University of Derby

An exploration of undergraduates' vocabulary size, academic word usage and academic achievement

Jennifer Marshall

Doctor of Education **2021**

Contents

List of Appendices

List of Figures and Tables

Figures

Tables

List of Abbreviations

AVL Academic Vocabulary List

AVST Academic Vocabulary Size Test

AWL Academic Word List

BAME Black, Asian and Minority Ethnic

BAWE British Academic Written English Corpus

BNC British National Corpus

CATSS Computer Adaptive Test of Size and Strength

COCA Corpus of Contemporary American English

DfE Department of Education

EAL English as Additional Language

EAP English for Academic Purposes

EFL English as a Foreign Language

ESL English as a Second Language

ESOL English for Speakers of Other Languages

ESP English for Specific Purposes

ESS Employers Skills Survey

GPC Grapheme–Phoneme Correspondence

GSL General Service List

HE Higher Education

K1 First 1,000 most frequent words in English

LD Lexical Density

LFP Lexical Frequency Profile

LO Lexical Originality

LS Lexical Sophistication

MWE Multi-word Expression

NAWL New Academic Word List

NEET Not in Education, Employment or Training

NGSL New General Service List

NZ New Zealand

PPVT Peabody Picture Vocabulary Test

PVLT Productive Vocabulary Levels Test

SAT Scholastic Aptitude Test

SATS Standard Assessment Tests

SLA Second Language Acquisition

SSP Systematic Synthetic Phonics

TTR Type–Token Ratio

UG Undergraduate Student

UK United Kingdom

USA United States of America

VLT Vocabulary Levels Test

WP Widening Participation Agenda

Abstract

The aim of this research was to explore undergraduates' receptive and productive knowledge of vocabulary and how these relate to academic achievement in order to improve educational outcomes for students. The research was divided into two phases. In the first phase, an estimate of UG students' receptive vocabulary knowledge was measured by administering a vocabulary size test devised by Goulden *et al* (1990) to 389 undergraduates in one HE institution. In Phase II, an estimate of UG students' productive vocabulary knowledge was measured by examining the lexical richness of 41 UG students' written work, totalling 369 assignments, using AntWordProfiler software. Both phases tested for correlations between vocabulary knowledge and academic achievement measured using degree classifications in Phase I and marks on assignments in Phase II.

Key findings from Phase I suggest undergraduates have an estimated vocabulary size of around 11,000 words. The study also found that students' vocabulary sizes change between Stages 1 and 2. It was hypothesised that vocabulary sizes would play a role in academic achievement as measured by degree classifications. However, no correlation between vocabulary sizes and academic achievement was found and there is no difference between the mean vocabulary size of students predicted a first [($\bar{x}$) = 11,521] and those who were predicted an upper second-class [($\bar{x}$) = 11,312]; lower second-class [($\bar{x}$) = 11,450] and a third class [($\bar{x}$) = 9,833] degree.

Key findings from Phase II indicate that most of the words that students use in their writing are low frequency or K1 Words (the first 1,000 most frequent words in English). However, the proportion of K1 words used decreases from Stage 1 to 3 and student writers are using more academic words from the New Academic Word List (NAWL) in Stage 2 as compared to Stages 1 and 3. It was also found that as students progress through their degree, they use more words from the New General Service List (NGSL) 2 and 3 in addition to the NAWL in Stage 2. Furthermore, students used more 'off-list' words in Stage 3 compared to Stages 1 and 2. This suggests that students do experience change in their academic vocabulary while at university.

No correlations between marks on students' assignments and the percentage of academic words used from the NAWL were found. However, correlations were found with the general

English word lists. A medium, negative correlation was found between marks on assignments and K1 words in Stage 1 and a small correlation in Stage 2. There is also a small positive correlation between the usage of K2 words used in student assignments and marks in Stage 1 but not in Stage 3. In terms of the NGSL 3, no correlations were found in Stages 1 and 3 in relation to marks but there is a small positive correlation in Stage 2.

This indicates that students achieving higher marks use fewer high frequency and more low frequency words in their assignments. Although these correlations are not strong enough predictors, they still could have some consequence for student achievement; it implies that students need to write with a degree of lexical richness to achieve higher marks.

By capturing UG students' vocabulary knowledge, this research contributes to our current understanding of students' vocabulary size, types of vocabulary needed for degree success and methods of measuring vocabulary to better inform practitioners in HE as well as those interested in vocabulary research.

Acknowledgements

I would like to take this opportunity to thank all the people who have helped me through this doctoral journey. First, I would like to thank my supervisors, Neil Radford, Barbara MacMahon and Melanie Pope for their endless feedback and encouragement throughout the project. Their support has been invaluable.

To my colleagues, particularly Vanessa Cottle, who have been there for me, always listening and ready to offer a kind word of praise or reassurance.

To my family, particularly Andy, Shelby and Ellie who have endured endless hours of me at the computer, your patience and love have been unwavering, steadfast and resolute. I dedicate my perseverance in completing this thesis to you. And finally, to my mom, a lifelong educator, who I admire greatly, you inspired me into teaching and instilled in me a love of learning.

Chapter 1: Introduction

1.1 Context for the study

> We are surrounded by a vast wealth of words and they profoundly affect our lives – words we use and receive, hear and speak. From the cradle to the dinner table, the classroom to the boardroom, our wealth of words can determine our status in life. (Quigley, 2018, p. 1)

The importance of knowing words cannot be understated. This chapter will establish the case for investigating the role of language, vocabulary specifically, in higher education. It will explain why there is a need to conduct research on undergraduate students' vocabulary sizes based on the importance of vocabulary in relation to employability, social mobility, and academic achievement. Finally, it will establish how this piece of research makes an original contribution to knowledge.

This piece of research began because of the researcher's experience working as a lecturer in Education with undergraduate students who were native speakers of English. It became apparent that many students at all stages of study had questions in sessions mainly about vocabulary rather than content. Students asked about the meaning of words such as *implication* and *domestic violence*. Firstly, it had to be established if there really was a language issue and if so to what extent. This raised other questions around whether a limited vocabulary affected students accessing knowledge, the curriculum and academic texts. The researcher further wanted to investigate how the consequences of a limited vocabulary might manifest themselves in terms of their written work and academic achievement. In an endeavour to explore these concerns to improve the situation, a research project was devised and carried out in two phases.

The first phase used a vocabulary test measuring receptive vocabulary (words that you understand when you hear or read them) devised by Goulden *et al* (1990) and also used by Treffers-Daller and Milton (2013). The test is discussed in Chapter 2. A main rationale for using it was to provide a comparison from the findings of this research to previous research, albeit in different contexts. The second phase examines students' productive vocabulary (words that you can recall and use in speech and writing) through investigating the language students use in their written

work (fully discussed in Chapter 2). A broader rationale for investigating vocabulary knowledge in higher education (HE) is determined in the next section.

1.2 Rationale

1.2.1 Vocabulary, communication and employability

Good language skills are essential in higher education for several reasons. Firstly, language is at the heart of communication. Communication skills consist of reading, writing, listening and speaking (Worth, 2009). Vocabulary is the component that connects all four of these skills (Alqahtani, 2015). Vocabulary is, thus, key to effective communication and communicative competence (Coady & Huckin, 1997; Alqahtani, 2015). Communicative competence consists of sub-competences: grammatical (ability to apply grammar rules, spelling, pronunciation and vocabulary), socio-linguistic (ability to use language in different social situations and modes such as informal/formal or academic) and strategic (the ability to negotiate meaning) (Canale & Swain, 1980). To negotiate meaning in the higher education (HE) context, knowledge of vocabulary is paramount for successful communication.

Communication is complex and the need to effectively communicate for a variety of purposes in different contexts is an essential graduate attribute (see University of Bradford, 2018; University of Hertfordshire, 2018; University of Sheffield, 2018). Effective communication skills are also sought by employers in graduate recruits (Prospects, 2017; Inside Careers, 2018) and employability is high on universities' agenda. Preparing students for the world of work after graduation is seen as an essential outcome of any degree (Hewitt, 2020). Fostering excellent communication skills in students is key to this preparation.

Worth (2009) reports that we spend 9% of our communication time in the workplace writing, 16% reading, 30% talking and 45% listening. The question remains whether universities are preparing students with these skills and the necessary vocabulary to communicate effectively. 'Employers have complained consistently for many years that universities do not equip students with the skills they need for the workplace' (Hewitt, 2020, p. 28). The DfE (2018) *Employers Skills Survey* (ESS) suggests this may be the case. The ESS surveyed over 87,000 employers on their skills challenges both within their existing workforces and when recruiting. In general, it found that the most common reported skill shortage was a lack of specialist knowledge needed to perform the role (DfE, 2018). They also reported that a lack of skills in reading and

understanding instructions and reports increased from 26% in 2015 to 33% in 2017 and a lack of writing skills increased from 25% in 2015 to 28% in 2017 (DfE, 2018).

The UK government commissioned a review of post-18 education and the subsequent report known as the Augar Report states, 'Many graduates are working in non-graduate jobs and some employers report dissatisfaction with graduate skills. Both higher technical and craft skills are in short supply with long-standing skills gaps in strategic sectors such as engineering, IT and digital' (Augar Report, 2019, p. 15). This corresponds to previous surveys conducted by The British Chambers of Commerce (2014) and YouGov (2013). Around 3,000 businesses in the UK participated in The British Chambers of Commerce's 2014 *Workforce Survey*. The survey revealed that a clear majority of them (92%) identified skills shortages in leadership and management, planning and organisation, languages, computer literacy and creativity. Furthermore, YouGov (2013) asked 635 employers about whether graduates were work ready. The survey found employers felt that 17% of graduates were not prepared for work at all and were lacking in basic attributes such as teamwork, communication, punctuality and the ability to cope under pressure (Nathan, 2013).

These surveys are useful for universities to identify potential gaps in the curriculum and think about how they are equipping young people for the world of work. However, it may be possible that they have a very narrow definition of what graduate skills consist of and what it means to be work ready. There are potentially other qualities that they have not identified which make an individual employable such as motivation, valuing people and being respectful. Nonetheless, LinkedIn (2019) surveyed 2,000 business leaders and found the top four soft skills employers are looking for are leadership, communication, collaboration and time management. This supports the earlier point that universities must prepare students for the world of work by helping to advance their communication skills.

Through an appraisal of the literature on research pertaining to the communication skills of undergraduates, most studies that assess these skills refer to medical education (e.g. Lee *et al*, 2018; Sanson-Fisher *et al*, 2018; Salamonson *et al*, 2019). However, a recent study in engineering focused on how game-based learning may enhance communication skills (Bodner & Clark, 2017). The rationale for that study was based on the idea '… that engineering graduates are not well versed in the areas of written and oral communication when they graduate and transition into industrial-

based positions' (Bodner & Clark, 2017, p. 24). This supports the ESS and YouGov studies mentioned previously and provides further evidence for the need to investigate vocabulary in order to improve the communication skills of undergraduates.

1.2.2 Education, vocabulary and social mobility

Having established that vocabulary is a key component of effective communication and therefore employability, the following section will argue that education and specifically learning vocabulary is vital for social mobility and reducing educational inequalities for those from disadvantaged backgrounds. Firstly, it is important to determine whether the education system in the UK is equitable. The UK has some of the lowest levels of social mobility amongst high income countries (Sutton Trust, 2014). In 2015/16, disadvantaged children (from low-income communities) were six times as likely to be recorded as not in education, employment or training (NEET) after taking their GCSEs as those from better-off families (6% compared with 1%) (Fair Education Alliance, 2018). Furthermore, young people from wealthy backgrounds are more than twice (2.1 times) as likely to go to university than their peers on Free School Meals (FSM) and young people from disadvantaged backgrounds are nearly ten times less likely to go to a top university (Fair Education Alliance, 2018). This is not good for social mobility and the literature suggests that education, particularly higher education, is a crucial factor in improving it (Haveman & Smeeding 2006; Universities UK, 2016).

One way of resolving the issue is through what is known as the widening participation (WP) agenda. Widening participation in higher education has been a major component of UK educational policy for several decades (see Kettley, 2007) and is concerned with increasing the opportunities and uptake in HE of individuals from disadvantaged groups (e.g. low socio-economic) and students who are traditionally underrepresented (e.g. the disabled). The majority of WP students attend post-1992 institutions (French, 2013).

A key question pertains to the challenges that WP may bring to those institutions. There are significant attainment gaps in higher education for Black, Asian and Minority Ethnic (BAME) students. Universities UK (2018a) identified that while there has been a 50% increase in the number of BAME undergraduates in England between 2007 and 2016, there is disparity in the attainment of students from BAME backgrounds once they get to university: 'While 78% of white students who

graduated last year ended up qualifying with a first or a 2:1, 66% of Asian students achieved the same, and just 53% of black students' (Universities UK, 2018a). The Equality Challenge Unit (ECU, 2018) believe there are serious implications for these students as many graduate-level jobs and post-graduate courses have a minimum entry requirement of a 2:1 degree or above.

One solution to the BAME attainment gap can be found in the decolonisation of the curriculum movement (Universities UK, 2019). Charles (2019) believes that Keele University's definition of decolonising the curriculum is the most comprehensive:

> Decolonization involves identifying colonial systems, structures and relationships, and working to challenge those systems. It is not 'integration' or simply the token inclusion of the intellectual achievements of non-white cultures. Rather, it involves a paradigm shift from a culture of exclusion and denial to the making of space for other political philosophies and knowledge systems. It's a culture shift to think more widely about why common knowledge is what it is, and in so doing adjusting cultural perceptions and power relations in real and significant ways. (Keele University, 2019)

At the heart of this definition lies the notion that the curriculum in HE is too white and is dominated by western attitudes, narratives and practices. Keele University (2019) believe that 'Black and Minority Ethnic (BME) students are the first to feel the effects of a white curriculum' and list a number of ways to decolonise the curriculum. However, none of these specifically mention language. Language and culture are intrinsically linked. Looking at the role of language in helping to achieve decolonisation might address some of the challenges universities face with the WP agenda which in turn could potentially narrow the BAME attainment gap.

There is a lack of literature on the role of language in decolonisation, most likely because the decolonisation of the curriculum is a fairly new movement in the UK that began around 2015 (Charles, 2019). However, prior to this, the University of Plymouth devised a list of issues affecting WP in HE and at the top of the list was: 'lack of familiarity with the structures, processes and vocabulary of further and higher education with regard to, for example, researching options, making applications, funding and student support' (University of Plymouth, 2008, p. 4). It appears from this students may be disadvantaged by a lack of language before they even enter HE.

Hirsch (2013), based on Grade 12 of the Scholastic Aptitude Test (SAT) scores, states there is a positive correlation between a student's vocabulary size and the

probability they will graduate from higher education with an accompanying higher level of income. Thus, in order to reduce economic inequality and increase social mobility educators could look at vocabulary sizes as a proxy for a variety of educational attainments and abilities – not just in communication skills but also general knowledge of science, history and the arts (Hirsh, 2013).

Quigley (2018, p. 2) believes '[b]y closing the vocabulary gaps for children in our classrooms with their peers, we can offer them the vital academic tools for school success, alongside the capability to communicate with confidence in the world beyond the school gates'. The UK government's primary ambition for improving social mobility through education is to close the vocabulary the gap in early years (DfE, 2017). The vocabulary gap is where disadvantaged children fall behind (in some cases a full year and a half) their more affluent peers in early language development contributing to long-lasting effects on their social mobility (DfE, 2017). Therefore, it can be argued that knowing words or vocabulary is key to improving social mobility.

> In the years I've devoted to literacy, I have learnt what I should have always known – that nothing matters more than words. Our vocabulary allows us to interpret the world, to express ourselves with greater clarity, to show confidence, insight and perceptiveness. Words lie at the heart of our quest to narrow gaps between the advantaged and disadvantaged, to address social mobility. (Barton, 2018 in Quigley, 2018, p. i)

Hart and Risley (1995) conducted a study analysing the language children were exposed to in low-income families as compared to their more economically advantaged peers. They found that children from advantaged families had heard 30 million more words than their disadvantaged peers. The consequences were that by the age of 4, the spoken vocabulary sizes of children from disadvantaged families were much smaller. In the USA, their research triggered a number of initiatives in schools to try and reduce this gap (Golinkoff *et al*, 2019, Sperry, Miller & Sperry, 2019).

Hart and Risley's (1995) research highlighted the importance of language in lifelong success (Golinkoff *et al*, 2019). This notion is central to the UK government's action plan to improve social mobility by closing the vocabulary gap. The aim is to try and improve educational outcomes (attainment) for children from lower socio-economic households.

The vocabulary gap concept is not without its critics (Gorski, 2008; Blum & Riley, 2014; Alim & Paris, 2015; Sperry, Sperry & Miller, 2018). Much of the criticism centres around an underlying premise that children and their parents from poorer backgrounds are language deficient. This deficit approach not only apportions the blame on disadvantaged groups (especially parents for not talking enough to their children), but it also suggests that their language is inferior to more privileged peers (Alim & Paris, 2015). Moreover, recent research (e.g., Sperry, Sperry, & Miller, 2018) has shown that a vocabulary gap between rich and poor may not actually exist.

Further criticism of Hart and Risley's (1995) research stems from methodological issues. The main issue is they only recorded direct speech between parent and child, and not any indirect language that children may have taken part in (e.g., conversations between siblings) (Blum & Riley, 2014; Sperry, Sperry & Miller, 2018). Children learn language in a variety of ways and not exclusively by the number of words their parents speak to them (Blum & Riley, 2014). In a study analysing the language communication between parent and child in 60 low-income families, Hirsh-Pasek *et al* (2015) found that it is the quality of interaction between child and parent that matter more than the quantity of words children hear in terms of later language ability.

Despite these criticisms, Golinkoff *et al* (2019, p. 985) insist that 'denying the existence of the 30-million-word gap has serious consequences'. They argue the consequences associated with a vocabulary gap are differences in income, health care outcomes, high school completion rates, and employability. It could be more harmful to ignore the existence of a vocabulary gap than to accept it despite the criticism surrounding it. Given the potential impact language has on an individual's lifelong success, the notion of a vocabulary gap remains central to this research.

1.2.3 Widening participation, English language needs and academic literacy

Much of the research pertaining to widening participation (WP) deals with access, focusing on participation rates and barriers to entry rather than on students' experience in HE once they have arrived (Burke, 2017). There is very little research that addresses attainment gaps and the contributing factors such as language proficiency. However, a few studies in Australia have identified that the WP agenda has led to an increase in students with diverse educational, linguistic and cultural backgrounds and it is native English-speaking students that need English language

support as much as non-native English students (Dunworth, 2010; Murray 2012, 2013; Moore & Harrington, 2016). Murray (2013, p. 300) best summarises the issue by stating that English language support provision needs to include the 'non-traditional' native speakers of English whose language is at a mismatch with HE and 'not necessarily in keeping with the expectations of the academy, or indeed the workplace post-graduation'.

There is in fact, very little research on the language needs of native English-speaking students in a UK context – a gap this study aims to partially fill. Matthews *et al* (2016) report that several universities in Australia and New Zealand have implemented diagnostic testing of students' English language proficiency and academic literacy skills in order to help identify students at risk. This practice does not widely exist in UK higher education institutions.

Whilst language proficiency and academic literacy are often conflated, Murray (2012, pp. 235-236) distinguishes between the two: 'Proficiency can be defined as a language user's control of the formal and functional properties of language such that they are able to express and understand meaning accurately, fluently and appropriately according to context'. In other words, students need to have knowledge of grammar, syntax and vocabulary as well as reading, writing, listening and communication skills.

Leading on from language proficiency is academic literacy, which Matthews *et al* (2016, p. 445) define as '… an ability to understand and work with the ideas, languages and artefacts that make up the communication practices within educational institutions'. Murray (2012, p. 236) points out that academic literacy is not just a set of generic skills but is embedded in subject disciplines and 'recognises the specialised vocabularies, concepts and knowledges associated with particular disciplines, as well as distinctive patterns of meaning-making activity (genres, rhetorical structures, argument formulations, narrative devices etc.) and ways of contesting meaning'.

Academic literacy is important in higher education. Preece and Godfrey (2004, p. 6) contend that '… expertise in academic literacy practices is crucial to student success in higher education, and that these practices are fundamental for "graduateness" and the creation and communication of knowledge'. Furthermore, Ehren *et al* (2004) argue that advanced literacy skills are needed for young people to be productive

citizens and engage in the wider society. Academic literacy is essential for students both during their degree and after as well.

Weideman (2003, 2006, 2007, 2011; Van Dyk & Weideman, 2004; Van der Slik & Weideman, 2005, 2007, 2008; Weideman *et al*, 2016) has spent over a decade researching academic literacy and how to measure it. Weideman and Van Dyk (2014, p. ii) believe '[t]he ability to use language to meet the demands of tertiary education is called academic literacy'. Some of the components of academic literacy include: understanding a range of academic vocabulary in context; interpreting the use of metaphor and idiom in academic usage; perceived connotation; word play; and ambiguity (Weideman, 2007; Weideman & Van Dyk, 2014). Other components of academic literacy are concerned with elements of reading comprehension and writing. However, language plays the central role: 'It is clear then that the ability to use language for cognition and analysis becomes (or actually remains) critical as one's education progresses' (Weideman & Van Dyk, 2014, pp. ii). This interpretation is supported by Nagy and Townsend (2012, p. 103) who proclaim 'Theoretically, gains in academic vocabulary knowledge should inform gains in these larger realms of academic literacy'.

The concept that academic literacy is seen as *only* the ability to use language in HE is, arguably, somewhat limited. It can also be defined as 'being able to examine, refute, agree with, unpick or apply – to interact with the ideas of others and to do so in disciplinarily approved ways' (Stacey & Granville, 2009, p. 327) This links with Tardy (2005) who asserts that academic literacy goes beyond simply being able to read and write; for students to become active members in their field, they need to learn ways of thinking about, interacting with, and constructing knowledge specific to their disciplines. To think, interact and construct knowledge in a discipline, students need vocabulary specific to that subject. In other words, they need not only academic but subject specific vocabulary as well.

A key question is concerned with whether undergraduate students are academically literate. Ehren *et al* state that various studies have found that between '30 and 90% of students entering community college need remedial reading, writing, or mathematics courses' (2004, p. 683). Jameson's (2007) research shows that the number of high school graduates entering higher education with weak verbal abilities

and lacking in key literacy skills has substantially increased as a result of widening participation initiatives.

A recent study on the language used in assessment for widening participation students revealed that language was a barrier for these students: 'Findings highlighted important pragmatic and conceptual issues around the language of assessment' (Butcher *et al*, 2017, p. 27). More specifically, students reported they did not understand the language used in assessment guidance and they were unfamiliar with the 'type' of language experienced in HE. As a result, the researchers argued the language used for assessment should be more transparent and inclusive which could aid in retaining students from groups underrepresented in HE.

Vocabulary is important in both language proficiency and academic literacy for undergraduates *during* their degree course. Vocabulary knowledge as previously mentioned is important *before* students arrive in terms of researching degree options and making university applications. Vocabulary knowledge is important *after* they complete their degree as communication skills are essential to employability. The question remains as to what we know about the vocabulary knowledge of our students. If vocabulary is important in language proficiency, academic literacy and effective communication, but is a potential barrier to achievement as suggested by Butcher *et al* (2017), what kind of vocabulary do undergraduates have and need; how big a vocabulary is sufficient for study in HE and does this affect students' achievement? These questions will be explored in this research.

For general language use like reading a novel or newspaper, watching a movie, and taking part in a conversation, a person needs a receptive vocabulary size of 8,000–9,000 word families for comprehension of written text and a vocabulary of 6,000–7,000 for spoken text (transcribed speech) (Nation, 2006). Word families include inflected and derived forms of a word even if the part of speech is not the same (Bogaards & Laufer, 2004). For example, *add* is the base word but its family includes *added*, *adding*, *adds*, *addition*, *additional* and *additive*. Much of the research around receptive vocabulary sizes has been conducted in the field of English for Speakers of Other Languages (ESOL). For the purpose of this research, ESOL is used as an umbrella term to include English as a Foreign Language (EFL), English as an Additional Language (EAL), English as a Second Language (ESL), and English for Academic Purposes (EAP) (e.g. Laufer & Nation,1995; Lan & MacGregor, 2010;

Zhang, 2013; Kurniawan, 2017; Sudarman & Chinokul, 2018; Tan & Goh, 2018). Research on native speakers of English is summarised in Table 1.1.

Table 1.1 Various estimates of the number of English words known by adults

Study	**Estimated number of known words**	**Sample**
Hartmann (1946)	215,000	US undergraduates
Nusbaum *et al* (1984)	14,400	US undergraduates
Goulden *et al* (1990)	17,200	NZ undergraduates
D'Anna *et al* (1991)	17,000	US undergraduates
Anderson and Nagy (1993)	40,000	US high school seniors
Zechmeister *et al* (1995)	12,000	US undergraduates
Treffers-Daller and Milton (2013)	9,800	UK undergraduates

Source: adapted from Brysbaert *et al*, 2016

Treffers-Daller and Milton (2013) is the only study to date conducted on UK undergraduates. They assessed the vocabulary size of 178 undergraduate students from three British universities (City University, Swansea University and the University of the West of England, Bristol). While this one study gives an insight into UG vocabulary sizes in a UK context, the sample was small and the participants were only selected from degree courses in Humanities at Swansea and UWE and a degree course in Speech and Language Therapy at City University. Thus, to add to this knowledge and given the importance of vocabulary knowledge before, during and after UG study, the first phase of this doctoral research sought to investigate the vocabulary sizes of UG students in a widening participation University in the East Midlands. The aim was to recognise how vocabulary might impact on students' achievement. The research set out to use a randomly selected, bigger sample compared to Treffers-Daller and Milton's (2013).

1.3 Original contribution to knowledge

This study makes an original contribution in the following ways. It contributes to existing research around receptive vocabulary sizes of UK undergraduates in Phase I. In Phase II, the productive vocabulary knowledge of native English speaking undergraduate students is investigated using Lexical Frequency Profiling or LFP.

Productive vocabulary knowledge is usually measured in two ways: controlled productive knowledge using Productive Vocabulary Levels Test (PVLT); or free productive knowledge where learners write an essay and the lexical richness of their writing is measured (Nizonkiza, 2016). Measuring lexical richness involves quantifying the degree to which a writer is using a varied and large vocabulary (Laufer & Nation, 1995). Laufer and Nation (1995) developed what is known as the Lexical Frequency Profile (LFP) which shows the percentage of words a learner uses at different vocabulary frequency levels in their writing. Studies using LFP mainly focus on ESOL students rather than native speakers (e.g., Meara & Fitzpatrick, 2000; Muncie, 2002; Morris & Cobb, 2004; Kojima & Yamashita, 2014; Lutviana *et al,* 2015). The use of LFP for a native speaker study makes this research unique.

In Phase II, software called AntWordProfiler was used to measure the lexical richness of undergraduates' written work from a cohort of students from an Education Studies Programme at all stages of study. It divides the text into five categories by frequency bands starting with the high frequency vocabulary. The first 3,000 most frequent words of English are derived from the New General Service List (NGSL) produced by Browne *et al* (2013) taken from a corpus of written English. The first band is the most frequent 1,000 words of English (for example, words such as *a, and, the, one, with, or*). The second band is the second most frequent 1,000 words of English, i.e. 1,001 to 2,000 (for example, words such as *empty, spelling, paste*). The third band consists of words from 2,001 to 3,000. Academic words from the New Academic Word List or NAWL are also identified. The NAWL contains 960 words that are the most frequent words found in academic texts across subjects and will be discussed in Chapter 2 (for example, words such as *affirm, rhetoric, sediment*). Finally, the remaining words, which are not found on the other lists, are labelled as 'off-list'.

Vocabulary profiling software such as AntWordProfiler was designed with second language learners in mind to help ascertain the vocabulary level and complexity of texts (Anthony, 2021). The software is mainly used to find out the difficulty level of texts and the vocabulary needed to understand those texts. In this study, conversely, it was used to analyse student's written work and their level of vocabulary rather than texts they need to read. Analysing data in this way contributes to the originality of the research. Other areas of original contribution were revealed in the findings around changes in vocabulary sizes, the use of words from the NGSL and NAWL in student writing, and links between vocabulary and academic achievement.

1.4 Summary

To summarise, vocabulary is important for UG students for many reasons; they need to understand the language of higher education when making decisions about where and what to study. Vocabulary is key to effective communication, which is a pivotal graduate employability skill. Vocabulary is vital for social mobility and reducing both educational and economic inequalities by closing the word gap between the advantaged and disadvantaged (including those students traditionally underrepresented in HE). While it is agreed that vocabulary plays a key role in academic literacy in HE, there is less known about what vocabulary knowledge students possess during their degree course. This study makes an original contribution to the field by investigating undergraduates' receptive and productive vocabulary knowledge. Lastly, it seeks to establish whether there are any correlations between this knowledge and academic achievement.

Chapter 2: Literature Review

2.1 Introduction

This chapter establishes why it is important to investigate vocabulary in higher education in order to provide a context for the interpretation of the data findings. It aims not only to contextualise the research questions being investigated but also to highlight how this project might impact on professional practice in education (Burgess *et al*, 2006). To do this, it will explore what is already known in the field about the significance of vocabulary in reading and writing; different types of vocabulary; what it means to 'know' a word; how to measure vocabulary; and the variety of tools used to analyse texts. While it primarily focuses on previous research conducted in these areas, it will also highlight gaps that exist and how this project potentially fills those gaps. The literature review provides the rationale for the study's research questions presented at the end.

After reviewing the literature, five key themes emerged and the chapter is arranged around these themes:

- Vocabulary and reading
- Vocabulary and academic writing
- Types of vocabulary
- Vocabulary knowledge and measurement
- Vocabulary and academic achievement

2.2 Literature base

The five key themes emerged after examining the literature on vocabulary from multiple disciplines and fields within education but mainly from applied linguistics. Trafford and Leshem (2008) suggest that doctoral researchers must 'engage with the literature' which intimates at a sense of personal ownership and choice over the sources consulted. Therefore, it must be reasoned why it was necessary to draw mainly from the discipline of applied linguistics.

First, arguably, education is not a discipline but a field of study; it can be viewed to be multidisciplinary because it draws from disciplines such as history, sociology and politics (McMillan & Weyers, 2009; Walkup, 2011). In addition, vocabulary research

can be found in multiple disciplines such as psychology (e.g. Hartmann,1946; O'Connor & Solity, 2020), applied linguistics (e.g. Schmitt, 2010; Nation, 2013; Schmitt & Schmitt, 2014; Gablasova, 2015) and corpus linguistics (e.g. Cobb, 2010; Chen, 2018). Research on vocabulary is also conducted in the field of reading (e.g. Chall *et al,* 1990; Carver, 1994; Lesaux & Kieffer, 2010; Olinghouse & Wilson, 2013) (see Figure 2.1).

