's duties as suggested by the learner's explanation that if a bride (determined to be a newly married wife through further probing rather than a bride per se) does not meet this expectation, then this could lead to possible misunderstandings between her and her husband. Since this experience carries a lot of significance in his cultural context, the learner makes syntagmatic links between 'wood' and 'tight'. Without data from stimulated reconstructions, it would have been impossible to understand the reasoning behind this learner's otherwise unusual lexical choices.

Similarly, for item number six on the WAT, Interviewee Three's output also provides evidence on the possible impact of cultural associations. The excerpt provided shows the learner's output for test item six, relating to the key word 'general'.

General

closed different usual whole	country idea reader street
Correct lexical choice	Incorrect lexical choice

The three correct choices suggest that the learner has considerable understanding of the lexical item *'general'* and indeed, discussion of this term during interviews attested to this knowledge. Therefore, the decision to associate the word *'street'* with the key word *'general'* was not expected from a learner whose written and verbal output strongly suggests that he is very comfortable with the meaning of this word.

The excerpt provided below captures the learner's rationale for this choice:

> *General, this word I use it a lot....I mean general, when you say general, that's not special. Special things, ... But if its general...that's mixed. ...not any special sample of the thing. Now...'general' and 'street'.....in my country, it's different. Real different. Many of the streets, we call them just general streets because they have no name. Not like England maybe each road has a name. So general, we use like that a lot.*

Evidently, the learner's lexical choice was heavily influenced by his specific cultural context, which one may argue, taught him to associate *'general'* with the word *'street'.* This suggests that it may also be possible that in the learner's L1, this lexical item will carry different associations.

6.2.2 Previous individual learning experiences

For lexical items to qualify into this category, the output produced during stimulated reconstructions had to directly link the choice made to some form of formal learning such as previous academic learning or independent study linked to a particular course of study. Interviewee One was the only participant educated to degree level (studied Law in France) so it is not surprising that some of the rationale that she gave for the lexical choices made was found to be related to her educational background. Relative to her peers, Interviewee One's lexical profile showed particular strengths in terms of her knowledge of words from the Academic Word List, although this knowledge seemed to be still developing as will be illustrated in the ensuing discussion. It is evident that the learner was able to tap into her academic background to build her understanding of AWL words, for example, her output for test item twenty-four was as shown below:

coherent

clear normal recent together	crime health speech theory

Correct lexical choice | Incorrect lexical choice

The highlighted lexical items are the only answers that the learner provided for this test item. When interviewed, the learner reported that none of the other options made sense to her and that she did not like guessing as she felt that this was a '*form of cheating*'. In terms of the actual answer provided, this is the rationale that she gave:

> *I don't know this word, I never use this word in England...but when I see it..... I mean, I think obviously it's the same like other word we use in France...you know if criminals go to court, maybe it's a really bad thing and then he don't want jail, he try something....I mean try to say something else to the person of law. Then the person will see if the story is not like, fit together.*

Looking at the two answers that the learner provided in the WAT ('*crime*' and '*theory*'), it is very difficult to rationalise why the learner would choose the two for the word '*coherent*' because one answer (theory) suggests good understanding of the key word, yet the other answer (crime) sounds far-fetched. However, what becomes evident from the interview excerpt is that the learner has used her educational background in Law to link the word '*coherent*' to '*crime*' in the sense of criminals trying to put together a coherent story to help them escape possible convictions. This provides evidence of the learner's emerging knowledge of the term but does not reveal

knowledge of the syntagmatic relationships under investigation. For the same key word, Interviewee Six produced the following output:

coherent

clear normal recent together	crime health speech theory

Correct lexical choice Incorrect lexical choice

Interviewee Six was the only student who obtained all correct answers for the key word '*coherent*' which the majority of the students found difficult. This led to the suspicion from the researcher that the learner may have guessed the answers to this item. However, the learner produced output that suggested very good understanding of the term:

> *X* (teacher's name) *bring this video and we watch it in the lesson time. It was good I enjoy it. When the video finish, X talk to us and he explain the new word, and I remember this word since from the day. Since I learn this word, I try to make my writing, and also talking, cohe.., coherent right?*

As evident from the learner's narrative, the student first came across the word *coherent* during one of the ESL lessons while on the course that he is currently on, and the word was explicitly taught to her (and her peers). Seemingly, her memory has retained this term and drawing from this resource, she was able to identify the correct responses whereas none of her peers cited this encounter.

6.2.3 Guessing

Section 2.3.2 identified that while guessing can be used as a well-informed strategy for dealing with difficult test items, it can also be used for dealing with completely unknown test items. Against these

perspectives, in the current study, responses that were categorised as guessing were those that involved considered reflection (lexical inferencing) as well as responses made on the basis of random guessing. A case in point is Interviewee Eight who produced the following Output for test item twenty-five:

dramatic

exciting official surprising worried	adventure change patient salary

Correct lexical choice Incorrect lexical choice

Interviewee Eight decided to choose all four of her answers from the left-hand side of the answer key despite clear written and verbal instructions provided at the beginning of the test instructing participants to choose at least one answer from each of the left and right hand boxes. On the basis of the rationale that the interviewee gives for the selection of her answers, it is evident that this was the learner's strategy for maximising chances of getting some marks for this question. The participant had the following to say:

> *I just guessed all. I choose four words same side, because I know some of it, some of those words, may one is right, one or two if I have good luck…. because I really don't, I don't know this word. I have no idea how to use, but you see now I got two marks. I think it's a good way, a good thing for the difficult question in test if I want to get marks, always I know I should try.*

Clearly then, guessing is a strategy that the participant used for dealing with unknown words. Instead of following the instructions that required lexical choices to be made from both the right and left hand columns, the participant realised that choosing words from the same side was a guaranteed way of getting at least one correct answer. When further probed about the use of this strategy, it became clear that the learner uses both inferencing and outright guessing as strategies for dealing with difficult items depending on the level of difficulty for a particular item:

Interviewer: *So, if you find a test question difficult, do you just guess in any way or you have a method, or a way of trying to work out the correct answer?*

Interviewee Eight: Really it depends. It depends on each test. If a very hard question like I get in this test, then I just guess. *If I see something, like something I know a little, I use it, so I get the answer this way, but truth is, I don't know this answer.*

The interview evidence presented above suggests that Interviewee Eight is comfortable using both random guessing and lexical inferencing as strategies for dealing with difficult test items. Similarly, Interviewees Three, Four and Six reported using guessing as a strategy during the WAT, hence the following excerpts which relate to Question 10 on the WAT:

Interviewer: *So, you mentioned that you just guessed that the word 'effort' can go with the word 'conscious'. Can you tell me more about how you actually go about guessing...I mean did you just pick any of the words given or did you have a way of kind of working out your guess?*

Interviewee Six: *Mhhhhh. I just look at the word and when I read it...conscious.... it remind me of what my teacher use*

to say to me sometimes...like you are a conscious student. Because she think I work really hard because I always I try, I try my best.

The excerpt from Interviewee Six suggests that the participant relied on perceived phonological similarities in trying to work out the meaning of the word. Therefore, although she puts her answer down to random guessing, it is evident that the learner used a phonologically-based inferencing strategy. As a result, the word 'conscious' was phonologically confused with the word 'conscientious' but nevertheless led to the correct choice of collocate.

Another interesting observation made during interviews with Interviewee Four relates to Question 19 on the WAT test administered in this study:

Interviewer: You gave some very good answers to the word 'independent'. But you also said that you had never heard or seen this word before. Can you tell me how you got your answers then?

Interviewee Four: Well, before you tell us that we must think about four answers maybe two and also two, or three one or one three, whatever we will find, but altogether four answers right? So I see for myself that I don't know the word independent, I never use it before, then I think what I must do because I want an answer to this one. So I look at the word because always my teacher say don't rush, always look at the word. So I am very happy because I look and I see the word dependent and I know that word, like this college they ask me if I have dependent, like my daughter so I---- that make me think that maybe independent child as well, but really, I am not sure because I never hear independent, just dependent, so I just guess really.

The interviewee claims that she guessed the meaning of the word independent but in fact, a close examination of the student's explanation as to how she arrived at her answer suggests that her prior

knowledge (knowledge of the word 'dependent') and her ability to analyse words as an inferencing strategy helped her to arrive at the correct meaning of the word 'independent' even though she had not come across it before. Therefore, her knowledge of the noun 'dependent' helps her make links with the adjective 'independent'.

What is notable about the students who adopted strategic guessing is that closer analysis of their input suggests that they adopted strategic guessing, but when questioned about their choices, the students generally believe that they made random guesses. This suggests that the learners are comfortable with both random guessing and inferencing as strategies for dealing with unknown vocabulary.

In contrast, three students (Student 1, Student 10 and Student 18) reported categorically that they would never guess because '…. that is cheating the exam' (Student 10) and '*it is wrong because it is not the honest answer if you guess … you do not know this answer, just you guess it'.* Student 1 highlighted that '*guessing… is not good; if I don't know the answer I just not say anything'*. Indeed, true to their narratives, no evidence of guessing was identified from their output from the stimulated reconstructions. The implications of these findings are considered in the Discussion chapter.

6.2.4 Partially developed knowledge

Lexical choices that were categorised as being influenced by partially developed knowledge are those that upon scrutiny, bore evidence to suggest that the learner had some knowledge about the lexical item, albeit partial. In light of the Literature reviewed in Section 2.2.1, these would most likely be words that are receptively known. The rationale for such words did not however resonate with the other three

categories already discussed. Therefore, in order for a lexical item to be classified as belonging to this category, the learner had to get at least one answer correct and/or produce a correct sentence using the word. This was to ensure that responses generated through guessing were not included in this category. Therefore, if a learner chose a correct answer but interview data suggested that the choice was a random guess, then such a lexical item would be excluded from this category as it would be considered to be an instance of *Guessing* rather than *Partially developed word knowledge.*

The requirement for students to produce a correct/meaningful sentence or indeed any example of usage of the term was inspired by insights provided by the Vocabulary Knowledge Scale (VKS) reviewed in Section 2.3.1. Recall from this review that one of the criticisms that the current research proffered about the VKS is that the scale seems to work on the assumption that if an individual has never seen a word, then that suggests they do not know the word, yet in light of Richards' (1976) framework of what it means to know a word, knowledge of what the word looks like is one out of many aspects of word knowledge. Thus, the requirement for the learners to provide an example of how a word might be used was a way of giving learners the opportunity to demonstrate their knowledge of words, no matter how imperfect. This was an important consideration in light of the Literature Reviewed in Section 2.4.2 which highlights vocabulary knowledge as incremental in nature. Therefore, the responses that fell into this category qualified on the basis that the interviewees were able to identify at least one correct answer and in addition, they were able to produce evidence (e.g example of usage in a sentence) to show that their answers were not randomly guessed. This category emerged as the most common out of the four, with the majority

(71%) of all the narratives produced by the interviewees relating to it. A case in point is the output produced by Interviewee Two for the key word 'convenient'.

Convenient

easy fresh near suitable	experience sound time vegetable

Correct lexical choice Incorrect lexical choice

The highlighted items are the lexical choices that Interviewee Two identified as the most relevant associations for the key word 'convenient'**.** As stipulated by the test guidelines, the student picked four lexical items which he felt were related to the key word. Out of the four words only two of them (suitable and time) were correct choices.

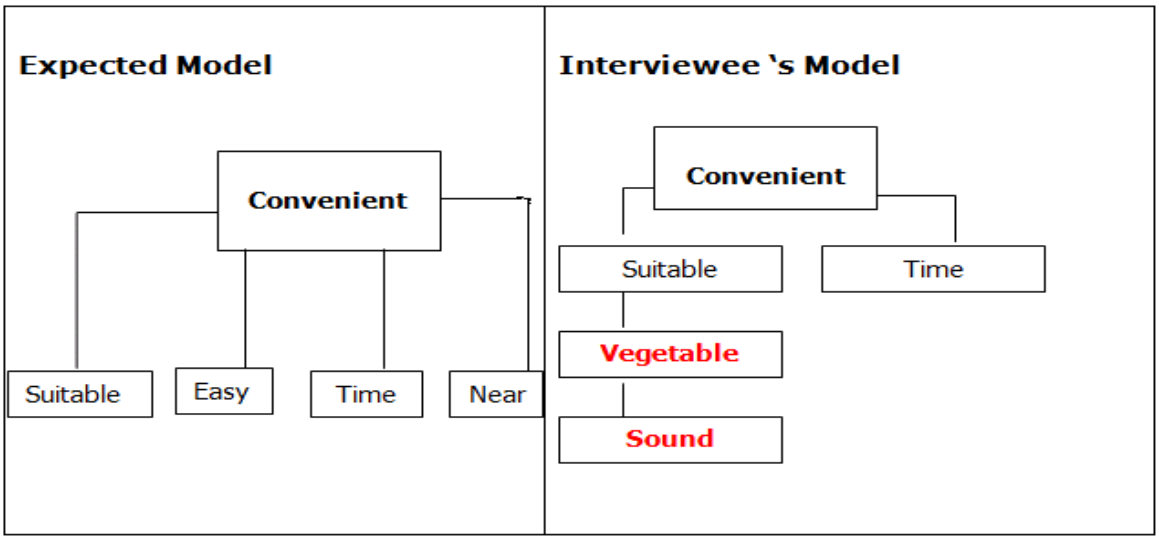
The learner was therefore interviewed on the other two (sound and vegetable) to determine his logic for the choices made. The rationale provided was as follows:

> *For me convenient, suitable, fit, those three words to me it's the same. So convenient that's suitable, that's fit. So I think may be suitable to feed another thing. Why? Because convenient, suitable, fit, it mean the same so I say convenient, may be vegetable because for example, maybe I can say convenient for feed someone who is vegetarian, which mean like, suitable for them.*

The narrative was interpreted as an example of a lexical choice impacted by partially developed knowledge. This is because the learner was able to produce two correct answers where the lexical items 'suitable' and 'time' were identified as being related to the key word 'convenient', suggesting that the learner possesses partial

knowledge of the word. However, it was also evident that the learner over-extended the meaning sense of the word 'convenient'. This is evident in the way in which the learner linked word associates for response items to the key words as shown in the Leaner Lexical Model (LLM) presented in Figure 6.1

Figure 6.1 Expected vs. learner model of word associations



The lines on Figure 6.1 show the links that the learner made between words. This means that in cases where the learner explained the meaning of a word by linking it directly with the key word, then the line joins the two words directly without any words in-between. However, if the learner links several words together based on their interpretations between or among particular lexical items, then the line joins all the words together following the learner's explanation. Therefore, Figure 6.1 above shows that according to this learner's logic, the word 'convenient' can mean 'suitable'. The learner then goes on to suggest that the word 'vegetable' can go with the word 'convenient' as in a 'suitable vegetable'. The learner also suggests that 'sound' can go with 'suitable' as in a 'suitable sound'. 'Vegetable' and 'sound' are

highlighted in red because they are incorrect lexical choices which the learner has linked to the key word 'convenient' through linking the meanings of these words with the word 'suitable'. When probed further about the relationship between 'convenient' and 'sound', the learner comments:

> *Sound....yes sound go all the time with convenient. Always I can say may be, just as example, may be this is not suitable sound from the radio. So that's why I choose this word because always I hear this word.*

Therefore, even though the learner clearly has some knowledge of the lexical item in question, it is evident that further exposure to the term is still required in order to help the learner anchor this term more securely into his mental lexicon. This would enable him to use the word more appropriately and be able to develop depth of knowledge for the lexical item.