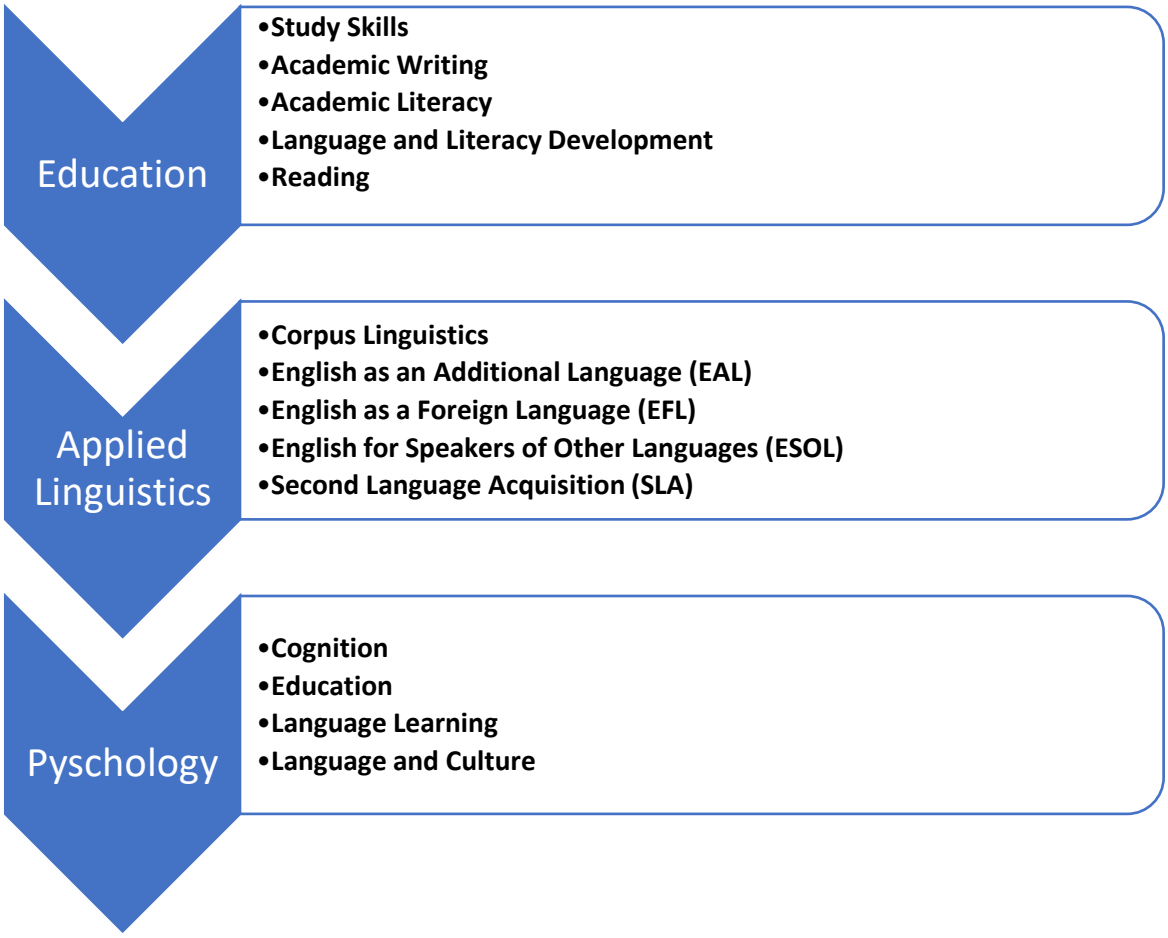


Figure 2.1 The literature around vocabulary

There are numerous attempts to define applied linguistics but there is no consensus as it means different things to different people (Davies, 2007). Hudson (2009) defines it as such:

> Applied linguistics (AL) provides the theoretical and descriptive foundations for the investigation and solution of language-related problems, especially those of language education (first-language, second-language and foreign language teaching and learning), but also problems of translation and interpretation, lexicography, forensic linguistics and (perhaps) clinical linguistics … The main distinguishing characteristic of AL is its concern with professional activities

> whose aim is to solve 'real-world' language-based problems, which means that research touches on a particularly wide range of issues – psychological, pedagogical, social, political and economic as well as linguistic. As a consequence, AL research tends to be interdisciplinary.

As outlined in Chapter 1, this research is borne out of a perceived 'real-word' vocabulary issue from the researcher's own professional practice, a key characteristic of AL according to the definition. It seeks to investigate this issue with first language UG students and to provide viable solutions in order to improve educational outcomes. It can be reasoned that AL offers the most scope and knowledge base to conduct this study. The first theme identified is concerned with reading in education with a particular focus on vocabulary.

2.3 Reading in compulsory education

The findings from a meta-analysis of 300 studies shows that successful early literacy development correlates with later literacy achievement (Shanahan & Lonigan, 2010). Learning to read and write involves all aspects of language structure and use: phonology, graphology, vocabulary and grammar (Crystal, 2020), but it is vocabulary that is perhaps the strongest determinant of reading success (Biemiller, 2003). Successful reading requires two things: being able to identify written words but also knowing what those words mean (Biemiller, 2012a). This section will focus on the role of vocabulary in reading from primary through to tertiary education.

In part, the English national curriculum aims to equip school age children to:

- read easily, fluently and with good understanding;
- develop the habit of reading widely and often, for both pleasure and information;
- acquire a wide vocabulary, an understanding of grammar and knowledge of linguistic conventions for reading, writing and spoken language (DfE, 2014a).

More specifically, in Key Stages 1 and 2 children are expected to be skilled in word reading, which involves both the speedy working out of the pronunciation of unfamiliar printed words (decoding) and the speedy recognition of familiar printed words (DfE, 2014a). Learning to read written texts is not the same as learning to understand them (Biemiller, 2012a). Therefore, pupils are expected to comprehend what they read. Government guidelines state,

> Good comprehension draws from linguistic knowledge (in particular of vocabulary and grammar) and on knowledge of the world. Comprehension skills develop through pupils' experience of high-quality discussion with the teacher, as well as from reading and discussing a range of stories, poems and non-fiction. All pupils must be encouraged to read widely across both fiction and non-fiction to develop their knowledge of themselves and the world they live in, to establish an appreciation and love of reading, and to gain knowledge across the curriculum. Reading widely and often increases pupils' vocabulary because they encounter words they would rarely hear or use in everyday speech. (DfE, 2014a)

The government guidelines demonstrate the importance of vocabulary and reading in the curriculum. Studies have shown a strong positive correlation between vocabulary knowledge and reading comprehension (Chall *et al*, 1990; Lesaux & Kieffer, 2010; Mancilla-Martinez & Lesaux, 2010). Biemiller (2012b) found that vocabulary is a weak predictor of First Grade (Year 2 in England) reading achievement in the United States (US), but it is a much stronger predictor of Fourth Grade (Year 5 in England) reading achievement. Paige *et al* (2019) found that around 66% of all Fourth-Grade elementary school students are unable to read proficiently (reading at less than adequate levels on national reading achievement tests) due to poor comprehension. This could be due, in part, to the 'vocabulary gap' (Quigley, 2018) as discussed in Chapter 1, and by the Seventh or Eighth Grade (Years 8 and 9), vocabulary is the main predictor of reading success (Lesaux & Kieffer, 2010).

In the UK, the situation is similar but not quite as acute as Paige *et al* (2019) suggest. Nationally, Standard Assessment Tests (SATS) are administered at the end of Key Stage 1 in May of Year 2 (age 7) and at the end of Key Stage 2 in May of Year 6 (age 11). In 2019, in the reading test, 73% of pupils reached the expected standard in the Year 6 SATs (DfE, 2019). However, 51% of disadvantaged pupils reached the expected standard in all of reading, writing and maths combined compared to 71% of all other pupils (DfE, 2019). This is significant as 30% of pupils at the end of Key Stage 2 were classed as disadvantaged (defined as those who were registered as eligible for free school meals at any point in the last six years, children looked after by a local authority or have left local authority care in England and Wales through adoption, a special guardianship order, a residence order or a child arrangement order – DfE, 2019).

The Picture Vocabulary Size Test is a receptive vocabulary size test aimed at young pre-literate native speakers up to eight years old as well as young non-native speakers of English (Anthony & Nation, 2017). Research using the test in three New Zealand schools on 8-year-old children indicates significant differences in vocabulary sizes. The children from the school located in a low socio-economic area were reported to have a mean vocabulary size of 4,540 words while those from a higher socio-economic area had a mean vocabulary size of 5,056 words, almost a full year ahead (Nation & Coxhead, 2021). Given that vocabulary knowledge has been identified as a strong predictor of reading comprehension, this puts children from a lower socio-economic background at a serious disadvantage in becoming successful readers.

The repercussions are considerable as the reading gap widens over time (DfE, 2015). In 2018, 120,000 disadvantaged students transitioned from primary to secondary school below the expected standard for reading and based on previous cohorts only 1 in 10 are expected to achieve passes in English and Maths at GCSE (EEF, 2019). It can be reasoned that the reading gap starts in primary school and notably, from the Third Grade (Year 4) on, the main limiting factor for the majority of children is vocabulary, not reading mechanics (decoding print into words) (Biemiller, 2012b).

The national curriculum stipulates that pupils need to be able to decode and recognise words in printed texts (DfE, 2014a). Therefore, the UK government has mandated the teaching of systematic synthetic phonics (SSP) in the early years to facilitate reading in children (DfE, 2015; Ofsted, 2019). As a result, many schools buy in phonics reading schemes (see DfE, 2014b).

> Reading schemes consist of a series of specially-written books that are sequenced and graded according to their level of difficulty. They are based on the premise that written English is largely irregular and difficult to learn, and therefore needs to be simplified for beginning readers. As a result schemes contain a restricted, carefully controlled vocabulary consisting of high-frequency and phonically regular words that are introduced gradually and repeated frequently. (Solity & Vousden, 2009, p. 470)

The books described in the quote above are often called decodable books. A main advantage of using these books is that they provide children with the opportunity to practise independently what they have learned in the classroom (Castles *et al*, 2018). One such reading scheme states: ‘Phonics reading books … are written so that

children can decode all the words using their phonic knowledge rather than using other clues or seeking help' (The Reading Chest, 2020). Children taught through SSP learn how to link phonemes (sounds) to graphemes (individual letters or groups of letters) and blend them together (synthesise) (National Literacy Trust, 2017).

However, there are a number of disadvantages to this approach. O'Connor and Solity (2020) identified two main problems with SSP reading schemes. Firstly, children will only be learning Grapheme–Phoneme Correspondence (GPC) that enables them to progress through the phonically decodable texts linked to the scheme, but that seldom occur in written English beyond the scheme. The other issue pertains to multiple mappings. Multiple mappings are where a grapheme represents two or more phonemes. For example, the word *coin* has nine different ways it can be pronounced. The letter 'c' can be /k/ as in *can* or /s/ as in *city*. The letter 'o' can be /ɒ/ as in *hot*, /ʌ/ as in love, /əʊ/ as in go. The letter 'i' can be pronounced /aɪ/ as in *time*, /I/ as *ship*, /i/ as in *taxi.* Finally, 'n' has one pronunciation as in /n/ *nice*. The only way children can know the correct pronunciation of the word *coin* is by being familiar with and knowing the meaning of the word.

A focus on decoding in primary schools rather than on word knowledge and comprehension has potentially contributed to a reading gap for some pupils as they reach secondary school (Solity & Vousden, 2009). At the end of compulsory education, this gap is acute for certain ethnic groups. 43.2% of all pupils in England got a 'strong pass' (grade 5 or above) in GCSE English and Maths in the 2018 to 2019 school year compared with 6% of White Gypsy and Roma pupils, 13.9% of Irish travellers; 26.5% of Black Caribbean pupils achieved a 'strong pass' (GovUK, 2020). This could be attributed to the 'Matthew effect' where greater initial knowledge, in this case vocabulary knowledge, contributes to greater learning (e.g. more vocabulary knowledge which in turns aids comprehension). More advanced learners make greater gains than those with less knowledge who are likely to fall further and further behind (Webb & Nation, 2017). There is much evidence in both the L1 and L2 literature to support this view (Stanovich, 1986; Duff *et al*, 2015; Webb & Chang, 2015).

Despite this gap, students in secondary school are expected to read and critically evaluate a range of classic literature and extended literary non-fiction, such as essays, reviews and journalism (DfE, 2014a). Even as early as primary school, 'students are expected to learn new information from content-area texts, so failure to

understand the academic language of those texts can be a serious obstacle in their accessing information' (Snow & Uccelli, 2009, p.112). There are large, individual differences in vocabulary knowledge when pupils enter primary school; this typically extends into children's schooling years (Duff *et al*, 2015). One solution, offered as a key recommendation from the Education Endowment Foundation (EEF, 2019), is to provide targeted vocabulary instructed in every subject in secondary schools. Students need to learn about the language–content relationship in each discipline area (e.g. Biology, Maths, Chemistry and so forth) in order to achieve (Schleppegrell & O'Hallaron, 2011).

To summarise, the literature reviewed in this section clearly demonstrates the importance that vocabulary knowledge plays in reading in compulsory education and student success. A wide range of vocabulary is needed in schooling to understand the variety of texts pupils encounter. A key question remains whether schools are preparing children to meet the demands of the required reading as they progress from primary to secondary and continue to higher education.

2.4 Reading in higher education

Vocabulary continues to play a key role in reading in the higher education context. Almost all activities in higher education involve some amount of reading (Gunobgunob-Mirasol, 2019). Students need reading skills to access the curriculum and engage in scholarly activity on multiple levels (writing, listening and speaking). Bharuthram (2012) makes the case that reading is one of the most important academic tasks encountered by students in higher education and thus necessitates a need to teach reading across the curriculum in higher education. This is because they are exposed to a number of texts and textbooks that require independent reading. 'At this level they are expected to comprehend what they read so that they can analyse, critique, evaluate and synthesize information from various sources' (Bharuthram, 2012, p. 205).

The importance of the role vocabulary plays in comprehension cannot be understated. In a textbook for practitioners in HE, Leamnson (1999) argues that students have major deficiencies in language use and understanding which affects their ability and motivation to read. The findings from a recent study involving 121 UG students (L1/L2 not specified) in the Philippines, found a link between reading motivation and vocabulary size suggesting that an increase in student motivation to read may result in an increase in vocabulary size (Gunobgunob-Mirasol, 2019).

Vocabulary is pivotal to comprehension, as previously argued, and is necessary before information can be analysed, evaluated and synthesised. You cannot analyse what you do not understand. Perhaps students are not motivated to read because they lack the necessary vocabulary to comprehend assigned texts on their courses and conduct independent reading.

Colombo and Prior (2016, p. 115) report that in Latin America students 'do not understand what they read and they cannot write properly'. Furthermore, they remark that there is a widespread concern in the Uruguayan higher education system in particular, pertaining to first year students who face challenges in identifying the main idea in readings and producing their own written work due to lack of vocabulary knowledge.

It must be acknowledged that vocabulary is not the *only* thing students need to be able to understand texts. In addition to vocabulary (e.g. Oslund *et al,* 2018), reading comprehension research focuses on other factors such as decoding , as previously discussed (e.g. Wang *et al*, 2018), working memory (e.g. Yeari, 2017), cognitive strategies (e.g. Oakley, 2011) and instructional procedures/strategies (e.g. Fasih *et al*, 2018).

Some studies have focused particularly on morphological awareness (e.g. Nagy *et al*, 2006) and syntactic awareness (Bowey, 1986), as well as metacognitive strategies (Nergis, 2013). Guo *et al* (2011) examined the relationship between vocabulary knowledge, morphological awareness, syntactic awareness and reading comprehension in English-speaking adults. They found that morphological and syntactic awareness contribute to reading comprehension through their impact on vocabulary knowledge, meaning higher levels of both are linked to greater vocabulary knowledge. Vocabulary knowledge made its own contribution to reading comprehension beyond those of morphologic awareness and syntactic awareness suggesting that it should play a key role in explaining individual differences in reading comprehension (Guo *et al*, 2011). In other words, this study indicates that vocabulary knowledge may be considered as the primary factor in influencing one's ability to read and understand texts.

Guo *et al*'s (2011) research is of particular interest because data was collected from 155 undergraduates and graduate students (all native speakers of English) enrolled in three different higher education institutions in the USA. The contextual focus was in higher education so the findings are relevant to this study. However, one of the

limitations of the study is the sample – the majority were female (72.9%) and white (76.1%) which indicates the findings are not entirely representative. The test used in the research also posed further limitations. For example, morphological awareness was measured using a modified Grammatical Application Wug Test by Berko (1958) and syntactic awareness was assessed by the Syntactic Awareness Questionnaire (SAQ) developed by Layton *et al* (1998). Both of these tests were originally designed to be taken by children. In fact, the authors acknowledge that the adults who participated in the study found most of the test items that measured syntactic awareness were relatively easy.

There is further evidence of the prime importance of vocabulary in reading comprehension in the literature. Davis (1944, p. 186) outlined nine skills basic to reading comprehension:

> 1) Knowledge of word meanings; 2) Ability to select the appropriate meaning for a word or phrase in the light of its particular contextual setting; 3) Ability to follow the organization of a passage and to identify antecedents and references in it; 4) Ability to select the main thought of a passage; 5) Ability to answer questions that are specifically answered in a passage; 6) Ability to answer questions that are answered in a passage but not in the words in which the question is asked; 7) Ability to draw inferences from a passage about its contents; 8) Ability to recognize the literary devices used in a passage and to determine its tone and mood; 9) Ability to determine a writer's purpose, intent, and point of view, i.e., to draw inferences about a writer.

Davis (1944) created a 240-item multiple-choice test to obtain the intercorrelations of scores of these nine skills. His sample was large (421) and taken from first year UG students, so relevant to this study. The research found that knowledge of word meanings correlated strongly with the eight other components (except for item 4 which was small) maintaining the notion that '… knowledge of word meanings is basic to the measurement of all the other skills, since to read at all one has to recognize words and understand their meanings …' (Davis, 1944, p.186). Moreover, Davis's (1942, 1944, 1968) work has been hugely influential as many other studies have used his tests for further research on reading comprehension (e.g. Thurstone, 1946; Spearritt, 1972; Smith & Zinc, 1977).

In higher education, reading is a valuable tool in evaluating information, problem solving and producing new or original work. Reading is a complex skill involving

multiple factors. Catts and Kamhi (2017) summarise the argument for a multidimensional view of reading comprehension that involves an interplay between the reader, the text and the task or purpose of reading. Firstly, '[t]he reader brings a set of cognitive-linguistic abilities, motivations, interests, and background knowledge to the task of reading' (Catts & Kamhi, 2017, p. 1). As a result of the widening participation agenda mentioned in Chapter 1 and the reading gap in Section 2.3, students enter higher education with a range of qualifications, for example: A-levels, B-Techs, International Baccalaureate Diploma Programme, and Access to HE Diploma. Therefore, students inevitably come to UG study with a range of knowledge and skill sets leaving some more prepared than others. Motivation will also vary (see Prowse & Delbridge, 2013) as well as interest in the subject they are studying (Skatova & Ferguson, 2014). If background knowledge, motivation and interest are low for the subject one is studying, it can be assumed that this may potentially have a negative impact on their ability to comprehend what they are reading which in turn can affect achievement.

Reading comprehension can also be affected by not only vocabulary but also prior knowledge of the content of a text (Stahl & Jacobson, 1986). 'Text refers to a set of variables that includes such factors as genre, topic, complexity, and form (e.g., electronic vs. paper), each of which imposes its own set of constraints/challenges on comprehension' (Catts & Kamhi, 2017, p.1). Ernawati and Ardi (2017) consider academic texts difficult to read. There are several academic reading genres where students must be familiar with the content of the texts (e.g., case studies, journal articles, argumentative essays, research reports and so forth). This familiarity will aid comprehension (Sunderland & Spiegel, 2009). From personal experience, students are often not familiar or lack confidence with the report genre and struggle when asked to write one. Genres will be further explored in Section 2.5.1.

Students not only need to be familiar with academic genres but also need to read with a purpose. 'Reading activity refers to the task or purpose of reading, which could include reading to learn, reading to identify a specific fact, reading to evaluate the strengths and weaknesses of an argument, or reading a narrative for pleasure' (Catts & Kamhi, 2017, p. 2). Anecdotally, some students simply will not engage with an assigned reading if they do not see a purpose to it. Moreover, comprehension is likely to be negatively affected if students read aimlessly without a goal and understanding of why that particular text is important (Hyland, 2006).

The position that you 'read' for a degree is widely accepted amongst the academic community. However, '[a] common source of frustration for college instructors is getting their students to read' (Hatteberg & Steffi, 2013, p. 346). Despite this, there has been very little research conducted on the reading habits of students in this context (Nadelson *et al*, 2013).

Research pertaining to the reading compliance (completing the assigned reading given in a course) of undergraduates has revealed that it is low and as a result negatively affects scholarly performance (Sharma *et al*, 2013). Cressman (2018) concluded that less than a third of students complete the assigned reading before coming to class. Brost and Bradley (2006, p. 101) believe that students are unprepared for reading in HE: 'The idea is that students don't read because they can't read; at least, they do not read well enough for the kind of texts most faculty assign'. A study conducted in the USA on the reading compliance of university students found that while 46% of students were reading compliant, only one out of two were able to demonstrate a basic level of comprehension of the assigned reading (Hoeft, 2012). The major limitation of Hoeft's study is that it did not explore intrinsic factors to do with reading comprehension. It simply looked at external drivers such as one's work schedule or family, placing emphasis on reading (or not placing emphasis on reading in the case of non-compliance). Even though this study is not exploring reading compliance as objective, reading is an integral part of UG study and examining internal drivers and motivating factors may shed light on why students are not engaged fully with it.

Furthermore, the act of reading interacts '… within a sociocultural context that involves factors such as where the reading occurs (school, home), how much support there is (individual or group activity), and what cultural value is placed on reading by students' family, peers, and racial-ethnic group' (Catts & Kamhi, 2017, p. 2). The living environment can have a positive and negative impact on reading attitudes (Morni & Sahari, 2013). Morni and Sahari (2013) found that parental influence was important in fostering good reading habits in undergraduates. In the home, parents who encouraged reading and exposed their children to a wide range of reading material had a positive effect. Some students, particularly those from disadvantaged backgrounds, may not have had this level of support before coming to university (DfE, 2017). Furthermore, the physical space where reading occurs can also have an impact on comprehension; student accommodation can often be noisy and not conducive to concentrating (The Student Room, 2017).

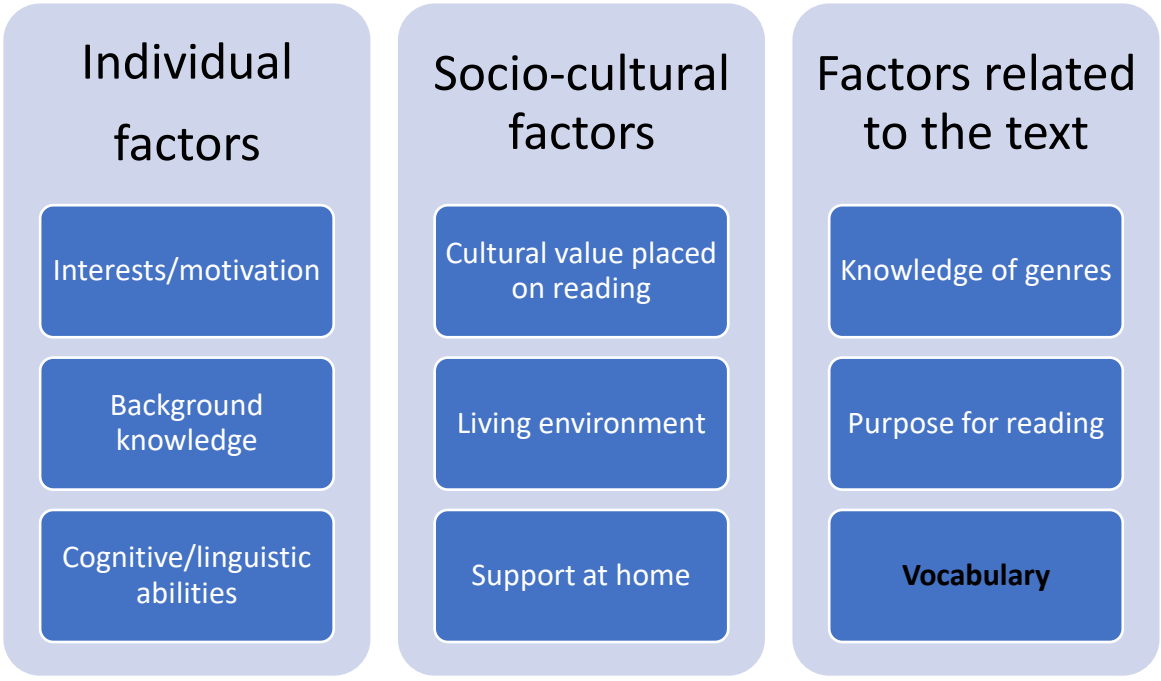


Figure 2.2 Selected factors affecting reading comprehension

Sources: Laflamme (1997); Leamnson (1999); Morni and Sahari (2013); Prowse and Delbridge (2013); Skatova and Ferguson (2014); Catts & Kamhi (2017)

If vocabulary plays such a key role in reading comprehension, then educators, arguably, should incorporate vocabulary activities into the curriculum at all levels of education for students to access it effectively. This is particularly important for students as they must 'read' for their degree and vocabulary instruction may provide some degree of motivation for reading compliance. 'Many students entering college do not make satisfactory progress because they have not learned to read advantageously and comprehensively' (Blake & Dearborn, 1935). This suggests if students are to achieve in higher education then vocabulary is key to reading comprehension which will enable student success.

2.5 Academic writing

The previous sections established the importance of vocabulary in reading; this section will argue it is vital in writing as well. Writing is an important skill for undergraduates as most of the assessment in higher education is done through the medium of writing (Fairburn & Winch, 2011). The importance of writing is further supported by a study on faculty views on the nature and instruction of academic writing in HE. It concluded that it was crucial for student's future career success, particularly as a communication tool (Zhu, 2004).

Moreover, the University of Essex (2008, p. ii) produced a comprehensive guide on how to improve academic writing and it states 'Writing is at the very heart of academic life. Good writing makes a good student'. Furthermore, academic staff at the University of Essex claim that essay writing and reasoning are the two most important skills for success, and it is the former that students most often lacked. Goodson (2017), drawing from research on writing in universities, argues that academics (students, faculty, research staff and even administrators) write for a living and their future success (e.g. grades, promotions, presentations to professional groups, funding for research projects and so forth) will depend on how well they write. Clearly, then, writing well has implications for student success.

2.5.1 Vocabulary and writing

'Selection of vocabulary is considered an important part of the writing process' (Olinghouse & Leaird, 2009, p.546). In higher education, academics and students need to express complex ideas in their writing and vocabulary knowledge is pivotal to this. In fact, it has been argued that 'Knowledge of vocabulary is obviously a prerequisite for writing' (Li & Schmitt, 2009, p.85). Furthermore, Maskor and Bharudin (2016, p. 265) contend that 'In relation with the theoretical ideas, vocabulary knowledge is the key factor to ensure that written production is understandable and meaningful'. Zhai (2016) maintains that inappropriate vocabulary use leads to inferior writing quality. Moreover, Wilkins (1972, pp.111-112) writes '… without grammar very little can be conveyed, without vocabulary nothing can be conveyed'. However, Laufer (2013) believes that there is often a gap between what we want to communicate in terms of content and what we can communicate because of our limited vocabulary. Thus, it can be challenging if our lexical repertoire is small and particularly if knowledge of academic vocabulary is narrow.

Academic discourse has its own specialised vocabulary which students need to know. The concept of academic vocabulary will be discussed and operationalised in Section 2.7.3 This section will closely examine the specific role vocabulary plays in academic writing as the genre has its own characteristics including forms of expression and vocabulary (Oliver, 2013). Typically, academic texts consist of 9% academic vocabulary (Schmitt, 2010; Nation, 2013). Furthermore, Nation (2013, p. 315) states academic vocabulary is 'overwhelmingly Graeco-Latin and is not easy to learn because words refer to abstract ideas, they are infrequent and their forms do not reveal their meaning'. Examples include words such as *aggregate*, *consequent*,

denote, innovate and *phenomenon* (Coxhead, 2000). Moreover, academic language can be a barrier that students need to overcome (Corson, 1995). Academic tasks such as reviewing, analysing, evaluating, critiquing and so forth require academic vocabulary (Nation, 2013). Thus, it is possible to conclude that in order to write successfully in higher education knowledge of academic vocabulary is essential.

Durrant and Benchley (2019) conducted a study which investigated how children's use of vocabulary in writing changes as they progress through compulsory schooling. They found that younger children in their writing repeat more high frequency verbs and adjectives and low frequency nouns. They also showed a preference for fiction-like vocabulary over academic-like vocabulary. From Years 2-11, this changed: the children made greater use of academic vocabulary in their writing and used less 'literary' or fiction type vocabulary in their 'non-literary' writing, demonstrating a sense of register appropriateness.

Not only is selecting the most appropriate word important in terms of meaning and form, but also the word has to suit the particular context in which it is being used. Schmitt and Schmitt (2020, p. 59) use the term register to 'describe the stylistic variations that make each word more or less appropriate for certain language situations or language purposes'. In other words, register accounts for the varieties of a language according to use (Chiu, 1972) and context is key. Register includes the ability to use both grammar and vocabulary in context. Read (2000) believes that there needs to be more work done on defining exactly what a register is and how different types of register can be classified. This is evident in recent research – for example, it can be argued that studies focusing on academic register have not adequately conceptualised the term and have defined it very broadly. Academic register has been described as the language used in academic settings or schooling which uses highly precise, technical and specialised vocabulary (see Bahr *et al,* 2020) or seen to exist along an informal/formal continuum (see Galloway *et al*, 2015; Larsson & Kaatari, 2020).

Chiu (1972) describes six types of register variations and two in particular apply to HE: field of discourse and mode of discourse. Field of discourse refers to the purpose of the language in use (e.g. technical, medical English or non-technical, English used for greetings) and mode of discourse refers to speaking or writing. Field of discourse has been linked to the term genre. A genre is 'an expected style of discourse, which determines appropriate language use' (Schmitt & Schmitt, 2020, p. 61). Researchers

classify writing genres in a number of ways which makes comparing findings problematic: ‘literary’ or ‘non-literary’ (e.g. Durrant & Brenchley, 2019); ‘personal, factual, and analytical’ (e.g. Schleppegrell, 2004) and ‘narrative, persuasive and informative’ (Olinghouse & Wilson, 2013). ‘Non-literary’ (e.g. non-fiction), ‘factual or analytical’ and ‘persuasive or informative’ could fall under the broad umbrella of academic genres (see University of Sydney, 2020).

Relevant to this research is the British Academic Written English Corpus (BAWE) which contains just under 3,000 good-standard student assignments (6,506,995 words) from across four broad disciplinary areas (Arts and Humanities, Social Sciences, Life Sciences and Physical Sciences) and across four levels of study (undergraduate and taught Masters level), with 30 main disciplines represented (Coventry University, 2020). Nesi and Gardner (2012) were able to identify and describe 13 genre families from the BAWE (see Table 3.9 and Section 3.16). Their categories will provide the conceptual framework for the genres identified in this research due to the same target population being investigated, namely UG students.

Nesi and Gardner also mapped the genre families to five register dimensions providing linguistic support for their genre classifications. For example, they identified that the persuasive register was most prominent in the genre ‘Problem Questions’. ‘Overt expressions of persuasion or argument are seen at the positive end of this dimension where infinitives (hope to go), prediction modals (will, would, shall), suasive verbs (command, insist, propose), conditional subordination (if you want) and necessity modals (must, should, have to) are significant’ (Nesi & Gardner, 2012, p. 46). This suggests that student writers need to use certain vocabulary and grammar in accordance with the genre of the assignment.

Register is measured differently by researchers too. Fang *et al* (2020) place a strong emphasis on grammar in their operationalised definition of academic register; they identified 11 lexical and grammatical features only two of which included vocabulary: specialised terminology and general academic vocabulary. Olinghouse and Wilson (2013) operationalised register as the proportion of words in a text of Latinate origin as compared to Germanic origin and VocabProfile (see Cobb, 2017) was used to determine textual register via calculating the proportion of words of Latinate origin.

While most guidance for students on academic register from universities focuses on the informal/formal continuum, it tends to place a more or less equal emphasis on both vocabulary and grammar (University of Bradford, 2020; University of Sheffield,

2020; University of York, 2020). The University of York (2020) summarises academic register as formal language which is impersonal and precise, uses sentences which are often long and carefully constructed, and contains vocabulary which is specialised and complex. Academic writing tends to use the passive voice and compound noun phrases, for example, Austere twentieth-century French existentialism (the University of Bradford, 2020). Furthermore, contractions (e.g. don't) should not be used in academic writing (University of Bradford, 2020; University of York, 2020).

Academic discourse needs to be coherent or make sense (UEfAP, 2021). This can be achieved by using internal cohesion techniques which help the reader navigate around an expanse of discourse (Harmer, 2015). These can include lexical cohesion (repeating words or phrases to link paragraphs); anaphoric reference (referring back to things already mentioned by using a pronoun); and grammatical cohesion (e.g. keeping to the same verb tense) (Harmer, 2015; Valenzuela, 2020). Another technique is to use conjunctions or linking words such as *but, so* and *however* (Valenzuela, 2020). These are sometimes called discourse markers or cohesive devices and are important for connecting our ideas and parts of a text together. Walsh (2010, p. 339) avows that 'One of the most important features of good writing is the way the writer signposts and links the argument for the reader'. The most common linking word is *therefore* but this tends to be overused (Walsh, 2010).

Walsh (2010, p. 338) also found a number of other common problems in L2 student academic writing:

- an under- or over-use of discourse markers (e.g. in addition);
- a lack of attention to cohesion and coherence;
- not using an adequate range of vocabulary (detectable through type-token ratio counts, for example);
- improper use of academic conventions, such as citing and referencing, etc.
- not paying attention to the audience.

It is clear from the literature that students need to pay attention to academic register but they also need to ensure their writing is coherent and cohesive as well. Vocabulary plays a key role in achieving both academic register and cohesion. However, use of vocabulary and appropriate academic register is not reflected in the

marking criteria used in this research. The undergraduate marking descriptors for a pass (40-49%) (University of Derby, 2020) state:

> A sound standard of work; a fair level of critical analysis and evaluation; little evidence of original thinking or originality; adequately researched; a sound standard of presentation; ideas fairly clear and coherent, some significant mis-understandings and errors; some weakness in style or presentation but satisfactory overall.

And 'For the highest marks (90-100%): an exceptional standard of work illustrating thorough and in-depth understanding, communicated with exceptional authority' (University of Derby, 2020). It can be argued that while vocabulary is not explicit in the marking criteria, students must demonstrate in-depth knowledge of the subject of which vocabulary plays a vital role. To conclude, vocabulary is an important feature of academic register, maintaining cohesion and demonstrating subject knowledge, yet students' work is not necessary overtly marked according to the vocabulary they use as per the marking criteria.

2.6 What is vocabulary? The concept of vocabulary

Having established the importance of vocabulary in reading and writing, this section will outline the concept of vocabulary. Vocabulary can be subdivided in many ways and the following section will discuss the nature of vocabulary and how it is categorised as the central concept of the research. Vocabulary is the connection between meaning and form in the shape of words (Schmitt & Schmitt, 2020). The total stock of meaningful words in a language is known as a lexicon. Linguists refer to these words as lexemes or lexical items (Crystal, 2020). There are various kinds of lexical items, but individual words are considered the most basic type (Lewis, 1993; Hatami, 2014).

In comparison, there are multi-word items which consist of multiple words but function in meaning as a single unit (e.g. *out of breath* meaning breathing quickly). Similarly, multi-word expressions (MWEs) can be defined as 'linguistic objects formed by two or more words that behave like a "unit" or "chunk" in that they display some formal or functional idiosyncratic properties with respect to free word combinations (i.e., normal phrases)' (Masini, 2019). The term MWEs can largely be found in the fields of phraseology (e.g. Mitkov, 2016), corpus linguistics (e.g. Chen, 2018) and natural language processing (e.g. Tanabe *et al*, 2014) amongst others (Masini, 2019).

There are many types of multi-word units, but formulaic language is a general term which covers the various types (Schmitt & Schmitt, 2020). While these multi-word units can be subcategorised, Lewis (1993) notes that any categorisation will involve overlap as well as marginal cases. The plethora of terminology also makes it difficult to refer to research findings across research areas because of the ways in which each is defined (Wray, 2002).

There are many terms associated with formulaic language which include collocations, idioms, set phrases, phrasal expressions, phrasal verbs, lexical bundles, fixed and variable expressions (Wray, 2002; Schmitt & Schmitt, 2020). Research on multi-word items is still in the initial stages and the majority of vocabulary studies have focused on single words until recently (Webb, 2020). The unit of analysis for this study is at word level so it is not necessary to define the various definitions associated with formulaic or MWEs. It is therefore pertinent to categorise different types of words.

Words that convey meaning such as nouns (e.g., *house*) or verbs (e.g. *run*) are content words (Webb & Nation, 2017). Learners need to acquire both the form and the meaning (Brezina & Gablasova, 2015) but not all vocabulary has a meaning/form connection as some words perform a grammatical function and are called function or grammatical words (e.g., *of, the, and*) (Schmitt & Schmitt, 2020). While there are relatively fewer function words, Schmitt and Schmitt (2020) estimate there are between 150-300, with the word *the* being the most frequent in general English.

2.7 Types of vocabulary

2.7.1 General vocabulary

Vocabulary can be found in a variety of different contexts and can be categorised according to the three main types: general; academic; and subject specific or technical (Schmitt & Schmitt, 2020). ‘General vocabulary is the term that is used to describe vocabulary that is useful across a wide range of topics and contexts, in both speech and writing’ (Schmitt & Schmitt, 2020, p. 7). General vocabulary, as described by Schmitt (2010, p. 75) is considered to be ‘the higher-frequency vocabulary necessary to achieve a basic functionality with a language’ but ‘there are no agreed limits for which vocabulary this might include, as the notion itself is rather vague’. General vocabulary words include content words such as *cat*, *house*, *park* and function words such as *above*, *can*, and *with*. Paquot (2010, p. 10) refers to these words as ‘core’ or ‘basic’ or ‘nuclear’. However, Schmitt and Schmitt (2020) believe that word frequency is only one of the factors that make vocabulary core (the most

important words to start learning first in English) and that ‘usefulness’ to the learner is perhaps more important.

Nonetheless, word frequency is still a good indicator of the relative value of a word. Higher frequency words have greater value as they are more likely to be needed in communication than less frequent words (Webb & Nation, 2017). They are also more likely to be learned since they are encountered more than low frequency words (Vilkaitė-Lozdienė & Schmitt, 2020). The best-known list associated with high frequency words is the General Service List (GSL) devised by West (1953) (Paquot, 2010; Schmitt, 2010; Nation, 2013; Brezina & Gablasova, 2015; Webb & Nation 2017). For a comprehensive account of the history of the GSL see Gilner (2011). The GSL is a list of around 2,000 of the most frequent word families in English taken from a corpus of written English. Of the 2,000-word families on the list, around 165 are function words and the rest content words (Nation, 2013). However, for teaching and learning purposes, Schmitt and Schmitt (2014) argued that the category for high frequency be extended to the 3,000-word level and the low frequency boundary be lowered to 9,000 in addition to labelling the category between 3,000 and 9,000 as mid-frequency.

Main criticisms of the GSL centre on the selection criteria used to create the list (Schmitt, 2010) and its outdatedness (most of the texts were published before 1930) (Gardner, 2013; Browne, 2014; Brezina & Gablasova, 2015). Based on these critiques, there have been more recent attempts to update the GSL by creating new core general vocabulary word lists (see Browne, 2014; Brezina & Gablasova, 2015; Dang & Webb, 2016). Browne (2014) created the NGSL (New General Service List) by expanding the corpus (273 million words) as compared to the original (about 2.5 million words) and including more recent texts. Brezina and Gablasova (2015) have created a new General Service Word list (new-GSL) which is a list of core English vocabulary that contains the 2,500 most frequent and general words of the English language based on four language corpora, the total size over 12 billion running words. The new-GSL is different from other general English vocabulary lists because it is based on four corpora while the other lists are based on one corpus. It is based on lemmas as the organising principle and distinguishes between word classes (*go* as a verb or noun) (Brezina, 2017). In addition to general vocabulary, students in higher education need subject specific terminology.

2.7.2 Technical/subject specific vocabulary

Technical words are those that are closely related and common to a topic or subject area but not necessarily elsewhere (Nation, 2013). Schmitt (2010, p. 77) defines technical words or phrases '[as] those which are recognizably specific to a particular field' and believes this vocabulary 'is essential to understanding discourse in a field'. Similarly, Schmitt and Schmitt (2020, p.8) define technical vocabulary as '… the jargon that is specific to particular domains (e.g. business, medicine, chemistry) and that represents concepts and ideas specific to those domains (*ledger, scalpel*, *catalyst*)'. Technical vocabulary is very field specific so can also be referred to as specialised vocabulary (Schmitt & Schmitt, 2020).

It is important to note that there is no consensus on how to define technical vocabulary (Chung & Nation, 2003; Ha & Hyland, 2017). The difficulty in deciding which words are technical and which are not has meant that there is very little known about this type of vocabulary and very few studies have been conducted compared to academic vocabulary (Chung & Nation, 2004). Further to this, Gablasova (2014) reports that there is very little attention given in the literature to the acquisition of technical vocabulary from academic learning.

'Technical vocabulary is subject related, occurs in a specialist domain, and is part of a system of subject knowledge' (Chung & Nation, 2004, p. 252); it differs from subject area to subject area (Nation, 2001). Technical words are low frequency words over a range of texts but can appear frequently within a specific text or number of texts related to a particular discipline (Nation & Kyongho, 1995; Webb & Nation, 2017). Nagy and Townsend (2012, p. 91) called these words 'discipline-specific academic words' while Gablasova (2015) has referred to these words as simply 'disciplinary vocabulary'. A knowledge and understanding of technical vocabulary are essential in order to access disciplinary or subject knowledge (Schmitt, 2010) as many of the key concepts are represented in this vocabulary (Schmitt & Schmitt 2020). Therefore, for the purpose of this research, the term subject specific vocabulary will be used instead of technical for this reason.

Sutarsyah *et al* (1994) believe subject-specific words make up around 10% of the words in academic texts but Gillet (2012) suggests a figure of about 5%. However, Chung and Nation (2003) found in their research that almost one out of every three words in an anatomy text (31.2%) was a technical word. They also examined an Applied Linguistics text and found that 20.6% of the running words in that text were classified as technical. From this data, it is reasonable to assume that there are a

large number of subject-specific words within each discipline area. Dictionaries used in discipline areas such as economics, geography or electronics usually contain around 1,000 entries (Nation, 2013).

The importance of subject specific vocabulary cannot be underestimated. ‘For all students, understanding and being able to use specialist terminology is essential’ (Mole & Peacock, 2005) and students will benefit from the ability to overtly use and define discipline-specific vocabulary (Carroll & Ryan, 2005). Carroll and Appleton (2007, p. 78 in Jones and Brown, 2007) also argue that ‘Support and guidance that are carefully organised and specific to the discipline are widely welcomed by [International] students and, increasingly, expected by many’. Lastly, ‘Students will appreciate help with discipline specific vocabulary in the form of glossaries and explanations on-the-spot when they arrive’ (Carroll, 2002, p. 2).

Because of the importance of subject specific vocabulary, an Education Studies glossary was created within the researcher’s own institution as a result of a previous research project (see Marshall, 2009) that examined the attitudes and perceptions of ESOL (English for Speakers of Other Languages) students towards their English language learning needs while studying at degree level. A key finding was that 77% of respondents expected the university to provide a list of technical words and definitions for each subject. Hence, in a previous research project the researcher created a glossary based on the need for international students to access the language used in Education Studies.

Schmitt (2010) states there are two main ways of identifying technical vocabulary: using the intuition of experts in the field; or analysing the discourse from a subject specific corpus. In terms of using the intuition of experts, Chung and Nation (2003, p. 103) identified three ways: using a rating scale where words are classified as being technical or non-technical; using a technical dictionary compiled by subject specialists; and ‘by making use of clues that the most relevant specialist, the actual writer of the text, used to mark the words considered to be important for the message of the text, since when new terms are introduced in a text, the writers deliberately provide contextual clues to help readers manage new terminology’. Another way of identifying technical terms is to compare word frequencies in a technical text with those in a different corpus – ‘[t]echnical words should be much more frequent in the technical corpus’ (Chung & Nation, 2003, p. 103).

Hyland (2006, p. 11) believes, in general, subject specialists neither have the expertise nor the desire to teach disciplinary literacy and argues that, '[r]arely do lecturers have a clear understanding of the role that language plays in their discipline or the time to develop this understanding in their students'. Hyland (2006, p. 11) also adds that 'they [lecturers] are often too busy to address language issues in any detail and rarely have the background, training or understanding to offer a great deal of assistance'. Although Hyland is referring to international students, the same can be said for all students regardless of their linguistic background. Furthermore, Crème and Lea (1997, p14) suggest that academics can be 'so engrossed in their subject that they seem to forget that they need to explain their discipline as a particular way of constructing knowledge about the world to their students'.

In summary, both Hyland (2006) and Crème and Lea (1997) maintain the view that lecturers do not understand the role language plays in their relevant disciplines nor are they engaged in language-related issues. This observation supports the researcher's own experience as an EAP and ESP lecturer in two UK universities for over ten years. A key question was always whose responsibility is it to teach language in a subject – a language specialist or a subject specialist.

Cowan (1974, p. 391) stated it is not necessarily the responsibility of the English teacher to teach subject specialist words but that the learning of this vocabulary '… is an automatic consequence of studying the discipline which uses them'. More recent research, however, seems to suggest there is a greater awareness of discipline specific literacy and subject specialists are addressing language issues. A number of studies have focused on improving discipline specific language competency in particular (e.g. Hillege *et al*, 2014; Nguyen *et al*, 2015; Green & Lambert, 2018; Wu & Carmichael, 2018). While these studies are an important indicator that some institutions are taking measures to embed language instruction within disciplines, most of the research is overseas (e.g. Australia, Singapore, Hong Kong and the USA) so caution is needed. There is very little evidence in the literature to suggest this approach is becoming widespread practice in the UK.

2.7.3 Academic vocabulary

The concept of an academic vocabulary has a long history in the field of English for academic or specific purposes (EAP or ESP) (Hyland & Tse, 2007). A definition of

academic vocabulary is not straightforward, and it is often conflated by researchers and theorists with subject specific or technical vocabulary. It has been referred to in the literature as 'sub-technical' (Webb & Nation, 2017), 'semi-technical' or 'specialized nontechnical lexis' (Hyland & Tse, 2007) and 'discourse-organising vocabulary' (Paquot, 2010). However, it is generally accepted that subject specific and academic vocabulary are deemed to be different. Academic vocabulary has been previously described as 'semi-technical' and defined as 'formal, context-independent words with a high frequency and/or wide range of occurrence across scientific disciplines, not usually found in basic general English courses; words with high frequency across scientific disciplines' (Farrell, 1990, p. 11). Academic words are not only found in scientific disciplines but the humanities as well, so a broader definition is necessary.

Paquot (2010, p. 9) refers to academic vocabulary as '… a set of lexical items that are not core words but which are relatively frequent in academic texts'. More specifically, Coxhead (2000, p. 218) describes academic words as '… lexical items [that] occur frequently and uniformly across a wide range of academic material but are not among the first 2,000 words of English as given in the GSL'. Considering Coxhead (2000), Paquot (2010) and Hyland and Tse (2007), academic vocabulary is identified by elimination and which categories it does not belong to. In other words, academic vocabulary is not core, not in the GSL and not technical either.

Over the years, there have been several academic word lists produced by researchers (Campion & Elley, 1971; Xue & Nation, 1984; Browne *et al,* 2013; Gardner & Davies, 2014). Word lists provide an efficient way of organising vocabulary learning and are often derived from word frequency (Webb & Nation, 2017; Schmitt & Schmitt, 2020). Word lists should also cover a range of texts and consideration given to the counting unit, for example lemmas or word families, as the latter could increase the learning load (Schmitt & Schmitt, 2020).

Coxhead (2000) compiled an academic word list from a corpus of 3.5 million running words of written academic text, excluding the first 2,000 words from the GSL. 'The AWL contains 570-word families that account for approximately 10.0% of the total words (tokens) in academic texts but only 1.4% of the total words in a fiction collection of the same size' (Coxhead, 2000, p. 213). She divided the 570 words into 10 sub-lists where sub-list 1 contains the most frequent words in the AWL, sub-list 2 contains

the next most frequent words, and so forth. There are 60 families in each sub-list except for sub-list 10 which has 30.

A word family contains a base/root word and its inflections and derivations while a lemma is a base word and its inflections together (McEnery & Hardie, 2012; Webb & Nation, 2017; Schmitt & Schmitt, 2020). 'The notion of a headword (as found in the dictionary) is generally equivalent to that of a lemma' (McEnery & Hardie, 2012, p. 245). For example, the word *alter* is a lemma and includes its inflections *altered, altering,* and *alters*. This is also called a lemma set where the lemma and its inflections belong to the same word class (Gardner, 2007). When the part of speech changes by adding a prefix or suffix this is called a derivation, e.g. *alterable, alteration.* The word family *alter* contains *alterable, alteration, alterations, altered, altering, alternate, alternating, alters, unalterable, unaltered.*

Coxhead (2000) based her selection of word families on three key criteria:

1. The word family could not be in the GSL (West, 1953) as previously mentioned
2. The word family had to occur at least 10 times in the four discipline areas used in her corpus (Arts, Law, Commerce and Science) and at least 15 in the 28 subject areas within the disciplines (e.g. education, constitutional law, economics and biology)
3. The word family had to occur at least 100 times in the corpus.

'Frequency was considered secondary to range because a word count based mainly on frequency would have been biased by longer texts and topic-related words' (Coxhead, 2000, p. 221). Furthermore, the corpus included 414 texts from academic journal articles, textbooks, book chapters, laboratory manuals and from other corpora.

A more recent list, the Academic Vocabulary List (AVL) was produced by Gardner and Davies (2014) based on two key criticisms of the AWL; the use of word families to determine word frequencies, and the relationship of the AWL with the General Service List (GSL) (West, 1953). The use of word families is problematic as members of the word family may not share the same core meaning. For example, the word *bulk* has different meanings depending on the discipline; in business it means to buy/sell/order in large amounts, or to store loose and not wrapped separately (*Cambridge Dictionary*, 2020). In science, when *bulk* is combined with *density*, it means the weight of soil in a given volume (Soil Quality, 2020). These different meanings are highlighted when the words are used in different disciplines (Hyland & Tse, 2007; Gardner & Davies, 2014). Learners also need to understand

morphology/word derivations (e.g. *unalterable*) which in terms of learning usually comes after acquiring inflections (e.g. *altered*) (see Gardner, 2007). Therefore, Gardner and Davies (2014) argue that lemmas not word families are a better choice for word lists.

The second criticism pertains to the AWL and its relationship to the GSL, firstly being outdated and no longer representative of high frequency words with many words on the AWL actually being high frequency even though they have not been described as such (Gardner & Davies, 2014). Secondly, there are many high frequency academic words such as *market* that were excluded from the AWL as they were found in the GSL (Gardner & Davies, 2014; see also Cobb, 2010).

Borne out of these criticisms of the AWL, the AVL was created and derived from a 120-million-word academic sub corpus of the 425-million-word Corpus of Contemporary American English (COCA). Using lemmas and not word families for the list, it contains 3,000 words ranked according to their frequency in academic discourse. It has a 14% coverage of written academic texts (Gardner & Davies, 2014). However, the list contains many higher frequency words that students may already know (e.g. *group*) and because it is extensive, teachers and students will need to determine which words to prioritise which may take some effort (Webb & Nation, 2017).

Durrant (2016) conducted a study on the usefulness of the AVL in university student writing by exploring its coverage from texts taken from the BAWE. It was found that as students progressed through their degree there was a corresponding increase in the word usage from the AVL. However, Durrant (2016) also found significant variation across text types and disciplines. In addition, around half the words on the AVL are used very little and a small number of core items (427) could be found across 90% of disciplines. Durrant (2016) concludes that the AVL may be more relevant to some student writers than others and while a generic productive academic vocabulary exists, it is smaller in scope than the full AVL.

Browne *et al* (2013) created the New Academic Word List (NAWL) to work in conjunction with the NGSL discussed in Section 2.7.1. The NAWL excludes general high frequency words like the AWL but unlike the AWL the counting unit is the lemma rather than the word family. The final list contains 960 words and has around a 6% coverage. In combination with the NGSL, both lists have a reported coverage of 92% of the academic corpus used to create the NAWL. A recent study using the NAWL

for lexical frequency profiling of L2 university pre-sessional students revealed that the NAWL accounted for 5% of the words in a sample of essays (n = 118) that scored in the upper quartile (Higginbotham & Reid, 2019). A further discussion of these two lists can be found in Section 3.13.2.

Of the three lists discussed (AWL, AVL and NAWL), the AWL is the most widely used in English language teaching and has inspired a wide range of pedagogic materials, including textbooks, vocabulary tests and dictionaries (Paquot, 2010). Others such as Liu and Han (2015), Green and Lambert (2018) and Quigley (2018) support the view that the AWL is one of the most influential academic word lists to date. Knowledge of these academic words is important for students when reading a wide range of texts but also when writing, as mentioned previously. The AWL can also help in vocabulary goal setting particularly when students need to focus on which words are most important to learn.

Nevertheless, there has been criticism of the AWL in the literature. Paquot (2005) believes that the AWL is suitable for reading comprehension and receptive purposes but not productive ones such as writing or speaking as learners' needs for academic writing are not the same as for academic reading. The corpus used to create the AWL focused on a limited number of discipline areas with 50% of the corpus being derived from Law and Commerce (Durrant, 2014). Hyland and Tse (2007, p. 249) found that 'many items are considerably underrepresented in particular fields or disciplines' and 'the AWL seems to be most useful to students in computer sciences, where 16% of the words are covered by the list, and least useful to students in biology, with only 6.2% coverage'. As a result, they argued that '… the AWL might not be as general as it was intended to be and, more importantly, questions the widely held assumption that students need a single core vocabulary for academic study' (Hyland & Tse, 2007, p. 235). The AWL has also been criticised for excluding core vocabulary (as identified by the GSL) and the words on the list are not as infrequent and specialised as previously thought (Masrai & Milton, 2018). This point is particularly pertinent as students do need more than a single core vocabulary of academic words; they need to know subject-specific words in the context of their discipline as well (discussed in Section 2.7.2). Durrant (2014) argues that students see insufficient knowledge of subject specialist vocabulary as more pressing than academic vocabulary. Despite these views, AWL still remains the most influential academic word list to date (Galal, 2020).

2.8 Knowing a word

Having defined types of vocabulary words, this section will discuss aspects of word knowledge. There is no simple answer or consensus to the question, 'What does it mean to know a word?' (Pignot-Shahov, 2012). There are many elements to knowing a word and there are various degrees to which one can know a word (Meara, 1990). According to Schmitt (2017), knowledge of a word begins with making the connection between its form and meaning, for example, the meaning of *puppy* is a young dog. Word knowledge also goes beyond the form and meaning connection in terms of being able to recognise and use it, so the receptive/productive distinction is an important one in terms of what is involved in knowing a word. Receptive knowledge is the knowledge required to listen or read while productive knowledge is required for speaking and writing (Webb & Nation, 2017). It is generally accepted that receptive word knowledge usually precedes productive and that learners' receptive knowledge is greater than productive, but the process and relationship between the two remains unclear (Schmitt, 2010, 2019; Schmitt & Schmitt, 2020). Nation (2013) provides a comprehensive view on what is involved in knowing a word which encompasses all three aspects: form, meaning and use (see Table 2.1)

Table 2.1 What is involved in knowing a word

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

Source: Nation, 2013, p. 49

Nation's (2013) conceptualisation of word knowledge is the most referred to by vocabulary scholars due to its comprehensiveness (González Fernández & Schmitt, 2020). But, González Fernández and Schmitt (2020) point out this breadth also makes it very difficult to apply in research as measuring all the components is practically impossible in a single study. As a result, most vocabulary research focuses on a single component (e.g. Laufer & Nation, 1995; Laufer & Goldstein, 2004) and there are limited studies which focus on multiple components (e.g. Schmitt, 1998;

Webb, 2005; González Fernández & Schmitt, 2020). Moreover, the list does not state what the connection is between the components which leaves questions as to how each of them contribute to vocabulary knowledge (González Fernández & Schmitt, 2020) and in which order they are learned (Schmitt, 2019). Finally, Barclay and Schmitt (2019, p. 803) believe Nation's (2013) framework is 'aspirational' rather than 'prescriptive' as it describes all aspects of vocabulary knowledge rather than what is needed to use a word appropriately.

Vocabulary researchers have also disagreed as to whether the receptive/productive concept is dichotomous (separate and distinct) or operates on a continuum. Melka Teichroew (1982, p. 21) believes that learners gradually move from receptive to productive knowledge:

> The notion of familiarity (or knowledge) of a word could be represented as being a line, a continuum starting, roughly, with the first stage of recognition, passing through various intermediary points and finishing near productive knowledge, with productive knowledge itself composed of several stages or phases (e.g., knowing a range of meanings for a word or the knowing of collocations or idioms could represent a more advanced stage of production compared to simply knowing a single meaning of a word which has many).

Critics of the continuum such as Meara (1996) argue, by definition, a continuum implies that a component of word knowledge varies continuously from receptive to productive but exactly what varies and how this variance works is not clear. First pointed out by Meara in 1983, there is a lack of a general theory of vocabulary acquisition due to the fact that researchers have not focused on the interrelationship between word knowledge components (González Fernández & Schmitt, 2020). In order to understand how learners acquire vocabulary, it is necessary to measure the various components of word knowledge. But the main issue pertaining to reception and production knowledge of vocabulary rest with how they are measured (Schmitt, 2010).

2.9 How do we measure receptive lexical knowledge?

Vocabulary can also be described in terms of breadth or size (number of known words operationalised as knowledge of the form-meaning connection) and depth or quality (how well words are known) (Schmitt, 2014). Most measurement and discussion of vocabulary have focused on size or receptive vocabulary (Schmitt, 2010). A number of tests have been created to measure receptive vocabulary knowledge which have

been used for both teaching and learning as well as research purposes. Most vocabulary tests measure the form/meaning link and ask participants to either recall (supply) the word or meaning or recognise (select) the word or meaning (Schmitt, 2010). A few of the tests used in vocabulary research are listed in Table 2.2 and a brief description of each will be followed by a critical discussion.

To begin, the Peabody Picture Vocabulary Test (PPVT) was designed for younger, native English-speaking children before they began reading and writing (Dunn & Dunn, 1997). In this test, words are read aloud, and the participant must choose the corresponding image that represents the word. The target audience of the PPVT is young children so unsuitable for HE students. Moreover, this test (PPVT) requires each individual to be assessed one at a time which would be lengthy to administer.

Table 2.2 Selected vocabulary tests measuring receptive knowledge

Author	**Test name**	**Task/target audience L1 or L2**	**Word knowledge component measured**
Dunn and Dunn (1959, 1997)	Peabody Picture Vocabulary Test (PPVT)	Listen to word and point to picture which represents the meaning/L1	Meaning/ recognition
Goulden *et al* (1990)	Vocabulary size tests	Tick the words you know and provide a definition for the last five words you ticked/L1	Form meaning link/recall
Nation (1990) revised by Schmitt *et al* (2011)	Vocabulary Levels Test (VLT)	Match the word to its meaning/L2	Form meaning link/ recognition
Laufer and Goldstein (2004)	Computer Adaptive Test of Size and Strength (CATSS)	Four formats involving different forms of L1/L2 translation/L2	Form meaning link/recognition and recall
Nation and Beglar (2007	The Vocabulary Size Test (VST)	Multiple choice (choose the definition of a word)/L2	Form meaning link/recognition
McLean and Kramer (2015)	The New Vocabulary Levels Test (NVLT)	Multiple choice (as above)/L2	Form meaning link/recognition

Goulden *et al*'s (1990) test is based on *Webster's Third New International Dictionary* (1961) and updates, and includes a representative sample of the 25,000 most frequent words in Thorndike and Lorge's (1944) frequency lists. The test contains 250 words including five sub-tests where 10 words are selected from each of the first 5,000-word bands in this list. According to Goulden *et al* (1990), the reason they

chose this dictionary was because it is the largest non-historical dictionary of English, containing over 450,000 words. The method uses a random sample of words that proportionately represent the dictionary. Participants go through the sample and mark the words they know (i.e. they have seen it before and can express at least one meaning of it). This is a usual format for vocabulary tests because it is quick (Schmitt & Schmitt, 2020) and often called checklist tests (Meara & Buxton, 1987). An advantage to this type of test over multiple choice is that it is possible to test very large number of items rather than a small sample (Meara & Buxton, 1987). To estimate the participant's vocabulary size, the number of known words from the sample is multiplied by the ratio that the sample of words bears to the total number of words in the dictionary, in this case 500. However, a major limitation of the test is the fact that it is self-reporting (relying on test takers to assess their lexical knowledge) rather than verifiable evidence (Read, 2000).