Another example of the impact of partially developed word knowledge on the choices that learners made is provided by Interviewee Eight in response to Item 2 of the WAT. An extract from the learner's output for the key word *'bright'* is presented below:

bright

clever famous happy shining	colour hand poem taste

Correct lexical choice Incorrect lexical choice

As can be seen from the above output from Interviewee Eight, the learner was able to identify three correct associates of the key word 'bright'. However, the third choice (poem) was unexpected so it was

one of the items that the learner was interviewed on. The rationale provided by the learner was as follows:

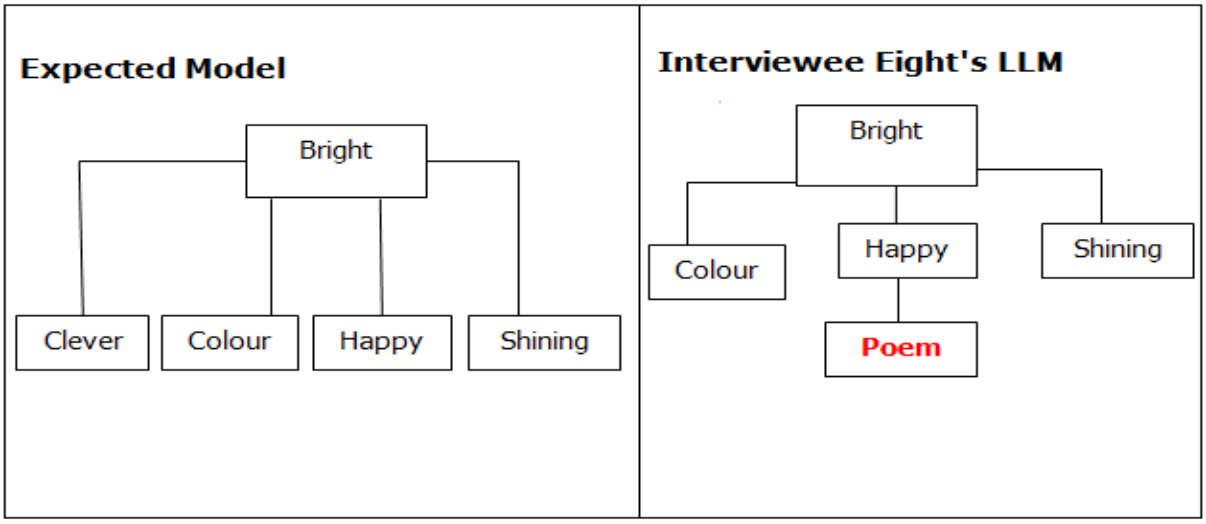
> *Because…it is very interesting poem. Because like 'bright day' people are happy, so bright poem, may be its about wedding or birthday party.*

The learner exhibits collocational gaps in her vocabulary knowledge through the rationale given. However, her logic is not far-fetched. Even though she does not choose 'happy' which is one of the expected answers, the logic behind the learner's choice is based around the concept of happiness which she has evidently managed to identify in the word 'bright', hence her reference to 'bright day, people are happy'. However, it would appear that as with Interviewee One, the learner has assumed that paradigmatic relations of subordinate words are automatically related to key words in a syntagmatic manner. Thus, when probed further as to whether the word poem goes with 'happy' or 'bright', the learner's comments suggests the aforementioned assumption that she has made:

> *Well, what I can say, it's the two words. Because it's happy thing or bright day isn't it?*

Clearly, the learner's model is flawed in as far as the collocational relations between the key word and given words are concerned, but the fact that the learner is able to identify three other associates correctly is a clear indication that she possesses substantial knowledge of this word but there are still gaps of knowledge in her lexicon particularly in relation to collocational knowledge, hence the lexical confusion evident. Figure 6.2 below summarises the learner model compared to the expected model of word associations for the key word 'bright'.

Figure 6.2 Expected vs. learner model of word associations



Further evidence highlighting the impact of insufficient word knowledge is provided by Interviewee Five. This learner produced, among others, interesting output for the keyword 'conscious' as presented below:

conscious

awake healthy knowing laughing	face decision effort student
Correct lexical choice	Incorrect lexical choice

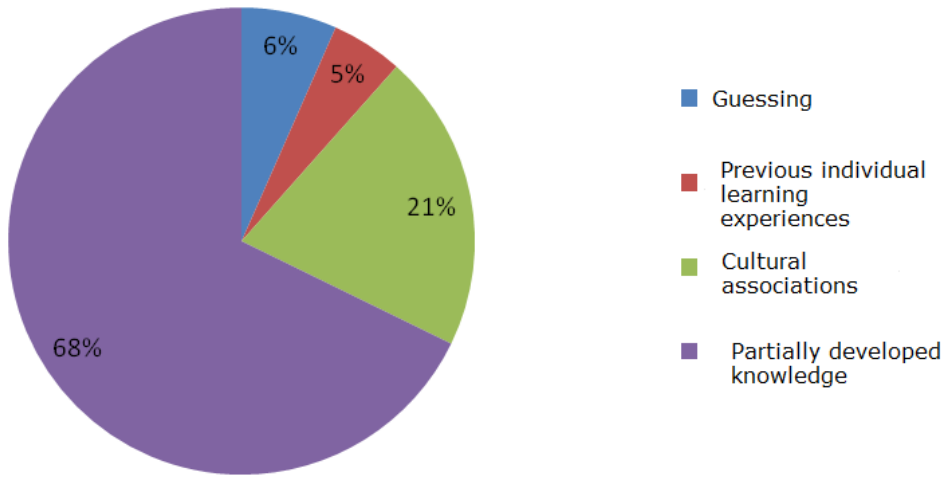
As can be seen from the learner output presented above, the student produced two correct and two incorrect matches for the key word 'conscious'. When asked about the logic behind her lexical choices, the learner had the following to say as part of her commentary:

> *Sometime I think people say a student is conscious.....consciousness student....he like,…he like…, he likes college work and all..., Oh I don't know, my pronunciation! I thought that's the one in the question so I thought yes, consciousness student so I put student...I mean consciousness, is that a word?'*

In the example given above, the student seems to have modest knowledge of the key word 'conscious' hence she seems to confuse it with 'conscientious' as suggested by the excerpt above where the learner provides a reasonable explanation of a conscientious student despite having successfully provided two correct paradigmatic relations of the key word 'conscious 'i.e. 'awake' and 'knowing'. This suggests that knowledge of this term is not yet sufficiently developed to enable the learner to differentiate successfully between the two words conscious and conscientious. Thus, the learner seems to have both concepts in her mind – being conscious and being conscientious but they both seem to be underdeveloped hence the evident lexical confusion and overlap in the learner's conceptualisation of the two terms. This instance was therefore classified as another example of the impact of insufficiently developed knowledge affecting the learner's lexical output.

Figure 6.3 provides an overview of the proportion of responses related to each of the four factors based on responses from interviewees. These were derived on the basis of the analysis of learners' responses as discussed above:

Figure 6.3 Overview of rationale behind incorrect lexical choices made by the ESL students



As Figure 6.3 shows, the learners' commentaries suggested that partially developed knowledge was the main contributing factor towards incorrect answers produced during the WAT. As such, close to three quarters of learners' commentaries fell into this category. This was followed by the Cultural Associations category. In this category, learners' commentaries suggested that factors to do with their cultural backgrounds may have shaped just over a fifth of their responses to particular lexical items. The two least ranking categories represented responses which were judged to be influenced by guessing and those influenced by learners' previous educational backgrounds respectively.

Partially developed vocabulary knowledge was a significant factor not only when looking at the group as a whole but also at an individual level. In every single learner's narrative, evidence was found to suggest that the majority of the incorrect lexical choices that the learners made were influenced by partially developed knowledge about target words.

While the highest factor impacting on the lexical choices that the ESL students made was Partially Developed Knowledge, it is interesting to note the variability that was evident among the participants. Therefore, while Partially Developed Knowledge accounted for the majority of factors overall (Figure 6.3) at an individual level, the contribution varied amongst the participants. For example, for Interviewee Two, this factor was noted in 14 instances, yet for Interviewee Four, the same factor was noted only four times.

For Cultural Associations, the highest occurrence was five times (Interviewee Two), yet for Interviewee Three, this figure was down to

just one occurrence. For Personal Individual Learning Experiences, the maximum occurrence recorded was from Interviewee One (three instances) yet this factor was not observed at all from a number of participants (Interviewees Two, Three, Six and Seven). Consequently, when all factors were aggregated, the contributions made by each student towards each factor showed variability. Therefore, although there were some points of commonality that allowed for the categorisation of the evidence obtained from stimulated reconstructions, there was also notable variability in terms of the extent to which each factor impacted on each of the learners' lexical choices. These findings are given further consideration in the Discussion chapter.

On the basis of the analysis of data from the sample from this study, it seems likely that a number of influences were at play when learners made their lexical choices. These included the use of guessing as a strategy for dealing with unknown vocabulary; the impact of cultural and educational backgrounds as well as the degree to which different lexical items are known. It is possible that a different sample might generate a different range of key factors but the study provides tentative results suggesting that it would be worthwhile future investigation to explore in depth, the impact of individual factors in shaping the type of knowledge that students may possess about different lexical items.

6.3 Chapter summary

Chapter 6 presented a nuanced analysis of learners' lexical profiles juxtaposed with findings obtained from interview narratives. In terms of the rationale behind the choice of particular lexical items during

the WAT, evidence from stimulated reconstructions suggested that partially-developed vocabulary knowledge, cultural differences, educational backgrounds and the impact of guessing were the major influences identified by learners as explanatory factors / rationale to the lexical decisions they made. The chapter concluded by presenting a synthesis of findings from written language output and semi-structured interviews as a way of synthesising findings in order to obtain more comprehensive insights about each learner's L2 trajectory.

Chapter Seven:

Discussion

7.0 Chapter introduction

Chapters 5 and 6 presented findings obtained from the investigation guided by the five research questions that form the core of this study. This chapter draws inferences and provides interpretations of these findings within theoretical and empirical frameworks relevant to L2 learning. To corroborate the interpretation of the findings, the chapter draws from both quantitative and qualitative evidence presented in Chapters 5 and 6. As reported Chapter 5, statistical analyses revealed some significant correlations between learners' size and depth of vocabulary knowledge as well as between vocabulary knowledge and written language skills. The tracing of individual trajectories, triangulated with data from stimulated reconstructions/semi-structured interviews revealed further insights about the ESL students' profiles including the variability that emerged from the analysis of individual ESL learners' profiles. These findings suggest that second language processes such as the development of lexical knowledge are complex and therefore, the chapter adopts a DST perspective for the interpretation of the findings.

7.1 The interplay between vocabulary size and depth among ESL learners investigated in the study

Table 5.2 highlighted that at an individual category level, the correlations obtained between individual VLT categories and WAT scores for the ESL students investigated in the current study did not reveal any significant relations. This provides empirical support for the conceptual distinction between size and depth of vocabulary knowledge (Akbarian, 2010). In other words, if size and depth represent the same type of vocabulary knowledge as argued by Vermeer (2001), then the expectation would be that of very strong correlations between all aspects of vocabulary size and all aspects of vocabulary depth which

was not the case when VLT across different frequency band. As was highlighted in Section 2.2.8, the size versus depth conceptualisation of vocabulary knowledge has had some critique and therefore, these results make a contribution towards this debate by providing supporting empirical evidence to suggest that vocabulary knowledge can be conceptualised through the size and depth dimensions.

However, it is noteworthy that the findings from this study (i.e. the correlations found between vocabulary size and depth) contradict those of Vermeer (2001), As was highlighted in Section 2.2.8, on the basis of high correlations found between size and depth scores among kindergarten children, Vermeer (2001) concluded that there is no conceptual difference between breadth and depth of vocabulary knowledge. This study argues that the different results may relate to the way in which Vermeer's (2001) study conceptualised vocabulary size and depth. Recall that two main issues were identified in Section 2.2.8, which, it was considered, could have impacted on Vermeer's findings. The first issue related to very similar conceptualisations of size and depth in Vermeer's study and the second issue related to the fact that the study participants were kindergarten children who can be expected to possess different word knowledge for lexical items because of their cognitive stages of development. On the basis of these two main issues, the current study argues that it is possible that Vermeer's study may have obtained different results if the size and depth dimensions were more distinct and the participants were adults. In the current study, the lack of strong correlations between vocabulary size and depth found in this study, and supported by empirical studies such as Read and Nurweni's (1999) study, suggest that the two are conceptually different but interconnected. If there was no conceptual difference, then the expectation would be that there

would be very strong correlations between the size and depth dimensions across different vocabulary frequency bands.

Another noteworthy finding from the analysis of correlations between vocabulary size and depth was that while individual VLT frequency bands were not found to be associated with depth of vocabulary knowledge, when total scores for VLT were computed against WAT scores, significant correlations (rho=.49) were obtained. However, although the correlation between TotalVLT score and the WAT scores reached significance level, the correlation obtained was of medium strength as it was below .05 (Cohen, 1988). To that effect, it can be argued that the results from the ESL students investigated in the current study provide support to the view that size and depth are very likely to be closely related (Milton, 2009) but the distinction between the two concepts is worth investigation because these two do not completely overlap (Bogaards and Laufer, 2004; Haastrup and Henriksen, 2000).

While the findings must be treated cautiously because of the small sample size involved in the current study, the results corroborate findings from other studies, which also found modest or no correlations between size and depth. A case in point is Nurweni and Read's (1999) study (see Section 2.2.8) which found that the relationship between size and depth varied according to proficiency levels among the university undergraduates that they investigated. The strength of the correlation slightly increased as proficiency increased so that the strongest correlations between the size and depth dimensions were recorded within the high-performance group while the lowest proficiency group did not record any meaningful correlations. Since the participants in the study were university students, it seems reasonable to argue that their English proficiency levels would have been

higher than the proficiency level of the ESL students investigated in the current study. The lack of strong correlations between the vocabulary size and depth of ESL learners in the current study is therefore in line with Nurweni and Read's (1999) findings.

Nurweni and Read (1999) hypothesized that guessing may have affected the performance of their low proficiency students, particularly since the WAT utilised in the study was the receptive version. Utilising the MMR approach adopted in the current study, the study was able to go beyond Nurweni and Read's (1999) hypothesis and qualitatively explore empirical evidence based on the rationale provided by the learners themselves about the lexical choices they made. As highlighted in Figure 6.3, in the current study, guessing accounted for only 2% of the lexical decisions made by each of the four learners who used this strategy during the WAT. Nonetheless, although this figure may be relatively small, it represents a strategy that was used by four out of the eight interviewees in the study. In other words, 50% of the students used this strategy but only to a limited extent compared to the other three factors (Cultural association, previous Individual Learning Experiences and Partially developed knowledge). It was also noted in Section 6.2.3 that two other students reported not using guessing as a strategy because they felt it would be dishonest to do so. The conscious decision to use guessing as a strategy supports literature (see for example O'Malley and Chamot, 1994; Clarke and Nation, 1980) which suggests that guessing is a widely recognized L2 strategy which can sometimes threaten the validity of a test (see Section 2.3.2). However, the conscious decision by some learners to avoid the use of guessing as a strategy suggests variability in the use of this strategy amongst the ESL students interviewed in the current study. This variability opens up further avenues for in-

vestigation as it may be linked to personal and/or cultural value systems, as was the case with the two students in this study. This has practical pedagogic implications particularly in terms of teaching and learning approaches that may effectively support this learner's vocabulary development.

These findings have important pedagogical and research implications. In pedagogical contexts, the distinction has implications on the teaching and learning of vocabulary. As highlighted in Section 2.2.8 of the Literature Review chapter, this conceptualisation is still a contested area. In the current study, Section 1.1.2 highlighted that lexical deficiencies have been identified by both learners and their teachers as one of the most challenging yet important areas of L2 development. The distinction between vocabulary size and depth adds an important dimension to this situation. This is because while learners may realize their lexical deficiencies in terms of the number of words that they need to 'know', findings from the current study raise the important question as to whether the learners' concept of what it means to know a word is comprehensive enough to encourage them to move beyond the development of vocabulary size, to the development of depth of knowledge. In the current study, this question is particularly justified in light of interview evidence which suggests students' resistance against the idea of deepening vocabulary knowledge of words that they already know, in favour of developing a larger vocabulary size. Indeed, for the ESL learners in the current study, lexical knowledge came out as almost synonymous with vocabulary size where the majority of the students considered known words to be those that they knew in terms of the form-meaning link. However, as will be further explored in Section 7.2, as the results presented in Table 5.4 show, 26% variance in learners' writing was

explained by TotalVLT. Similarly, 25% and 34% variance in learners' writing (within the Organisation and Task categories respectively) was explained by depth of vocabulary knowledge.

The results corroborate current thinking suggesting that deep word knowledge is likely to promote the effectiveness of the process of activating and accessing lexical resources for receptive and productive use (Stæhr, 2008).

From a DST perspective, the significant but modest correlations found between vocabulary size and depth would suggest a complex interplay among multiple factors. As Svalberg (2016, p.10) notes, in the teaching and learning context, 'the fact that agents are typically people, that is, volitional agents, adds to the complexity' which is inherent in trying to understand the dynamic interaction between agents and/or their factors. In the context of the current study, this implies that while size and depth are important factors in understanding the interplay between vocabulary knowledge and written language skills for the ESL students, it can be expected that the relationship will be further mediated by other factors. As will be demonstrated in the following sections, interview results succeeded in shedding some light into some of the factors that may be at play.

7.2 The interplay between vocabulary size and written language ability

Section 2.2.1 differentiated between receptive and productive vocabulary knowledge. Drawing on this distinction, it can be argued that writing requires well-developed productive knowledge, as the learners need to activate and retrieve lexical knowledge in order to compose text. The VLT test used in this study was the receptive version where lexical items were presented and learners only needed to rec-

ognise the lexical items and corresponding answers without having to actively retrieve items from their own lexicon. To this end, the fact that for the ESL learners in this study, significant correlations were found between size and writing (rho=.52 and rho =.51 for the Task and TotalVLT categories respectively) despite the slightly different dimensions of linguistic knowledge measured, lends further support to the key role that vocabulary knowledge plays in facilitating effective execution of other linguistic skills such as writing (Stæhr,2008).

In addition, writing is a complex process mediated not only by the availability of lexical resources, but also other cognitive abilities such as the ability to organise text (see Section 2.7.2). This suggests that the mastery of writing skills can be expected to be contingent upon multiple factors. To that end, the .51 Spearman correlation coefficient (26% variance) obtained between writing and TotalVLT scores and the .52 Spearman correlation coefficient (27% variance) obtained between vocabulary size and the Task category of the writing assessment scale are in fact noteworthy since they suggest the interaction of manifold factors that lead to the emergence of this complex skill. These findings are well aligned with Rosmawati's (2013) model of the multiple factors that interact within dynamic systems (see Figure 2.12). In the current study, the 26% and 27% variance explained by vocabulary size in the writing of ESL learners, as well as the 34% and 25% variance from vocabulary depth variables highlight that writing is a non-linear process. In non-linear processes, 'the relevant variables are not related one to another according to strict proportionality' (Keller, 2008, p.6), hence the differential variability noted between writing and different aspects of vocabulary knowledge. Therefore, in line with Rosmawati's (2013) conceptualization of a complex system (Section 2.8.2), findings from the current study suggest that the L2

learner is a complex system with multiple interconnected elements which interact during the L2 learning processes.

In the current study, the synthesis of interview findings and written assessment results relating to vocabulary size and depth (see Table 6.2) identified that one of the features of some of the learners who obtained higher scores across the assessments administered in this study was their willingness to engage in independent learning outside class including engagement in free writing activities. This suggests that, independent learning might have given the learners exposure to language development opportunities, which is essential for language development in general and vocabulary learning in particular, to take place (Nation, 2001) , an important resource for the effective functioning of the L2 development from a DST perspective (Larsen-Freeman, 2006). As was noted in section 2.2.9, previous studies have found correlations between ESL students' written language proficiency and measures of linguistic knowledge such as lexical diversity. The results from the current study thus corroborate such studies which suggest that lexical proficiency plays an important role in the process of producing good quality text (see for example, Crossley and McNamara, 2011; Jarvis et al, 2003; Engber, 1995).

However, since in the current study, vocabulary was measured within the *Language* sub-scale, the expectation was that out of the three sub-scales (*Task, Organisation and Language*), the *Language* subscale would produce strong and significant correlations with VLT measures, particularly within the higher frequency bands. This was not the case. In fact, although significant correlations were found between TotalVLT and total Writing scores, no significant correlations were found between any of the vocabulary size categories and the

Language category of the writing assessment scale. The unexpected results suggest that for the ESL students investigated in the current study, higher order skills such as the ability to fulfil expectations of the task at hand (*Task)* may have played an important role towards the overall quality of writing (Silva, 2009; Kobayashi and Rinnert, 2008). What is particularly interesting is that a similar trend was noted with regards to depth of vocabulary knowledge. This category (Task) produced significant correlations with WAT scores alongside the significant correlations that were obtained for the organisation category. Therefore, the next section extends discussion on the role of task fulfilment in writing as highlighted by findings from the current study.

7.3 The interplay between depth of vocabulary knowledge and writing

The current study found vocabulary depth to be associated with the ability to effectively execute the higher order skills of generating relevant content (rho=.58) and organizing the text well at sentence and paragraph level (r=.50), thus achieving greater cohesion and coherence for the written task overall (see Section 5.1.4). This suggests that effortful word production is likely to make it difficult for writers to attend to higher order aspects of the writing process such as the production and organisation of relevant content (Flower and Hayes, 2008). This forms the basis of the view that lexical proficiency is an important factor contributing to writing proficiency (Schoonen et al., 2011; Snellings et al., 2004), although as has already been argued, this should not be taken to suggest a linear relationship between the two variables. Indeed, from the results obtained from the tracing of individual trajectories in this particular study, (see Table 5.9) it was determined that for the majority of the ESL learners

investigated in the current study, (12 out of 18) students revealed a developmental trajectory where mastery of VLTK2 words could be expected to be associated with higher depth of vocabulary knowledge scores, which were in turn associated with higher free writing score. However, this is not to say that the relationship was based on strict proportionality (Keller, 2008) amongst different variables. For example, S5 got full marks on the VLTK2 assessment but went on to get LLA scores on the WAT and Writing assessments at 40% and 45% respectively. In contrast, S17 only got a marginal HLA score on the VLTK2 assessment (12 out of 18) yet went on to get much higher marks at 65% and 49% on the WAT and free writing assessments respectively. This highlights that writing is a complex process as the unpredictability demonstrated in the findings illustrates.

On the basis of the significant correlations found between vocabulary knowledge and writing, it can be argued that if the skills required for effective writing are not sufficiently developed, for example, lexical knowledge and/or content knowledge related to the topic that the writer is tasked to write about, then cognitive capacity is likely to be superseded leading to the production of poor quality text as explained by the inhibition theory (van Gelderen et al., 2010). This highlights the impact of underdeveloped lexico-grammatical skills on the effective production of text (see Section 2.7.2). As a result of the cognitive load presented by dysfluency at the lexico-grammatical levels, less proficient language users may fail to be oriented to text-level skills in writing such as the generation of relevant and appropriate content and ensuring the production of coherent text. In the current study, it would appear that the ESL learners who had a more readily available lexical repertoire had more cognitive resources

available to use at different stages of the writing process as reflected by the higher correlations obtained between vocabulary depth and written language skills compared to vocabulary size and written language skills.

The relationships found between the *Task* and *Organisation* scores on the one hand, and vocabulary size and depth on the other, suggest that ESL students who had larger and better developed lexical repertoires were also more capable of producing relevant and more coherent text compared to their peers. A related finding was obtained from the higher proficiency group (ICLE) where knowledge of vocabulary from the Offlist category was the only category which produced significant correlations with the Organisation, Language and Total Writing scores from VocabProfile (see Table 5.7). This suggests that the larger and more developed mental lexicons enhance the effective execution of other cognitive processes such as composing text. Lexical ability is likely to have freed cognitive resources that could then be used to better fulfil the task and organisation requirements of the free writing task (Crossley and McNamara, 2011).

7.4 The interplay between ESL and ICLE learners' VocabProfiles and the quality of their written compositions

As was highlighted Table 5.6, ESL students' compositions were positively impacted by knowledge of the first 2000 most common English words with at least 30% variance in writing (Task category) explained by knowledge of VocabProfK2 words. According to Laufer and Nation (1995), knowing the first 2000 words is likely to enhance reading, writing, speaking and listening skills for L2 students. The authors suggested that in fact, the first 1000 words are an essential

condition for the development of written expression (Laufer and Nation, 1995) since they include the majority of the function words as well as most of the basic lexis. Therefore, the correlations found between this category of words and ESL students' writing is well-aligned to the literature (Nation, 2001).

The positive association found between ESL student's writing and the most frequently occurring words in the English language lends support to earlier findings in this investigation where VLT scores for ESL students were found to be associated with the ability to fulfil Task requirements as well as overall writing skills, thus suggesting that these words have a positive impact on the development of writing skills. Since both tools (Vocabprofile and VLT) obtained somewhat similar results with regards to the association between words from this category and writing skills, put together, the associations identified, (association between VocabprofileK2 and Task for ESL; VLTK2 and Task as well as VLTK2 and overall written language ability) support the view that Vocabprofile has the ability to provide insights about learners' written language skills.

Since VLT is a standardised measure of vocabulary knowledge (Laufer and Nation, 1995), it is noteworthy that Vocabprofile produced results that are aligned with those produced by VLT. The ability to fulfil task requirements is an example of discourse level elements operating during the writing process (Uccelli et al., 2013). To that extent, these findings lend further support to written language ability as a highly complex process (Friginal, et al, 2014). For the ESL learners investigated in the current study, it was evident that beyond the mastery of lexis, proficient writing requires the interaction of discourse level elements such as the understanding of task

requirements, hence it was not just the overall writing score that was positively correlated to the TotalVLT score, but correlations were also observed within the Task and organisation categories. At an individual category level, Task was the only category that produced significant correlations with the TotalVLT score, thus highlighting the positive role of vocabulary knowledge in helping learners achieve higher order written language skills.

A contrasting picture emerged in terms of the association between VocabProf2K words and ICLE students' written language scores. For this group of learners, no significant correlations were obtained between VocabProfK2 words and any of the writing ability categories. Instead, for this higher proficiency group, it was the Offlist word category (VocabProfOfflist) which significantly correlated with writing. Indeed, 10% variance in writing was explained by VocabProfOfflist scores (rho = .32) for the ICLE group (see Table 5.7). In addition, the Language score and the overall written language ability score obtained significant correlations with VocabProfOfflist words (.53 and .47 as per Table 5.6). This suggests that rare/low frequency vocabulary has an impact on the quality of written language productions.

The importance of rare vocabulary noted from the statistical analysis was backed up by data from interviews where evidence from seven out of the eight interviewees suggested that the learners believed that rare vocabulary is key to the development of linguistic skills (see Table 6.2). It was therefore quite interesting to note that for the learners investigated in this study, use of words from the *Offlist* category was a significant distinguishing feature between lower proficiency and higher proficiency learners. As highlighted in Section 2.5.4, the Offlist category contains words that cannot be classified into any

of the specific word frequency categories, for example, beyond the 25000-word frequency level offered by the VocabProfile tool. This would suggest very little utility of such words, yet in the current study, these were found to produce significant correlations with written language ability. This finding is well aligned to insights obtained from semi-structured interviews as will be further explored in this chapter.

7.5 Variability in the interplay between vocabulary size, depth and written language skills

The interplay between vocabulary size and depth; vocabulary size and writing skills; and vocabulary depth and writing skills are at the centre of the current investigation. The main findings in this regard were the significant correlations that were found when written language was computed against vocabulary size and depth respectively. From the tracing of individual trajectories, one of the noteworthy findings related to both the trends and the variability that characterised the development of key linguistic skills amongst ESL learners (see Section 5.2). The variability noted resonates with the notion of L2 learning as a complex adaptive process (see Section 2.8.2). As De Bot, et al. (2007, p.7) note, DST 'recognizes the crucial role of interaction of a multitude of variables at different levels'; hence its relevance to the interpretation of the findings from this study. The following section utilises DST as a means of understanding the variability noted in the current study and highlights the relevance of specific DST features that resonate with findings from the study.

7.5.1 Interconnectedness

One of the features that characterize complex systems is the interconnectedness of their sub-systems (see Section 2.8.2). From the re-

sults obtained from stimulated reconstructions / semi structured interviews in the current study, it became apparent that the interplay between vocabulary knowledge and written language proficiency could not be comprehensively explored through a linear approach that relied solely on the end products (statistical outputs) from each of the variables. Instead, qualitative data obtained from the stimulated recalls / semi-structured interviews as well as the tracing of individual trajectories provided a richer source for the exploration of possible multiple interrelationships.

As was highlighted in the results chapters, the stimulated reconstructions revealed that learners' lexical choices were affected by factors related to cultural and previous individual learning experiences; the learners' willingness to use guessing as a strategy for dealing with unknown vocabulary; as well as the extent to which a particular lexical item was known (partially developed lexical knowledge); albeit to different degrees so that some factors were had higher rates of occurrence among some learners than others. Since 'complex systems are sets of interacting variables' (De Bot et al. 2005, p.116), the interaction of the aforementioned factors in the current study suggests that each learner's developing linguistic knowledge can be considered to be a dynamically complex system. This finding is aligned to De Bot et al. 's (2007) when they state that:

> *... a language learner is regarded as a dynamic subsystem within a social system with a great number of interacting internal dynamic sub-subsystems, which function within a multitude of other external dynamic systems. The learner has his/her own cognitive ecosystem consisting of intentionality, cognition, intelligence, motivation, aptitude, L1, L2 and so on. The cognitive ecosystem in turn is related to the degree of exposure to language, maturity, level of education, and so on, which*

> *in turn is related to the social system, consisting of the environment with which the individual interacts* (De Bot et al., 2007, p.14)

Indeed, in the current study, stimulated reconstructions / semi-structured interviews highlighted a number of factors that impact on the learner such as those related to cognitions, their motivations, their social context as well as their willingness to engage in independent learning.

Section 2.7.2 identified writing in the L2 as a complex process which involves the orchestration of multiple skills including effective utilization of lexical resources. This suggests that lexical knowledge is a resource during the writing process. The findings from the current study therefore suggest that L2 development is a dynamic system akin to the one presented in Figure 2.12. In the current study, results obtained from both qualitative and quantitative data suggest multiple factors that may be at play, therefore making it difficult to predict outcomes as was identified in the case of learners' trajectories which detracted from expected trends.

7.5.2 Non-linearity

Complex adaptive systems are non-linear because not only are they sensitive to initial conditions but they are also context-dependent (Section 2.8.2; Larsen-Freeman, 2012; van Geert and Steenbeek, 2005). They may exhibit relative stability but this stability tends to be a preferred but unpredictable state (Burns and Knox, 2011). In the current study (see Section 5.2), a case in point is that of the five (out of the six) students who achieved HLA scores for writing (S1, S2, S13, S14 and S16).

Looking at their VLT and WAT scores, one could conclude that a high VLT score as well as a high WAT score are part of the requisite initial conditions for the development of writing skills. However, one student's profile (S6) provided evidence that this is not always a predictable outcome because this student obtained an HLA for *Writing* and VLT despite failing to obtain an HLA for WAT. Both VLT and WAT scores for the other five students in this category were high, suggesting that for these particular learners, these were part of the initial conditions associated with well-developed written language skills.

In addition, when results from written assessments and interviews were synthesised (Table 6.2), Student 6's profile came out as being peculiar from another perspective. The student displayed good metacognitive knowledge about the importance of both size and depth of vocabulary knowledge in the development of linguistic skill; and this knowledge was supported by a commitment to expand vocabulary size and depth through engaging in independent learning activities such as doing language exercises and reading books. Since this profile was similar to that of the highest achieving student across the board (Student 16), Student 6 could be expected to perform highly for Writing. However, Student 6 obtained high marks for VLT and Writing but not WAT. Therefore, there was an irregularity between the learner's beliefs and practices on one hand, and his linguistic profile on the other; and this irregularity could not be easily explained in the context of the current study. This variability and seemingly contradictory results can be taken to illustrate the complexity of the interplay between the learner's vocabulary knowledge and written language skills. However, it also has to be acknowledged that students may provide socially desirable answers,

which may be hardly related to their actual language learning practices but rather, to their desire to give answers that they deemed the researcher expected (See Section 3.5.4).

Nonetheless, on the basis of the literature on writing process theories reviewed in Section 2.7.2, the process of writing involves the simultaneous orchestration of multiple skills such as generating content and ensuring cohesion in text. The interaction of all of these factors can mediate the outcome, leading to unpredictable outcomes (Smith and Thelen, 1994). This highlights that it is not viable to treat second language processes as mechanical systems where different elements are simply processed to produce a particular outcome as suggested by the majority of the ESL learners' commentaries which highlighted the belief that possessing a large vocabulary size would lead to high scores in writing assessments. Instead, what seems more tenable is a dynamic perspective which views second language learning as an ecological or dynamically complex system characterised by emergent patterns of behaviour which are a result of the interaction of constituent elements over time (Burns and Knox, 2011; Kramsch, 2002). In the current study, both the learners' written language skills and their lexical knowledge can be considered to be emergent behaviours and the unpredictable outcomes where predictability might otherwise have been expected (as in the case of Student 6) may have been related to the various affordances provided by each learner's context.

7.5.3 Context dependence

As was highlighted in Table 5.9, Student 9 was the only one who managed to obtain an HLA for WAT after obtaining an LLA for VLT. This was in contrast to a majority (11 out of 13 students) who fol-

lowed the trajectory of HLA for VLT followed by HLA for WAT. On the basis of informal notes taken during feedback sessions, it was noted that Student 9 has an English-speaking partner who does not speak any other languages. One of the key studies on variability highlighted in the Literature Review chapter is that of Churchill (2007).

Churchill's (2007) study presents a DST perspective to learning a word (see Section 2.8.2). The case of Student 9 resonates with findings from that study. Churchill (2007) found that affordances from his context (which included written and oral Japanese vocabulary input from the particular environment where he was situated (i.e. the various interactions that he had with medical staff while he was admitted in hospital) impacted positively on his learning of Japanese vocabulary. A similar argument seems applicable in the current study. Student 9 is a learner whose environment readily presented ample opportunities for incidental learning to take place in the form of authentic linguistic input from the learner's partner. This would have contributed to the development of a seemingly atypical profile because context is an important part of development (Larsen-Freeman, 1997). As the learner herself confessed, she started watching television before she could '*even understand two English word at one time'* because '*my partner is English, together we watch English TV only'.* This suggests a conducive environment for Student 9's vocabulary learning.

As can be expected, oral incidental exposure such as the one experienced by Student 9 is unlikely to impact directly on the development of writing skills because the complexities involved in developing writing skills are unlikely to be addressed through incidental verbal input. In dynamic complex systems, resources are a key part of develop-

ment so without resources, a system is likely to be less effective (Burns and Knox, 2011). In the case of Student 9, the unexpected lack of development related to writing skills could be related to the relative lack of resources to facilitate the development of writing compared to the resources available to this particular learner in relation to the development of her L2 verbal skills.

It can therefore be concluded that the affordances offered by an English-speaking partner might explain the high WAT score which is not accompanied by similarly developed writing skills.

S13 provides another example of variability that resonates with DST perspectives. As noted in Chapter 6, S13 is one of the students who obtained HLA scores in all written assessments (see Table 5.9). However, this level of attainment is very inconsistent with the learner's language practices and cognitions as reflected by data from semi-structured interviews (see Table 6.2). Evidently, the learner shows very little engagement and limited positive behaviours associated with L2 learning such as the willingness to engage in independent study. In contrast, Student 10 who exhibited positive language practices and cognitions (on the basis of semi-structured interviews summarised in Table 6.2) obtained LLA scores across the board, a result which was unexpected. None of the evidence collected during semi-structured interviews / stimulated recalls was found to be explanatory of the unpredictable results obtained by the two learners (Student 13 and Student 10). However, as Burns and Knox (2011,p.7) observe, DST 'embraces contradiction and unpredictability' as these are at the core of the theory. Therefore, the cases of Student 10 and Student 13 further attest to the relevance of DST perspectives to the current investigation as they highlight that it is

'impossible to extract and measure single factors that contribute to SLA because they all interact' (De Bot et al., 2005, p.76).