The Vocabulary Levels Test (VLT) was first created by Nation in 1983 and tests vocabulary at four frequency levels (2,000, 3,000, 5,000 and 10,000) and Coxhead's (2000) AWL. It has since been updated by Schmitt *et al* (2001) and more recently by Webb *et al* (2017). It uses a matching format (word to the meaning) (Nation, 1990). While some studies have combined the frequency levels to generate an overall vocabulary size figure, it was not really designed to do this (Schmitt, 2010). Because the VLT tests one's knowledge of words at frequency levels, it works well as an instrument for diagnostic and placement purposes (Nation & Beglar, 2007). Moreover, it can identify lexical weaknesses at a particular frequency level and help teachers place students quickly into ability groups based on their vocabulary knowledge (Kremmel & Schmitt, 2018).

However, a major limitation of the VLT pointed out in the literature is the lack of testing of the first 1,000-word frequency level (Webb & Sasao, 2013; McLean & Kramer; 2015). These words are important since they have the greatest value to learners because of their impact on comprehension and production. McLean and Kramer (2015) also point out that previous versions of the VLT sampled target words from outdated frequency lists (e.g. from West's (1953) General Service List) and the VLT format poses a number of problems such as lack of item independence and student difficulty understanding the format. To overcome some of these limitations, Webb *et al* (2017) updated the VLT to include testing at the 1,000 word frequency level and word lists were taken from the British National Corpus (BNC) and the Corpus of

Contemporary American English (COCA) so that items on the test were a better reflection of current English.

The Computer Adaptive Test of Size and Strength (CATSS) (Laufer & Goldstein, 2004) is aimed at second language learners and involves testing vocabulary size (number of words the learner knows) and the strength of the word form-meaning link. Strength was determined using a hierarchy of difficulty categorisation from passive recognition (easiest) to active recall (hardest). The test uses a multiple-item format and various permutations of L1/L2 translations. For example, users are given the meaning of a word and they supply the form in L2 (active recall) but also, they are given the form and choose the L1 meaning from a multiple-choice format (passive recognition). Two other test items involve supplying the form in L1 (passive recall) of a given meaning and choosing the L2 multiple choice definition (active recognition) of a given form. Test of receptive vocabulary knowledge such as the VLT can only indicate whether the form-meaning link is known rather than the degree to which they can understand or use it (He, 2019). The strength of the CATSS test is that it provides the degree to which the form-meaning link is made rather than simply rating words as known or unknown (Schmitt & Schmitt, 2020).

Nation and Beglar (2007) devised a test which measures a learner's vocabulary size from the first 1,000 to the 14th 1,000-word families of English. The format is multiple choice where the correct definition is chosen from a list. Multiple-choice formats are the most popular question type in standardised tests of vocabulary (Anderson & Freebody, 1981; Read, 2000). Here is a sample item from the fifth 1,000-word level.

1. miniature: It is a miniature.
   a a very small thing of its kind
   b an instrument for looking at very small objects
   c a very small living creature
   d a small line to join letters in handwriting

To choose the correct meaning, a developed understanding of the word is needed as the example sentence is very vague. If this is not the case, it could lead to a certain amount of guessing. Guessing of answers on multiple choice tests lowers the reliability of the test as there is some probability that a guessed answer may be correct (Bereby-Meyer *et al*, 2002).

The final test to be considered was the New Vocabulary Levels Test (NVLT) (McLean & Kramer, 2015). It tests knowledge of English words from the first five 1,000-word

frequency levels of the British National Corpus (a collection of texts with over 100 million words) and the Academic Word List (AWL) (Coxhead, 2000). While the test could provide some insight into the degree of mastery of the most frequent words of English, it does not provide an overall estimate of an individual's vocabulary size. In addition, the test is quite lengthy with 150 items. McLean and Kramer (2015) state it can be completed in 30 minutes, but with 150 items, in reality, this would take a bit longer.

The tests in Table 2.2 can be categorised as either 'size' or 'levels' tests. Schmitt and Schmitt (2020, p. 29) aver that 'while most vocabulary tests measure size in one way or another…', there is a key difference between size and levels tests which must be noted as it is important in interpreting test scores. Levels test measures vocabulary at frequency levels as discussed in Section 2.7.1 and tests such as the VLT and NVLT indicate how many words are known within a particular set of items (Webb & Sasao, 2013). 'The primary purpose of a levels test … is to estimate learners' mastery of the most frequent vocabulary in the hope of assigning appropriate learning materials' (McLean & Kramer, 2015, p. 2). For example, the VLT produces a learner's knowledge of vocabulary at different word frequency levels as opposed to a test of overall vocabulary size (Schmitt, 2010). 'The greatest value of the VLT is that it indicates at which word frequency level students should focus their learning' (Webb *et al*, 2017, p. 33). Although vocabulary size tests may take samples from frequency lists, they estimate how many words, lemmas or word families learners know (Schmitt & Schmitt, 2020). These scores are reported globally rather than at certain frequency levels. Like levels tests, size tests can be used for diagnostic purposes but also as a proxy for language proficiency in general (Kremmel & Schmitt, 2016).

Most of the vocabulary tests discussed so far are aimed at L2 learners rather than L1 or native speakers, with the exception of Goulden *et al* (1990) and the PPVT. This project is interested in native speaker's vocabulary sizes in the context of HE so it is important to examine this test and other similar ones closer.

2.9.1 Dictionary sampling methods

As previously stated, receptive vocabulary knowledge pertains to knowing and understanding the meaning of a word when reading and listening but not necessarily being able to use it when speaking or writing (Schmitt, 2010; Maskor & Bharudin, 2016; Webb & Nation, 2017). Research into estimating receptive vocabulary sizes of native speakers has traditionally used dictionary sampling methods (Hartmann, 1941;

Goulden *et al*, 1990; D'Anna *et al*, 1991). This type of investigation is well established and according to Nation (1993), one of the earliest published studies into vocabulary size was conducted by Kirkpatrick (1891) using this method. The method uses a random sample of words that proportionately represent the dictionary. Participants go through the sample and mark the words they know. To estimate the participant's vocabulary size, the number of known words from the sample is multiplied by the ratio that the sample of words bears to the total number of words in the dictionary.

2.9.2 Limitations of vocabulary testing using dictionary sampling methods

Because dictionary sampling methods for estimating vocabulary sizes have been used over a significant number of years by several researchers, the benefits and challenges of this particular type of data collection tool have been discussed at length in the literature (e.g. Thorndike, 1924; Hartmann, 1941; Lorge & Chall, 1963; Dupuy, 1974; Goulden *et al*, 1990; Nation, 1993). Lorge and Chall (1963, p. 147) write, 'One of the oldest problems in educational research has been the estimation of vocabulary sizes of children and adults'.

The main disadvantage of using dictionary sampling methods is that estimates can vary widely (see Table 2.3) (Lorge & Chall, 1963; D'Anna *et al*, 1991; Nation, 1993). They vary widely for a number of reasons, but the main reason appears to be the size and nature of the dictionary; the specific sampling techniques used; the criteria for testing if a word is known; and definitions of what constitutes a word (D'Anna *et al*, 1991). When selecting words from the dictionary for a sample, the criteria for what to include or exclude is critical in determining the sample size and whether it is biased. Considerations include whether words should be included if they are not a main entry (e.g., derivatives or compounds) or are technical, slang, foreign or archaic (Anderson & Freebody, 1981). Additionally, attention must be given to whether abbreviations, names of persons, geographical name places, other names prefixes, and suffixes be counted as words (Goulden *et al*, 1990). Finally, it is important to consider whether base words, inflections and derivations should be counted as separate words. If so, this can lead to overestimations (e.g. Kirkpatrick, 1891) and inflate vocabulary sizes. Studies (e.g. Goulden *et al*, 1990; D'Anna *et al*, 1991) which report lower estimates use word families (a base word, its inflection and derivation as the counting unit) (Nation, 2013; Treffers-Daller & Milton, 2013).

Table 2.3 Estimates of vocabulary sizes of college/university graduates

Author	**Number of words**
Kirkpatrick (1891)	Between 20,000 & 100,000
Hartmann (1946)	215, 040
Goulden *et al* (1990)	17,200
D'Anna *et al* (1991)	16,785
*Brysbaert *et al* (2016)	11,100
Treffers-Daller & Milton (2013)	11,000

*Note: this study did not use dictionary sampling methods

There are also issues with what it means to know a word, as discussed in Section 2.8. There are many aspects of vocabulary knowledge: pronunciation; spelling; grammar; collocation; denotation (e.g. '*a cat*' denotes a kind of animal); connotation; appropriateness and word formation (Nation, 2013; Mykhailiuk, 2016) and dictionary sampling techniques typically test the form-meaning link. Hartmann (1941, p. 353) states, 'Another serious limitation of the common procedure is that it makes little if any allowance for qualitative variations in the precision of the definitions. One can "know" anything very slightly or exceedingly well'. Henriksen (1999) conceptualised this knowledge into three aspects: partial to precise knowledge of word meaning; depth of knowledge; and receptive to productive use ability. But, 'it is almost impossible to blend all of these into a single holistic view of vocabulary, but separately, each of these facets provides a useful framework for conceptualization' (Schmitt & Schmitt, 2020, p. 34). Therefore, predictably, most vocabulary tests based on dictionary sampling methods only ask for partial, receptive knowledge of words to be demonstrated.

Goulden *et al* (1990) use a simple yes/no format in their test as previously stated. Respondents put a tick next to each word they know (defined by having seen the word before and being able to express at least one meaning of it). They can do this using a variety of forms: sentences, synonyms, drawings or diagrams. But a key consideration is the fact that the test uses self-reporting procedures and Goulden *et al* (1990) ask whether it is possible to determine whether respondents have a tendency to overrate their knowledge of words. Other researchers (e.g. Brysbaert *et al*, 2016) have used a number of non-words in their lists and if respondents say that

they know a number of them then this can give an indication that they overrate their vocabulary knowledge (scores can be adjusted accordingly). However, Golden *et al* (1990, p. 354) state that 'This procedure has been used in studies with schoolchildren and second language learners, but such a check may not be necessary in the case of adult native speakers'. This suggests that school children and second language learners are more likely to overrate their vocabulary knowledge than adults.

A further test based on dictionary sampling methods which needs to be discussed is that used by Zechmeister *et al* (1995). As previously mentioned, Zechmeister *et al* (1995) reported that first-year college students were able to recognize the meanings of about 12,000 words. This is even higher than the mean number of words reported by Treffers-Daller & Milton (2013) for first years (9,756 words) in their research which used Goulden *et al*'s test. However, a direct comparison of results is difficult as the content and format of the test Zechmeister *et al* (1995) used varied from Goulden *et al's* (1990). Firstly, Zechmeister *et al* (1995) used a much smaller dictionary (the Oxford American Dictionary or OAD) and the number of entries counted in the OAD dictionary sample was 26,901 (similar to Brysbaert *et al*'s, 2016) compared to that of Goulden *et al*'s (1990) which included 54,241 entries or word families. Moreover, the OAD does not contain many low frequency words (Zechmeister *et al,* 1995). So in the case of Zechmeister *et al* (1995), it would be reasonable to assume their result would be lower than that of Goulden *et al* (1990) but similar to Brysbaert *et al* (2016) which is in fact true. However, Zechmeister *et al* (1995) claim that because the OAD does not contain many low frequency words and there are a smaller number of entries their results are likely to be underestimated. If this is the case, their findings could actually be similar to that of Goulden *et al* (1990) with an average vocabulary size of 17,200 words.

Another issue with Zechmeister *et al*'s (1995) research is the format of the test. They used a multiple-choice test which as we have seen in this Section is a common type of test for measuring vocabulary sizes. As previously mentioned, the multiple-choice format could lead to a certain amount of guessing which lowers the reliability of the test as there is some probability that a guessed answer may be correct (Bereby-Meyer *et al*, 2002). Further to this Wesche and Paribakht (1996, cited in Read, 2000) point out other limitations of the multiple-choice format such as participants knowing another meaning of the target word but not the one being tested; items may test learners' knowledge of distractors rather than their ability to identify an exact meaning

of the target word; and the format usually only permits a limited sampling of the learners' vocabulary (usually around 25 items in total).

The final test to mention is Brysbaert *et al* (2016). Although based on word frequency lists, it is worth discussing as the target audience was native speakers of English. They conducted a large-scale crowdsourcing experiment where they tested 221,268 individuals using a test similar to Goulden *et al* (1990). The tests are similar in task with the participants simply having to indicate whether they know a word or not. However, it is important to note that there are differences between them as well. About half of the words on Brysbaert *et al*'s (2016) test consist of non-words (33 in total on the test) alongside 66 actual words. Participants are punished if they indicated they knew a non-word. This contrasts with Goulden *et al*'s (1990) test which did not contain any non-words making the total number of words on the tests different. However, a test's reliability is not necessarily dependent on its length (Angoff, 1953).

A final and more significant difference worth noting is that Goulden *et al*'s (1990) test is based on 54,241-word families and Brysbaert *et al*'s (2016) test on 18,269-word families. As mentioned in Section 2.9.1 testing which involves dictionary sampling methods takes the proportion of the words correctly answered on the test as the proportion of the words they know in the dictionary (or word list in the case of Brysbaert *et al*, 2016). Goulden *et al*'s (1990) test is based on more word families so their test could potentially lead to an overestimation of vocabulary sizes as compared to Brysbaert *et al* (2016). This might partially explain why Goulden *et al* (1990) found the average vocabulary size of a native English-speaking university graduate to be 17,200 words as compared to 11,100 in Brysbaert *et al*'s (2016) research. Brysbaert *et al* (2016) acknowledge this as a limitation of their test but also state they are confident that any increase in the number of word families would not change much in terms of the words known by participants.

2.10 Vocabulary sizes and lexical coverage

Having critiqued the tests used to measure vocabulary sizes, it is important to consider the relationship between the latter and comprehension. What size of vocabulary do undergraduates have and need in order to access the discipline? The foci for this doctoral research are vocabulary sizes, and academic words used by native English-speaking undergraduates. Extensive searches using key terms such as 'vocabulary sizes' and 'undergraduates' on The British Education Index, the University's Library Plus and Google Scholar have produced only one other study

(Treffers-Daller & Milton, 2013) similar to this project which is discussed later in the Methodology chapter. Treffers-Daller and Milton (2013) reported that the average first year undergraduate has a vocabulary size of 10,034-word families and by the time they reach the third year, they know approximately 11,167-word families. This suggests that students increase their vocabulary size by around 400-500 words per year (Treffers-Daller & Milton, 2013). These findings must be interpreted with caution though, as the data was not collected with the same participants in each year so any notions of growth could be misleading. A university graduate will have an average vocabulary size of around 20,000 words according to Goulden *et al* (1990). But as Schmitt and Schmitt (2020) indicate, it is more likely that the average native English speaker knows roughly between 10,000 and 13,000-word families.

2.10.1 Lexical coverage

How many words do children, adolescents and adults need to know to read successfully? Lexical coverage is the percentage of known words in spoken discourse or written texts (Webb & Nation, 2017). Much of the research on lexical coverage has been done in the context of L2 speakers of English (Laufer, 1989, 1992; Hu & Nation, 2000; Nation, 2006; Schmitt *et al*, 2011) and estimates have varied. Originally, it was thought that around 3,000-word families were needed to comprehend authentic texts which would give 95% lexical coverage (Laufer, 1989,1992). Later research conducted by Laufer and Ravenhorst-Kalovski (2010) established a minimum threshold (the ability to read with some guidance) of 95% text coverage, including proper nouns, which requires L2 learners to have a vocabulary size between 4,000-to-5,000-word families. An optimal threshold (the ability to read independently) of 98% text coverage requires knowledge of 8,000-word families (Laufer & Ravenhorst-Kalovski, 2010). Hu and Nation (2000) found that 98%–99% of the words in texts needed to be known before adequate comprehension was possible. Following on from these figures, Nation (2006) found that native speakers need a vocabulary size of around 8,000 to 9,000-word families (base word plus its derivations, e.g. *wide*, *widen*, *widely*) to read widely. Similarly, Schmitt *et al* (2011) concluded that 98% coverage was necessary for readers of academic texts. On the other hand, Carver (1994) conducted research on both primary school students and graduates and found that native English-speaking students needed to know nearly 100% of the words in a text for comprehension to occur. However, even if there are a few unknown words, people can usually understand speech or writing therefore 100% coverage is not usually necessary (Schmitt *et al*, 2017).

Hsu's (2011, 2014) research suggests that students need knowledge of around 5,000-word families to access discipline specific texts. Hsu (2014) examined the vocabulary load of engineering textbooks and found that students needed to know the most frequent 5,000-word families plus proper nouns, apparent compounds and abbreviations which would command 95% lexical coverage of an engineering textbook to ensure adequate comprehension. In a previous study on the vocabulary threshold needed by English as a Foreign Language (EFL) learners in Business, Hsu (2011) found that students need to know the most frequent 5,000-word families in order to adequately comprehend texts. Even though this is in the context of business, a key consideration is whether native speakers have the necessary vocabulary size to know and understand these 5,000 word families? Hsu (2011) also recommends that further research is needed on the lexical needs of students and their perception of business specialist texts in relation to vocabulary and reading. The same research is needed not only for EFL students in higher education but also for native speakers as many students face challenges learning technical vocabulary (Gablasova, 2015).

2.11 Vocabulary and academic achievement

This chapter has so far shown the importance of vocabulary in reading and writing in HE with a focus on academic and subject-specific words. There is an increasing body of research around academic language competence and students' success in school (Nagy & Townsend, 2012). There is much evidence from the literature to suggest there is a link between vocabulary knowledge/size and academic achievement (Smith *et al*, 1991; Treffers-Daller & Milton, 2013; Bleses *et al*, 2016; Schuth *et al*, 2017; Masrai & Milton, 2018). Schuth *et al* (2017) found that children's academic vocabulary knowledge predicted school performance in four subjects, controlling for age, gender, language background, and nonverbal cognitive abilities. These results confirm the assertion that already in primary school academic language proficiency significantly influences success. Csomay and Prades (2018) conducted a study with English as a Second Language (ESL) university students in the USA and found a significant relationship between academic vocabulary use and essay scores in some text types, but not all. These findings, to some extent, strengthen Quigley's (2018) argument that the solution to closing attainment gaps in schools is to increase students' vocabulary knowledge and size.

Masrai and Milton (2018) devised a new checklist test, The Academic Vocabulary Size Test (AVST), to measure the written receptive academic vocabulary knowledge

of non-native speakers of English of the 570 academic word families presented in the AWL (Coxhead, 2000). The AVST comprises 114 items divided into six equal frequency bands, each including a sample of 19 words. The words form a difficulty continuum from the highest frequency words in the AWL to the lowest and it is assumed that higher frequency words will be learned more easily than lower-frequency words. The test includes a further 19 control words (beyond the 25,000 word level in Thorndike and Lorge's (1944) word list and used in Goulden *et al*'s test) to adjust for guesswork when calculating the final score (Masrai & Milton, 2018).

They found there was a strong positive correlation between academic vocabulary size measured using the AVST, overall vocabulary size measured using the XK-Lex, and learners' academic performance as measured by GPA (Grade Point Average). XK-Lex tests knowledge of the most frequently occurring 10,000 words in English and presents an estimate of the overall breadth knowledge of this vocabulary (see Masrai & Milton, 2012). Furthermore, Treffers-Daller and Milton (2013) also reported a link between vocabulary size (measured in word families) and achievement at two universities, Swansea University and UWE Bristol (Table 2.4).

Table 2.4 Vocabulary size of first year students and degree classification: Swansea University and UWE Bristol

Degree class	Mean vocab score Swansea University	Mean vocab score UWE Bristol
1	10,618 words	11,766
2:1	9,952 words	10,300
2:2	8,839 words	10,060
3	5,950 words	6,900

Source: Treffers-Daller and Milton (2013)

It is clear from the figures in Table 2.4 that the bigger a students' vocabulary size the higher the degree classification they earn. Research on academic literacy highlights the role that written literacy skills play in student progression (Scouller *et al*, 2008). However, there is no evidence to suggest whether this is a direct causal link. Further to this, Nation and Coxhead (2021) believe more research is needed on using data

from vocabulary size tests to predict education performance. If vocabulary plays such a key role in students' learning and achievement, then strategies should be employed to facilitate this.

2.12 Vocabulary and age

Psychologists have long been interested in the effects of aging on language and cognition (Jones & Conrad, 1933; Gold *et al*, 1995; Schaie,1996; Bowles *et al*, 2005; Boaz *et al*, 2015). This body of research seems to suggest a consensus that individuals keep learning new words throughout their lifetime, vocabulary knowledge is stable in adults and there is even vocabulary growth throughout adulthood. The following quote sums up this development:

> Vocabulary knowledge is one of the few cognitive skills that remain relatively intact over adulthood. Unlike most cognitive abilities, which peak when a person is around the age of 20 and then decline with age, vocabulary knowledge seems to peak around age 50 or possibly later, and decline only slowly, if at all, into old age. (Bowles *et al*, 2005, p. 234)

In a summary of the literature, Boaz *et al* (2015) indicate that vocabulary scores are taken to reflect a long-term memory quality that is learned and accumulated with age and, though other language skills have been found to deteriorate with aging (e.g., sentence complexity), vocabulary usually remains stable from adulthood or even increases with age (see Richards, 1976; Gold *et al*, 1995; Schaie, 1996). However, Bowles *et al* (2005) point out that there are a few issues with the research on vocabulary and age. Firstly, age trends and vocabulary knowledge is at least in part attributable to cohort effects where actual declines in vocabulary (e.g. as a result in the decline in reading) are mitigated by an increase in schooling which makes vocabulary knowledge appear stable. Furthermore, vocabulary knowledge is not generally studied in depth and there may be different age trends for different aspects of vocabulary knowledge (Bowles *et al*, 2005). This is a key limitation of vocabulary testing in general and discussed in the next section. Despite this, vocabulary tests are still the most widely used method in estimating vocabulary sizes (e.g. Kirkpatrick, 1891; Hartmann, 1946; Goulden *et al*, 1990; D'Anna *et al*, 1991; Nation & Beglar, 2007; Treffers-Daller & Milton, 2013; Brysbaert *et al*, 2016).

Brysbaert *et al* (2016) found that between the ages of 20 and 60, the average person learns 6,000 extra base words or about one new base word every two days. They also found that the knowledge of words increases not only with age but education as

well. Their explanation is that older people have come across more words than younger people so know more of them. This has significant implications for students at university as it suggests that undergraduates could potentially increase their vocabulary as they progress year on year.

2.13 Weaknesses in receptive vocabulary size testing

Vocabulary size tests are discrete and context independent for the most part, measuring breadth not depth. While these tests can give a more representative picture, albeit superficial, of the overall state of a person's vocabulary, examining the quality of vocabulary knowledge would require an in-depth inquiry of a learner's knowledge of a limited number of words (Read, 2000). Language knowledge does not exist in isolation and is complex (as seen in Table 2.1). Read (2000, p. 116) supports the view that 'vocabulary should always be assessed in context' but that 'the more we contextualise the assessment of vocabulary, the less clear it may be to what extent it is vocabulary knowledge that is influencing the test-takers' performance'.

Schmitt *et al* (2020) argue that a major limitation of the tests mentioned in Table 2.2 centres on inadequate validation evidence. They call for test developers to specify the test's purpose and audience (both the learners and educational context) as well as the particular aspects of vocabulary knowledge being measured, and the way test scores should be interpreted. Other issues pertain to what various item formats tell us about the underlying vocabulary knowledge of learners (Kremmel & Schmitt, 2016). For example, if a learner answers an item correctly, it does not necessarily demonstrate the learners' ability to use the target word in reading, writing, listening or speaking. Furthermore, the depth of word knowledge (e.g. do they know the item's word-family members, collocations and can they use it in the appropriate context) cannot be ascertained from answering the target item correctly. 'These two facets are directly related to our ability to interpret vocabulary size test scores in any meaningful way, because a raw size score without knowing what the learner can do with that amount of vocabulary is of little use' (Kremmel & Schmitt, 2016).

2.14 Measuring productive vocabulary knowledge

Because of the limitations in measuring receptive vocabulary, another method is to assess one's productive vocabulary knowledge. Productive vocabulary knowledge refers to not only knowing/understanding a word but also being able to use it in speech and writing (Schmitt, 2010; Maskor & Bharudin, 2016). Findings from several studies suggest that individuals have a larger receptive vocabulary than productive

but there is no consensus on the figures (Schmitt, 2010). Productive vocabulary knowledge is usually measured in two ways: controlled productive knowledge using Productive Vocabulary Levels Test (PVLT); or free productive knowledge where learners write an essay and the lexical richness of their writing is measured (Nizonkiza, 2016).

The PVLT is a version of the VLT mentioned in Section 2.9. Designed by Laufer and Nation (1995, 1999), it is a form/recall version of the VLT. The test contains missing words or gap fills, for example, 'This sweater is too tight. It needs to be stret____'. The missing words have one to six letters given as a clue. This could make some answers more obvious than others and there is no research on the effect that these different prompts may have (Schmitt, 2010). The fact that a clue is given makes the answer highly constrained which does not make it a test of truly productive word knowledge (Kojima & Yamashita, 2014). Furthermore, the test does not measure overall vocabulary size. 'Further, it is not productive in the sense that it does not require examinees to produce lexical items in the course of their spoken or written output' (Schmitt, 2010, p. 205). However, the PVLT can identify lexical weaknesses at a particular frequency level and help teachers place students quickly into ability groups based on their vocabulary knowledge (Kremmel & Schmitt, 2018).

There are other ways to measure vocabulary knowledge which are more context dependent and involve analysing learners' output, usually written (Schmitt & Schmitt, 2020). Most frequently measured vocabulary constructs in research analysing writing include diversity, maturity, content vocabulary (e.g. subject specific), academic vocabulary, and register (Olinghouse & Wilson, 2013). An advantage to using these measurements is that statistical analysis can easily be performed but conversely, they do not convey whether words are used accurately or appropriately (Schmitt & Schmitt, 2020).

Read (2020) believes that a possible measure of writing quality is through assessing the number of errors found in a text. Fewer errors with word choice, word form and register equate to a better-quality text. Looking at word usage errors in a text cannot be done solely through the use of computers but requires the researcher to analyse and code the data themselves (for example, see Faisal *et al,* 2017).

There are several ways to measure productive vocabulary knowledge that make use of computer software and statistics. One approach is to look at the lexical richness (or diversity) of students' written work. Measuring lexical richness involves quantifying

the degree to which a writer is using a varied and large vocabulary (Laufer & Nation, 1995). However, this approach does not indicate whether the words are being used appropriately or accurately; it counts word usage and cannot measure quality (Schmitt, 2010).

According to Kojima & Yamashita (2014) the most common approach to measuring lexical richness is based on the ratio of different words (types) to total number of words (tokens), known as the Type-Token Ratio (TTR). The number of tokens is the same as the total number of word forms, meaning individual words are counted each time they are used while the number of types is the total number of different word forms so that a word which is repeated is only counted once (Read, 2000). The TTR is sometimes referred to as lexical variation or lexical diversity and shows the range of words used in the output often referred to as 'range of expression' (Read, 2000). The closer the ratio is to 1 (or 100%), the more varied the vocabulary is, but the statistic is not directly comparable between corpora of different sizes (McEnery & Hardie, 2012). Olinghouse and Wilson (2013) refer to this as the breadth of words used in a text and is probably the most common vocabulary construct included in past writing research. More proficient writers have a larger vocabulary which enables them to use other words (e.g., synonyms) to avoid repetition so would have greater lexical variation (Read, 2000).

However, the main issue with this approach is its sensitivity to text length as when texts get longer the less likely it is that new words (types) will occur and certain words will start to be repeated; high-frequency words will be repeated more often as compared to low-frequency words (McCarthy & Jarvis 2007). Because lexical variation does not distinguish between the high frequency and low frequency words someone who knows more low frequency words could possibly show the same number of word types as someone who knows mainly high frequency words (Laufer & Nation, 1995).

Although writing about spoken output, another factor that affects the number of types produced which can be applied to written texts is the variety of different topics covered (Nation & Coxhead, 2021). The variety of topics can lead to doubling the number of types even if the number of tokens is the same. The diversity of topic strongly affects the nature of the output, but in naturalistic studies, it is very difficult to control for this (Nation & Coxhead, 2021).

Other measures of lexical richness are lexical originality (LO), lexical density (LD) and lexical sophistication (LS) and Lexical Frequency Profile (Laufer & Nation, 1995). Lexical Originality (LO) is the percentage of words in a given piece of writing that are used by one particular writer and no one else in the group. In other words, it measures the learner's performance relative to the group in which the composition was written so if the group changes, so does the LO, rendering the LO of a particular composition as unstable (Laufer & Nation, 1995).

Lexical density (LD) can be measured as the proportion of lexical words (content words) as opposed to function words (non-lexical) to the total number of words in a text (Engber, 1995). 'It can give some indication of the *content load* of the output and is commonly used to place text along the spoken to written continuum' (Schmitt & Schmitt, 2020, p. 225). Lexical words are nouns, adjectives, verbs and adverbs while function words do not provide any additional meaning, such as the preposition *to*. Morris and Cobb (2004) define function words as any word belonging to the closed set of words playing a grammatical role such as articles, prepositions, auxiliaries, pronouns and so forth (see Section 2.6). However, the validity of LD is challenged by the fact that a text may have quite sophisticated syntax (e.g. more participial phrases) so fewer function words which influence the density (Laufer & Nation, 1995).