7.6. Rationale behind learners' lexical choices: learners' perspectives

A major finding in terms of the rationale provided by learners for the lexical choices that they made in the WAT is that differential conceptual models were evident amongst the learners interviewed. In other words, different learners had different ways of constructing meanings of given lexical items as was demonstrated by the different LLMs produced by the learners. Nonetheless, these could be categorised mainly into individual differences such as the quality of the learners' lexical repertoires, learners' backgrounds, including their educational and cultural backgrounds as well as the use of guessing as a strategy for dealing with unknown vocabulary. Out of these, the most common reason for learners' incorrect lexical choices was found to be partially developed lexical knowledge of lexical items. This category represented words where the learner clearly had some kind of knowledge about a particular lexical item, no matter how superficial (see Section 6.2.4). This finding supports the view that vocabulary knowledge is a multidimensional construct which develops incrementally and is therefore not a yes or no phenomenon (Nation, 2001; Read, 2000; Schmitt, 2010). Henriksen's (1999) word knowledge model, which ranges from partial to precise knowledge, seems to capture this. As Read (2000) points out, word knowledge is a process under constant development for native and non-native speakers alike.

The second most cited reason for word choice related to cultural associations. It has been suggested that the influence of the L1

mental lexicon on the L2 is stronger at lower levels of proficiency (Söderman, 1993), a view that finds support in findings from the current study since the interviewees in the study were all low proficiency students. From the data obtained from stimulated reconstructions/ semi structured interviews (see Section 6.2), it is evident that the adult L2 learners in the study already possessed highly developed conceptual and experiential knowledge based on their L1 and experiences of the world around them. This provided a schema for the understanding and interpretation of vocabulary items. Unfortunately, as Wolter (2006, p.742) notes, although 'the L1 lexical/conceptual structure is useful for building L2 lexical networks,… it will also sometimes provide learners with misinformation about acceptable combinations of L2 words'. This was evident in the LLMs presented in Section 6.2.4 where learners' experiences had a noticeable impact on the types of links made between/among given lexical items hence the incongruent conceptual models produced by learners. As was highlighted in the Literature Review chapter, one of the key principles of DST is that dynamic systems are sensitive to initial conditions. It would appear that the students' various backgrounds contributed towards the initial conditions for ESL learning, thus ultimately impacting on the LLMs they produced for the lexical items investigated. An important insight to be gained from these perspectives provided by stimulated reconstructions is that developing vocabulary knowledge can be a highly-individualised experience which can be mediated by multiple factors, hence the importance of interrogating learner perspectives as these are likely to enhance understanding of the development of vocabulary knowledge from the learners' perspectives thereby potentially enhancing developments in pedagogy.

7.7 Variability on the factors influencing lexical decisions made by learners during the WAT.

From a DST perspectives, internal and external resources are a key requisite for development to take place as it is the availability of resources that ensures the continued functioning of the system (De Bot et al., 2007). In the current study, this phenomenon may be related to the profiles of Student 1, Student 8 and Student 13 as some of their rationale for lexical choices could be related to their educational backgrounds. Since Student 1 was the only participant in the current study who has an undergraduate degree / university level qualification (although the medium of instruction for the qualification was French), the rationale that she provided for some of the lexical choices that she made could be linked to this educational background.

This background can be considered to be a resource within Student 1's dynamic system. Student 13 was able to demonstrate awareness of academic vocabulary and link it to an activity that the ESL class did with their tutor. The student demonstrated memory of academic vocabulary from an ESL session which was not cited by any of the other students in the study. To that extent, the student's educational experiences seem to have had a positive impact on the development of the learner's vocabulary knowledge. Student 8 cited independent reading within the ESL context as a source of knowledge for the academic words that the learner knew. Thus, it would seem that the students' vocabulary knowledge was positively affected by their educational backgrounds as external resources.

This leads to the conclusion that previous individual learning experiences such as educational backgrounds are likely to be contributing

factors to learners' strategies and ultimately, successful development of vocabulary size and depth.

However, what was also noted in Table 6.2 is variability in terms of what could be considered to be part of the three learners' internal resources (De Bot, 2007). Student 13 in particular is a learner with relatively low motivation and therefore demonstrated low engagement levels within the ESL learning context compared to her peers. This distinguishes this student from the other two students who reported lexical decisions influenced by educational factors. This learner's trajectory may be explained by the proposition that dynamic systems are characterised not only by interconnectedness amongst subsystems but also compensatory relations between or among different resources or sub-systems (De Bot et al. 2007). In the case of Student 13, the learner was able to utilise an internal resource in the form of an evidently effective memory, in order to compensate for what appears to be lack of motivation to invest time outside class to reinforce and extend learning (see Section 6.2).

This may partially explain the learner's ability to recall information which may be related to the HLA score that the learner obtained across all written assessments. From the aggregation of factors cited as impacting on lexical choices that the ESL learners made during the WAT, the impact of each factor showed variation within the group (see Figure 6.3). It can be argued that the students' profiles show relative similarity in the sense that the overall picture is that of a group of ESL learners whose lexical choices were impacted the most by partial knowledge of lexical items; followed by cultural factors; impact of educational experiences; and finally, the use of guessing as a strategy impacting on learners' choices.

However, the differential impact of these factors on individual learners' choices illustrates that differences could easily be concealed when data is averaged (Larsen-Freeman, 2012; MacIntyre, 2012).

7.8 Chapter Summary

This chapter has highlighted theoretical interpretations related to the findings obtained from both the quantitative and qualitative findings from the current study. The relevance of DST perspectives is highlighted and the theory provided a framework for reconciling findings from the quantitative and qualitative findings in the study, thereby helping to bring to the fore, the importance of variability. Such variability could otherwise be hidden by the treatment of students only as groups who produce average rather than individual data. This is particularly important in the context of classroom based studies such as the current one because such an approach can produce insights that could assist teachers in understanding individual learner differences and therefore enhance the learning experience for the learners. The DST approach adopted also recognises the complexity of L2 processes by highlighting the interaction of multiple-factors in L2 developmental processes such as vocabulary knowledge and written language ability.

Chapter Eight:

Conclusion

8.0 Chapter Introduction

This chapter offers an overview of the study and captures key arguments presented in this thesis. On this basis, the chapter allows for the evaluation of the extent to which the study achieved its objectives as outlined in the introductory chapter. The chapter begins by presenting a summary statement of the issues at the centre of this investigation. This paves the way for a review of the key research findings, insights generated from them as well as the main contributions to knowledge. The chapter finishes by acknowledging limitations of the study, some of which lay foundation to recommendations for future research.

8.1 Summary of research focus

The introductory chapter highlighted the dissonance between some ESL learners' lexicogrammatical profiles and their ultimate linguistic performances as one of the key motivations behind this study. It was noted that some learners' overall language proficiency scores did not predict written language ability as might have been expected. This led to an interest in the investigation of the interplay between lexical knowledge and written language skills for ESL learners. Therefore, using an MMR approach grounded on a pragmatic epistemological stance, this study explored the interplay between vocabulary size, vocabulary depth and written language skills among L2 learners. While the small scale of the study does not allow for any grand claims to be made, the study has generated empirical evidence in support of the importance of vocabulary depth as a key aspect of lexical development and consequently generated some empirical evidence in support of the conceptual distinction between size and depth. Findings from the study also highlight the importance of not only group aver-

ages in L2 research but also, variability within groups and by so doing, highlight the relevance of DST approaches in the investigation of L2 processes. The findings from the study not only make contributions at a theoretical and methodological level, but also have practical implications in the field of Applied Linguistics as will be highlighted in this chapter.

8.2 The Study's contribution to knowledge

As Nation (2013, p.262) observes 'of all the four skills ...writing is the one where we know the least about the relationship between the skill and vocabulary knowledge'. While there now exists a substantial body of literature related to the interplay between different linguistic features and written language skill, the study utilised an innovative approach where performance and introspective data were corroborated with the tracing of individual learner developmental trajectories framed within a DST approach. The study revealed that the interplay between vocabulary knowledge and written language skill cannot be neatly summed up using quantitative data only. Instead, the MMR approach adopted for data collection and revealed variation within and among the ESL learners investigated. Such variation could easily be hidden behind averages/correlation data so it was through the triangulation of quantitative and qualitative data that the complexities of the interplay between vocabulary knowledge and written language ability were unveiled. To that extent, the study makes a contribution at the methodological level by drawing upon DST theory as a lens for viewing and interpreting findings from the study. The application of DST in the field of Applied Linguistics is fairly new (De Bot et al., 2007). Consequently, the current study contributes towards filling a methodological gap by providing an empirically based account of the complex nature of the relationship between vocabulary knowledge

and written language ability. The findings suggest that L2 learners can be construed as complex dynamic systems in line with DST models such as the one provided by Rosmawati (2013).

8.3 Pedagogical Implications

As highlighted in Section 6.2.3, some ESL students in the current study highlighted their teacher's efforts in trying to encourage them to learn more about words that they had come across before. In essence, the teacher was encouraging the students to develop not only the number of words that they knew, but also, the depth of knowledge for those words. However, what is concerning for an Applied Linguist is that, in most cases, the ESL students in the current study reported that they did not agree with their teacher's view regarding the importance of developing depth of vocabulary knowledge as they felt that this would be taking away from the time and effort that could be invested in developing vocabulary size. This view was supported by interview findings where students highlighted vocabulary size as their ultimate, and admitted that they were unwilling to invest time and effort into activities designed to help with the development of depth of vocabulary knowledge. This suggests that for the learners investigated in the current study, it is worth raising awareness of the importance of both size and depth of vocabulary knowledge (see Section 5.1.2) so that students can embrace positive beliefs about the importance of depth of knowledge.

In the current study, the WAT was used as an assessment tool for the evaluation of learners' depth of vocabulary knowledge. However, in light of the current findings relating to the apathy of ESL learners towards enhancing depth of lexical knowledge, the teacher could identify key words that are not well recycled in the particular units

that they teach to their students. They could then use such words as key words for their learners to set up WATs that are specifically tailored to their students' needs.

This could facilitate the development of depth of knowledge in a way that is relevant to specific learners' needs. Indeed, one of the criticisms of WATs is that sometimes such tests are derived from wordlists that were created several years ago and may therefore not be able to fully meet the needs of 21st century L2 learners (see Section 2.3.2). Therefore, devising WATs that are specifically designed for the needs of particular learners in a particular context could be one way of developing opportunities that facilitate the development of vocabulary depth. Such activities are likely to be more motivating and engaging to learners.

In addition, in the current study, illuminating findings about learners' individual lexical trajectories were obtained through the use of stimulated reconstructions as a follow up to the WAT. This further supports the view being advocated here that although WATs are most commonly used as tools for assessing depth of knowledge (Zareva and Wolter, 2012) they can also be used as a teaching and learning tools in tandem with the AfL approaches (see Section 2.7.4). The teacher can utilize WATs to obtain information about individual learners' mental lexicons as was the case in the current study. They could then set learning targets aligned to the needs of the learners. This would help teachers address any flawed lexical conceptions such as the ones that were reflected in the LLMs obtained from ESL students in the current study (see Section 6.2.3).

A challenge that may emanate from teachers' endeavours to supplement teaching materials as suggested in this discussion relates to time-constraints. However, what is being advocated here is for teachers to supplement teaching materials in the context of their own learners' needs and not a complete abandonment of commercial resources (Allen, 2015). Another way of mitigating against time pressure can be the use of technology.

The current study utilised Vocabprofile for the analysis of learners' lexical knowledge and found evidence which supports the utility of the tool for assessing written texts for vocabulary use at different frequency bands. This suggests that automatized tools such as this can be utilised for analysing not only ESL learners' writing but also, written text such as reading passages from course books. This can provide tutors with information about lexical coverage of high frequency words which could also provide useful insights about the sufficiency or otherwise of the coverage of key vocabulary which would, in turn, help tutors with planning for supplementary work to develop both size and depth of learners' vocabulary knowledge. Teachers need to evaluate teaching resources and keep up to date with any new research informed teaching methodologies to ensure that they are best suited to facilitate not just acquisition of new vocabulary but also, development of depth of knowledge.

8.3.1 The need to prioritise learners' specific needs to guide teaching and learning.

Laufer et al., (2005) advise that it is important to pay attention to semantic differences at an inter-lingual level as lexical errors may stem from such differences. For example, L1 words with several alternatives in English, English words with several unrelated transla-

tions in the L1, or English words with no L1 equivalents can cause particular difficulty (Laufer et al., 2005). In the current study, this recommendation resonates with the findings from cultural associations that students made with some lexical items during the WAT, thus resulting in LLMs which differed from expected lexical models (See Section 6.2.1). It was noted that although the LLMs showed some themes that allowed the output to be categorised into four categories, there was variation both within each student's output (so that not all LLMs per student were influenced by the same factor) and between students (so that different students' output was affected by different factors). This implies that there is need for teachers to prioritise and optimise individualised learning opportunities. This would help fully address students' lexical needs. This could involve innovation through assessment so that assessment is, as far as possible, integrated with teaching and learning in line with AfL perspectives (see Section 2.7.4) In the context of the current study, marking and providing summative written feedback would not have revealed the kinds of insights that were obtained herein about learners' individual lexical trajectories. However, it was through an individualised approach where I teased out students' views/rationale behind their lexical choices that individual perspectives and LLMs emerged. Such an approach is likely to require careful integration into the ESL curriculum as the individualised approach is likely to be time consuming. However, if teachers focus only on a limited number of words identified as important for particular students, then this is potentially a beneficial way of integrating teaching with assessment and helping learners expand their knowledge of particular words.

As an example, it was highlighted in Section 2.7.4 that as part of their ESL programme, the ESL students investigated in the current

study sit mid-course tests which are basically summative tests as the grades contribute a percentage towards the final examination score. Looking at the insights obtained from the current study, it seems likely that these students would benefit more from dynamic assessment rather than mid-term summative assessments considering the stress associated with such tests on one hand, and the fact that the only feedback that the students receive from this assessment is a grade. In addition, since some learners highlighted lack of prompt feedback as a demotivating factor with regards to engaging in independent learning activities such as writing, a particularly relevant way of addressing learners' individual needs would be the use of computerised assessment tools such as the VocabProfile tool as part of AfL. Students could be trained on how to input their own free writing onto the VocabProfile tool and generate output about their own lexical use. Since VocabProfile breaks down lexis in a text into different frequency bands, students could meaningfully track their own lexical development and start targeting words from a particular frequency band. They could even input text from a range of other written sources such as their textbooks in order to check lexical use from different frequency bands. This could also help them identify target words to learn and practise independently.

Other than the challenges that dynamic assessment might present to ESL teachers, it is also important to consider how dynamic assessment is likely to be received by the learners themselves as attitudes can impact on individuals' ultimate behaviours which can then affect outcomes. In the current study, individual feedback was well-received and highly valued by participants. This supports the widely cited view that personalised approaches to learning are likely to pro-

duce positive outcomes in a learning and teaching environment (Becker et al., 2007). This makes it reasonable to expect ESL learners to welcome such a development. At the institution where the ESL learners were based, students' progress to HE courses and at this stage, they start receiving personalised language support through a range of techniques including online independent learning and assessments. Since L2 learning is a complex process, it seems that it would be more beneficial for learners to start getting access to personalised learning opportunities and support while they are doing preparatory courses such as the ESL programme that the students were on. This is likely to allow teachers the opportunity to help students build skills relevant to their own needs which would be a welcome development considering the variability noted in this study. It would also supplement the traditional formative and summative assessment, thus increasing opportunities for the dynamic assessment *for* learning (AfL) rather than assessment *of* learning (AoL).

8.3.2 The need to help students develop effective strategy use

Section 2.3.2 highlighted that learning strategies such as guessing can facilitate autonomous learning (O'Malley and Chamot, 1994) and are therefore an essential tool for active learning as research suggests that effective use of such strategies can enhance achievement in L2 learning. In the current study, it was noted that while some learners were comfortable and effective in using guessing as a strategy for dealing with unknown vocabulary knowledge, some had fundamental philosophical objections to the approach. This implies that variability is an important part of the L2 learning process, and that cultural factors should constitute an important consideration in any classroom situation because understanding variability can help teachers develop appropriate strategies for supporting learning and devel-

oping appropriate learning opportunities which are differentiated according to the needs of their learners. For example, since literature suggests that lexical inferencing can lead to lexical development; and results from the current study highlighted potential benefits of using lexical inferencing as a strategy for dealing with unknown vocabulary, this study recommends that there is need to sensitise learners from different backgrounds on the value of lexical inferencing as a legitimate strategy. Following in the steps of Fraser (1999), students can be explicitly trained on what lexical inferencing is; why it is useful; and when, where, and how to use it Fraser (1999). The teacher could demonstrate how lexical inferencing might work in practice, for example, through morphological activities such as identifying prefixes and suffixes in words.

Such awareness may help learners go beyond their initial apprehensions about lexical inferencing as a strategy. In addition, in the literature, guessing and lexical inferencing are used interchangeably to refer to the same strategy of using contextual clues to deal with unknown vocabulary.

It is possible that the use of the term 'guessing' might bear negative associations for some learners so it might be helpful for course books and teachers to use the term 'lexical inferencing' instead to encourage learners to utilize this strategy as it seems to have potential learning benefits.

8.3.3 Implications for ESL learners

For the ESL students investigated in the current study, it would appear that it is important for learners to develop autonomy in their vocabulary learning. Considering the sheer amount of words that a second language learner needs in order to function effectively in the

L2, it is clear that learning L2 vocabulary presents a considerable challenge. In addition, since the course book is an integral part of many ESL programmes, it is essential that students are supported towards utilising the most appropriate books particularly when they engage in independent learning activities. These could provide opportunities for the development of both size and depth of vocabulary knowledge, particularly in light of the interview findings that students rely, to a very large extent, on course books for their independent learning. The additional need to develop not just vocabulary size but depth as well suggests that the role of autonomous learning cannot be over emphasized.

Research suggests that increased autonomy can lead to higher levels of intrinsic motivation, and this can ultimately lead to greater achievement (Tremblay and Gardner, 1995). In the current study, VocabProfile was identified as a tool that can be used to encourage learners to develop autonomy in their written language activities that can aid the development of vocabulary knowledge as learners can self-assess and get instant feedback on their vocabulary use in a particular piece of writing. This encourage and enhance the development of size and depth of vocabulary knowledge.

8.4 Generalisability

The triangulated approach adopted in this investigation enhanced the study's ability to provide more in-depth insights and comprehensive answers to the research questions. This provided a basis for naturalistic rather than scientific generalisability (Melrose, 2009). It is from this perspective that the contributions made by this study and their practical implications are considered. The study utilised data obtained from 18 primary participants and 40 samples of data

obtained from the ICLE corpus. This is a relatively small subject group since one of the subgroups (the ESL group) comprises less than 30 participants (Lowie and Seton, 2013). Consequently, it was not the aim of the study to establish norms in the form of scientific generalisability. Instead, the study offers opportunities for readers to obtain contextualised insights which can facilitate naturalistic generalisations in the form of readers drawing their own links from insights in this study, should they find them to resonate with their own contexts (Melrose, 2009). The study also provides empirical bases for further investigation of the interplay between vocabulary knowledge and written language skills among ESL learners.

8.5 Limitations of the study

8.5.1 Self-reported data

The stimulated reconstructions / semi-structured interviews provided qualitative data for the analysis of learners' trajectories. While this is an important part of the study as it allowed for the tracing of variability, a criticism can be levied at the methodological level, given the fact that self -reported data may not be as accurate as objectively captured data. This is because the participants may report what they believe the researcher wants to know. As such, the findings in this study mat have been affected by the social desirability effect especially considering that the researcher was a teacher at the institution where the research was carried out. However, steps were taken towards mitigating against this self-reported data since the researcher ensured that all the students who took part in the study were in classes taught by the researcher's colleagues rather than the researcher herself.

8.5.2 Sample Size

Section 3.3.2. highlighted that the original plan was to collect data from a larger sample of students for the quantitative part of the study (72 students in total were targeted as the study sample). However, upon entering the field for data collection, permission to access some of the participants was withdrawn at the last minute, thus reducing the sample size significantly. In the end, 18 ESL students whose written compositions were analysed alongside 40 ICLE corpus samples comprised the study sample, which gave a total of 58 written essays. However, a triangulated approach was adopted both for data collection where each of the 18 participants completed three written assessments (VLT, WAT and free writing) and out of these, eight students also took part in stimulated reconstructions/semi-structured interviews). This ensured that deeper insights could be obtained from the relatively small sample. In addition, in recognition of the relatively small sample involved, the study sought naturalistic rather than scientific generalisability of its findings (Melrose, 2009). Therefore, it is hoped that readers will gain insight by reflecting on the details and descriptions presented in this study and then determining the relevance of findings from this study to their own contexts. This should open new and extended avenues for future research in vocabulary studies. Furthermore, owing to the multiple data sets that the study utilised from the participants, the numbers were deemed adequate for addressing the research questions in the current study. However, a larger sample and more precisely comparable data would allow for more conclusive results so this is an area for development in future studies. In addition, for reasons beyond the researcher's control, not all participants who completed written assessments were interviewed. Interviewing more students might have

provided more evidence for factors that were identified as rationale for lexical choices that students made during WAT activities. It is also possible that additional insights might have emanated from the analysis of a wider base of student narratives.

8.6 Recommendations for future studies

As Larsen-Freeman (2012) notes, DST is a fairly new approach to studies of L2 development. On the basis of findings from this study which suggested a complex interplay between lexis and writing, it is recommended that the field needs more studies investigating the interplay between vocabulary knowledge through the adoption of new methodologies such as those offered by DST. This would open avenues for extending insights about L2 developmental processes. This study also noted that students investigated in this study did not always appreciate the value of depth of knowledge in the development of vocabulary knowledge. This suggests that studies aimed at exploring learners' metacognitive knowledge and understanding how metacognition can be best achieved with particular reference to the development of depth of vocabulary knowledge would be beneficial. However, such studies need to be built on the foundation of further studies to clarify the concepts of vocabulary size and depth. Owing to its small sample size, this study has only provided tentative evidence to suggest that size and depth may be useful conceptualisations. Nonetheless, further studies are required in order to make any grand claims about the understanding and development of size and depth of vocabulary knowledge.

8.7 Final remarks

Through an MMR approach, the study highlighted the complexities involved in what might appear at first sight, to be a simple linear rela-

tionship between linguistic variables. However, the study highlighted that while the relationship between vocabulary knowledge and writing skills may provide a window into the learners' linguistic abilities, the relationship is by no means a straightforward one. This is because a number of factors can impact on it including the very nature of vocabulary knowledge as a complex phenomenon as well as learners' varied backgrounds and the affordances for vocabulary knowledge that they each identify and utilise in any learning context. Therefore, through quantitative and qualitative data obtained from L2 learners, the study identified the L2 learner as a complex dynamic system who is impacted upon by multiple factors, leading to the complexity of processes that the learner goes through during the language learning process.

The study concludes that depth of vocabulary knowledge should be given as much attention as vocabulary size because of its potential role in the development of key linguistic skills such as the ability to compose written text. This can only be achieved through teachers and researchers working together to shape and inform pedagogy resulting in principled teaching of vocabulary as opposed to a vocabulary-takes-care-of-itself approach. However, from the findings of the current study, it is clear that even with the best teachers and the most highly motivated ESL learners, the interplay between written language skills and lexical knowledge is unlikely to take a linear route. This is because of the interaction of multiple factors involved during the composing process on the one hand, and lexical development on the other hand. To that extent, the interplay between vocabulary knowledge and written language skills remains at best, described as a complex and dynamic one, hence the need for further DST studies in this domain.

REFERENCES

Agdam, S. J., and Sadeghi, K. (2014). Two formats of word association tasks : a study of depth of word knowledge. *English Language Teaching*, *7*(10), 1–12.

Aitchison, J. (2012). *Words in the Mind*. Oxford: Wiley-Blackwell.

Akbarian, I. (2010). The relationship between vocabulary size and depth for ESP / EAP learners. *System*, *38*(3), 391–401.

Albretchtsen, D., Haastrup, K., and Henriksen, B. (2008). Vocabulary and Writing in a First and Second Language:Process and Development. Palgrave: Macmillan.

Alderson, J. C. (2005). Diagnosing Foreign Language Proficiency. Applied Linguistics. London: Continuum.

Alderson, J.C., Figueras, N., Kuijper, H., Nold, G., Takala, S. and Tardieu, C. (2004). The develop-ment of specifications for item development and classification within The Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Reading and Listening: Final report of The Dutch CEF Construct Project.

Allen, C. (2015). Marriages of convenience? Teachers and coursebooks in the digital age. *ELT Journal*, *69*(3), 249–263.

Altenberg, B. and Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied linguistics*, *22*(2), 173-195.

Altenberg, B. (1998). On the phraseology of spoken English: the evidence of recurrent word combinations. In A. P. Cowie (Ed.), Phraseology: theory, analysis and applications (pp. 101–122). Oxford: Oxford University Press.

Amrein, A.L. and Berliner, D.C. (2002). High-stakes testing and student learning. Education policy analysis archives, 10, p.18.

Anderson, R., and Freebody, P. (1981). Vocabulary knowledge. In I. Guthrie (Ed.), *Comprehension and Teaching : Research Reviews* (pp. 77–117). Newark: International Reading Association.

Babbie, E. (2007). The Practice of Social Research, Belmont, CA: Thomson Learning.

Bacha, N. (2001). Writing evaluation: What can analytic versus holistic essay scoring tell us? *System*, *29*(3), 371–383.

Bachman, L.F. (1996). Language testing in practice. Oxford University Press:Oxford.

Baddeley, A. (1990). *Human memory*. London: Lawrence Erlbaum Associates.

Bailey, K.M. (1996). Working for washback: A review of the washback concept in language testing. Language testing, 13(3), pp.257-279.

Baker, P. (2006). *Using corpora in discourse analysis*. London: A and C Black.

Baker, T. (1994). *Doing Social Research* (2nd ed.). New York: McGraw-Hill.

Barkaoui, K. (2007). Rating scale impact on EFL essay marking: A mixed-method study. *Assessing Writing*, *12*(2), 86–107.

Becker, K., Kehoe, J., and Tennent, B. (2007). Impact of personalised learning styles on online delivery and assessment. *Campus-Wide Information Systems*, *24*(2), 105–119.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M.H., Croft, W., Ellis, N.C., Holland, J., Ke, J., Larsen-Freeman, D. and Schoenemann, T., (2009). Language is a complex adaptive system: Position paper. *Language learning*, *59*(1), 1-26.

Beglar, D., (2000). Estimating vocabulary size. *JALT Testing and Evaluation SIG Newsletter*, *4*(1), 2-4.

Beglar, D., and Hunt, A. (1999). Revising and validating the 2000 Word Level and University Word Level Vocabulary Tests. *Language Testing*, *16*(2), 131–162.

Bergman, M. M. (2008). The straw men of the qualitative-quantitative divide and their influence on mixed methods research. In M. M. Bergman (Ed.), Advances in mixed methods research (pp. 11-21). London: Sage.

Bergman, M. M. (2012). The good, the bad, and the ugly in Mixed Methods Research and Design. *Journal of Mixed Methods Research*, *5*(4), 271–275.

Biesta, G. (2010). Pragmatism and the philosophical foundations of mixed methods research. In A. Tashakkori and C. B. Teddlie (Eds.), *SAGE Handbook of Mixed Methods in Social Behavioural Research* (pp. 95–118). Thousand Oaks:London.

Black, P. and Wiliam, D. (2006). The reliability of assessments. Assessment and learning, pp.119-131.

Bogaards, P., Laufer, B., 2004. Vocabulary in a Second Language. John Benjamins Publishing Company, Amsterdam.

Bowles, M. (2010). *The think-aloud controversy in second language research*. London: Routledge.

Bracken, S. (2010). Discussing the importance of ontology and epistemology awareness in practitioner research. *Worcester Journal of Learning and Teaching*, (4), 1-9.

Broad, B. (2003). *What We Really Value: Beyong Rubrics in Teaching and Assessing Writing*. Logan: USU Press.

Bruton, A. (2009). The Vocabulary Knowledge Scale: A critical analysis. *Language Assessment Quarterly*, *6*(4), 288–297.

Bryman, A. (2008). Why do researchers integrate/combine/mesh/blend/mix/merge/fuse quantitative and qualitative research? In M. M. Bergman (Ed.), *Advances in Mixed Methods Research Theories and Applications* (pp. 87–100). London: Sage.

Burns, R.P. and Burns, R. (2008). *Business research methods and statistics using SPSS*. London: Sage.

Burns, A., and Knox, J. (2011). Classrooms as Complex Adaptive Systems. *Classrooms as Complex Adaptive Systems*, *15*(1), 1–25.

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

Carter, R. (1998). Vocabulary: Applied linguistics perspectives (2nd ed.), London: Routledge

Caspi, T. and Lowie, W. (2014). The dynamics of L2 vocabulary development: A case study of receptive and productive knowledge. *Revista Brasileira de Linguística Aplicada*, *13*(2), 437-462.

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

Chen, Y.H. and Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. Language Learning and Technology, 14(2), pp.30-49.

Chenoweth, N., and Hayes, J. R. (2001). Fluency in writing: generating text in L1 and L2. *Written Communication*, *18*(1), 80–98.

Churchill, E. (2007). A dynamic systems account of learning a word: From ecology to form relations. *Applied Linguistics*, *29*(3), 339–358.

Cilliers, P. (2001). Boundaries, hierarchies and networks in complex systems. *International Journal of Innovation Management*, *5*(2), 135–147.

Clandinin, D. and Connelly, F. (2000). *Narrative Inquiry:Experience and story in Qualitative Research*. San Francisco: Jossey-Bass.

Clarke, D.F. and Nation, I.S.P. (1980). Guessing the meanings of words from context: Strategy and techniques. System, 8(3), pp.211-220.

Cobb, T. (1995). Imported tests: Analysing the task. Paper presented at TESOL (Arabia). Al-Ain, United Arab Emirates, March.

Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. Computer Assisted Language Learning, 12(4), 345-360.

Cobb, T. (2007) Why and how to use frequency lists to learn words. Compleat Lexical Tutor. Available at: http://www.lextutor.ca/research/ (Accessed:12 April, 2017)

Cobb, T. (2013). Frequency 2.0: Incorporating homoforms and multiword units in pedagogical frequency lists. In Laufer, B.,

Bardel, C., and Lindquist,C. (eds.) L2 Vocabulary, Acquisition, Knowledge and Use: New perspectives on assessment and corpus analysis, p.79.Colchester:EUROSLA, pp.57-79.

Cobb, T. (2017) Email Communication, 16 January.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.

Cohen, K., Manion, L., and Morrison, K. (2009). *Research Methods in Education*. London: Routledge.

Colby-Kelly, C., and Turner, C. (2007). AFL research in the L2 classroom and evidence of usefulness:Taking formative assessment to the next level. *The Canadian Modern Language Review*, *64*(1), 9–37.

Collins, K. M. T. (2010). Advanced sampling designs in mixed research: Current practices and emerging trends in the social and behavioral sciences. In A. Tashakkori and C. Teddlie (Eds.), *SAGE Handbook of Mixed Methods in Social Behavioral Research* (pp. 353–378). Sage.

Council of Europe (2001), Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press.

Coxhead, A., Nation, I. S. (2001). The specialised vocabulary of English for academic purposes. In *Research Perspectives on English for Academic Purposes.* (pp. 252–267). Cambridge: Cambridge University Press.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–38.

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., and Hanson, W. E. (2003). Advanced mixed methods research designs. In A.Tashakkori and C.Teddlie (Eds.), Handbook of Mixed Methods in Social and Behavioral Research (pp. 209–240). Thousand Oaks, CA: Sage

Cronin, V. (2002). The syntagmatic-paradigmatic shift and reading development. *Child Language*, *29*(1), 189–204.

Crossley, S.A. and McNamara, D.S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second

language writing. International Journal of Continuing Engineering Education and Life Long Learning, 21(2-3), pp.170-191.

Crossley, S.A. and McNamara, D.S. (2012). Predicting second language writing proficiency: the roles of cohesion and linguistic sophistication. Journal of Research in Reading, 35(2), pp.115-135.

Crotty, M. (1988). *The Foundations of Social Research: Meaning and Perspective in the Research Process*. London: Sage.

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

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

Davidson, D. J., Indefrey, P., and Gullberg, M. (2008). Words that second language learners are likely to hear, read, and use. *Bilingualism: Language and Cognition*, *11*(01), 133–146.

Davies, A., Brown, C. Elder, K. Hill, T. Lumley, and McNamara, T. (1999). Dictionary of Language Testing, Studies in Language Testing, Vol. 7. Cambridge: Cambridge University Press.

De Bot, K., Lowie, W., and Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition*, *10*(01), 7.

De Bot, K., Paribakht, T., and Wesche, M. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition: evidence from ESL reading. *Studies in Second Language Acquisition*, *19*(3), 309–329.

De Bot, K., Verspoor, M., and Lowie, W. (2005). Dynamic Systems Theory and Applied Linguistics: the ultimate "so what"? *International Journal of Applied Linguistics*, *15*(1), 116–118.

De Laine, M. (2000). *Fieldwork, participation and practice: Ethics and dilemmas in qualitative research*. London: Sage.

De Larios, J.R., Manchón, R., Murphy, L. and Marín, J. (2008). The foreign language writer's strategic behaviour in the allocation of time to writing processes. Journal of Second Language Writing,

17(1), pp.30-47.

Denzin, N. K., and Lincoln, Y. S. (1994). *Handbook of Qualitative Research*. *Qualitative Inquiry*. London: Sage.

Dornyei. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.

Dornyei, Z. (2000). Motivation in action: Towards a process-oriented conceptualisation of student motivation. *British Journal of Educational Psychology,* 70(4),519-38.

Dornyei, Z. (2014). Researching complex dynamic systems: "Retrodictive qualitative modelling" in the language classroom. *Language Teaching*, *47*(1), 80–91.

Dornyei, Z., Henry, A. and MacIntyre, P.D. eds., (2015). Motivational dynamics in language learning (Vol. 81). Bristol, Multilingual Matters.

East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing*, *14*(2), 88–115.

Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, *25*(2), 155–185.