Durrant and Brenchley (2019, p. 1929) state 'Researchers rarely state exactly what they mean by the term [lexical sophistication] …' so a clear definition is important. 'Lexical sophistication is the number of "sophisticated" words in the output ÷ the total number of tokens in the output' (Schmitt & Schmitt, 2020, p. 225). Another way of defining Lexical sophistication (LS) is the ratio of lexical words above a certain grade level or 'advanced' to the total number of words (Engber, 1995; Šišková, 2012). Olinghouse and Wilson (2013) refer to this construct as vocabulary maturity. However, what is considered 'advanced' varies according to the researcher rendering this an unstable measure according to Laufer and Nation (1995). Schmitt and Schmitt (2020) advise caution when defining what vocabulary is to be considered 'advanced' as this should be done in the context of the students (background, proficiency levels and learning goals). More recent studies have used word frequencies based on a reference corpus as guidelines for what is considered sophisticated; sophisticated words occur less frequently (e.g. *octogenarians*) while frequent words (e.g. *people*) are deemed less so (Kyle & Crossley, 2016; Kyle *et al*, 2018). Because academic language is found less frequently in general corpora, the number of academic words

used in a text can be used as a measure of lexical sophistication (Kyle & Crossley, 2016).

To mitigate some of the issues with these measures, Laufer & Nation (1995) developed what is known as the Lexical Frequency Profile (LFP) which shows the percentage of words a learner uses at different vocabulary frequency levels in their writing. The frequency levels are derived from The General Service List (GSL) of English words (West, 1953); and the AWL (Coxhead, 2000) (see Section 2.7.3). LFP uses software called vocabulary profilers which divide the words in the text into four frequency levels:

1. List of the most frequent 1,000-word families
2. List of the second most frequent1,000-word families
3. Academic Word List
4. Words that do not appear on the other lists or 'off-list' (Cobb, 2017).

Studies using LFP mainly focus on ESOL students rather than native speakers (e.g., Meara & Fitzpatrick, 2000; Muncie, 2002; Morris & Cobb, 2004; Kojima & Yamashita, 2014; Lutviana *et al,* 2015; Higginbotham & Reid, 2019). Schmitt (2010) reports that less proficient learners of English generate texts mostly consisting of the highest frequency vocabulary (first 1,000) with very little in the lower frequency bands and the AWL; in contrast, more advanced learners would use more lower frequency words. Laufer and Nation (1995) conducted a study using LFP analysing two pieces of writing (around 300 words) on 20 Israeli first year first semester UGs and 23 Israeli first year second semester university students. They found that 79.6% of the first semester students' work consisted of words in the first 1,000 band, 6.8% in the second 1,000, 8% from the AWL and 6.1% not on any list or 'off-list'. The second semester students' work comprised of 75.5% in the first 1,000, 6.1% in the second 1,000, 9.1% from the AWL and 8.1% 'off-list'. Arguably, these figures need to be read with caution as they are derived from only two pieces of work with quite a short word count.

In an updated version of Laufer and Nation's (1995) study, Higginbotham and Reid (2019) used LFP on 472 advanced second language learners 2,000-word essays (using both older and newer word lists – the GSL, the AWL as well as the NGSL and the NAWL) – and correlated results with a range of language proficiency tests (receptive) and essay scores (productive). They found that the essay and overall

scores generally correlated significantly with the LFP ratios. The essay scores correlated negatively with the high-frequency words and positively with the academic and off-list words. ‘In general, students that used a lot of high-frequency (GSL1) words in their essays tended to get lower grades overall and lower grades for their essays. In contrast, the students who used a higher proportion of academic (and off-list) words tended to get higher scores’ (Higginbotham & Reid, 2019, p. 18). These findings support Schmitt’s (2010) assertion that lower proficiency students rely mainly on high frequency words in their writing.

A further consideration of using the AWL in LFP is the list is not based on frequency and the words vary greatly in frequency, so the profile cannot be viewed in a sequential order (Schmitt, 2010). For example, words like *analyse* are more frequent than *invoke* (Coxhead, 2000) so a profile may contain more high frequency academic words than low. There are some technical issues with the way the texts are processed in terms of how errors are dealt with, the way in which proper nouns are counted and how formulaic phrases are treated (Meara, 2005; Smith, 2005). Schmitt and Schmitt (2020) believe that in addition to any analysis using frequency profiles, an evaluation needs to be conducted to see if the words have been used appropriately. Finally, some researchers (e.g. Smith, 2005; Schmitt, 2010) believe that LFP has problems showing vocabulary improvements as it is not detailed enough (for example, it does not show variety of frequent words). Despite these criticisms, Laufer and Nation (1995) claim that LFP:

- is a reliable and valid measure of lexical use in writing
- provides similar stable results for two pieces of writing by the same person
- discriminates between learners of different proficiency levels
- correlates well with an independent measure of vocabulary knowledge
- focuses directly on lexis, putting aside at least to some degree the influence of grammar
- a useful diagnostic tool as well as a sensitive research tool.

When evaluating vocabulary knowledge in written text, the type of measurement used is important. Olinghouse and Wilson (2013) conducted a study with native speaking American fifth graders which examined the link between measures of productive vocabulary (academic words, register, elaboration, content vocabulary, maturity, diversity) and the quality of students’ written work in three genres (story, persuasive and informative). They found that for story text, vocabulary diversity was a unique

predictor, while for persuasive text, content words and register were unique predictors. Finally, for informative text, content words were the strongest unique predictor, although maturity was also a unique predictor. These findings suggest that genre plays a key role in vocabulary use and quality is partly determined by the measures used. However, Read (2000) argues that it is very difficult to determine how large a role vocabulary ability plays in the overall quality of student writing without also giving learners a battery of tests to assess other aspects of their language knowledge as well.

2.15 Tools used to evaluate productive vocabulary

In applied linguistics, corpus linguistics is an approach that has become one of the dominant methods used to analyse language today (Anthony, 2013). Essentially, corpus linguistics is a scientific method of language analysis (Brezina, 2018a). There are various tools available for analysing language output in written texts; however, their functionality can be different. It is important to assess software tools for functionality as this largely prescribes what research methods are available to a researcher (Anthony, 2013).

Berberich and Keiber (2020) have compiled a list of 245 tools used in corpus analysis. A corpus is a collection of texts that is designed to be representative of some aspect of language (e.g. learner corpora represent the language of L2 users) (Webb & Nation, 2017). In other words, a corpus usually represents a sample of language or a small subset of the language production of interest (Brezina, 2018a). Concordancers are computer software programmes that allow for the quick analysis of texts and corpora. Concordancers are similar to one another in terms of their core functionality which is to retrieve concordances (a specific sequence of characters of any length – a word, part of a word or a phrase); frequency lists, collocations and keyword analysis (McEnery & Hardie, 2012). It is beyond the scope of this research to review all the tools available but will focus on a select few based on the goals of this research instead.

The following tools have been used in recent corpus research: AntWord Profiler (e.g. Csomay & Prades, 2018); Compleat Lexical Tutor (e.g. Morris & Cobb, 2004); Docuscope (e.g. Miller & Pessoa, 2018); #Lancsbox (Shi, 2018); MonoConc Pro (e.g. Gablasova & Brezina, 2018); Wordsmith Tools 5.0 (e.g. Chen, 2018) so are worthy of consideration. Throughout the literature review, key ideas in vocabulary research have centred around word frequency, types of vocabulary and measurement. It can

therefore be reasoned that is important to evaluate available tools in the context of these constructs in terms of functionality.

Concordance software is able to analyse large amounts of data very quickly and is useful for finding patterns of language in texts (Schmitt & Schmitt, 2020). There are several concordancers that can be either run online or downloaded onto a personal computer. The strengths and limitations of these are summarised in the Table 2.5.

Table 2.5 Software for analysing texts

Name of tool	**Strengths**	**Limitations**
AntConc/AntWord Profiler	Windows based, free, support group on Facebook, provides a range of functionality. AntWord Profiler can analyse multiple texts in batches. An entire corpus can be uploaded and analysed in one. Runs on numerous operating systems. Both are free to download and have support groups on social media.	AntConc are lemmatised but not PoS (part of speech) tagged, website contains tutorials on how to use it.
Compleat Lexical Tutor	Provides a range of tools for analysing vocabulary mainly, free, support group on Facebook, user friendly, continually updated.	Concordance file size is a max 60k words, links to the BAWE. VocabProfile (VP Classic) has a maximum file size of 35,000 words. It does not identify PoS nor group word forms into lemmas. It also does not recognise compound words. Key words in text will be repeated but will be referred to in different ways (e.g. Bob Ross = the artist = He).

Docuscope	Useful for rhetorical analysis, allows the user to interact visually with the data.	Not free, lack of user-friendliness of the original DocuScope interface for educational purposes. The original tool requires experience and training to generate reports.
#Lancsbox	Provides a wide range of analysis functions, produces word lists, uses TreeTagger (a tool for annotating text with PoS and lemma information), visualises language data such as collocation graphs; uses Windows and all major operating systems. Files can be word documents.	Interface is not appealing but is user-friendly; existing corpora does not include BAWE.
Sketch Engine	Contains 500 ready-to-use corpora, provides a range of functionality	Not free, create your own corpus up to a billion words, links to the BAWE.
WordSmith Tools	Fast, Windows based, produces word lists, key word lists and word clusters	Not free.

Sources: Weisser, 2016; Brezina, 2018b; Helberg *et al*, 2018; Corino and Onesti, 2019

Corino and Onesti (2019, p. 5) write 'LancsBox is certainly the most user-friendly and it offers some advantages such as the search for semantic categories … As for Sketch Engine, the querying possibilities are far more developed, but teachers and students do not need such a professional level of search'. Limitations of DocuScope Classroom are that it may classify "false positives" (rhetorical patterns that are not genre-relevant) which generate noise and skew the statistical data (Helberg *et al*, 2018). Weisser (2016) believes that AntConc is the best free concordance available which could be why Corino and Onesti (2019) state it is the most popular.

When analysing the vocabulary load of texts, there are three main programmes: Range Program (formerly called VORDS), FVORDS and VocabProfile (discussed in

the previous section) which shows how much and what vocabulary occurs in a particular text or group of texts (see Nation & Heatley, 1994). VocabProfile is a much simpler web-based version of Range and can be found on the website Compleat Lexical Tutor (Lextutor) (Heatly *et al*, 2002; Cobb, 2017). Both the Range and AntWordProfiler have to be downloaded which could be seen as a disadvantage whereas VocabProfile is available on the web (Lextutor). Range, VocabProfile and AntWordProfiler are all free.

The Range programme cannot distinguish between homographs and homonyms like *Smith* (the family name) and *smith* (blacksmith), and *March* (the month) and *march* (as soldiers do) (Nation, 2006). Proper nouns and nouns are not distinguished in Range and are counted in the same family and as the same type. How proper nouns are dealt with makes a significant difference to an output profile (Cobb, 2010). If proper nouns are treated in the same way as nouns, it could lead to overestimation of the number of high frequency words. VocabProfile has the option to recategorise proper nouns into the 1K frequency band or put into the off-list category. In this study, proper nouns were listed in the off-list category and not in the 1K band to give a more realistic account of the types of words students use. Students are expected to reference literature in their work and as a result their writing should contain many authors' names which are classified as proper nouns. Counting these names in the 1K frequency band would have given an unrealistic account of the number and types of words students use so VocabProfile was chosen over Range.

There are a few disadvantages of VocabProfile. It does not allow for several texts to be input at the same time nor keep track of which texts are contributing to which parts of the profile (Cobb, 2017). It also cannot manage large texts and has a size limit of about 200,000 characters/35,000 words. Some larger texts move fairly slowly through the server-side processing. Conversely, Cobb's (1997) entire PhD thesis established the validity of an earlier version of Lextutor. Cobb's work is based on Laufer and Nation's (1995) study where they validate vocab profiling as a research instrument for measuring the lexical richness of texts. Finally, other studies (e.g. Meara & Fitzpatrick, 2000; Morris & Cobb, 2004) have used VocabProfile to assess productive vocabulary in a second language and to predict the academic and pedagogic performance of TESL trainees.

VocabProfile breaks texts down by word frequencies in the language at large, not in the text itself. As previously mentioned, the vocab profiler is based on Laufer and

Nation's Lexical Frequency Profiler (see Laufer & Nation,1995) and was created to measure lexical richness by calculating the percentage of words in a text that fall into the most frequent words in the English language (1–1,000 band or K1) and the second most frequent words (1,001–2,000 band or K2), academic words, and less frequent words not found in any of those lists (Laufer & Nation, 1995).

AntWordProfiler has two tools: the main tool is for vocabulary profiling while the other is for editing. The vocabulary profiling tool generates vocabulary statistic and frequency information about a corpus of texts loaded into the program (Antony, 2021). It compares the files against a set of vocabulary level lists that can be uploaded. It has a number of extra features such as a thesaurus option and is also compatible with AntConc; for example, token settings can be matched. There is an additional editing tool where the file can be edited and the immediate effect of that can be seen on the vocabulary level of the words and the text can be seen. The main advantage of AntWordProfiler is that it gives the user greater control in terms of analysing individual texts or the corpus as a whole.

A final vocabulary analysis tool that needs to be mentioned is LancsLex, offered by Lancaster University. This is a free, online tool that analyses texts with the new GSL discussed in Section 2.7.1. LancsLex is a lexical tool that analyses the lexical coverage and complexity of texts by comparing it to the new GSL (Brezina, 2017). A benefit of this tool is that it allows the user to decide whether to include proper nouns and numbers in the calculation of text coverage. However, a major limitation is there is a limit to the size of the text that can be analysed (50,000 characters) so a corpus cannot be uploaded.

Despite the number of tools available to analyse texts, concordancers in general have limitations and analysing corpora does not answer all research questions in linguistics (McEnery & Hardie, 2012). All these tools require much preparation of the data before any analysis can be done (Antony, 2013); this is time consuming as texts often need to be cleaned and reformatted before uploading. While concordancers can help to identify patterns in language, a key limitation is they cannot ascertain why those patterns occur. They also cannot give any indication if a word is used correctly. A finer grained analysis incorporating word usage in texts may be necessary for this but also to glean a richer, more nuanced understanding.

Another key consideration is whether the corpus is balanced or representative. A balanced corpus usually covers a range of text categories but currently there is no

reliable scientific measure of corpus balance, so any claim of balance is an ‘act of faith rather than a statement of fact’ (McEnery *et al*, 2006). However, very specialised corpora do not need to be large as more general corpora to yield reliable results as they are more likely to represent a particular register or genre (Koester, 2010).

2.16 Conclusion

After reviewing the literature, vocabulary plays a key role in both reading and writing, essential skills for success in higher education. It is evident from the review that there is a clear gap in the research surrounding university students, native speaking undergraduates, and their language needs within the context of their disciplines. The review highlighted several issues with testing receptive vocabulary sizes but despite this no validated alternative has been established in the field. Several key questions persist around native speaker’s vocabulary sizes and whether they can predict educational performance (Nation & Coxhead, 2021). Equally, research measuring productive vocabulary knowledge has typically focused on L2 speakers rather than native. This potentially has important implications for UG students that need investigating. What vocabulary knowledge, both receptive and productive, do students possess and does this change during their undergraduate journey? Key questions remain surrounding their receptive vocabulary size, their productive knowledge of general, academic and subject-specific vocabulary and how this links to their academic achievement. Building on existing quantitative research, this piece of research asks the following questions in order to make an original contribution on UK native speakers of English and language in higher education. The following questions will be researched around receptive knowledge:

1. What is the average receptive vocabulary size of English-speaking undergraduate students?

2. What types of words do students not understand?

3. Is there is a link between vocabulary sizes and academic achievement (measured by expected final degree classifications)?

And around productive knowledge:

4. What is the level of vocabulary richness in written assignments of monolingual and bilingual (using two or three languages including English routinely in the home) English speaking undergraduate students using lexical frequency profiling?

5. What are the changes, if any, in word usage from the NGSL 1, NGSL 2, NGSL 3, NAWL and ‘off-list’ in student assignments between the stages of study?

6. Is there a correlation between the percentage of academic words from the NAWL in students’ written work and assignment scores?

Chapter 3: Methodology

This chapter outlines and justifies the methodological approach that was taken in this research project including data collection and analysis, reliability, validity and sampling techniques. Ethical considerations are also addressed as well as positionality, potential for bias and reflexivity. As there were two phases to the project, the first phase is addressed initially followed by the second.

3.1 A quantitative approach

This section will justify why a quantitative design was employed in this research project. Quantitative research is associated with the philosophy of positivism. Positivists believe we can study social reality objectively by applying the procedures and techniques employed by natural scientists. Hence, most positivist researchers adopt the scientific method where individuals begin with a theory or hypothesis and then collect data to either support or refute the theory (Creswell, 2009). Quantitative research involves testing out a hypothesis or theory objectively by examining the relationship among variables (Creswell, 2014; Paltridge & Phakiti, 2015).

Punch (2005) states that it is typically directed at theory verification rather than theory generating. This is also known as deductive research since we draw logical conclusions from the evidence (Newby, 2010). As this research is quantitative and involves theory/hypothesis testing, it is reasonable to construct a hypothesis based on logic and reasoning sometimes referred to as the hypothetico-deductive model or H-DM (Teddlie & Tashakkori, 2009). This is a model used by quantitative researchers that involves creating a hypothesis based on a theory or past experience, deducing the observable consequences that must occur if the hypothesis is true and testing of the hypothesis by collecting new data and using statistical analysis to check for statistical significance of the finding (Teddlie & Tashakkori, 2009).

In other words, the H-DM entails producing ideas and devising ways to test them. In this particular project, while the research questions are not necessarily framed in a hypothesis type statement, there are several hypothesis-like assumptions that are being tested. Firstly, as result of being asked a number of times about the meaning of high frequency vocabulary in sessions (e.g. implication), it might be expected that

the vocabulary sizes of students at the University are small as compared to those found in the literature review. Testing and estimating the vocabulary sizes of students, investigated this intuition. Secondly, the research also tests the assumption that there is a link between vocabulary size and academic achievement. The research aims to examine whether students who have a larger vocabulary size tend to have better achievement (measured by marks on their assignments) and whether the converse is true (smaller vocabulary sizes correlate to lower marks).

In the first phase of the research, the theory (hypothesis) posited in this project after reviewing the literature is that the receptive vocabulary size of undergraduates is between 10,000 and 11,000-word families and that there is a link between vocabulary size and academic achievement. The hypothesis in the second phase of this research was based on the first phase; it was conjectured that productive knowledge of vocabulary in terms of vocabulary richness would be low, but students would experience some change in their academic vocabulary during their studies. These hypotheses were then tested in this research project.

Positivists believe knowledge is based on careful observation and measurement of the objective reality that exists 'out there' in the world (Creswell, 2009, p. 7). This is in accordance with the objectivist epistemology, and it means data (or evidence) is collected that is numeric and measurable. However, the truth and reality can change as new knowledge develops; this is recognised by the researcher.

Two key characteristics of quantitative research is that it is objective by avoiding personal opinion, values and judgements and it also uses objective measures or tests to extract variables (Paltridge & Phakiti, 2015). Making inferences about variables and generalising findings is another key characteristic of quantitative research (Paltridge & Phakiti, 2015).

In Phase I, vocabulary sizes (receptive) were measured by using a test based on dictionary sampling methods (Goulden *et al*, 1990). The results are quantitative as it gives an estimate to the number of words an individual knows. In Phase II, productive vocabulary was measured by using vocabulary profiling software called AntWordProfiler (Anthony, 2021) and the results were again numeric. It is in the spirit of positivism that this research was undertaken by being objective through testing, measuring and estimating the words students know and use in their writing.

Measurement in quantitative research requires the data to be analysed through numeric and statistical means (Paltridge & Phakiti, 2015). A qualitative approach was

rejected since this research involved measuring vocabulary sizes and identifying word frequencies, both generating numeric data. The organisation and analysis of the data in the project was drawn from the principles found in the field of corpus linguistics which uses the scientific method. Corpus linguistics is essentially a quantitative methodology which identifies word frequencies and patterns in language data. Statistics are essential in helping to analyse and make sense of this data (Brezina, 2018a).

3.2 Rationale for a two-phase research project

The literature review revealed that there is disagreement among vocabulary researchers as to whether the receptive/productive concept is dichotomous or operates on a continuum. Students' receptive and productive vocabulary needs are not equal either (Malmström *et al*, 2018). Added to this, González Fernández and Schmitt (2020) assert that measuring all the components of word knowledge – form, meaning and use as per Nation's (2013) conceptualisation – is practically impossible in a single study. As a result, how these components interrelate is largely unknown. Taking these factors into account and for the purpose of this research, receptive and productive vocabulary knowledge are treated separately.

The project was conducted in two separate phases (see Figure 3.1). In the first phase, an estimate of UG students' receptive vocabulary knowledge was measured by testing vocabulary sizes using a test devised by Goulden *et al* (1990). This test measures the form/meaning link component of word knowledge. In Phase II, an estimate of UG students' productive vocabulary knowledge was measured by examining the lexical richness of UG students' written work using AntWordProfiler. Lexical richness or diversity demonstrates the breadth of words used in a text (Olinghouse & Wilson, 2013) which corresponds to the word component 'use' (where, when, and how often can we use this word?) in Nation's (2013)'s conceptualisation.

In other words, Phase I examines how much vocabulary students know, and Phase II investigates what students can do with that knowledge. By looking at both the receptive (form/meaning link) and productive (use) vocabulary knowledge of UG students, a more comprehensive and accurate overview could be determined to meet the overall aim of the project which was to explore undergraduates' vocabulary knowledge. In support of this, Schmitt (2010) believes to gain an understanding of overall vocabulary knowledge a researcher should use both a receptive and

productive measure. Ideally, this would be conducted with the same participants; however, this was not possible in this instance.

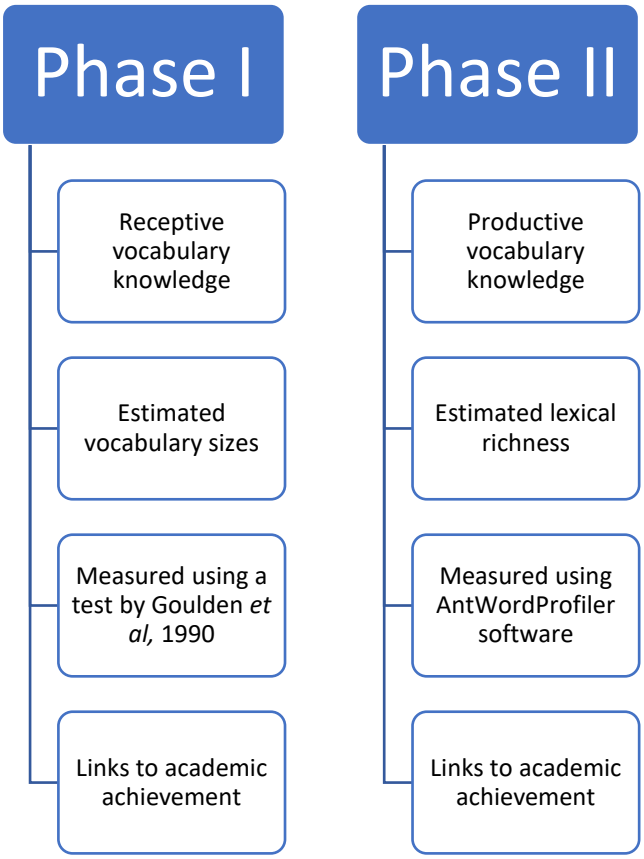


Figure 3.1 Phases of the research

3.3 Phase I methods of data collection: Having the right tools

3.3.1 The vocabulary size test

Nation (2013) believes a recent interest in how children's vocabularies grow has resulted in a corresponding interest in the vocabulary size of native speakers of English. Because vocabulary size tests estimate the total number of words a person knows, they are useful when comparing groups of students and measuring long-term vocabulary growth (McLean & Kramer, 2015). Another reason is to find out what vocabulary knowledge is needed for reading (Nation & Coxhead, 2021). Furthermore, Nation and Coxhead (2021, p. 5) emphasise that 'vocabulary size is central to

language use …'. In addition, 'the vocabulary size of native speakers of English is of interest to language teachers because it provides one kind of goal for learners of English as a second or foreign language' (Nation, 2013, p. 522). Finally, as discussed in the literature review, most measurements and discussions of vocabulary have centred on size or receptive vocabulary (Schmitt, 2010). For these reasons, in Phase I, the research project focused on vocabulary sizes of native speakers of English.

Several tests have been created to measure receptive vocabulary knowledge which were critiqued in Section 2.9. 'There is no commonly accepted standardised test of English vocabulary' (Schmitt & Schmitt, 2020, p. 221), resulting in the various estimates of vocabulary size seen in Table 2.3. This leaves a question as to what kind of test should be used to measure vocabulary sizes of native speakers.

Goulden *et al*'s (1990) test was chosen for this research because most of the vocabulary tests discussed in Section 2.9 are aimed at L2 learners rather than L1 or native speakers. Goulden *et al* (1990) originally used their test on undergraduates providing further results from which to interpret the data. Because this test asks participants to recall the meaning of words, there is less guessing involved as compared to multiple choice (recognition) tests such as Nation and Beglar's (2007). As discussed in Section 2.9.1, research into estimating receptive vocabulary sizes of native speakers has traditionally used dictionary sampling methods (Hartmann, 1941; Goulden *et al*, 1990; D'Anna *et al*, 1991; Treffers-Daller & Milton, 2013). Additionally, other tests based on frequency levels or levels tests do not report a global score which would not give an overall estimate of the number of words students know. Finally, Goulden *et al*'s (1990) test is also easy to administer and cost effective since it uses a simple yes/no format, allowing for a greater number of students to be tested.

A further rationale for choosing the vocabulary test devised by Goulden *et al* (1990) pertains to replication. One of the issues with vocabulary research is that few studies are replicated (Schmitt, 2010). As a result, there have been recent efforts to actively promote and facilitate replication studies in second language research (Marsden *et al*, 2018). This lack of replication is problematic as it makes it difficult to confirm and refine results which means key information in the field is often based on findings from single studies (Schmitt, 2010). In pure replication studies the same type of participants with the same instrument are used to see if the results are reliable. In approximate or partial replication research, modifications are made to two or more elements of the original study (Porte & McManus, 2019). In Phase I, a partial

replication of a similar study – Treffers-Daller & Milton (2013) – was conducted in order to compare results and confirm whether UG students have small vocabulary sizes.

Based on Porte and McManus's (2019) considerations for selecting studies to replicate, the reasons for choosing Treffers-Daller & Milton's (2013) study are as follows:

1. **The general topic of the original paper is one that continues to generate much debate.** As seen in the literature review, estimating vocabulary sizes of both adults and children has a rich history of research dating from Kirkpatrick (1891) to more recent studies such as Brysbaert *et al,* 2016. The topic generates as much interest today (Nation, 2013; Nation & Coxhead, 2021) as it did over a hundred years ago.
2. **The original paper's findings are not consistent with previous or subsequent work in the area.** Estimates of vocabulary sizes of university graduates have varied greatly from 215,040 words (Hartmann, 1946) to 11,000 (Treffers-Daller & Milton, 2013).
3. **The original paper continues to be cited in publications.** There are no metrics available for Treffers-Daller and Milton's (2013) study from the publisher's website. However, Google Scholar indicates that their research article, particularly their findings, has been cited by 99 authors to date (for example, Brysbaert *et al,* 2016; Masrai & Milton, 2017; Quigley, 2018 amongst others).
4. **The original study identified limitations.** Treffers-Daller and Milton (2013, p. 168) state that 'The sample in this study is comparatively small so a repetition of the study with a larger and more diverse sample of monolingual university students is sure to be useful'. The total sample in Treffers-Daller and Milton's (2013) study was 161 participants while this study tested a total of 389 which is more than double the size.
5. **The journal in which the paper was first published is a prestigious one.** It can be argued that prestigious journals are those that are peer reviewed where the integrity of the research is scrutinised by experts in the field which ensures the article's quality (Elsevier, 2021). This enables researchers to build on what could be considered an already sound and valid piece of work (Porte & McManus, 2019). Treffers-Daller and Milton (2013) published in *Applied Linguistics Review* (ALR), an international, peer-reviewed journal which

'serves as a testing ground for the articulation of original ideas and approaches in the study of real-world issues in which language plays a crucial role. ALR brings together critical reflections of current debates and new theoretical and empirical research' (De Gruyter, 2021). The latter statement not only supports point 1 above but also seeks to establish itself as an important publisher of scholarly work. The 5-year impact factor (number of times the journal is cited over a 5-year period) for this journal is 2.31 (De Gruyter, 2021). An impact factor of 3 is considered good and the average score is less than 1 (SCI Journal, 2021). We can, therefore, determine that a score of 2.31 is 'good' and ALR is a quality journal.

6. **The participants chosen/assigned in the original study were very similar to those the researcher works with.** The participants in Treffers-Daller and Milton's study (2013) were first, second- and third-year undergraduates, the same target audience of this research and the groups the researcher was teaching at the time this project was devised.
7. **Effect size data is not presented or is not convincing.** The effect size was not reported in the original study. Effect size is important in indicating how large any differences are between variables (Brezina, 2018a). The effect size is reported in this study.

A partial replication of Treffers-Daller and Milton (2013)'s study was implemented in Phase I as this research deviated from their study in a number of ways. Goulden *et al*'s (1990) sample was divided into 250 words which were sub-divided into five tests in addition to a list of 221 words that were unlikely to be known. Treffers-Daller and Milton (2013) used test five of the sub-tests as well as the list of words that were unlikely to be known for all their participants. In this research, all five tests were used to ensure the entire sample of words from the dictionary were represented, as well as the list of unlikely-to-be-known words.

Other key differences in methodology were related to the sample size and the courses students were recruited from as well the universities. Treffers-Daller and Milton's (2013) participants were drawn from UG degree courses in Humanities at Swansea and UWE and one course in Speech and Language Therapy at City University. This research recruited from UG degree courses in American Studies, Information Technology & Management, Biology, Child and Youth Studies, Computer Forensic Investigation, Computer Science, Construction Management and Property Development, Education Studies, Graphic Design Law, Popular Music with Music

Technology, and Sociology at one East Midlands university. The total sample in Treffers-Daller and Milton's (2013) study was 161 participants while this study tested a total of 389.

In addition to the vocabulary tests, Treffers-Daller and Milton (2013) asked students to complete a short questionnaire about the volume of reading they engaged in, to allow comparison between subjects. That questionnaire was not administered in this piece of research. Students were tested in the second semester of their study year whereas Treffers-Daller and Milton (2013) tested in the first term.