Eldridge, J., and Neufield, S. (2009). The graded reader is dead, long live the electronic reader. *Reading*, *9*(2), 224–244.

Ellis, N. (1998). Emergentism, connectionism and language learning. *language learning*, *48*(4), 631–664.

Ellis, N. C., and Larsen-Freeman, D. (2006). Language emergence: implications for Applied Linguistics : introduction to the special issue. *Applied Linguistics*, *27*(4), 558–589.

Ellis, N.C., Simpson-Vlach, R. and Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. TESOL *Quarterly*, *42*(3), pp.375-396.

Engber, C. (1995) 'The relationship of lexical proficiency to the quality of ESL compositions', Journal of Second Language Writing, Vol. 4, pp.139–155

Entwinstle, D., Forsyth, D., and Muuss, R. (1964). The syntagmatic-paradigmatic shift in children's word associations. *Journal of Verbal Learning and Behaviour*, *3*(1), 19–29.

Entwisle, D. (1966). *Word associations of young children*. Baltimore: Johns Hopkins Press.

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.

Fani, T. and Ghaemi, F. (2011). Implications of Vygotsky's zone of proximal development (ZPD) in teacher education: ZPTD and self-scaffolding. *Procedia-Social and Behavioral Sciences*, *29*, 1549-1554.

Faerch, C. (1987). Lexis in composition: a performance analysis of Swedish learners' written English. *Studies in Second Language Acquisition*, *9*(02), 254-256.

Feldman, D. (2004). Piaget's stages: the unfinished symphony of cognitive development. *New Ideas in Psychology*, *22*(3), 175–23.

Fielding, N. G. (2012). Triangulation and mixed methods designs: Data integration with new research technologies. *Journal of Mixed Methods Research*, *6*(2), 124–136.

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

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

Fitzpatrick, T. and Izura, C. (2011). Word Association in L1 and L2. *Studies in Second Language Acquisition*, *33*(03), 373–398.

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911.

Flower, L. and Hayes, J. R. (2008). A cognitive process theory of writing. *College Composition and Communication*, *32*(4), 365–387.

Francis, B. (1993). A corpus-driven approach to grammar:Principles, methods and examples. In Baker, M., Francis, G., Tognini-Bonelli, E. (Eds.), Text and technology (pp137-156). Amsterdam:Benjamins.

Friginal, E., Li, M. and Weigle, S. C. (2014). Revisiting multiple profiles of learner compositions: A comparison of highly rated NS and NNS essays. *Journal of Second Language Writing*, *23*, 1–16.

Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A Critical Survey. *Applied Linguistics*, *28*(2), 241–265.

Gardner, D. (2008). Vocabulary recycling in children's authentic reading materials: A corpus- based investigation of narrow reading. *Reading in a Foreign Language*, *20*(1), 92–122.

Gass, S.M. and Mackey, A. (2000). *Stimulated recall methodology in second language research*. London: Routledge.

Goulden, N. (1992). Theory and vocabulary for communication assessments. *Communication Education*, *41*(3), 258–269.

Goulden, N. (1994). Relationship of analytical and holistic methods to raters' scores for speeches. *The Journal of Research and Development in Education*, *27*, 73–82.

Grabe, W. (1991). Current developments in second langauge research. *TESOL Quarterly*, *25*(3), 375–406.

Granger, S. (2003). The International Corpus of Learner English: A new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly*, *37*(3), 538–546.

Granger, S., Dagneaux, E., Meunier, F. and Paquot, M. (2009). International Corpus of Learner English: version 2. Presses Universataires de Louvian, Belgium.

Gray, J. (2010). The Construction of English: Culture, Consumerism and Promotion in the ELT Global Coursebook. *Palgrave, Macmillan*.

Gregg, K.R. (2010). Review article: Shallow draughts: Larsen-Freeman and Cameron on complexity. Second Language Research, 26(4), pp.549-560.

Grice, J.W. (2004). Bridging the idiographic-nomothetic divide in ratings of self and others on the Big Five. *Journal of Personality*, *72*(2), 203-241.

Guba, E. G. (1990). The alternative paradigm dialog. In E. G. Guba (Ed.), *The paradigm dialog*, Vol. 1 (17–27). London: Sage.

Guba, E. G., and Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin and Y. S. Lincoln (Eds.), *Handbook of qualitative research*, Vol. 2 (105–117). London : Sage.

Guba, E. G., and Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin and Y. S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research*, 3rd Vol., (pp. 191–215). London: Sage.

Gutiérrez-Bermúdez, J. F. (2014). An exercise in course-book evaluation: Strengths, weaknesses, and recommendations regarding New English File: Elementary. *Latin American Journal of Content and Language Integrated Learning*, *7*(1), 98–111.

Haastrup, K. and Henriksen, B. (2000). Vocabulary acquisition : acquiring depth of knowledge through network building, *10*(2), 221-240.

Haastrup, K. (1991). *Lexical Inferencing Procedures or Talking about Words*. Tübingen: Gunter Naar.

Hale, G. (1992). Effects of amount of time allowed on the Test of Written English. *ETS Research Report Series*, *1*, 1–35.

Halliday, M.A.K. (1961). Categories of the theory of grammar. Word, 17(3), pp.241-292.

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

Halliday, M.A.K. and Matthiessen, C.M. (2014). An introduction to functional grammar. London: Routledge.

Halliday, M.A.K. and Webster, J. (2002). On grammar: Collected

works of MAK Halliday (Vol. 1). London: Continuum.

Hamp-Lyons, L. and Henning, G. (1991). Communicative writing profiles: An investigation of the transferability of a multiple-trait scoring instrument across ESL writing assessment contexts. Language Learning, 41(3), pp.337-373.

Harrits, G. (2011). More Than Method? A discussion of paradigm differences within mixed methods research. *Journal of Mixed Methods Research*, *5*(2), 150–166.

Hashemi, M. R. (2012). Reflections on mixing methods in Applied Linguistics Research. *Applied Linguistics*, *33*(2), 206–212.

Hasko, V. (2013). Capturing the dynamics of second language development via learner corpus research: A very long engagement. *The Modern Language Journal*, *97*(S1), 1–10.

Hawkey, R., and Barker, F. (2004). Developing a common scale for the assessment of writing. *Assessing Writing*, *9*(2), 122–159.

Hayes, J. (1996). A new framework for understanding cognition and affect in writing. In C. Levey and L. Ransdell (Eds.), *The Science of Writing: Theories, Methods, Individual Differences and Application* (pp. 1–27). Hillsdale: Erlbaum.

Hedge, T. (2001). *Writing*. Oxford: Oxford University Press.

Hellstrom, T. (2008). New generalisability concepts for Social Science or old wine in new bottles? *Quality and Quantity*, *42*(3), 321–337.

Henderson, L. and Tallman, J. (2006). Stimulated Recall and Mental Models. The Scarecrow Press, Lanham, MD.

Hennink, A., Hutter, M., and Bailey, I. (2011). *Qualitative Research Methods*. London: Sage.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, *21*(02), 303-317.

Higginbotham, G. (2010). Individual learner profiles from word association tests: The effect of word frequency. *System*, *38*(3), 379–390.

Hirose, K., and Sasaki, M. (1994). Explanatory variables for Japanese

students' expository writing in English: An exploratory study. *Journal of Second Language Writing*.

Honeyfield, J. (1977). Word frequency and the importance of context in vocabulary learning. *RELC Journal*, *8*(2), 35–42.

Howe, K. R. (1988). Against the quantitative-qualitative incompatibility thesis or dogmas die hard. *Educational Researcher*, *17*(8), 10–16.

Hu, G. (2002). Potential cultural resistance to pedagogical imports : The case of communicative language teaching in China. *Language, Culture and Curriculum, 15*(2), 93–105.

Hu, H., and Nation, I. (2000). Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language*, *13*, 403–30.

Huang, J. (2008). How accurate are ESL students' holistic writing scores on large-scale assessments? A generalizability theory approach. *Assessing Writing*, *13*(3), 201–218.

Hulstijn, I. H., and Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *51*(3), 539–558.

Hutchinson, T. and Torres, E. (1994). The textbook as agent of change. *ELT journal*, *48*(4), 315-328.

Hyland, K., and Tse, P. (2007). Is There an "Academic Vocabulary"? *TESOL Quarterly*, *41*(2), 235–253.

Hyltenstam, K. (1988). Lexical characteristics of near-native second-language learners of Swedish. *Journal of Multilingual and Multicultural Development*, *9*, (1) 67–84.

Jacobs, H., Zingraf, S., Hartfiel, V., and Hughey, J. (1981). *Testing ESL Composition: A Practical Approach*. Rowley, MA: Newbury House.

Jacobson, S. and Wood, F. (2006). Lessons learned from a very small pilot study. *Online Journal of Rural Nursing and Healthcare*, *6*(2), 18–28.

Jarvis, S., Grant, L., Bikowski, D., and Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of*

Second Language Writing, *12*(4), 377–403.

Jiang, Y. and Dewaele, J.M. (2015). What lies bubbling beneath the surface? A longitudinal perspective on fluctuations of ideal and Ought-to L2 self among Chinese learners of English. *International Review of Applied Linguistics in Language Teaching*, *53*(3), pp.331-354.

Jin, W. (2001) 'A quantitative study of cohesion in Chinese graduate students' writing: variations across genres and proficiency levels', Paper presented at the Symposium on Second Language Writing, September, Purdue University, West Lafayette, Indiana

Johnson, R. B., Onwuegbuzie, A. J., and Turner, L. A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, *1*(2), 112–133.

Johnson, E. and Jenkins, J. (2009). Formative and summative assessment. *Psychology of classroom learning: An encyclopedia. Detroit: Macmillan Reference*.

Johnson, R.B. and Onwuegbuzie, A.J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, *33*(7), 14-26.

Kaivanpanah, S. and Moghaddam, M.S. (2012). Knowledge sources in EFL learners' lexical inferencing across reading proficiency levels. *RELC Journal*, *43*(3), 373-391.

Kent, G., and Rosanoff, A. (1910). A study of association in insanity. *American Journal of Psychiatry*, *67*(1), 37–96.

Kezar, A. (2000). The importance of pilot studies: Beginning the hermeneutic circle. *Research in Higher Education*, *41*(3), 385–400.

Kobayashi, H., and Rinnert, C. (2008). Task response and text construction across L1 and L2 writing. *Journal of Second Language Writing*, *17*(1), 7–29.

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, *21*(4), 390–403.

Kramsch, C. (2002). How can we tell the dancer from the dance? In C. Kramsch (Ed.), *Language acquisition and language socialization: Ecological perspectives* (pp. 1–30). London:

Continuum.

Kremmel, B. and 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), pp.377-392.

Kuiper, K., Van Egmond, M.E., Kempen, G. and Sprenger, S. (2007). Slipping on superlemmas: Multi-word lexical items in speech production. *The Mental Lexicon*, *2*(3),313-357.

Kvale, S. (2006). Dominance through interviews and dialogues. *Qualitative Inquiry*, *12*(3), pp.480-500.

Larsen-Freeman, D. (1997). Chaos/Complexity Science and Second Language Acquisition. *Applied Linguistics*, *18*(2), 141–165.

Larsen-Freeman, D. (2003). Teaching language: From grammar to grammaring. Boston: Heinle and Heinle

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, *27*(4), 590–619.

Larsen-Freeman, D. (2012). Complex, dynamic systems: A new transdisciplinary theme for Applied Linguistics? *Language Teaching*, *45*(02), 202–214.

Larsen-Freeman, D., and Cameron, L. (2008). Research methodology on language development from a complex systems perspective. *The Modern Language Journal*, *92*(2), 200–213.

Larsen-Freeman, D., and Long, M. (1991). *An Introduction to Second Language Acquisition Research*. New York: Longman.

Laufer, B. (1997). The lexical plight in second langauge reading. In *Second Language Vocabulary Acquisition* (pp. 20–34). Cambridge: Cambridge University Press.

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

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

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

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

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

Lee, I. (2011). Bringing innovation to EFL writing through a focus on assessment for learning. *Innovation in Language Learning and Teaching*, *5*(1), 19–33.

Leech, G. (1992). 100 million words of English: the British National Corpus (BNC). Language Research, 28(1), pp.1-13.

Leech, N. L., Dellinger, A. B., Brannagan, K. B., and Tanaka, H. (2010). Evaluating mixed research studies: A mixed methods approach. *Journal of Mixed Methods Research*, *4*(1), 17–31.

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

Leung, C. (2007). Dynamic Assessment: Assessment *for* and *as* Teaching? *Language Assessment Quarterly*, *4*(3), 257–278.

Leung, C., and Mohan, B. (2004). Teacher formative assessment and talk in classroom contexts. *Language Testing*, *21*(3), 335–359.

Lewis, P., Thornhill, A., and Saunders, M. (2009). *Research Methods for Business Students* (5th ed.). Essex, UK: Pearson Education Inc.

Light, R. (2001). Making the most of college. Cambridge, MA: Harvard University Press.

Lincoln, Y., and Guba, E. (1985). *Naturalistic Inquiry*. London: Sage Publications.

Lindgren, E. and Sullivan, K.P. (2003). Stimulated recall as a trigger for increasing noticing and language awareness in the L2 writing classroom: A case study of two young female writers. *Language Awareness*, *12*(3-4), 172-186.

Linnarud, M. (1986). *Lexis in composition: A performance analysis of Swedish learners' written English*. Lund: CWK Gleerup.

Liu, E.T. and Shaw, P.M. (2001). Investigating learner vocabulary: A possible approach to looking at EFL/ESL learners' qualitative knowledge of the word. *IRAL*, *39*(3),171-194.

Lowie, W. and Seton, B. (2013). *Essential Statistis for Applied Linguistics*. NewYork: Palgrave Macmillan.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *Modern Language Journal*, *96*(2), 190–208.

MacIntyre, P. D. (2012). The idiodynamic method: A closer look at the dynamics of communication traits. *Communication Research Reports*, *29*(4), 361–367.

Macintyre, P. D., and Legatto, J. J. (2010). A dynamic system approach to willingness to communicate: Developing an idiodynamic method to capture rapidly changing affect. *Applied Linguistics*, *32*(2), 149–171.

Malvern, D.D., Richards, B.J., Chipere, N. and Durán, P. (2004). Lexical diversity and language development. *Hampshire: Palgrave Macmillan*.

Manchon, R. (2009). Individual differences in foreign language learning : *RESLA*, *22*, 245–268.

Mantle-Bromley, C. (1995). Positive attitudes and realistic beliefs : Links to proficiency. *The Modern Language Journal*, *79*(3), 372–386.

Martinez, R., and Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, (1991), 1–23.

Masny, D. (1997). Linguistic awareness and writing: Exploring the Relationship with Language Awareness. *Language Awareness*, *6*(2), 105–18.

Matsuda, P.K. (1997). Contrastive rhetoric in context: A dynamic model of L2 writing. Journal of Second Language Writing, 6(1), 45–60

Maxwell, J. A. (2010). Using numbers in qualitative research. *Qualitative Inquiry*, *16*(6), 475–482.

McCarthy, P., and Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing*, *24*, 459–488.

McGrath, I. (2006). Teachers' and learners' images for coursebooks. *ELT Journal*, *60*(2), 171–180.

McGrath, I. (2013). *Teaching materials and the role of EFL/ESL teachers: Theory versus practice*. London, England: Bloomsbury.

McNamara, D.S., Crossley, S.A. and McCarthy, P.M. (2010). Linguistic features of writing quality, Written Communication, Vol. 27, pp.57–86.

McNamara, D.S., Kintsch, E., Songer, N.B. and Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. Cognition and instruction, 14(1), pp.1-43.

Meara, P. (1980). Vocabulary acquisition: A neglected area of Language learning. Language Teaching and Linguistics, 15(4), 221-246.

Meara, P. (1983). Word associations in a foreign language. *Nottingham Linguistics Circular*, *11.*(2), 29–38.

Meara, P. (1996). The dimensions of lexical competence. In Brown, G., Malmkjaer, K., and Williams, J. (Eds.). Perfomance and competence in second language acquisition (pp. 35-53). Cambridge: Cambridge University Press.

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

Meara, P. and Alcoy, J.C.O.(2010). Words as species: An alternative approach to estimating productive vocabulary size. Reading in a Foreign Language, 22(1), p.222.

Meara, P., and Wolter, B. (2004). V_Links: Beyond vocabulary depth. In D. Albrechtsen, K. Haastrup, and B. Henriksen (Eds.), *Writing and vocabulary in foreign language acquisition* (Vol. 4, pp. 85–96). John Benjamins Pub. Co.

Melka, F. (1997). Receptive versus productive aspects of vocabulary.

In N. Schmitt and M. Mccarthy (Eds.), *Vocabulary: Description, Acquistion and Pedagogy.* New York: Cambridge University Press.

Melrose, S. (2009). Naturalistic Generalization. In A. Mills, G. Durepos, and E. Wiebe (Eds.), *Encyclopedia of Case Study Research.* Thousand Oaks, C.A: Sage.

Menezes, V. (2013). Second language acquisition : Reconciling theories. *Open Journal of Applied Sciences*, *3*(7), 404–412.

Meier, A.M. and Vogt, F. (2015). The potential of stimulated recall for investigating self-regulation processes in inquiry learning with primary school students. Perspectives in Science, 5, pp.45-53

Mertens, D. M., and Hesse-Biber, S. (2012). Triangulation and mixed methods research: Provocative positions. *Journal of Mixed Methods Research*, *6*(2), 75–79.

Milton, J., 2009. Measuring Second Language Vocabulary Acquisition. Multilingual Matters, Bristol, England.

Milton, J., and Fitzpatrick, T. (2014). Deconstructing vocabulary knowledge. In J. Milton and T. Fitzpatrick (Eds.), *Dimensions of Vocabulary Knowledge* (pp. 1–12). Basingstoke: Palgrave Macmillan.

Milton, J. and Treffers-Daller, J. (2013) Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review*, *4*(1), pp.151-172.

Mitchell, K. and Anderson, J. (1986). Reliability of holistic scoring for the MCAT essay. Educational and psychological measurement, 46(3), pp.771-775.

Mitchell, R., and Myles, F. (2004). *Second Language Learning Theories* (2nd ed.) London: Arnold.

Moore-Hart, M. A. (1995). The effects of multicultural links on reading and writing performance and cultural awareness of fourth and fifth graders. *Computers in Human Behavior*, *11*(3-4), 391 – 410.

Morgan, A., (2007). Using video-stimulated recall to understand young children's perceptions of learning in classroom settings. European Early Childhood Education Research Journal, 15(2), pp.213-226.

Morgan, B., and Oberdeck, L. (1930). Active and passive vocabulary. In Bagster-Collins (Ed.), *Studies in Modern Language Teaching* (pp. 213–221). Basingstoke: Macmillan.

Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, *1*(1), 48–76.

Morris, L. and Cobb, T., 2004. Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. System, 32(1), pp.75-87.

Morris, M.W., Leung, K., Ames, D. and Lickel, B. (1999). Views from inside and outside: Integrating emic and etic insights about culture and justice judgment. Academy of Management Review, 24(4), pp.781-796.

Muncie, J. (2002). Process writing and vocabulary development: comparing Lexical Frequency Profiles across drafts. *System*, *30*(2), 225–235.

Munoz, A. P. and Alvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, *27*(1), 33–49.

Namei, S. (2004). Bilingual lexical development: A Persian-Swedish word association study. *International Journal of Applied Linguistics*, *14*(3), 363–388.

Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 Learners ' lexical inferencing use and success strategy. *The Modern Language Journal*, *90*(3), 387–401.

Nation, I. S.P., Waring, R. (1997). Vocabulary size, text coverage and word lists. In Schmitt, N. and McCarthy (Ed.), *Vocabulary: Description, Acquistion and Pedagogy.* Cambridge: Cambridge University Press.

Nation, I.S.P.(1990) Teaching and Learning Vocabulary. *New York: Heinle and Heinle*

Nation, I. S.P . (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, I. S.P . and Beglar, D. (2007). A vocabulary size test. *The*

Language Teacher, *31*(7), 9 – 13.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening ?, *63*(1), 59–81.

Nation, P. and Coxhead, A. (2014). Vocabulary size research at Victoria University of Wellington, New Zealand. Language Teaching, 47(03), pp.398-403.

Nation, I.S.P. and Heatley, A., (1996). VocabProfile, word and range: Programs for processing text. LALS, Victoria University of Wellington.

Nazmia, F., and Sima, T. (2004). L2 Reading proficiency and lexical inferencing by university EFL Learners L2 reading proficiency and lexical inferencing by university EFL Learners, *61*(2), 225–249.

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

Nesselroade, J., Gerstorf, D., Hardy, S. A., and Ram, N. (2007). Idiographic filters for psychological constructs. *Measurement*, *5*(4), 217–235.

Neuman, W. L. (2000). *Social Research Methods:Qualitative and Quantitative Research Methods* (4th ed.). Boston: Allyn and Bacon.

Neuman, W. L. (2006). *Social Research Methods: Qualitative and Quantitative Approaches*. New York: Pearson.

Nissen, H. B., and Henriksen, B. (2006). Word class influence on word association test results. *International Journal of Applied Linguistics*, *16*(3), 389–407.

Nunan, D., (2001). English as a global language. *TESOL quarterly*, *35*(4), pp.605-606.

Nunes, M. B., Martins, J. T., Zhou, L., Alajamy, M., and Al-mamari, S. (2010). Contextual sensitivity in grounded theory : The role of pilot studies, *8*(2), 73–84.

Nurweni, A. and Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes*, *18*(2), 161-175.

Olinghouse, N.G. and Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. Reading and Writing, 26(1), pp.45-65.

O'Loughlin, R. (2012). Tuning in to vocabulary frequency in course books. *RELC Journal*, *43*(2), 255–269.

O'Malley, J. M., and Chamot, A. U. (1994). *Learning Strategies in Second Language Acquisition*. Cambridge: Cambridge University Press.

Ohta, A. (2005). Interlanguage pragmatics in the zone of proximal development. *System*, *33*(3), 503–517.

O'reilly, T. and McNamara, D.S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. Discourse processes, 43(2), pp.121-152.

Orr, J. (1991). Piaget's theory of cognitive development may be useful in deciding what to teach and how to teach it. *Nurse Education Today*, *11*(1), 65–69.

Paas, F., and Ayres, P. (2014). Cognitive load theory: A broader view on the role of memory in learning and education. *Educational Psychology Review*, *26*(2), 191–195.

Pascarella, E., and Terenzini, P. (1991). *How college affects students: Findings and insights from twenty years of research.* San Francisco: Jossey-Bass.

Pilar, M., and Llach, A. (2009). Examining the relationship between receptive vocabulary size and written skills of primary school learners, 31 (*1)*, 129–147.

Powers, D., and Fowles, M. (1997). Effects of applying different time limits to a proposed GRE writing test. *ETS Research Report Series*, *2*, 1–19.

Prescott, P., and Soeken, K. (1989). The potential uses of pilot work. *Nursing Research*, *38*(1), 60.

Punch, K. F. (2005). *Introduction to Social Research: Quantitative and Qualitative Approaches*. *London*. Sage.

Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language learning*, *52*(3), 513-536.

Qian, D. D., and Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, *21*(1), 28–52.

Read, J. (1988). Measuring vocabulary knowledge of second language learners. *RELC Journal*, *19*(2), 12–25.

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. Kunnan (Ed.), Validation in language assessment (pp. 41-60). Mahwah, NJ: Lawrence Erlbaum.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2004). Plumbing the depths: how should the construct of vocabulary knowledge be defined? In P. Bogaard and B. Laufer (Eds.), *Vocabulary in a Second Language: Selection, Acquisition and Testing* (pp. 209–227). Philadelphia, P.A: John Benjamins.

Richards, C., and Schmidt, R. (2013). *Dictionary of Language Teaching and Applied Linguistics* (4th ed.). Oxon: Routledge.

Rosmawati, (2013). Second language developmental dynamics: How dynamic systems theory accounts for issues in second language learning. *The Australian Educational and Developmental Psychologist*, *31*(01), 66–80.

Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaker students. TESOL Quarterly, 22, 69-90.

Saunders, M. and Thornhill, A. (2009). *Research Methods for Business Students* (5th ed.). Essex: Pearson Education Ltd.

Schermer, V. L. (2012). Group-as-a-whole and complexity theories: areas of convergence. *Group Analysis*, *45*(3), 275–288.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, *13*, 206.

Schmitt, N. Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching: Plenary speech. *Language Teaching*, 1–20.

Schmitt, N. (2010). *A Vocabulary Research Manual*. Basingstoke: Palgrave Macmillan.

Schmitt, N., and Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 1–20.

Schmitt, N., and Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2. *Language Teaching*, *47*(4), 484–503.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*(4), 913-951.

Schmitt, N., Schmitt, D., and Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88.

Schmitt, N., Ng, J. W. C., and Garras, J. (2011). The Word Associates Format: Validation evidence. *Language Testing*, *28*(1), 105–126.

Schmitt, N., and Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, *36*(2), 145.

Schoonen, R., Van Gelderen, A., Stoel, R. D., Hulstijn, J., and De Glopper, K. (2011). Modeling the development of L1 and EFL writing proficiency of secondary school students. *Language learning*, *61*(1), 31-79.

Schumann, J.H. (1986). Research on the acculturation model for second language acquisition. *Journal of Multilingual and Multicultural Development*, *7*(5),379-392.

Seidman, I. (2015). *Interviewing as Qualitative Research: A Guide for Researchers in Education and the Social Sciences* (4th Ed.). New York: Teachers' College Press.

Sheldon, L. (1988). Evaluating ELT textbooks and materials. *ELT Journal*, *42*(4), 237–246.

Sheorey, R. (1986). Error perceptions of nativespeaking and non-

nativespeaking teachers ofESL. ELT journal, 40(4), pp.306-312.

Sieber, Joan E. (1998). Planning Ethically Responsible Research, in Bickman, L. and Rog, D.J., eds., *Handbook of Applied Social Research Methods*, Thousand Oaks, CA: Sage, 127-156.

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

Smith, L., and Thelen, E. (1993). *A dynamic systems approach to development.* Cambridge: The MIT Press.

Smith, R. (2005). The lexical frequency profile: Problems and uses. In *JALT2004 Conference Proceedings. Tokyo: JALT*.

Snellings, P., Van Gelderen, A. and De Glopper, K. (2004). Validating a test of second language written lexical retrieval: A new measure of fluency in written language production. *Language Testing*, *21*(2),174-201.

Soars, L., and Soars, J. (2005). *New Headway Student's Book*. *New headway student's book (Upper–Intermediate).* Oxford, UK: Oxford University Press.

Söderman, T. (1993). Word associations of foreign langauge learners and native speakers: The phenomenon of a shift in response type and its relevance for lexical development. In H. Ringbom (ed), *Near Native Proficiency in English* (pp. 91–182). Abo, Finland: Abo Akademi.

Soleimani, H., and Dabbaghi, A. (2012). Textbook evaluation: A reflection on the New Interchange Series. *International Journal of Research Studies in Language Learning*, *1*(2), 19–32.

Spoelman, M., and Verspoor, M. (2010). Dynamic Patterns in Development of Accuracy and Complexity: A Longitudinal Case Study in the Acquisition of Finnish. *Applied Linguistics*, *31*(4), 532–553.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, *36*(2), 139–152.

Stake, R. (2000). Case Studies. In N. Denzin and Y. Lincoln (Eds.), *Handbook of Qualitative Research* (2nd ed., pp. 435–454). Thousand oaks, CA: Sage Publications.

Stewart, J. and White, D. (2011) Estimating guessing effects on the vocabulary levels test for differing degrees of word knowledge. TESOL Quarterly, 45(2), pp.370-380.

Stolz, S., and Tiffany, J. (1972). The production of "child-like" word associations by adults to unfamiliar adjectives. *Journal of Verbal Learning and Behaviour*, *11*(1), 38–46.

Svalberg, A.M.L, and Askham, J. (2015). A dynamic perspective on student language teachers' different learning pathways in a collaborative context. In J. King (Ed.), *The dynamic interplay between context and the language learner* (pp. 172–193). London: Palgrave Macmillan.

Svalberg, A.M.L. (2007). Language awareness and language learning. *Language Teaching*, *40*, (4), 287–308.

Swan, M. (1997). The influence of the mother tongue on second language vocabulary acquisition and use. Vocabulary: Description, acquisition and pedagogy, pp.156-180.

Tashakkori, A., and Teddlie, C. (2003). *Handbook of Mixed Methods in Social and Behavioural Research.* Thousand Oaks: Sage.

Tashakkori, A. and Teddlie, C. (2008). Quality of inferences in mixed methods research: Calling for an integrative framework. In M. M. Bergman (Ed.) *Advances in mixed methods research*, (pp.101-119) Thousand Oaks: Sage.

Taylor, L. (2005). Washback and impact. *ELT Journal*, *59*(2), pp.154-155.

Teddlie, C., and Tashakkori, A. (2009). Foundations of Mixed Methods Research. Thousand Oaks: Sage.

Tashakkori, A. and Teddlie, C. (eds). (2010). *Sage Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks: Sage.

Teddlie, C., and Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research*, *1*(1), 77–100.

Thelen, E., and Smith, L. (1994). *A dynamic Systems Approach to the Development of Cognition and Development of Cognition and*

Action. Cambridge: MIT Press.

Tomlinson, B. and Masuhara, H. eds., (2011). *Research for materials development in language learning: Evidence for best practice*. Bloomsbury Publishing.

Tomlinson, B., and Masuhara, H. (2013). Adult coursebooks. *ELT Journal*, *67*(2), 233–249.

Tucker, G., (1999). The Lexicogrammar of Adjectives, Bloomsbury Publishing PLC, London. Available from: ProQuest Ebook Central. [12 February 2017].

Tucker, G. (2007). Between grammar and lexis: towards a systemic functional account of phraseology. In Hasan,R., Matthiessen, C., and Webster, J.(eds). 953–977.

Tweedie, F., and Baayen, R. (1998). How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities*, *32*(5), 323–352.

Uccelli, P., Dobbs, C. L., and Scott, J. (2013). Mastering Academic Language: Organization and Stance in the Persuasive Writing of High School Students. *Written Communication*, *30*, 36–62.

Underwood, C., Schmitt, N., and Calpin, A. (2004). The eyes have it : An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition , Processing and Use* (pp. 153–172). Philadelphia, P.A: John Benjamins.

Ure, J. (1971) Lexical density and register differentiation. In G.E Perren and J.L.M. Trim (eds.) *Applications of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics, Cambridge, 1969*. Cambridge: Cambridge University Press. pp.443-452.

Van Geert, P. (2007). Dynamic systems in second language learning: Some general methodological reflections. *Bilingualism: Language and Cognition*, *10*(01), 47.

Van Geert, P., and Steenbeek, H. (2005). Explaining after by before: Basic aspects of a dynamic systems approach to the study of development. *Developmental Review*, *25*(3-4), 408–442.

Van Gelderen, A., Oostdam, R., and van Schooten, E. (2010). Does

foreign language writing benefit from increased lexical fluency? Evidence from a classroom experiment. *Language Learning*, *61*(1), 281–321.

Van Teijlingen, E., and Hundley, V. (2002). The importance of pilot studies. *Nursing Standard*, *16*(40), 33–6.

Van Zeeland, H., and Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*(3), 609–624.

Vandergrift, L. (2002). It is nice to see that our predictions were right: Developing metacognition in L2 listening comprehension. *The Canadian Modern Language Review*, *58*(4), 555–575.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, *22*(2), 217–234.

Verspoor, M. (2012). Symposium: Dynamic systems/Complexity theory as a new approach to second language development. *Language Teaching*, *45*(04), 533–534.

Verspoor, M. H., and Schmitt, N. (2012). Language and the lexicon in SLA. In P. Robinson (Ed.), *The Routledge Encyclopedia of Second Language Acquisition.* (pp. 353 - 360). London: Routledge.

Victori, M. (1999). An analysis of writing knowledge in EFL composing: a case study of two effective and two less effective writers. *System*, *27*(4), 537–555.

Waring, R. (1997). A study of receptive and productive vocabulary. *Studies in Foreign Language and Literature*, *21* (1), 94–114.

Webb, S. (2008). Receptive and Productive Vocabulary Sizes of L2 Learners. *Studies in Second Language Acquisition*, *30*(01), 79–95.

Webb, S.A. and Sasao, Y. (2013). New directions in vocabulary testing. RELC Journal, 44(3), pp.263-277.

Weigle, S. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.

Wenden, A. (1998). Metacognitive knowledge and language learning. *Applied Linguistics*, *19*(4), 515–537.

Weir, C. (2005). Limitations of the Common European Framework for developing comparable ex-aminations and tests. Language Testing 22, 281–300

Wesche, M., and Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *Canadian Modern Language Review*, *53*(1), 13–40.

West, M. (1953). *A General Service List of English Words*. London: Green and Co. Ltd, Longman.

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing*, *21*(4), 321–331.

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. (2002). Assessing proficiency through word associations: is there still hope? *System*, *30*(3), 315–329.

Wolter, B. (2006). Lexical Network structures and L2 vocabulary acquisition: the role of L1 lexical/conceptual knowledge. *Applied Linguistics*, *27*(4), 741–747.

Wray, A. (2000). Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics*, *21*(4), 463–489.

Wray, A., and Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language and Communication*, *20*(1), 1–28.

Young, R. (1991). *Variation in interlanguage morphology*. New York: Peter Lang.

Yu, G. (2010). Lexical diversity in writing and speaking task performances. Applied linguistics, 31(2), pp.236-259.

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., and Wolter, B. (2012). The "promise" of three methods of word association analysis to L2 lexical research. *Second Language Research*, 28 (1), 41-67.

Van Zeeland, H. and Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. System, 41(3), pp.609-624.

Zhang, L. F. (2009). From conceptions of effective teachers to styles of teaching: Implications for Higher Education. *Learning and Individual Differences*, *19*(1), 113–118.

Zhang, X., and Lu, X. (2013). A longitudinal study of receptive vocabulary breadth knowledge growth and vocabulary fluency development. *Applied Linguistics*, 35(3), 283-304

Zhou, A. (2009). What adult ESL learners say about improving grammar and vocabulary in their writing for academic purposes. *Language Awareness*, *18*(1), 31–46.