In terms of how the test was administered in this research, participants were only given one test out of the possible five sub-tests. Each test was rotated to ensure a balance (i.e., first group received test one, second group, test two and the process was repeated). The participants went through the test and marked the words they knew (i.e., they had seen it before and could express at least one meaning). Then they found the last five words they ticked and to show they knew the meaning of them gave a synonym, definition or used it in a sentence or diagram. They checked their explanations in a separate handout with the words and definitions (see Appendix 1). If more than one of the explanations was not correct, they had to go back through the list, beginning with the sixth to last word they ticked. They had to write a meaning of that word and check it in the definition handout. They continued this until they had a sequence of four ticked words (which may have included some of the original checked five) that were explained correctly. To estimate the participant's vocabulary size, the number of known words from the 50-item test were multiplied by the ratio that the sample of words bears to the total number of words in the dictionary, in this case 500. If participants scored above 15,000 words they checked the list of words not likely to be known (see Appendix 1). For the words they knew on this list, they gave a synonym or definition as per the main test. Each word on this list represents 100 words. The score from the list of 'likely to be unknown' is added to the score from the test to give overall estimated vocabulary size.

3.3.2 Limitations of the test

All research has limitations. It is therefore important as a researcher to acknowledge those limitations in order to prevent any misinterpretation of the findings. Bryman (1988, p. 101) believes that 'quantitative research conveys a view of social reality that is static in that it tends to neglect the impact and role of change in social life'. In terms of vocabulary testing, this is true. By testing and therefore estimating the vocabulary

sizes of undergraduates at a particular point in time, it can only ever be a snapshot of their knowledge which could change. The knowledge from the test is very limited as the strength of the form/meaning link is not examined in depth. The test does not indicate what linguistic journey students have been on before entering university or potentially after. However, testing all year groups may give some indication where they are on their journeys. As pointed out in the literature review, a major limitation of the test is the fact that it is self-reporting (relying on test takers to assess their lexical knowledge) rather than verifiable evidence (Read, 2000). There was also a small element of self-assessment as the participants marked their own test. When checking their meaning with the word's definition, their interpretation could be either be too strict or too lenient.

3.3.3 Measuring academic achievement

Most studies that investigate the effects of a factor (e.g., interventions, peer learning) on academic achievement use grades as a means of measuring achievement (Carrell *et al*, 2009; Dancer *et al*, 2015; Veas *et al*, 2015; Vulperhorst *et al*, 2018). This study used expected degree classifications in Phase I to measure academic achievement as per Treffers-Daller and Milton (2013). Degree classifications are based on a weighted average of all the module marks in the second and third year of UG study. Because testing took place before students had finished their degrees, they were asked to provide their expected degree classification as the actual was not available.

It must be acknowledged that there are limitations to using expected degree classifications as a measure of achievement. Students were asked to self-report their average marks at the point of taking the test. There was no way of checking whether these corresponded to their actual final degree classification. Students in reality could have over- or under-estimated their scores.

3.3.4 Personal information questions

Before the vocabulary test began participants were given a small pre-test questionnaire (see Appendix 2). This was designed mainly to find out their expected degree classification, but it was also to explore the strategies they employ to deal with unknown words they encounter. Because vocabulary learning strategies do not answer the research questions, no analysis was undertaken.

3.4 Phase I: Piloting the test

A pilot study of the test and pre-test questionnaire was conducted before the main data collection phase in Phase I (see Table 3.1).

Table 3.1 Pilot questions for Phase I

Time frame	• Approximately how long will the vocabulary test take? • Is one month long enough for students to record unknown vocabulary in their diaries?
Language	• Are the instructions clear on both the test and diary?
Missing items	• Are there any missing items from the test and diary?
Reliability	• Is the vocabulary test reliable? E.g., equivalent forms' reliability to be conducted on the test.

The test was firstly piloted on colleagues and family in order to determine the time it would take participants to complete, which was approximately 20-25 minutes. It helped to identify typos in the instructions and potential questions that needed to be addressed during the administration of the test. It also confirmed there were no missing items. Finally, there were five sub-tests used in the research so a further pilot study was conducted with five undergraduate students which allowed the researcher to test for reliability. Participants took all five sub-tests, and their scores were correlated to see if the tests were equivalent. Reliability will be discussed further in Section 3.8.

3.5 Pilot: The diary

Receptive vocabulary knowledge is mainly ascertained through testing as discussed in Section 2.9. As previously stated, the second research question was to form an understanding of the specific words students came across in their studies in a naturalistic setting, for example from the readings, lectures and discussions with other students. This was important as receptive knowledge is the knowledge required to listen or read (Webb & Nation, 2017). Receptive knowledge involves being able to identify to some degree what a word looks and sounds like as well as the meaning

(Nation, 2013). Students were asked to self-report the words they did not recognise, had not seen before and did not know the meaning. It was also important to determine whether these words could be found on the NGSL or the NAWL to be discussed further in Section 3.13.2. The NGSL and the NAWL were used in the second phase of the research, and this gave some indication of the value of the word lists. In order to do this, students (15 in total) from a second-year module in education were asked to record in a diary voluntarily over a two week period the following: the unknown words they encountered, the date, a definition of what they might think the word means, and where they saw or heard the word (e.g. newspaper, textbook, lecture notes, seminar discussion).

By looking at these three items, the researcher wanted to investigate the following as a result of the literature review:

- Are the unknown words that students record found on the NGSL or the NAWL?
- Are there any specific unknown words that have the highest frequency? If so, what are they, where were they encountered the most?
- Which location of unknown words is most frequent? In other words, do students find the majority of unknown words in journal articles, for example?
- Which location of unknown words is least frequent?

There are three main ways in which diaries are used in social research: the diary as a method of data collection, the diary as a document, and the diary as a log of the researcher's activities (Bryman, 2008). In this research, I wanted to use the diary as a method of data collection and the type of diary that is being used in this research is what Elliott (1997) describes as a researcher-driven diary. Diaries are structured and participants log observations which act as a substitute for accurate scientific observation in situations where it is not possible for a researcher to be present. The diary is not open in the sense that participants record their thoughts, feelings and impressions in a free-flowing manner. It is based on personal observations, not opinions, so is positivist in nature.

In terms of language research, diaries have been used in a qualitative way to record students' and teachers' experiences of a second language (see Seliger & Shohamy,1989). Such diaries are very subjective as they are kept in order for participants to record their feelings during language lessons, attitudes to language learning and self-analysis on the language learning methods. There are very few quantitative studies using diaries in language research (see Schmitt, 2010).

On a pedagogical level, diaries can be used to heighten learners' awareness of their learning (Fry, 1988). A potential and unintentional benefit that could result from students recording unknown words is that it may increase their awareness of their own vocabulary knowledge. Some academics (e.g., Curtis, 2006) have stressed the importance of promoting word consciousness in the role of vocabulary instruction in adult basic education. When students have an awareness of and an interest in words, they make greater gains in growing their vocabulary (Anderson & Nagy, 1996 cited in Curtis, 2006). Recording unknown words in a diary may lead to an increased level of word consciousness in undergraduates who participate in this research project. In turn, they may find a renewed interest in learning new words which could ultimately lead to a growth in the level of their vocabulary.

There are, however, a number of limitations with using diaries in social research. Bryman (2008, p. 228) lists a number of problems:

- They tend to be more expensive than personal interviews (because of the cost associated with recruiting diarists and of checking that diaries are being properly completed).
- Diaries can suffer from a process of attrition, as people decide they have had enough of the task of completing a diary.
- This last point raises the possibility that diarists become less diligent over time about their record keeping.
- There is sometimes failure to record details sufficiently quickly, so that memory recall problems set in.

Self-reporting also needs to be a consideration with using diaries to collect unknown words. A key question is whether students will record *all* the words they do not know. It would not be possible to know if they did or did not. Another question is whether the diaries are verifiable and as Schmitt and Schmitt (2020, p. 35) indicate '… it is impossible to know how accurately learners can judge their own level of knowledge'. And as Bryman (2008) points out diarists can become less diligent over time about their record keeping. This could lead to a possible underestimation of the number of words students report.

After careful consideration, diaries were piloted with a group of 15 students. Only one student completed a diary in the 2-week period. This supports Bryman's (2008) findings as the pilot showed high attrition, low diligence and lack of detail on the part of the participants. Therefore, it was decided to truncate the approach and a

vocabulary record sheet was devised while still trying to maintain the essence of how diaries can contribute to our understanding of unknown words.

3.5.1 Vocabulary record sheet

Because of the limitations outlined above for the use of diary research, a vocabulary record sheet was devised (see Appendix 3) to increase participation. The vocabulary record sheet consisted of only two columns, one to identify the unknown word and another to record where it was encountered. This was important to know in case all the unknown words emanated from one source, say a newspaper.

The sample for this was taken from the students who participated in the vocabulary test. They were asked at the end of taking the test if they wanted to participate in keeping a record sheet for one week. There were 27 participants in all who volunteered, and they recorded 157 words in total. However, this sample is small for quantitative research as larger samples are more likely to be representative of the population (Gray, 2018). Caution was needed when analysing and reporting the findings as this limits the possibility of making generalisations and coherent connections.

3.6 Phase I: Handling the data

In quantitative research, identifying the type of variable that is being measured is important in the overall research design as this produces various kinds of data which determines the method of analysis. In the first part of the research, the two variables are an estimate of an individual's vocabulary size and their expected degree classification. Furthermore, vocabulary sheets were used to collect the unknown words students encounter in lectures, books, newspapers, speaking in seminars and so forth.

Numeric data can fall into four broad categories: nominal, ordinal, interval or ratio. The vocabulary test is neither a nominal or ordinal variable as nominal variables differentiate items based on placing them in a category and ordinal variables on ranking them. However, the vocabulary sheet data will be nominal as the unknown words will be placed into three categories based on the type of vocabulary discussed in the literature review: general English, academic English and subject specific English. The variable of degree classification is ordinal as degrees are ranked in order from first class to an ordinary PASS degree (e.g., 1st, 2:1, 2:2, 3rd, PASS).

It could be argued that the variables for the vocabulary test could be either interval or ratio. Black (1999) argues that the distinction between interval and ratio data is subtle since both can produce continuous data. However, the real question is whether an individual can score a zero on the vocabulary test. Interval data has no zero point (where the trait does not exist) and there is equal distance between intervals while ratio data has an absolute zero (Field, 2013).

Technically, an individual could score a zero on the vocabulary test, but this is not likely as the words at the beginning of the test are high frequency (e.g., *as*, *dog*, *pony*). Besides, it would be impossible for an undergraduate to have an estimated vocabulary size of zero. Moreover, Black (1999) gives the example that IQ scores are interval data because a score of zero would have no meaning. Lastly, Treffers-Daller and Milton (2013), using the same vocabulary test in their research, did not start at zero when reporting their findings for the number of words known. They categorised the data into frequency bands in widths of 5,000 starting from the number one thus indicating that a vocabulary size of zero does not exist.

Quantitative variables are divided into discrete and continuous and like interval and ratio variables the distinction between them is not always clear. Discrete variables can only take on certain values whereas continuous data can take on any value (Saunders & Brown, 2008). In terms of vocabulary size, it is not possible to know half a word so we could say the data is discrete. Klugh (1986, p. 12) states, 'If a test has 50 items, it would appear that we have a discrete variable; you can get 39 or 40 right but not 39.126 right. However, we usually treat test scores as if they were continuous variables'. Furthermore, Salkind (2014) gives test scores as an example of a continuous variable. It is possible, for example, to score 78.125% overall on a test (25 correct out of 32) or 50% (16 out of 32); in other words, any number of values on a test can be scored. Thus, the scores from the vocabulary test will be treated as continuous data.

As the first part of the research implemented a test to collect data, it is important to establish whether the data is parametric or non-parametric in order to apply the most appropriate statistical test. According to Cohen *et al* (2007, p. 414), 'Parametric tests are designed to represent the wide population, e.g., of a country or age group. They make assumptions about the wider population and the characteristics of that wider population, i.e., the parameters of abilities are known'. Parametric tests assume that there is a normal distribution of scores in the population producing a bell-shaped

Gaussian curve whereas non-parametric tests make no assumptions about the distribution or characteristics of the population.

Field (2013) contends that testing for normality is not always necessary due to the central limit theorem which states that when samples are large (above about 30), the sampling distribution will take the shape of a normal distribution regardless of the shape of the population from which the sample was taken. Furthermore, Field (2013) states that with sample sizes of 100 and over we get a sampling distribution that is more normal. The total sample size of this research was 389 participants with 130 in Stage 1, 139 in Stage 2 and 120 in Stage 3. It is possible to assume from the sample size that the data is normally distributed.

If the data is assumed to be normally distributed from the central limit theorem, then parametric tests can be applied. Further evidence for parametric testing can be found in the types of data collected. Nominal and ordinal data are considered to be non-parametric, while interval and ratio data are considered to be parametric data; non-parametric data are often derived from questionnaires and surveys (though these can also gain parametric data), while parametric data tend to be derived from experiments and tests (e.g., examination scores) (Cohen *et al*, 2007).

Applying Cohen *et al*'s (2007) rationale to this research means that the data produced by the vocabulary test can be analysed using parametric statistical analysis while the diaries will be subject to non-parametric analysis. Another assumption of parametric statistics is that the sample is large enough to be representative of the population (Salkind, 2014) and Gray (2018) suggests this number to be at least 30. The sample size for the test was 389 participants so this satisfies this assumption. A summary of the kind of data that was produced from Phase I of the research can be found in Table 3.2.

Table 3.2 Types of data collected in Phase I

Vocabulary test (Goulden *et al*, 1990)	**Vocabulary sheet**
Interval variables	Nominal variables
Continuous data	Discrete data
Parametric testing	Non-parametric testing

3.7 Data analysis

3.7.1 Parametric statistical analysis

Having established the type of data that was collected from the vocabulary test this section will discuss the method of analysis. Firstly, it was important to examine group characteristics, i.e., what is common within a group as well as differences between groups. Groups in this sense were stages of study (Stage 1, 2 and 3). Thus, the mean score for each year group was calculated as well as the standard deviation. It is assumed that the distribution of the data is normal so the standard deviation, which is a standardised measure of the dispersal of the scores or how far away from the mean/average each score is, was calculated for each stage (Cohen *et al*, 2007). 'A low score indicates that the scores cluster together while a high standard deviation indicates that the scores are widely dispersed' (Cohen *et al*, 2007, p. 512). The standard deviations (SD) of test scores in Stages 1, 2 and 3 were 2,674, 3,016 and 2,944 and the means were 10,070, 11,614, and 11,582 respectively (see Table 4.1). The sample standard deviation (SD) for Stages 1, 2 and 3 are at 27%, 35% and 34% of the mean. This indicates that the majority of scores are close to the mean, not widely dispersed, and the distribution of the data is likely to be normal.

SPSS was used to conduct an ANalysis Of Variance (ANOVA) to test the difference between means. This provided pairwise comparisons between vocabulary sizes from Stages 1 to 2, 2 to 3 and 1 to 3. ANOVA was also used to see if there were any differences between the tests (i.e., to establish that differences were on account of the test not the participant). *P*-values were calculated and if low (<0.05) then the two groups were statistically different. Finally, Boneferroni post-hoc tests and effect sizes were conducted to minimise Type I errors and to establish the size of any differences found in the data (Brezina, 2018a).

3.7.2 Non-parametric statistical analysis

The research also looked at whether there is a correlation between vocabulary sizes and academic achievement. This can be estimated by using a correlation coefficient. Spearman's rank correlation coefficient (r_s) was used since the data analysed was ordinal and interval. Spearman's (r_s) is used to calculate correlations for ranked and ordinal data (non-parametric) while Pearson's correlation coefficient is used for interval or ratio data (parametric) (Field, 2013). Pearson's can only be calculated if both variables are continuous (e.g., interval) (Salkind, 2014) while Spearman's can be used when one variable is ordinal and the other interval/ratio (Bryman, 2008). Because degree classifications are ordinal it was decided to use the non-parametric Spearman's. The sample size is also important when determining the outcome of

statistical tests. The larger the sample, a smaller statistic value is required to reach significance (Punch & Oancea, 2014). *P*-values were looked at to evaluate the significance of the value of the correlation coefficient. If *p* is less than 0.05 ($p < 0.05$) then the correlation is significant. Gray (2018) states the significance level is usually set at 5% or 1% but the former was selected due to the sample size as it is large enough but not in the thousands.

When using correlation coefficients such as Spearman's, issues can arise in their interpretation. Brown and Rogers rightly point out that correlation coefficients do not indicate causality. They state that 'Correlation coefficients are no more than estimates of the degree to which two sets of numbers are related' (2002, p. 190). Just because there is a strong relationship between two variables, it does not mean that one caused the other.

The raw data from the vocabulary sheet was used to create frequency tables in order to establish the types of words students do not know. Non-parametric statistical tests are used with frequency data (Foster *et al*, 2015) and are often referred to as distribution free tests (Burns, 2000) as they do not specify any conditions about the shape or character of the distribution of the population. In relation to the vocabulary sheet data, there were no assumptions being made and the shape or character of the distribution of the data was unknown. Moreover, purposive and convenience sampling means that the data cannot be assumed to have a normal distribution.

3.8 Reliability Phase I

'Reliability is generally taken to concern the extent which the same measurement technique or strategy produces the same result on different occasions …' (Hammersley, 2008, p. 43). There are a number of ways to evaluate whether a research instrument is reliable with consistent results from one occasion to the next. When a research instrument produces the same data time after time and any variations are a result of the variable being measured rather than the instrument itself it, the instrument is said to be reliable (Denscombe, 2007). As the proposed research involves testing individuals, establishing that the test is reliable is very important. If the vocabulary test being used in this research is not reliable then scores will vary, and the accuracy of the results would be questioned.

Reliability coefficients can be used to determine whether a test is reliable. In language testing, there are three main ways to ascertain reliability using these coefficients: test-retest, equivalent forms and the split-half method (Brown, 1998). The split-half

method determines whether all parts of a test are contributing equally to what is being measured. On the tests used in this study, the words increase in level of difficulty and taking the first half of the test and comparing it to the second half would not yield equal results, so the split-half method is inappropriate in this case. Goulden *et al*'s (1990) five sub-tests all contain alternative sets of items but are intended to measure the same construct and any one of them can be used to estimate vocabulary size equally. These tests are called parallel or equivalent forms. Equivalent forms reliability is approximated by giving one group of participants two equivalent tests (forms A and B of a test) and calculating the correlation coefficient between the scores on the form (Brown, 1998). This makes the equivalent forms reliability check the best way to determine whether the test will give consistent results. Parallel (or equivalent) forms are considered to be the best form of reliability by many (Rust & Golombok, 2014).

A small pilot of five participants (all UG students) was conducted to establish equivalent/parallel forms reliability. Each participant was given all five sub-tests and the correlation between all the scores was calculated in SPSS. Reliability is calculated by dividing the true variance by the common variance. In this case, it gives a common inter-item correlation of .898 and a reliability of scale score of .978. In other words, 98% of any variance in test scores is attributed to true score differences while 2% is due to measurement error. This means all five sub-tests are equivalent (see Table 3.3).

Table 3.3 Equivalent forms reliability data

Reliability statistics		**Inter-item correlation matrix**					
Common variance	32899600.000		**Test 1**	**Test 2**	**Test 3**	**Test 4**	**Test 5**
True variance	29550000.000	**Test 1**	1.000	.817	.909	.969	.977
Error variance	3349600.000	**Test 2**	.817	1.000	.941	.859	.873
Common inter-item correlation	.898	**Test 3**	.909	.941	1.000	.963	.947
Reliability of Scale	.978	**Test 4**	.969	.859	.963	1.000	.953
		Test 5	.977	.873	.947	.953	1.000

3.9 Validity Phase I

As previously stated, Schmitt *et al* (2020) argue that a major limitation of vocabulary testing is there is inadequate validation evidence for the majority of the vocabulary tests available (for example the VLT and VST). Goulden *et al*'s (1990) test is no exception. 'Validation of a test or measure involves a process of evidence accumulation over time …' (Rose *et al*, 2020, p. 71). There is no established validity for this test in the literature. However, as Burns (2000, p. 350) suggests 'the subject of validity is complex, controversial, and peculiarly important in research'. Validity in the broadest sense can be defined as 'the degree to which the research has measured/explored the phenomena that it set out to study' (McGrath & Coles, 2013, p. 238). When referring to particular data collection tools such as tests, validity refers specifically to that instrument and in the case of estimating vocabulary sizes based on testing, it forces the question, 'Does the test measure what it is supposed to?'.

There are several ways a researcher can determine whether their research instrument is valid. However, 'the issue of test validity is broader than that of reliability, and (for once) not especially amenable to statistical terms' (Tolmie *et al*, 2011, p. 149). Hence, the most appropriate way of determining whether the vocabulary test

used in this research is valid is to examine content validity. Burns (2000, p. 352) states that 'Content validation is guided by the question: "Is the substance or content of this measurement representative of the content or the universe of content of the property being measured'" and lastly, 'Content validity is most often determined on the basis of expert judgement'. The following section will examine to what extent Goulden *et al*'s (1990) test could be considered valid despite the fact there is little empirical evidence in the literature to support this.

To test validity in this study, it is important to ask whether the test accurately represents an individual's vocabulary size. As previously stated, the main issue with estimating the vocabulary size of individuals from tests based on dictionary sampling is that they may give an overestimation of the actual number of words (see Nation, 1993 and Zechmeister *et al*, 1995). Problems occur when the number of words in the dictionary is either too few (leading to underestimation) or contains a number of high frequency words which have several dictionary entries (leading to overestimation). Hartmann (1941, p. 353) wrote that the main difficulty lay in the sampling technique and that 'All vocabulary "estimates" made in recent years fit only the dictionaries upon which they were based'.

Nation (1993, pp. 31-35) believes that in order for tests based on dictionary sampling to be valid, in this case NOT over or underestimating vocabulary size, that certain procedures must be followed when devising them:

1. Choose a dictionary that is big enough to cover the known vocabulary of the people being investigated. For educated adults, a dictionary should contain at least 30,000 base words.
2. Use a reliable way of discovering the total number of entries in the dictionary.
3. Use explicit criteria for deciding and stating (a) what items will not be included in the count and (b) what will be regarded as members of a word family.
4. Use a sampling procedure that is not biased towards items which occupy more space and have more entries. (Item 3 above can be applied before, while or after the sampling is completed.)
5. Choose a sample that is large enough to allow an estimate of vocabulary size that can be given with a reasonable degree of confidence.

6. The sampling should be checked for the reliability of the application of the criteria for exclusion and inclusion of items.
7. The sample should be checked against a frequency list to make sure that there is no bias in the sampling towards high-frequency items.
8. In the written report of the study, describe clearly and explicitly how each of the previous seven procedures was followed in sufficient detail to allow replication of any or all of the procedures.

Nation (1993) has described Goulden *et al*'s (1990) study as methodologically sound as they adhered to a number of the principles as outlined above. They used *Webster's Third Dictionary* but supplemented it by using a recent addenda. With respect to Item 2 above, the total number of words used in the dictionary (267,000) was corroborated by a previous study (see Dupuy, 1974, p. 5). Examining the third point, Goulden *et al* (1990) excluded from their test proper words, compound words, derived words, and a variety of items classified as 'others'. In all they excluded 113,161 words from the test accounting for 42.4% of the total entries.

It would never be possible to measure the exact number of words (whether in their active or passive vocabulary) an individual knows so an estimate of receptive vocabulary size will always be the best possible method to formulate an understanding, if somewhat limited in accuracy.

The external validity of a research study is the extent to which its results apply to situations beyond the study itself – in other words, the extent to which the conclusions drawn can be generalised to other contexts' (Leedy & Ormrod, 2015, p. 105). Generalisability is not straightforward (Punch & Oancea, 2014) and not all quantitative research is automatically generalisable (Gilbert & Stoneman, 2016). In order for our research to be generalisable, the sample needs to be representative of the population (Bryman, 2008).

One way of ensuring the sample is representative is to use probability sampling (Bryman, 2008). Phase I used probability sampling but a key question is whether this is representative of the population. 'The population is the largest group of units that the cases actually used in the research could actually represent' (Gorard, 2013). Defining the population in this research is difficult in terms of how large the parameters are. Is it all undergraduates in the UK or England, or the Midlands or just the University? In 2016-2017 there were around 1.76 million undergraduates currently in the UK (Universities UK, 2018b). This is not a case study - the research

was only conducted in one institution in the East Midlands and the sample for the vocabulary test consisted of 389 participants. While this is not large enough to be generalisable to the UK UG population, the sample size is big enough to give us a valuable insight into the vocabulary sizes of UK HE students.

3.10 Phase I sample

Probability sampling is a significant feature of quantitative research (Gray, 2018). The participants for this study were chosen using a random sampling technique which is a key probability sampling method (Newby, 2010). Random sampling was used to ensure that each individual in the population had an equal chance of being selected and this makes it possible to generalise to the population (Creswell, 2014). All the single honours undergraduate programmes at one university from 2014-2015 (excluding a campus due to geographical location) were included in the population. Fourteen programmes out of 74 were selected at random out of the hat (see Appendix 4). It was not necessary to include joint honours programmes in the hat draw as students study two subjects on this programme as one undergraduate degree (e.g., Politics and Criminology). Selected programmes were not asked to differentiate their students between single honours or joint honours when they agreed to participate in the study so joint honour students would naturally occur in any random sample. In other words, if joint honours programmes had been included, there was a possibility that these students could have been included twice in the sample. A description of the sample is stated below, and the sample divided by stages of study can be seen in Table 3.4.

- 14 programmes were selected at random out of the population, but students from 11 single honours and 19 joint honours programmes participated and were represented in the sample.
- Test sample size (n) = 389 where the total undergraduate population was 13,595 (full and part-time) at the time the research was conducted in 2014.
- 27 participants recorded unknown words.
- Gender: Male 155 and Female 232 and 2 did not state.

Table 3.4 Participants by stage of study

	Frequency	**%**
Stage 1	130	33.4
Stage 2	139	35.7
Stage 3	120	30.8
Total	389	100.0

The sampling and testing in Phase I took a cross-sectional approach where data are collected at one point in time (Gray, 2018). Researchers collect data from one or more cohorts at a single point of time or within a short period of time and is often described as a snapshot of data collection (Paltridge & Phakiti, 2015). This method is typically quicker and less expensive than longitudinal research (Donley, 2012). A cross-sectional design was used in order to identify potential correlations between variables (vocabulary size and degree classifications), but any direct causation cannot be determined (Gray, 2018). Cross-sectional research does not indicate change at an individual level (UCL, 2021) and this was not the aim in Phase I. Testing at an individual level for changes in vocabulary size would have been time consuming and expensive as this would have needed to be done over a three-year period, the time it takes to typically complete an UG degree. The sample size would have been potentially much smaller as gaining access to the same set of participants over an extended period of time is a major challenge (Gray, 2018).

3.11 Observations over the four-week testing period in March 2014

The following observations were made during the administration of the test:

- Several students asked whether dyslexia would affect their scores.
- One student had the lecturer read the words out loud.
- Many students expressed verbally or with nods of heads that they find language difficult.
- Males seemed more competitive than females – they were trying to have the highest score.
- Students asked whether words were spelled correctly.
- Students asked whether words were made up on the test.
- A couple of students kept the word lists.
- Some students said they really enjoyed it.

Observations were not systemically recorded – in hindsight, this would have provided some valuable insight into vocabulary testing. Notes were taken during the test but if a structured observation sheet had been used, useful data such as how many times students asked about dyslexia or spelling, or made-up words could have been collected. This data could have given an indication of the factors that may impact on vocabulary test scores. However, from the unstructured observations during testing, it can be noted that language is important to students as they wanted to do well on the test, and they were very willing to engage in the testing process.

3.12 Research questions Phase II

As a reminder, the following research questions were explored in Phase II:

1. What is the level of vocabulary richness in written assignments of monolingual and bilingual (using two or three languages including English routinely in the home) English speaking undergraduate students using lexical frequency profiling?
2. What are the changes, if any, in word usage from the NGSL 1, NGSL 2, NGSL 3, NAWL and 'off-list' in student assignments between the stages of study?
3. Is there a correlation between the percentage of academic words from the NAWL in students' written work and assignment scores?

3.13 Data collection

3.13.1 AntWordProfiler 1.5.1

As discussed in the literature review (Section 2.15), the software AntWordProfiler, version 1.5.1 was chosen for this project because it gives the user greater control in terms of analysing individual texts or the corpus as a whole. The VocabProfiler tools on LexTutor do not allow for this functionality. The researcher also wanted to use freeware in order to make the research accessible to other researchers, students and teachers. The default word lists used on AntWordProfiler are the GSL and the AWL, but it also allows the user to upload their own word lists, another key advantage. AntWordProfiler is also compatible with AntConc, allowing for the possibility of further analysis and future research.

3.13.2 Word lists

The following word lists were considered for this research: the GSL (West, 1953); the AWL (Coxhead, 2000); the NGSL (Browne, 2014); the NAWL (Browne *et al,*

2013); new-GSL (Brezina & Gablasova, 2015) and the AVL (Gardner & Davies, 2014). It was decided to use the NGSL, and the NAWL and the following discusses the rational for this decision. As mentioned in the literature review, the GSL (West, 1953) is outdated as most of the texts in the corpus were published before 1930. Relatively frequent modern-day words such as *computer* and *internet* are not to be found on the GSL while words such as *schilling* and *footman* are listed (Brezina & Gablasova, 2015). The NGSL and new-GSL are much more up to date.

A corpus from which word lists are created needs to be both balanced and representative of the language they represent so it has been argued that the bigger the corpus the better (Nelson, 2010). Although the new-GSL is based on the largest corpus, Browne (2014) believes there to be an imbalance in the size of the four corpora used (e.g., 1 million words for the LOB compared to 12 billion for the EnTenTen 12) (see Table 3.5). However, McEnery and Hardie (2012) believe that balance, representativeness and comparability are only ideals and rarely achieved. It was therefore decided that these were not necessarily a key deciding factor in choosing the general English word list for this research.