APPENDICES

Appendix 1 Marking Guide

Task fulfilment	Relevance to given context; balance of argument, length, ability to answer the question; appropriacy of style; comprehensibility. *20%*
Organisation **of text and content**	Text organization; paragraph organization; cohesion, coherence; Control of cohesive devices. *20%*
Language control	Accuracy of spellings and grammar; sentence structure, depth of grammar; range and appropriacy of vocabulary used; spellings, punctuation and script layout appropriate for writing context *60%*

Appendix 2 Semi Structured Interview and Stimulated Reconstruction Guide

Interviewee…….

Interview guide	Researcher's Notes
Section 1 **Background information about interviews** i. Why are you learning English/what do you hope to achieve? *e.g. personal satisfaction, circumstances etc.* ii. Have you ever been taught other subjects in English? iii. Did you complete any qualifications in your own country? Was this in your own language or in English? iv. Do you have family/friends in the UK?	
Section 2 **Metacognitive questions related to size and depth** i. Which test did you find easiest/most difficult? Why? ii. The three tests completed were measures of **how many** words students know and **how much knowledge** students have about the words. Do you think it's important to measure both? Why or why not? iii. The three tests completed were measures of **how many** words you know and **how much knowledge** you have about the words (Explain the concept of vocabulary size and depth and check understanding before proceeding to ask the question) Which one do you think is more important for : • Writing skills • Reading skills • Speaking skills • Listening skills iv. What sort of things do you think you could do if you knew **more words**? v. What if you knew the words that you already know, **better?** How could that help you?	

Section 3 **Activities that learners engage in** i. How much **independent language work** do you do at home? – *probe for speaking, listening, reading, writing* ii. *Probe for details of what they /write/listen to and/or how they **practise** the four skills, if at all.*	
Section 4 **Questions directly related to WAT** i. Pick out words **incorrectly answered** on the WAT test ii. Probe for **rationale/influences/inspirations** behind choice of associations.	
Any other comments?	

Appendix 3 Word Association Tests (WAT)

Word Associates Test - 40 items - choose four per set (both boxes)

1. **beautiful**

☐ enjoyable ☐ expensive ☐ free ☐ loud	☐ education ☐ face ☐ music ☐ weather

2. **bright**

☐ clever ☐ famous ☐ happy ☐ shining	☐ colour ☐ hand ☐ poem ☐ taste

3. **calm**

☐ open ☐ quiet ☐ smooth ☐ tired	☐ cloth ☐ day ☐ light ☐ person

4. **natural**

☐ expected ☐ helpful ☐ real ☐ short	☐ foods ☐ neighbours ☐ parents ☐ songs

5. **fresh**

☐ another ☐ cool ☐ easy ☐ raw	☐ cotton ☐ heat ☐ language ☐ water

6. **general**

☐ closed ☐ different ☐ usual ☐ whole	☐ country ☐ idea ☐ reader ☐ street

7. **bare**

☐ empty ☐ heavy ☐ uncovered ☐ useful	☐ cupboard ☐ feet ☐ school ☐ tool

8. **acute**

☐ hidden ☐ often ☐ rich ☐ sharp	☐ angle ☐ hearing ☐ illness ☐ stones

9. **common**

☐ complete ☐ light ☐ ordinary ☐ shared	☐ boundary ☐ circle ☐ name ☐ party

10. **complex**

☐ angry ☐ difficult ☐ necessary ☐ sudden	☐ argument ☐ passengers ☐ patterns ☐ problem

11. **broad**

☐ full ☐ moving ☐ quiet ☐ wide	☐ night ☐ river ☐ shoulders ☐ smile

12. **conscious**

☐ awake ☐ healthy ☐ knowing ☐ laughing	☐ face ☐ decision ☐ effort ☐ student

13. **convenient**

☐ easy ☐ fresh ☐ near ☐ suitable	☐ experience ☐ sound ☐ time ☐ vegetable

14. **dense**

☐ crowded ☐ hot ☐ noisy ☐ thick	☐ forest ☐ handle ☐ smoke ☐ weather

15. **curious**

☐ helpful ☐ interested ☐ missing ☐ strange	☐ accident ☐ child ☐ computer ☐ steel

16. **distinct**

☐ clear ☐ famous ☐ separate ☐ true	☐ advantage ☐ meanings ☐ news ☐ parents

17. **dull**

☐ cloudy ☐ loud ☐ nice ☐ secret	☐ colour ☐ knife ☐ place ☐ rock

18. **direct**

☐ honest ☐ main ☐ straight ☐ wide	☐ fence ☐ flight ☐ heat ☐ river

19. **favorable**

☐ helpful ☐ legal ☐ possible ☐ positive	☐ habit ☐ response ☐ teacher ☐ weather

20. **secure**

☐ confident ☐ enjoyable ☐ fixed ☐ safe	☐ game ☐ job ☐ meal ☐ visitor

21. **tight**

☐ close ☐ rough ☐ uncomfortable ☐ wet	☐ bend ☐ pants ☐ surface ☐ wood

22. **violent**

☐ expected ☐ smelly ☐ strong ☐ unlucky	☐ anger ☐ death ☐ rubbish ☐ storm

23. **chronic**

☐ continuing ☐ local ☐ serious ☐ unplanned	☐ accident ☐ examination ☐ illness ☐ shortage

24. **compact**

☐ effective ☐ small ☐ solid ☐ use-ful	☐ group ☐ kitchen ☐ medicine ☐ string

25. **crude**

☐ clever ☐ fair ☐ rough ☐ valuable	☐ behaviour ☐ drawing ☐ oil ☐ trade

26. **domestic**

☐ home ☐ national ☐ regular ☐ smooth	☐ animal ☐ movement ☐ policy ☐ speed

27. **profound**

☐ bright ☐ deep ☐ exact ☐ great	☐ effect ☐ machine ☐ taste ☐ thought

28. **fertile**

☐ dark ☐ growing ☐ private ☐ special	☐ business ☐ egg ☐ mind ☐ soil

29. **formal**

☐ fast ☐ loud ☐ organised ☐ serious	☐ bomb ☐ education ☐ growth ☐ statement

30. **independent**

☐ changed ☐ equal ☐ important ☐ separate	☐ child ☐ country ☐ ideas ☐ prices

31. **original**

☐ careful ☐ closed ☐ first ☐ proud	☐ condition ☐ mind ☐ plan ☐ sister

32. **sensitive**

☐ feeling ☐ interesting ☐ sharp ☐ thick	☐ clothes ☐ instrument ☐ skin ☐ topic

33. **professional**

☐ paid ☐ public ☐ regular ☐ religious	☐ advice ☐ manner ☐ musician ☐ transport

34. **critical**

☐ clear ☐ dangerous ☐ important ☐ rough	☐ festival ☐ illness ☐ time ☐ water

35. **synthetic**

☐ artificial ☐ electronic ☐ expensive ☐ simple	☐ drug ☐ meal ☐ radio ☐ sound

36. **liberal**

☐ free ☐ moderate ☐ plenty ☐ valuable	☐ crops ☐ furniture ☐ parents ☐ transport

37. **dramatic**

☐ exciting ☐ official ☐ surprising ☐ worried	☐ adventure ☐ change ☐ patient ☐ salary

38. **conservative**

☐ cautious ☐ hopeful ☐ traditional ☐ warm	☐ clothes ☐ estimate ☐ meeting ☐ signal

39. **coherent**

☐ clear ☐ normal ☐ recent ☐ together	☐ crime ☐ health ☐ speech ☐ theory

40. **ample**

☐ heavy ☐ large ☐ plentiful ☐ windy	☐ amount ☐ climate ☐ feelings ☐ time

Appendix 4 Vocabulary Levels Test

Your name .....

Instructions

This is a vocabulary test. Please choose the right word to go with each meaning. Write the number of that word next to its meaning.

Here is an example

1. business		
2. clock		part of a house
3. horse		animal with four legs
4. pencil		something used for writing
5. shoe		
6. wall		

You answer it the following way.

1. business		
2. clock	6	part of a house
3. horse	3	animal with four legs
4. pencil	4	something used for writing
5. shoe		
6. wall		

Some words are in the test to make it more difficult. You do not have to find a meaning for those words. In the example above, these words are *business, clock, shoe*.

2000 level

1. original			1. apply		
2. private		complete	2. elect		choose by voting
3. royal		first	3. jump		become like water
4. slow		not public	4. manufacture		make
5. sorry			5. melt		
6. total			6. threaten		

1. blame			1. accident		
2. hide		keep away from sight	2. choice		having a high opinion of yourself
3. hit		have a bad effect on something	3. debt		something you must pay
4. invite		ask	4. fortune		loud, deep sound
5. pour			5. pride		
6. spoil			6. roar		

1. basket			1. birth		
2. crop		money paid regularly for doing a job	2. dust		being born
3. flesh		heat	3. operation		game
4. salary		meat	4. row		winning
5. temperature			5. sport		
6. thread			6. victory		

3000 Level

1. administration			1. bench		
2. angel		managing business and affairs	2. charity		part of a country
3. front		spirit who serves God	3. fort		help to the poor
4. herd		group of animals	4. jar		long seat
5. mate			5. mirror		
6. pond			6. province		

1. coach			1. marble		
2. darling		a thin, flat piece cut from something	2. palm		inner surface of your han
3. echo		person who is loved very much	3. ridge		excited feeling
4. interior		sound reflected back to you	4. scheme		plan
5. opera			5. statue		
6. slice			6. thrill		

1. discharge			1. annual		
2. encounter		use pictures or examples to show the meaning	2. blank		happening once a year
3. illustrate		meet	3. brilliant		certain
4. knit		throw up into the air	4. concealed		wild
5. prevail			5. definite		
6. toss			6. savage		

5000 Level

1. alcohol		1. circus	
2. apron	cloth worn in front to protect your clothes	2. jungle	speech given by a priest in a church
3. lure	stage of development	3. nomination	seat without a back or arms
4. mess	state of untidiness or dirtiness	4. sermon	musical instrument
5. phase		5. stool	
6. plank		6. trumpet	

1. apparatus		1. bruise	
2. compliment	set of instruments or machinery	2. exile	agreement using property as security for a debt
3. revenue	money received by the government	3. ledge	narrow shelf
4. scrap	expression of admiration	4. mortgage	dark place on your body caused by hitting
5. tile		5. shovel	
6. ward		6. switch	

1. blend		1. desolate	
2. devise	hold tightly in your arms	2. fragrant	good for your health
3. embroider	plan or invent	3. gloomy	sweet-smelling
4. hug	mix	4. profound	dark or sad
5. imply		5. radical	
6. paste		6. wholesome	

University Word List

1. affluence
2. axis
3. episode
4. innovation
5. precision
6. tissue

- introduction of a new thing
- one event in a series
- wealth

1. deficiency
2. magnitude
3. oscillation
4. prestige
5. sanction
6. specifition

- swinging from side to side
- respect
- lack

1. configuration
2. discourse
3. hypothesis
4. intersection
5. partisan
6. propensity

- shape
- speech
- theory

1. anonymous
2. indigenous
3. maternal
4. minimum
5. nutrient
6. modification

- without the writer's name
- least possible amount
- native

1. elementary
2. negative
3. static
4. random
5. reluctant
6. ultimate

- of the beginning stage
- not moving or changing
- final, furthest

1. coincide
2. coordinate
3. expel
4. frustrate
5. supplement
6. transfer

- prevent people from doing something they want to do
- add to
- send out by force

Appendix 5 Free Writing Activity

Instructions

1. Please read through the topics below and choose the one that you feel most comfortable writing about.
2. Now write about 250words about your chosen topic.
3. Please remember to write your name and the number of the topic that you have chosen in the box provided at the bottom of the page.

1. "Money is the cause of all evil". Do you agree?
2. The prison system is old fashioned. No society should punish its criminals by sending them to prison. It should help them become good citizens. What are your views about this statement?
3. The car is man's worst friend. Do you agree?
4. Some people prefer to live in a small town. Others prefer to live in a big city. Which place would you prefer to live in? Use specific reasons and examples to support your answer.
5. "When people succeed, it is because of hard work. Luck has nothing to do with success." Do you agree or disagree? Use specific reasons and examples to explain your position.
6. A woman's place is in the home. Do you agree?
7. Some people say Christmas is the best time of the year. What are your views?
8. Television commercials should be banned. Do you agree with this statement?
9. Children under the age of five years should not be allowed to watch television. Do you agree?

YOUR NAME:.....

TOPIC NUMBER:.....

This is the end of the activity, thank you very much for taking part!

Appendix 6 Information For Participants

Information for participants

You are being invited to take part in a research study. Before you decide whether to take part or not, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish.

What is the purpose of research?

The research is part of a PhD project on vocabulary learning so the information gained from you will help the researcher to understand English Language teaching and learning better.

Do I have to take part?

It is up to you to decide whether or not to take part but there is no obligation for you to do so. However, your participation will be very much valued as it will contribute to the success of this research project.

What information will I get before the study?
The researcher will describe the study to you. They will go through this information sheet and answer any questions you may have. If you decide to take part you will be asked to sign a consent form to show you have agreed to take part. However, you are free to change your mind at any time and you do not have to give a reason. If you decide to withdraw from the study you can also decide whether or not you want your data to be used.

What will happen to me if I take part?
If you are interested in taking part your teacher will tell you the day, time and venue for the tests. You will complete the tests which will take about 1hour 30 minutes. You will be given the scores for the tests and you will get the opportunity for one-to-one feedback about your writing if you wish. If you agree, part of these sessions will be recorded so that the researcher can have an accurate record of what you said. The recordings may be transcribed but any personal information such as your name, your date of birth, your address etc. will be removed. Any quotations used will not contain information that will allow you to be identified.

What are the possible benefits of taking part?
You may find the one-to-one feedback sessions useful for further developing your language skills but the scores on the test will not affect your course at all. All the activities will be done purely for purposes of the research.

I have more questions, who can I ask?
Please contact Sihle Ndlovu on telephone 07940421712 or email **sn173@le.ac.uk** .
Your English Language tutor is also happy for you to contact him with questions on how the research activities will be scheduled or any other relevant questions.

PARTICIPANT CONSENT FORM

Statement of Consent

1. I have received a briefing about the purpose of this research.
2. I understand that my participation is voluntary and that I can withdraw unconditionally at any time from taking part in this study.
3. I have been informed that data obtained from me, either through written exercises, one-to-one or group discussions, may be used as part of this research. However, all the information will be treated as highly confidential and my name or the names of other participants will not be mentioned on the research.
4. The overall findings may be submitted for publication in an academic source such as a journal, or presented at conferences.
5. All written work will be marked and returned to the students.
6. In addition, one-to-one feedback will be provided to those who would like to discuss their scores and get advice on how they can improve their writing.

If you would like a one-to-one feedback session, please tick ALL the times when you can be available to receive this feedback.

Day	Time		
Monday	9:00-10:00am ☐	1:00pm-2:00pm ☐	5:00-6:00pm ☐
Tuesday	11:00-12:00 ☐	1:00pm-2:00pm ☐	5:00-6:00pm ☐
Wednesday		1:00pm-2:00pm ☐	5:00-6:00pm ☐
Thursday	9:00-10:00am ☐	1:00pm-2:00pm ☐	5:00-6:00pm ☐
Please tick this box if you **DO NOT** require one-to-one feedback ☐			

7. I agree to take part in this research.

Student Name: _____ **Signature:** _____ **Date**: _____

Researcher: *Sihle Ndlovu*
PhD Student
School of Education
University of Leicester
Telephone 07940421712
Email sn173@le.ac.uk

Appendix 7 Lumley's Scale

Writing performance criteria and descriptors (Lumley, 2002)

	0	1	2	3	4	5
Task fulfilment and appropriacy (TFA)	No comprehensible English words. (Copied text should not be assessed.)	Text is entirely inappropriate to given context or predominantly incomprehensible although a few words or sentences are present.	Text relates poorly to given context and is only sporadically appropriate or comprehensible. Some appropriate vocabulary within restricted range.	Text relates in part to given context although there are considerable errors. Appropriate vocabulary used although there are considerable errors.	Text relates generally to given context with a few confusions of meaning. Vocabulary choices are generally effective although there are some inappropriacies.	Text relates well to given context. It is thoroughly appropriate and easily understood. Vocabulary choices are appropriate and effective.
Conventions of presentation (CoP)	Absence of presentation conventions (spelling, punctuation, script, layout).	Very poor command of conventions of presentation (spelling, punctuation, script, layout) or sentences are present.	Uncertain command of conventions of presentation (spelling, punctuation, script, layout).	Adequate command of conventions of presentation, with some inconsistencies.	Generally good command of conventions of presentation although one area (spelling, punctuation, script or layout) may be weaker.	All aspects of presentation conventions (spelling, punctuation, script or layout) are handled skilfully.
Cohesion and organization (C&O)	Neither cohesion nor organization.	Very disjointed with minimal organization.	Limited control of simple cohesive devices, little awareness of appropriate organization of ideas relevant to this task.	Simple cohesion is controlled but problems of over use or inappropriate choices occur; there is some awareness of appropriate organization of ideas relevant to this task.	Generally cohesive, though some problems may be noticed in this area; organization of ideas is mainly effective.	Text is cohesive and organization is clear and appropriate to task.
Grammatical control (GC)	No grammar is evident.	Poor control of grammatical structures within this context.	Some control of grammatical structures suitable for this context but errors dominate.	Fair control of grammatical structures within this context but with considerable errors.	Generally good control of grammatical structures suitable for this context with a few obtrusive errors.	Competent control of grammatical structures appropriate to the context with only unobtrusive errors.