Table 3.5 Overview of three general service word lists

	GSL (West, 1953)	**NGSL (Browne, 2014)**	**New-GSL (Brezina & Gablasova, 2015)**
Size of corpus	2.5 million running words, later expanded to 5 million	273 million running words	12 billion running words
Source of texts	Encyclopaedias, magazines, textbooks, novels, essays, biographies, books about science and poetry	CEC (Cambridge English Corpus)	The Lancaster-Oslo-Bergen Corpus (LOB), The British National Corpus (BNC), The BE06 Corpus of British English (BE06), and EnTenTen12
Size of list	2,000 words	2,800 high frequency words	2,494 lemmas
Selection criteria (Dang *et al*, 2020)	Frequency, ease of learning, necessity, cover, stylistic level and emotional neutrality	Frequency, dispersion, subjective judgment	Frequency, dispersion and distribution across language corpora
Counting unit	Word families	Modified lexeme (flemmas)	Lemmas
Organisation	1K and 2K list	1K, 2K and 3K lists	1 list
Coverage	84.2% in CEC corpus	92.34% in CEC corpus and 86% of the Academic CEC	80.1 and 81.7% of the text in the source corpora

A key factor when choosing word lists for either research or pedagogical purposes is the counting unit. The counting unit in vocabulary research is very important as seen in the discussion on the AWL versus the AVL in Section 2.7.3. The main counting units in vocabulary research are words, lemmas, flemmas and word families which can make comparisons between lists difficult. In the new-GSL (Brezina & Gablasova,

2015) the counting unit is the lemma and in the NGSL (Browne, 2014) the counting unit is the modified lexeme while the GSL uses word families. Browne (2014) describes the modified lexeme approach as counting the headword in all its various parts of speech and including all inflected forms. This is what is known as a flemma. A flemma is like a lemma but does not distinguish between word classes (Schmitt & Schmitt, 2020). For example, from the Level 1 list for the NGSL in both Compleat LexalTutor and AntWordProfiler, the noun lemma *book* is combined with the verb lemma *book* into one flemma which contains the headword *book* and its inflections *books, booked, booking* and *bookings.* In the available new-GSL list, the lemmas would require the corpus to be annotated or parsed before any analysis can be undertaken. The data would need to be POS tagged using the same tagger that Brezina and Gablasova (2015) used which was a main reason for excluding the new-GSL.

In order to measure the lexical richness of UG assignments using the concept of lexical frequency profiling (LFP), both general and academic English vocabulary need to be used in the research. Because of the criticisms of the AWL outlined in Section 2.7.3 the AVL was considered but not chosen. Using the AVL and the new-GSL in combination for this purpose was explored. However, there is significant overlap between the two. Using Text Lex Compare and the word type as the unit of comparison, there are 1,452 tokens (15%) and 609 types (13%) that are shared between the lists. As an example, words such as *use, effect* and *term* are included in both. This also calls into question how specialised the AVL is if words on a general English core list are included. For this reason, the AVL was discounted, and the NAWL included because the NGSL and the NAWL work together similar to the GSL and the AWL.

The NGSL includes a supplement list of days of the week, month and numbers (only written in full). It does not include letters of the alphabet or proper nouns. In the creation of the NGSL, the academic sub-corpora were excluded from their analysis of the CEC so academic words are not on this list – academic as defined by Paquot (2010) which is a set of lexical items that are relatively frequent in academic texts and are not core. In other words, there is no overlap between the NGSL and the NAWL since they have been designed to be used together. The NGSL used on AntWordProfiler contains the first 1,000 most frequent flemmas (1K), the second 1,000 most frequent flemmas (2K) and the third 1,000 most frequent flemmas (3K). The NAWL used on AntWordProfiler contains 960 frequently occurring flemmas,

alphabetised and not in order of frequency. Words not on either the NGSL, the supplement list or the NAWL are considered to be ‘off-list’. Off-list words include proper nouns, subject-specific words, general low frequency words (not on the NGSL), acronyms, numbers, and letters of the alphabet.

A major limitation of the NAWL is that no methodological information is available. However, it was designed to work in connection with the NGSL so it is possible to assume that the selection criteria for the words would be similar to that of the NGSL (Therova, 2020). Another limitation of both lists is the use of flemmas. The meanings of words are potentially lost along with the part of speech. The example of *book* mentioned earlier is a case in point. The noun *book* means a written or printed work consisting of pages while the verb *book* has several meanings, one of which means to buy a ticket in advance (Oxford University Press, 2021). Vocabulary profilers are not able to distinguish between homonyms. However, McLean’s (2017) research on Japanese University student’s knowledge of inflectional and derivational word family members showed that the flemma was more suitable as a counting unit than word families.

A way to identify whether word lists are useful, particularly for L2 learners is to consider the lexical coverage of the list (Dang *et al*, 2020). Word coverage varies depending on the corpus and texts used to compute the percentage, so caution is needed. Coverage also varies across disciplines (Hyland and Tse, 2007; Durrant, 2016). If looking at the figures presented in Table 3.5, then the NGSL could potentially provide more coverage than the GSL and new-GSL.

To summarise, the rationale for choosing the NGSL and the NAWL in this research pertains to the fact that these lists are up to date, based on large corpora, when combined have a 92% coverage of the academic CEC corpus, and use flemmas as the counting unit. The NGSL and the NAWL lists were obtained from Laurence Antony’s AntWordProfiler website. The lists needed to be cleaned as there were a few errors/issues with repetition and spelling (e.g., *rpertoire* in NAWL). The NGSL and the NAWL were designed to work with each other so there should not be any overlap. However, the words *founded, founding, founds* and *statistics* were listed in both the Level 3 NGSL and the NAWL. For the purpose of this research, they were removed from the Level 3 NGSL but remained on the NAWL. The rationale for this is that these words can also be found on the AWL (Coxhead, 2000).

Table 3.6 Overview of three academic word lists

	AWL (Coxhead, 2000)	**NAWL (Browne *et al*, 2013)**	**AVL (Gardner & Davies, 2014)**
Size of corpus	3.5 million running words	288 million running words	120 million running words
Source of texts	414 texts published between the 1960s and1990s	Academic CEC (Cambridge English Corpus) MICASE (Michigan Corpus of Academic Spoken), and BASE (British Academic Spoken English)	Academic sub-corpus of the COCA
Size of list	570-word families (3,110-word types)	960 words	3,015 lemmas (~2,000-word families)
Education discipline included	Yes	Not specified	Yes
Basis of creation	GSL	NGSL	-
Selection criteria	Specialised occurrence (i.e., exclusion of GSL items) Range Frequency	Not specified	Ratio Range Dispersion Discipline measure
Counting unit	Word families	Modified lexemes (flemmas)	Lemmas
Organisation	10 sub-lists	Single list	Single list

Coverage	10% of the source corpus	6% of the source corpus	14% of academic materials in both COCA (120 million+ words) and the British National Corpus (33 million+ words).

Source: adapted from Therova, 2020

Using AntWordProfiler, student assignments were analysed by looking at:

1. What percentage of the vocabulary consisted of academic word tokens from the NAWL.
2. What percentage of the vocabulary consisted of word tokens in the K1, 2 and 3 bands of the NGSL (hereafter known as NGSL 1, NGSL 2 and NGSL 3).
3. What percentage of the vocabulary consisted of 'off-list' word tokens.
4. Changes in word tokens from the NGSL 1, NGSL 2, NGSL 3, Supp, NAWL and 'off-list' between the stages of study.
5. Further analysis was conducted on the output from AntWordProfiler with six participants selected at random.

Tokens (a single occurrence of a word from each text) were used in the analysis of the corpus data. Different corpus tools (as discussed in Section 2.15) count tokens differently leading to a wide variation between them (Brezina & Temperley, 2017). Therefore, it is necessary to be clear as to how the tokens were counted in the specific tool used (Brezina, 2018a). AntWordProfiler relies on counting the tokens in a target file against so-called base word lists. In this research, the base word lists are the NGSL 1, NGSL 2, NGSL 3, Supp and NAWL. In Figure 3.2, these are labelled Level 1, 2, 3, 4, and 5, respectively. Level 0 is 'off-list'. Items not on any of the base word lists are labelled as 'off-list'.

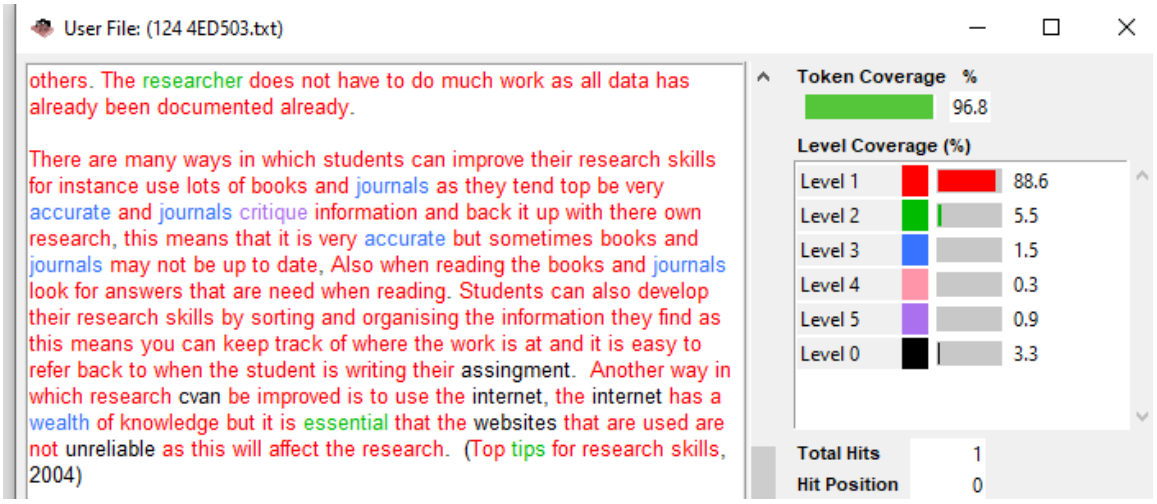


Figure 3.2 Example of output from AntWordProfiler

As seen in Figure 3.2, items not on the base word lists include numbers (unless written out) and spelling errors. Proper nouns are also considered as 'off-list'. Additionally, 'off-list' words such as *internet, website* and *unreliable*, arguably should be included in the definition of what constitutes a word token. However, because they are not on the base word lists, they are not defined as such, highlighting a problem with using word lists in vocabulary research. Dang (2020, p. 290) supports this by stating, 'One core issue in word list studies is what should be counted as a word'.

Tokens, a single occurrence of a word form in a text (Brezina, 2018a), were used in the analysis of the data rather than type/token ratios (TTR). Type/token ratios are sensitive to text length and decrease as the text is longer as more words are repeated; this ratio can only be used when comparing texts of the same length (Brezina, 2018a). The length of the student assignments increases between the stages of study making comparisons using TTR problematic. However, a limitation of analysing tokens only is that it does not account for the range of words used as some are repeated.

However, in the additional analysis of six cases, mainly types were used. A word type is classified as a unique word form in the text (Brezina, 2018a). A limitation of using word types stems from there is no distinction between forms with multiple grammatical functions or meanings (Brezina, 2018a). A further limitation is that using types can underestimate the frequency of word families and lemmas as words like *aspect* and *aspects* are treated separately.

3.13.3 Procedure

Three assignments from each year for each student were downloaded from the University's VLE and placed in a folder labelled with the student code rather than name. The assignments were cleaned ready for analysis. This meant removing the reference lists, any identifiable information such as names and student numbers. Front sheets with module codes, assignment questions and word counts were also removed. Each assignment was then converted to a plain text file. Each stage was analysed separately, and a word profile was created for each student. An example of the output produced by AntWordProfiler is shown in Figure 3.2. It can be seen what percent of the text consists of K1 word tokens (in red) from the NGSL 1, K2 word tokens (in green) from the NGSL 2, K3 word tokens (in blue) from the NGSL 3, supplementary words (in pink) from the Supp list, academic word tokens from the NAWL (in purple) and the 'off-list' word tokens (in black) in the box on the right.

3.14 Data analysis

The data in Phase II were non-parametric as a result of the purposive and convenience sampling method used. It cannot be assumed the distribution is normal and Motulsky (1995 cited in Gray, 2018) states that a parametric test should only be used if the researcher is sure that the population is normally distributed.

Descriptive statistics were calculated in SPSS. The median was used as a measure of central tendency because of the assumption that the dataset may not be normally distributed. The mean was not utilised because it is more suitable when the distribution is normal and it is influenced more by extreme data whereas outliers do not affect the median (Foster *et al*, 2015). The median showed the average percentage of high frequency, academic, subject-specific and 'off-list' words students used in each stage of study. The median was also used to calculate the interquartile range. The interquartile range was calculated to measure the spread of the data since it was assumed not to be normally distributed (Foster *et al*, 2015).

Thus, Wilcoxon rank tests were used instead of t-tests in Phase II as the data was non-parametric. This test compares the size as well as the differences between variables (Salkind, 2014) so is similar to the t-test. The Wilcoxon test indicated if there were any significant differences in the type of words students used between the stages.

Correlations between marks on the assignments and the NGSL 1, NGSL 2, NGSL 3, Supp, NAWL and 'off-list' words were calculated using Spearman's rank correlation coefficient (r_s) as the data is non-parametric. Spearman's rho (r_s) is used when parametric assumptions are violated (Brezina, 2018a). Spearman's rho (r_2) was calculated with a 99% confidence interval (p<.01) in order to reduce the risk of a Type I error.

3.15 Sampling method

Non-probability sampling was used in this phase of the study, i.e., convenience-sampling methods were employed. In this strategy, the sample is chosen because the researcher has easy access to participants (Cohen *et al*, 2007). Because access to a particular programme and the students' written work via the University's Virtual Learning Environment was available, this was the most logical sample to research. There is also an element of purposive sampling because the programme chosen for the study was selected purposefully on the basis that the cases (written work) would be information rich (Gray, 2018). This enabled a focus on students' use of vocabulary in one subject area. Disadvantages or limitations of this strategy are that the sample is potentially biased since participants are selected by the researcher (Gray, 2018) and caution is needed when generalising to the wider student population (Cohen *et al*, 2007). However, there were 41 participants who generated data from nine pieces of written work each, a total of 369. The size of the sample determines how much information we have about the population so logically the bigger the sample the more confident we can be about it being representative. Moreover, if the sample matches the characteristics of the population as closely as possible then the results of the study have a degree of generalisability (Salkind, 2014). While degree programmes will vary depending on subject content, an element of writing is involved in most higher education degrees. The results of this study can be generalised to other programmes similar in nature.

A longitudinal approach was taken in Phase II data collection. Longitudinal research is where the same aspect of information is collected from the same participants over a period of time (Paltridge & Phakiti, 2015). Moreover, the data are collected for two or more distinct time periods and analysis involves some comparison of data between these time periods (Menard, 2002 cited in Dörnyei, 2007). This allows the researcher to observe changes in learning/abilities and/or other cognitive development (Paltridge & Phakiti, 2015). These changes can be identified at the individual level (UCL, 2021).

Dörnyei (2007) states that the purpose of longitudinal research is to not only describe patterns of change but also explain causal relationships. The sample in Phase II consisted of BA (Hons) Education Studies (X300) students who completed their degree in the academic year 2015-2016. These students were chosen as access to their work was available on the VLE over a 3-year period from 2013-2014, 2014-2015 and 2015-2016. Phase II set out to find patterns of academic word use in an UG student corpus, comparing the data between years of study, as well as exploring links between the percentage of academic word tokens used from the NAWL in students' written work and their module marks.

The 2015-2016 cohort consisted of a total of 137 students. The sample excluded all students studying part-time as their period of study was significantly different. Also, many of the part-time students study at partner colleges so their experience would be different making comparisons difficult. Students studying on the SEND pathway were also excluded as well as JHS students. These study different modules so in order to ensure consistency and reliability only full-time students who were not on the SEND pathway and based at the main university campus were included in the sample (see Table 3.7 for a summary of the inclusion and exclusion criteria). A total of 41 students were eligible for the sample.

Table 3.7 Summary of inclusion and exclusion criteria of BA (Hons) Education Studies students in the sample

Inclusion criteria	**Exclusion criteria**
Full-time status	Part-time status
Started degree in 2013-2014 and finished in 2015-2016	Non-completers
No pathway (X300 only)	SEND pathway
No Erasmus exchange	Erasmus students
Based at Kedleston Road	Based in partner colleges
Study for three years	Direct entrants

The sample was taken from a cohort of students who started in Stage 1 from the academic year 2013-2014.

Table 3.8 Number of analysed scripts in the sample

Academic year	Stage	No. of possible assignments	No. of written assignments analysed per student	Sample size of individual students/no. scripts
2013-2014	1	6	3	41/123
2014-2015	2	6	3	41/123
2015-2016	3	5	3	41/123

For each student, by looking at three assignments per year means that 53% (or 9 out of 17 assignments) of their work was analysed for this research. The sample consisted of 41 students, meaning that 123 written assignments were analysed for each year giving a total of 369 assignments (see Table 3.8). Finally, six cases were selected at random and additional analysis of the output from AntWordProfiler was conducted.

3.16 Assignment selection

Research on academic writing focuses on broad genres such as academic journal articles/textbooks, doctoral theses or undergraduate student coursework and are largely recognised and unproblematic (Nesi & Gardner, 2012). This study focuses on the broad genre category of undergraduate student coursework. However, as discussed in Section 2.5.1 in the literature review, there is no agreed classification for defining genres and researchers operationalise them differently. Nesi and Gardner's (2012) research on families of genres in assessed UG student writing will provide the conceptual framework for the genres identified in this research due to the same target population being investigated, namely UG students. Their classification was derived from the BAWE corpus and included 2,761 assignments set by tutors in around 1,000 modules from 300-degree courses. It must be noted that assignments were not taken specifically from Education degree courses in the BAWE. However, assignments from the disciplines of Arts & Humanities as well as the Social Sciences were analysed. Arguably, the study of education is multi-disciplinary drawing from

subjects such as history, philosophy, economics, sociology and psychology and the genres identified from the BAWE corpus are therefore relevant.

The 13 genre families and their functions identified in Nesi and Gardner's (2012, p. 36) research are listed in Table 3.9.

Table 3.9 Genre families and function

Genre family	**Function**
1 Case study	To prepare for professional practice.
2 Critique	To develop powers of independent reasoning.
3 Design specification	To prepare for professional practice.
4 Empathy writing	To write for oneself and others.
5 Essay	To develop powers of independent reasoning.
6 Exercise	To demonstrate knowledge and understanding generally of basic skills and concepts. Usually short answers.
7 Explanation	To demonstrate knowledge and understanding but additionally expect students to explain how something works or functions. Longer answers than exercise.
8 Literature survey	To build research skills.
9 Methodology recount	To build research skills.
10 Narrative recount	To write for oneself and others.
11 Problem question	To prepare for professional practice.
12 Proposal	To prepare for professional practice.
13 Research report	To build research skills.

The student assignments selected for the study come from a naturally occurring classroom setting in HE. All modules were included in the study except Research Methods and Education in Context at Stage 2 thereby excluding the research report genre. Education in Context is highly individualised as the module is based on students' work placement experience. The assignment for this task is a reflective piece where students have to consider their personal and professional development which fall under the narrative genre. It was excluded as the individual nature of the assessment could lend itself to a variety of personalised language affecting the reliability of the results. Research Methods was excluded because this subject contains its own subject specific terminology which does not necessarily intersect with education. The assignment/genre is the research report. Modules with oral presentations were excluded due to slides not being uploaded on the VLE. However, a module with a poster presentation, namely the Sociology of Education was included as they could be found on the VLE. Students were asked to create an argument and provide evidence for it so this could be classed in the essay genre. Finally, the Independent Study at Stage 3 was excluded for the same reasons as Education in Context and Research Methods.

The Stage 3 cohort in 2015-2016 could take five optional modules so these were all included in the study. To summarise, the content of the modules ranged from contemporary debates, education and global inequality, education and wellbeing, learning and motivation, SEND and inclusion, understanding & managing behaviour to education and the arts. The writing genres for all these modules was an essay.

The Stage 2 2014-2015 modules ranged from curriculum frameworks, education and globalisation, educational alternatives, exploring disability, individual differences and learning, meeting emotional needs, specific learning difficulties to the sociology of education. The writing genre for these modules consisted of six essays and one academic poster.

The Stage 1 2013-2014 modules are all core with no options and included topics on inequality and education, comparative education, lifelong learning, study skills, psychology and teaching, learning and the curriculum. The writing genre included the critique genre (annotated bibliography and a report) and four essays. In all, the corpus consisted of only the essay and critique genre. Both have the function to develop powers of independent reasoning rendering the corpus stable to the effects of genre according to Nesi and Gardner's (2012) genre classifications. However, the topics

varied widely from SEND to behaviour management which did affect the overall number of word types (see Section 2.14) as well as the subject specific vocabulary used in each assignment and the number of 'off-list' words.

In addition to the assignments, the holistic assignment scores were collected. The student's work was scored according to the University's Marking criteria (Appendix 5) by independent raters. A minor limitation of this approach is that a second-rater opinion is missing.

3.17 Reliability Phase II

As previously mentioned, reliability refers to whether the measurement tool generates consistent results each time it is used (Frankfort-Nachmias & Leon-Guerrero, 2015). Quantitative/positivist research requires adherence to a procedure (Newby, 2010). This will help to generate consistent results each time. The same procedure was followed when the students' work was analysed. Checks were also conducted on every fifth folder to ensure they contained the correct assignments. As a result, errors such as typing in the wrong number or placing assignments in the wrong folder were able to be detected and rectified (see Figures 3.3 and 3.4)

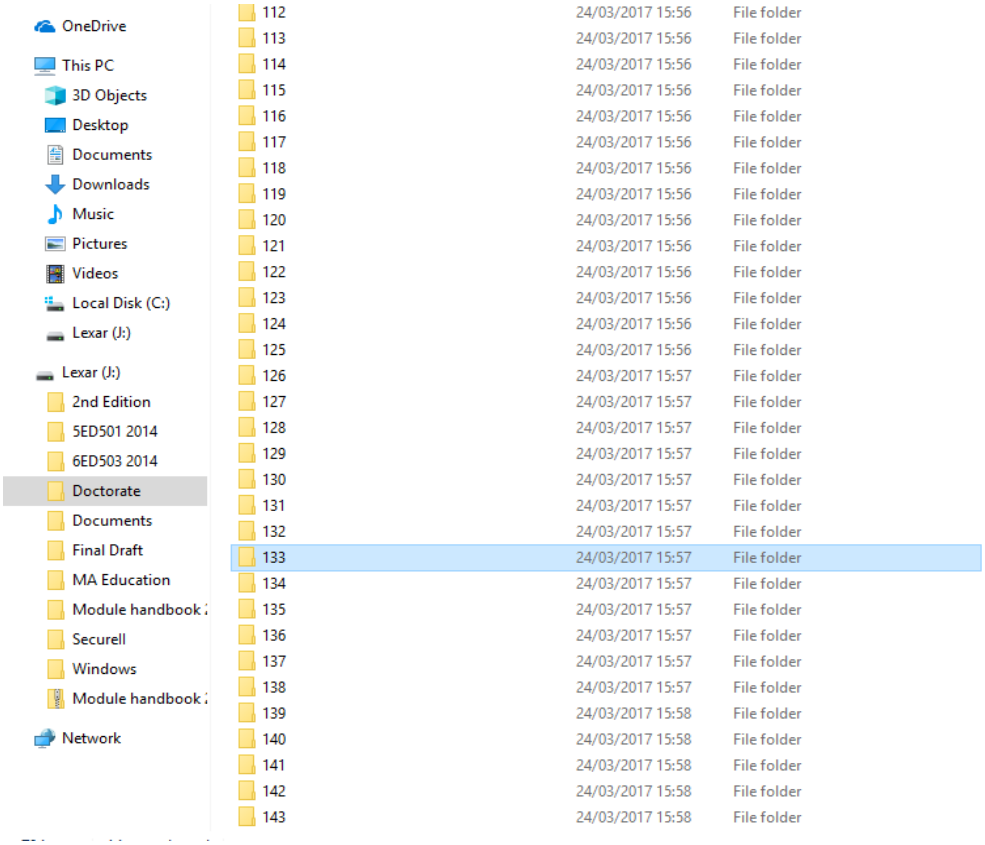


Figure 3.3 Participant folders

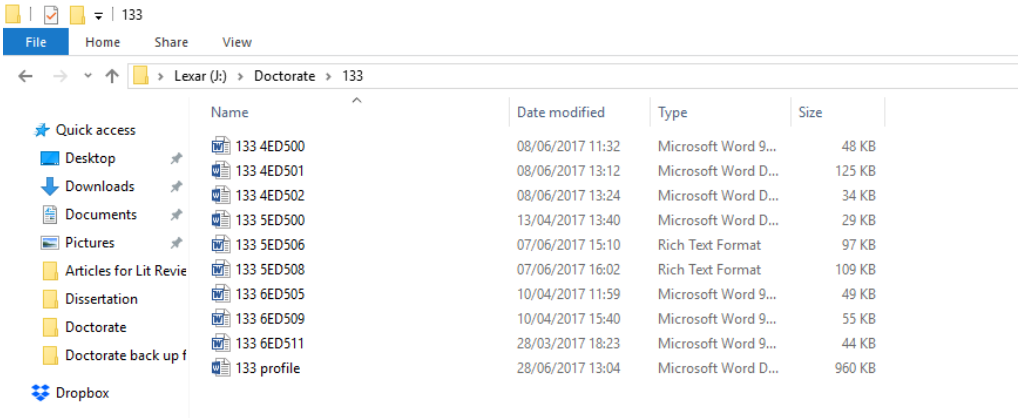


Figure 3.4 Individual folders with assignments and analysis (profile)

Gray (2018) suggests our results/data need to be stable. In other words, we should get the same results if we measured something on two separate occasions. Results in this study are deemed to be stable as not one but three assignments from each student in each stage of their studies were analysed.

3.18 Validity Phase II

As mentioned in Section 3.11, a research tool must measure what it is supposed to – it must have validity. Validity helps to establish whether accurate, meaningful and defensible conclusions can be made from the data (Leedy & Ormrod, 2015). Gorard (2013, p. 160) sees validity as 'the absence of errors' and 'about how convincing a claimed result is for a sceptical reader'. Accuracy is key to minimising errors but is also important for arriving at logical conclusions and making a convincing argument. The goal for this phase of the research was to be as accurate and error free as possible.

In order to minimise errors and acknowledge them, it is important to understand where and how they are most likely to occur. Common errors include inputting data incorrectly or the actual software not calculating accurately (Gorard, 2013). This can be mitigated in two ways. Firstly, by looking for outliers (e.g., extreme scores) in the data and secondly, checking whether it was inputted correctly. Outliers were found by examining scatter plots. These outliers were then checked to see if the data had been inputted correctly, which in this case, they had been.

Outliers may affect the data as extreme scores can bias the mean and Field (2016) asserts that if there is one or more it affects the data. In this research, there was only one outlier in a dataset at a time, which was not significant. Researchers have several options regarding how to handle outliers (see Barnett & Lewis, 1994). One option is

to remove the item from the data. In this case the outliers did not affect the mean because the sample was large enough, so it was decided to leave them in.

As previously mentioned, Gorard (2013, p. 161) states that statistical software is not always error free – '… there have been versions of IBM SPSS in which the results obtained for a calculation via the menu interface are different from the results obtained using the syntax procedure …'. He thus urges the use of later versions where all the 'bugs' have been 'ironed out' (Gorard, 2013, p. 161). SPSS has had a number of versions over the years – the latest is SPSS 25. Many recent course books use SPSS 21 (e.g., Frankfort-Nachmias & Leon-Guerrero, 2015). SPSS 24 was used in the analysis of this research, so it is fairly up to date and one can be confident that any previous issues have been resolved.

Using the wrong statistical test can also cause results to be invalid. However, careful consideration has been given to the tests used and a justification provided in Section 3.17, so it is safe to assume that the correct tests were used and therefore the results valid.

As previously mentioned, external validity is the extent to which research conclusions can be generalised (Leedy & Ormrod 2015) and in order for our research to be generalisable (external validity), the sample needs to be representative of the population (Bryman, 2008). To reiterate, one way of ensuring the sample is representative is to use probability sampling (Bryman, 2008). Phase II used non-probability sampling (purposive/convenience), so a key question is whether this is representative of the population. Due to the number of written scripts analysed (369) some degree of generalisability was viable and as a result there is to an extent a level of external validity.

3.19 Ethical considerations for Phase I and II

Ethical approval from the College of Education Research Ethics Committee was obtained for each phase of the study (see Appendix 6). There were no major ethical issues or challenges for this project. The College Ethics Committee gave no conditions but made a few recommendations which were acknowledged. For example, there was a concern that students might think that the research would influence their grades. This was resolved by including a line in the informed consent announcement which made clear that the research project will not in any way affect the grades of the assignments. The Committee also suggested using work that had

been previously submitted through Turnitin to ensure that all analysed work had not been plagiarised.

Ethical Guidelines for Educational Research (BERA, 2011, 2018) were consulted before during and after the collection of data.

3.19.1 Informed and gatekeeper consent

Creswell (2014) emphasises the importance of seeking gatekeeper approval to gain access to participants and permit the research to be undertaken. The gatekeepers for Phase I were the programme leaders and module leaders who allowed access into classes. Only one gatekeeper consent was needed for Phase II, and this was obtained from the Deputy Head of Department who was responsible for the programme where the students' work was being analysed (see Appendix 7).

Informed consent was sought and obtained for each individual participant. A signed sheet was collected for Phase I but in Phase II informed consent was acquired via the VLE. The approach was slightly different as students were given the option NOT to participate. Only one student contacted the researcher stating that they did not wish to participate. It was therefore assumed the remaining cohort had given consent to take part. The 'opt-out' approach was taken in order to maximise the sample size; 'opt-in' approaches tend to have lower response rates. It is important to note that the research was conducted before the General Data Protection Regulation (GDPR) 2018 became law. GDPR (2018) now requires participants to positively or actively 'opt-in' so it must be acknowledged that this 'opt-out' approach would not be acceptable under current legislation.

3.19.2 Confidentiality and anonymity

Each student was given a number code and all data was stored separately from this information. No names were used at any point in the analysis of the data.

3.19.3 Harm and benefit

According to BERA (2018) guidelines ethical research should avoid all physical and emotional harm to participants. Gray (2018) emphasises the importance of going beyond the notion of avoiding harm, arguing that research should aim to benefit participants. It is hoped that as a result of this study an understanding of the vocabulary use of undergraduate students in particular will be gained. The study also explored the links between vocabulary size and academic achievement.

Understanding the role language plays in achievement is important in ensuring the academic success of all students.

3.20 Timescale for whole project

Figure 3.5 Project timescale

The project began in September 2013 and in December 2014 dissemination of preliminary findings were reported at the Society for Research into Higher Education in Wales via a poster and presentation (see Figure 3.5).

3.21 Positionality

Punch and Oancea (2014, p. 50) contend that 'There is no such thing as a position-free project'. Every researcher no matter how objective one claims to be will inhabit a position within the research. Positionality refers to how researchers' social and ideological background, biography, prior experiences, attitudes and values influence what and how research is conducted (Light, 2010 cited in Scott & Watt, 2010). In other words, researchers must consider not only where they are located within the research in terms of being an 'insider' but also how their personal beliefs and experience or personal biography influences the decisions they make. Who we are is fundamental to the research process. Our class, race, gender and life experiences, both personal and professional are bound to shape how we do things. Potentially this can pose a number of challenges for the researcher.

To overcome these challenges, researchers need to be reflexive. Although reflexivity is often associated with qualitative research (Creswell, 2014), it is important in understanding our position within all research. Gilbert and Stoneman (2016, p. 514) define reflexivity as

> …a style of research whereby one addresses how the research process affects the results. It requires precision about the analytical methods and data collection procedures used and emphasises the researcher's own assumptions and beliefs through explicit statements of how the researcher's very presence affects what he or she is investigating.

Being reflexive has helped mitigate bias throughout the research process. For the researcher, they had to recognise that their prior experience teaching only international students English language for 16 years shaped their thinking and, in some ways, influenced their understanding of home student needs. The researcher expected native speakers of English to have fewer problems with vocabulary than non-native native speakers. This was not a fair and realistic expectation. The researcher chose the topic for this study based on their experience with home students. As a result, they had to discard some of their prior beliefs about which students had good vocabularies (home or overseas) and try to be as objective as possible. This is one of the reasons why the researcher chose quantitative research; they wanted to find a truth about what vocabularies students have.

The researcher's middle class and educational backgrounds affected their vocabulary knowledge and use of English. They had access to resources (books, good schooling, private tuition). The researcher's parents read and encouraged them to read. Learning vocabulary in primary school was a key component of the curriculum. Essentially, the researcher grew up in a rich linguistic environment. This partially enabled them to be successful in school and higher education. The literature supports this notion; there is, potentially, a vocabulary gap between rich and poor which affects both educational attainment and failures (Quigley, 2018). Having a language 'head start' so to speak could influence how the researcher interprets the findings. In other words, not everyone has access to such an environment, and caution needed to be applied – a 'small' vocabulary for an individual might be a huge achievement for them. While socio-economic status was not looked at as part of the study, it was necessary for the findings not to be based on the researcher's own background.

Conversely, having a good grounding in language and vocabulary is beneficial to the research. It has brought a level of knowledge and understanding which has been built upon in the project. The researcher is also hugely interested in vocabulary. This interest has been key to sustaining motivation throughout the project. Finally, having a good vocabulary has given the researcher access to the literature. It has enabled them to process complex ideas and understand the research in the field.

The professional doctoral researcher has a dual position: they are both a worker (practitioner) within an organisation and a researcher. This is often referred to in the literature as an 'insider researcher' (Workman, 2007). Furthermore, professional doctorates may even require individuals to have more than two roles: a professional or practitioner, a researcher and a student. Thus, the location of the research within professional practice can pose tensions as well as benefits.

A significant benefit of being an insider is having knowledge of the organisation that the research is taking place in and its participants. Other benefits include ease of access and greater rapport with participants (Burgess *et al*, 2006). Researching in your own practice can also be less time-consuming and therefore cheaper (Burgess *et al*, 2006).

The benefits outlined above of being an 'insider' researcher for this study were evident. The researcher was able to gain access to a wide range of participants for the vocabulary test. They were also able to gain access to three years' worth of students' work on the programme they were teaching. Someone outside the university would not have had this access due to privacy and data protection policies. Researching within the workplace saved time and expense in terms of travelling to and from sites.

However, there are a number of challenges that insider researchers face. Research conducted on colleagues can be problematic if asking them to disclose information that could change the nature of the relationship (Burgess *et al*, 2006) between the researcher and the research participant. Burgess *et al* (2006, p.36) outline other difficulties such as '… being too familiar, taking things for granted, displaying bias towards the informants and finding it hard to ask questions about shared events and experiences'.

Originally, the researcher wanted to investigate whether the language used in lectures created a barrier for learners. This could have potentially 'exposed' colleagues and their practice in a negative way as well as positive. The idea was

discarded after considering the ethics behind it. The researcher did not interview students about their language as they knew they would have difficulty in being neutral; another reason why the quantitative route was chosen.

Another significant dilemma or challenge is whether 'insiders' can achieve a degree of critical distance (Drake & Heath, 2011). Drake and Heath (2011, p.20) argue that 'insiders often have assumptions and ideas about what they expect to find out, and on the basis of experience as a practitioner, they may also have a theoretical stance before beginning their doctoral project'. Achieving critical distance is extremely complex for practitioners (Drake & Heath, 2011) and takes a great deal of self-awareness. By choosing the quantitative paradigm to frame this research and looking for a 'truth' and not necessarily the 'truth' helped achieve the critical distance associated with insider research.

3.22 Summary

Newby (2010) contends that in order for our research to have credibility and the results to be accepted, our research processes must be transparent. The aim of this chapter was to be as clear as possible about the methodological decisions and justifications that were made throughout the project. This chapter has provided a rationale for using a quantitative research design as the research is concerned with measuring students' receptive and productive vocabulary knowledge. The way the research was conducted, including procedures for administering the vocabulary test and collecting students' written work, ensured its reliability and validity. The sample and sampling strategy were detailed which further aided the reliability and validity of the research but also contributed to the assurance of transparency. The type of data collected was explored; this provided a justification for how it was analysed through the statistical tests applied. Ethical considerations were addressed along with researcher positionality and potential bias arising from this position. The transparency of the research was established in this chapter and the presentation of the findings in the next chapter will endeavour to be as transparent as well.

Chapter 4: Presentation and Analysis of Data Findings

4.1 Introduction

This chapter presents an overview and summary of the key findings. The discussion of these findings in relation to the literature will be presented in Chapter 5. The analysed findings are organised around the research questions for each phase and key themes presented in the literature review. At the end of each phase of the research, there is a summary of the key findings. In order to obtain results from the raw data some analysis was required.

4.2 Data findings Phase I

Restating the research questions for Phase I is necessary in order to place the findings in the context of the study's overall aim. The questions were:

1. What is the average receptive vocabulary size of English-speaking undergraduate students?

2. What types of words do students not understand?

3. Is there is a link between vocabulary size and academic achievement (measured by expected final degree classification)?

4.2.1 Research question 1: Vocabulary sizes

Table 4.1 displays the mean vocabulary sizes of students in Stages 1, 2 and 3. It also shows the overall mean for all three stages and standard deviation (SD). The standard deviation measures how participants' vocabulary size varies from the sample mean ($\bar{x}$) (Gilbert & Stoneman, 2016). As argued in Section 3.8, due to the central limit theorem, the sampling distribution will take the shape of a normal distribution regardless of the shape of the population from which the sample was taken. In a normal distribution, the percentage of cases will fall between the mean and ± 1 standard deviation or 34% to the right and left of the mean, totalling 68% (Frankfort-Nachmias & Leon-Guerrero, 2015). The sample standard deviation (SD) for Stage 1 is 27% of the mean which is small; this indicates that the majority of scores are close to the mean. Another way of interpreting the standard deviation, approximately 68% of the students in Stage 1 had a test score of between 7,396 and 12,744. Similarly, the sample standard deviation (SD) for Stage 2 is 35%, and for

Stage 3 is 34%. In Stage 2, 68% of the students scored between 8,598 and 14,630 and in Stage 3 between 8,638 and 14,526. When combining all the stages, the majority of scores fall between 8,123 and 14,053.

Table 4.1 Mean ($\bar{x}$) vocabulary sizes of undergraduates

Stage of study	Sample size (n)	Min/max scores	Mean number of words ($\bar{x}$)	Standard deviation (SD)	Square root of sample size $\sqrt{n}$	Standard error $\frac{\sigma}{\sqrt{n}}$
1	130	5,500/18,800	10,070.00	2,673.58	11.40	234.52
2	139	5,000/21,500	11,614.39	3,015.99	11.79	255.80
3	120	4,000/18,500	11,582.96	2,943.75	10.95	268.84
Overall	389	4,000/21,500	11,088.26	2,965.04	19.72	150.36

The standard error of the mean is used to determine whether a sample mean is different from the population due to chance (sampling variability) or whether the sample is genuinely quite different from the population (Foster *et al*, 2015). A large standard error (relative to the sample mean) means there is a lot of variability between the means of different samples and a small standard error indicates the sample mean is similar to the population or in other words, our sample mean is likely to reflect our population mean (Field, 2016). The standard error for all three stages was small relative to the mean so it is fair to say that our sample results are likely to reflect the population.

A significant question that arose was whether there was any change in vocabulary sizes between the stages of UG study. This was ascertained by comparing the means in each stage, so a one-way ANOVA (ANalysis Of Variance) was conducted to see if there was a statistically significant difference between the mean vocabulary sizes of Stage 1, 2 and 3 participants. Because ANOVA identifies statistically significant differences between groups but not exactly which group (Brezina, 2018a), a Bronferroni post-hoc test was also conducted in order to identify the specific group.

There was a statistically significant difference between groups as determined by ANOVA: $F(2,386) = 12.182$, $p = 0.000$. The Bronferroni tests confirmed that the difference in mean vocabulary scores were statistically significant between Stages 1 and 2, 1 and 3 but not between Stages 2 and 3.

P-values are the probability values which determine statistical significance in hypothesis testing. A Type 1 error is the probability associated with rejecting a null hypothesis when it is true (Frankfort-Nachmias & Leon-Guerrero, 2015). The risk of making a Type 1 error can be controlled by setting our significance levels at .05 or 5% or even lower at .01 or 1%. In other words, there is 5% or 1% chance that the null hypothesis is in fact true. In this research a significant level was set at 5% unless otherwise stated. This means there was a 5% chance of rejecting the null hypothesis when it could in fact be correct 5 times out of 100. To interpret *p*-values, if the *p*-value is greater than 5% then we accept the null hypothesis; if it is less than 5%, we reject it.

Based on a .05 significance level, there were significant differences in the mean vocabulary scores between Stages 1 and 2 ($p<.001$), 1 and 3 ($p<.001$) but not between Stages 2 and 3 ($p = 1.00$). The effect size (*r)* was calculated to determine how large this difference was to establish its practical importance. If (*r*) is between 0.1 and 0.3 the effect size is small; between 0.3 and 0.5 is medium; and 0.5 and above is large (Brezina, 2018a). Between Stages 1 and 2 (*r)* = 0.26 and between Stages 1 and 3, the same. This indicates a small to medium effect. This suggests that UG students experience some change in their vocabulary sizes between the first and second year of study but plateau in the third year. Figure 4.1 shows the mean vocabulary size of UG students depending on their stage of study. From the table a difference in vocabulary sizes between Stage 1 and Stage 2 students of around 1,500 words and virtually no difference between Stages 2 and 3 can be seen.

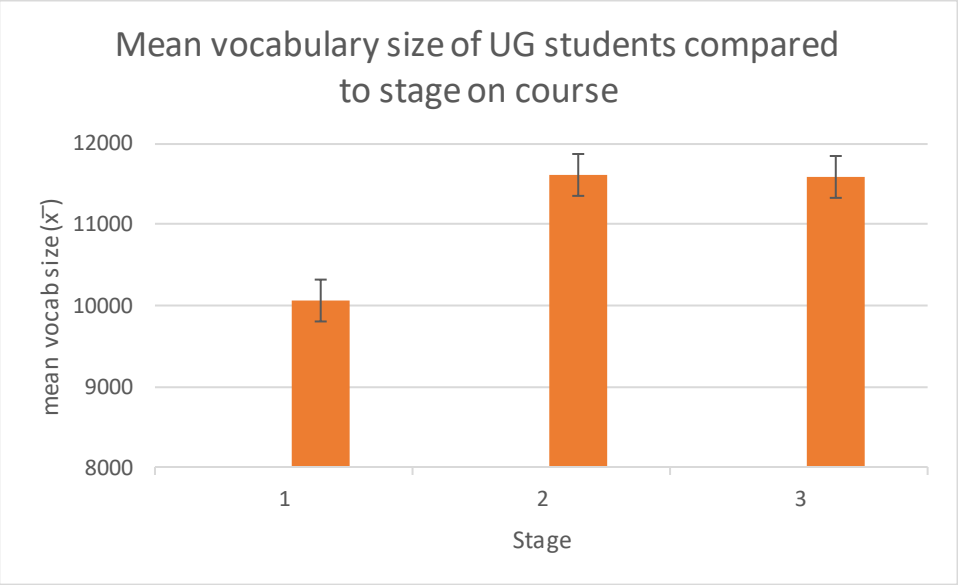


Figure 4.1 Mean ($\bar{x}$) vocabulary size with standard error bars to show significant differences compared to stage on course

ANOVA was also conducted to compare the mean test scores based on degree classification (see Table 4.2). There was a statistically significant difference between groups as determined by ANOVA $F(3,375) = 2.867$, $p = .037$. The *F* score is the ratio of variation between groups divided by the variation within groups, so a large *F* value indicates more variation between groups rather than within groups (Salkind, 2014). While $p < 0.05$, an *F* score of 2 could indicate there is very little difference in the means and vocabulary sizes have negligible effect on degree classifications. Bronferroni tests confirmed that the difference in mean vocabulary scores by degree classifications were not statistically significant. The importance of conducting a post-hoc test is illustrated here as there is a probability of a falsely positive result or Type I error with multiple testing which a post-hoc test can correct (Brezina, 2018a).

Table 4.2 Mean ($\bar{x}$) vocabulary sizes of students by expected degree classification

Degree classification	Sample size (n)	Min/max scores	Mean ($\bar{x}$)	Standard Deviation (SD)
1st	71	10,844/12,198	11,521.13	2,860.71
2:1	224	10,922/11,701	11,311.61	2,955.23
2:2	69	9,678/11,221	10,449.78	3,211.99
3rd	15	8,637/11,030	9,833.33	2,160.25
Total	379	10,834/11,437	11,135.45	2,982.61

To put the overall mean vocabulary score in context, it is compared to the results of vocabulary testing from other research (see Table 4.3). The estimated vocabulary size of undergraduates from this research is significantly smaller than the findings from Hartmann (1946), Goulden *et al* (1990), D'Anna *et al* (1991), and Anderson and Nagy (1993) which ranged from 215,000 words to 17,000 but closer to the findings of Nusbaum *et al* (1984), Zechmeister *et al* (1995), Treffers-Daller and Milton (2013) which ranged from around 14,400 words to as low as 9,800. It is important to note here that the vocabulary tests used were different except in Goulden *et al* (1990) and Treffers-Daller and Milton (2013).

Table 4.3 Various estimates of the number of English words known by adults adapted from Brysbaert *et al* (2016) and compared to the current study

Study	**Estimate**	**Sample**
Hartmann (1946)	215,000	US undergraduates
Anderson and Nagy (1993)	40,000	US high school seniors
Goulden *et al* (1990)	17,200	US undergraduates
D'Anna *et al* (1991)	17,000	US undergraduates
Nusbaum *et al* (1984)	14,400	US undergraduates
Zechmeister *et al* (1995)	12,000	US undergraduates
Treffers-Daller and Milton (2013)	9,800	UK first year undergraduates
Current study	**11,088**	**UK undergraduates**

Table 4.4 shows that sample sizes in previous research in a UK context were much smaller than in this study. For example, Treffers-Daller and Milton only had 18 participants in their sample for Stage 3 whereas this study had 120 participants thus making the findings of this study more generalisable and reliable.

Table 4.4 Sample size (n) comparison with Treffers-Daller and Milton (2013)

Stage	**Sample size (n) this research**	**Sample size (n) (Treffers-Daller & Milton, 2013)**
1	130	113
2	139	30
3	120	18
Total	389	161

The mean vocabulary size of Stage 1, 2 and 3 students at an East Midlands University (EMU) was compared to the mean vocabulary size found in Treffers-Daller and Milton's (2013) study using the same vocabulary test (Table 4.5).

Treffers-Daller and Milton's (2013) sample included three universities. Of those, students in their first year at UWE Bristol have a similar vocabulary size to those at the East Midlands University but again there is very little difference between all four institutions (Table 4.6).

Table 4.5 Mean vocabulary sizes in stages compared with Treffers-Daller and Milton (2013)

Stage	**Mean ($\bar{x}$) vocab size current study (EMU)**	**Standard Deviation (SD) Current study (EMU)**	**Mean ($\bar{x}$) vocab size Treffers-Daller & Milton (2013)**	**Standard Deviation (SD) Treffers-Daller & Milton (2013)**
1	10,070	2,673.56	9,756	1,976.06
2	11,614	3,015.99	9,793	1,598.72
3	11,582	2,943.75	10,856	1,961.25

Table 4.6 Mean ($\bar{x}$) scores per university for Stage 1 compared with Treffers-Daller and Milton (2013)

University	**Sample size**	**Mean ($\bar{x}$) vocab score**	**Standard Deviation (SD)**
UWE Bristol	21	10,090	1,866.23
City University	40	9,900	1,450.86
Swansea University	52	9,510	2,605.97
For all three above	113	9,756	1,076.06
Current study University (EMU)	130	10,070	2,673.56

4.2.2 Research question 2: What types of words do students not understand?

From the vocabulary record sheets (n = 27), the words students did not know appear in Table 4.7. These words include *ramifications*, *inculcate*, *heuristic*, *ethnocentrism*, *discourse*, *apathetic*, *infallible* and *didactic*. An analysis of the words from the vocabulary record sheet revealed that out of 157 total words, only 9 words and one prefix (neo) (around 10% of the total) were on the NAWL (e.g. *discourse*), 1.2% from the NGSL 1, and 1.2% from the NGSL 3 and the rest were mainly low frequency words (e.g. *edification*) not on any of the lists or 'off-list'. A further analysis of the low frequency words revealed that many of them were highly technical or subject specific

(e.g., *aetiology*, *phablet* and *proselytism*). A few of the low frequency words were slang (e.g., *bae*, *cotch*, *lecky* and *mandem*). It is also important to note that this task has highlighted a limitation of lexical frequency in that it cannot identify fixed expressions as one unit of meaning, for example, *doing a big un,* so has separated the words out in the analysis.

The NAWL offers 6% coverage as discussed in Section 3.13.2. and with 10% of the total words students recorded being as unknown on the NAWL, it could be perceived that the list offers a reasonable level of representation. Conversely, the NGSL has not offered the same level of coverage at 2.4% but given the majority of unknown words were derived from a higher education this may be the reason (see Table 4.7).

Table 4.7 Unknown words recorded by frequency (*f*) all stages from vocabulary record sheet (words highlighted are from NGSL 1, NGSL 2, NGSL 3 and NAWL)

Word	(*f*)	Word	(*f*)	Word	(*f*)	Word	(*f*)
acquiesce	1	empathy	1	moratorium	1	squee	1
adolescent	1	empiricism	1	myriad	1	stochastic	1
advocate	1			natch	1	subsequent	1
aetiology	1	encyclopadism	1	neo-liberal	1	subsequential	1
affiliation	2	epidemiology	1	norse	1	succinct	1
affluent	3	ethnocentrism	3	nuanced	1	syncretism	1
ambiguity	1	ethnographic	1	ontological	1	tautological	1
antecedents	1	exacerbated	1	opacity	1	teething	1
apathetic	6	expedient	1	orita	1	teleology	1
aspersion	1	fallible	1	osmosis	1	theology	1
assimilate	2	food baby	1	paradigm	1	thwart	1
autonomy	1	gleaned	1	pedagogy	1	totalitarian	1
bae	1	habitus	1	perpetuate	3	toting	1
bare	1	heuristic	1	phablet	1	tranquil	1
belligerent	1	hight	1	philogogy	1	unambiguous	1
biggie	1	hinder	1	philological	1	unequivocally	1
bitcoin	1	hypostatising	1	phylactery	1	unilinear	1
BOBFOC	1	ideological	1	plethora	1	verbatim	1
coagulate	1	ideology	1	prevaricate	1	volexional	1
co-located	1	imperialism	4	promulgation	1	wag one	1
concurrent	1	incarcerated	1	proselytism	1		
conflatations	1	incremental	1	proximal	1		
congruent	6	inculcate	1	putative	1		
				pygmalion	1		
convergent		inculcated		effect			
co-operate	1	infallible	1	ramifications	1		
cotch	1	irreconcilable	1	remit	1		
cutural	1		1				
imperialism		juxtaposition		repudiation			
dejected	5	kernal	1	roiling	1		
dekka	1	kinesthetic	1				
didactic	3	laissez-faire	1	schism	1		
discourse	3	latent	1	semantic	1		
discursive	1	lecky	1	serene	1		
doing a big	1		1		1		
un		mandem		simonious			
edification	1	melolagnia	1	simulacrum	1		
elucidate	1	merk	1	simultaneously	1		
emanating	1	microcosm	1	skillet	1		

4.2.3 Research question 3: Is there a correlation between vocabulary sizes and academic achievement?

The third research question in Phase I was to explore whether there is a link between vocabulary size and academic achievement (measured by expected final degree classification). Spearman's rank correlation coefficient (r_s) was calculated and there

is no statistically significant correlation between vocabulary sizes and academic achievement (r_s = .059, p = 0.255). Comparing the mean vocabulary score with expected degree classification (see Table 4.2) using ANOVA also supports this finding. Correlations between expected degree classifications and vocabulary sizes were also explored at Stages 1, 2 and 3 and none were found (see Table 4.8). Based on this evidence, it can be concluded that there is no link between vocabulary sizes and academic achievement from this data.

Table 4.8 *P*-values and correlation coefficients generated from Spearman's rho (r_s) between expected degree classifications and vocabulary sizes

Stage 1	**Stage 2**	**Stage 3**	**All three stages**
r_s = -.098, p = .276	r_s = -.052, p = .547	r_s = -.160, p = .084	r_s = .059, p = 0.255

*Please note four missing degree classifications for Stages 1 and 2 and two missing degree classifications for Stage 3

4.2.4 Summary of key findings for Phase I

Undergraduates have a much smaller vocabulary size than one would expect compared to previous research discussed in the literature review but not compared to findings from Treffers-Daller and Milton (2013). There are statistically significant differences in the mean vocabulary size of students between Stage 1 ($\bar{x}$)=10,070 and Stage 2 ($\bar{x}$)= 11,614. The effect size (*r)* = 0.26 indicates this difference is small suggesting that students' vocabulary sizes change between Stages 1 and 2 but not to a substantial extent. It was hypothesised that vocabulary sizes would play a role in academic achievement. In other words, a larger vocabulary size would link to a higher degree classification. However, there is no difference between the mean vocabulary size of students predicted a first (($\bar{x}$)= 11,521) and those who were predicted a 2:1 (($\bar{x}$)= 11,312); 2:2 (($\bar{x}$)= 11,450) and a 3rd (($\bar{x}$)= 9,833), which supports the finding from Research question 3 that there is no correlation between vocabulary size and academic achievement.

Lastly, from the vocabulary record sheets the words students reported they did not know were a mixture of academic but mainly 'off-list' words which were subject-specific or technical (not related to education). With 10% of the total words students recorded as unknown found on the NAWL, it could be argued that the list offers

reasonable representation of academic words to be more closely examined in Phase II.

4.3 Data findings Phase II

It is pertinent to restate the research questions to contextualise the findings presented in this section. This section is organised in the order of the questions:

1. What is the level of vocabulary richness in written assignments of monolingual and bilingual (using two or three languages including English routinely in the home) English speaking undergraduate students using lexical frequency profiling?
2. What are the changes, if any, in word usage from the NGSL 1, NGSL 2, NGSL 3, NAWL and 'off-list' in student assignments between the stages of study?
3. Is there a correlation between the percentage of academic words from the NAWL in students' written work and assignment scores?

4.3.1 Research questions 1 and 2

In Phase II, there were 41 undergraduate research participants. Three assignments from each participant in each stage of study were analysed in AntWordProfiler. The word count limit for Stage 1 assignments was 3,000, 3,500 for Stage 2 and 4,000 for Stage 3. For each student nine essays were analysed. In all, 1,044,013 running words (or word tokens) were analysed for this phase of the research. The median percentage and Interquartile Range (IQR) of text coverage (TC) (how much of the texts were covered by the words on the list), high frequency words (NGSL 1, the most frequent flemmas 1-1,000), the second 1,000 most frequent flemmas (NGSL 2) and the third 1,000 most frequent flemmas (NGSL 3), the supplementary list (Supp), the NAWL and 'off-list' words (not on any of the other lists) were calculated and can be seen in Table 4.9. The Interquartile range (IQR) was calculated to measure the spread of the data since the median was used a measure of central tendency (Foster *et al,* 2015). The IQR shows what the range or spread is for the middle 50% of observations and is a non-parametric equivalent to standard deviation. Large SDs (relative to the mean) and wide IQRs suggest a substantial amount of variation (Gablasova *et al*, 2017). Since the IQR is small for each item in comparison to the median, we can say that the data is not widely dispersed.

Table 4.9 Median (*Mdn)* percentage, minimum and maximum percentages, Standard Deviation (SD) and Interquartile Range (IQR) of word tokens from NGSL 1, NGSL 2, NGSL 3, Supplementary list, the NAWL and 'off-list' in student assignments for Stages 1, 2 and 3.

		Stage 1	**Stage 2**	**Stage 3**
TC	Median %	93.20	92.30	90.80
Min/max		86.30/98.70	82.10/96.30	85.10/95.50
SD		2.30	2.54	2.34
IQR		91.40-94.40	90.40-93.40	89.38-92.38
NGSL 1	Median%	81.50	80.60	79.20
Min/max		73.80/88.90	63.60/88.70	68.4/92
SD		3.30	4.50	5.51
IQR		78.80-83.50	76.70-82.68	76.43-79.20
NGSL 2	Median%	6.50	6.70	6.90
Min/max		3.50/10.20	4.5/10	3/11.30
SD		1.20	1.40	1.40
IQR		5.80-7.30	5.80-7.70	6.10-8.08
NGSL 3	Median%	2.45	2.35	3.30
Min/max		0.90/5.70	1.20/5.50	1.20/6.40
SD		0.97	0.78	0.98
IQR		2.00-3.20	2.00-3.00	2.48-3.60
Supp	Median%	0.30	0.20	0.20
Min/max		0.00/1.40	0.00/0.50	0.00/0.70
SD		0.28	0.14	0.14
IQR		0.20-0.50	0.10-0.40	0.10-0.030
NAWL	Median%	2.00	2.50	2.30
Min/max		0.70/4.70	1.20/4.40	1.20/4.30
SD		0.84	0.67	0.75
IQR		1.50-2.80	2.10-3.13	1.90-2.90
Off-list	Median%	7.10	7.30	9.20
Min/max		2.60/13.60	3.70/17.30	4.60/14.90
SD		2.20	2.50	2.20
IQR		5.60-8.60	6.70-9.58	7.70-10.70

From the data in Table 4.9, students in Stages 1, 2 and 3 use high frequency words (81.50%, 80.60% and 79.20%) from the NGSL 1 list the most in their written work and words from the supplementary list the least (0.3%, 0.2% and 0.2%). Around 2.0% of their work contains academic words from the NAWL and between 7-9% of their words cannot be found in any list or are 'off list'. Just under 7% of the words students use come from the NGSL 2 list and just under 3% are from the NGSL list 3.

All the lists combined (NGSL 1, NGSL 2, NGSL 3, Supp, and NAWL) cover around 92% of the word tokens contained in student assignments for Stages 1, 2 and 3. Correspondingly, around 8% of words used in student assignments are not found in any of the lists and are 'off-list'.

4.3.2 Differences between stages – Wilcoxon tests

This section reports on the use of statistical tests to identify whether any observed differences from analysing the median were statistically significant. The null hypothesis is there is no difference in the median between the stages. The same guidelines for interpreting *p*-values outlined in Section 4.2 for Phase I are applied here and the following is a reminder how they were used to analyse the results.

- A small *p*-value ($p < 0.05$) indicated strong evidence against the null hypothesis, so the null hypothesis was rejected.
- A large *p*-value ($p > 0.05$) indicates weak evidence against the null hypothesis, so the null hypothesis was not rejected.

Wilcoxon tests were conducted to see if there were any statistically significant differences between the usage of high frequency words from the NGSL 1, and less frequent words from the NGSL 2 and 3 as well as from the supplementary list, NAWL, and 'off-list' words between Stages 1, 2 and 3.

Firstly, in terms of NGSL 1 usage, in all three stages Wilcoxon test indicated that the difference in the median was not statistically significant (see Figure 4.2). The implications of this will be further discussed in Chapter 5.

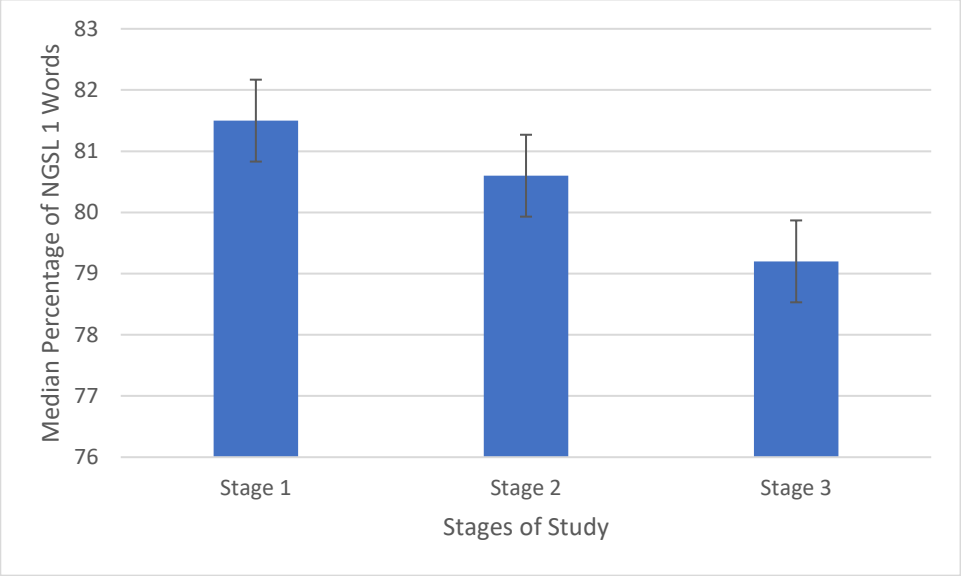


Figure 4.2 Median percentage of NGSL 1 words used in students' assignments compared in stages with standard error bars

In Figure 4.3, there was a statistically significant effect of the stages of study on the words used from the NGSL 2 list in students' assignments between Stages 1 (*Mdn* = 6.50, *SD* = 1.20) and 3 (*Mdn* = 6.90, *SD* = 1.40), *p* =.009. The size of the effect is large, *r* = -6.50, 95% CI [-1.10, -.150] which can be seen in Figure 4.3.

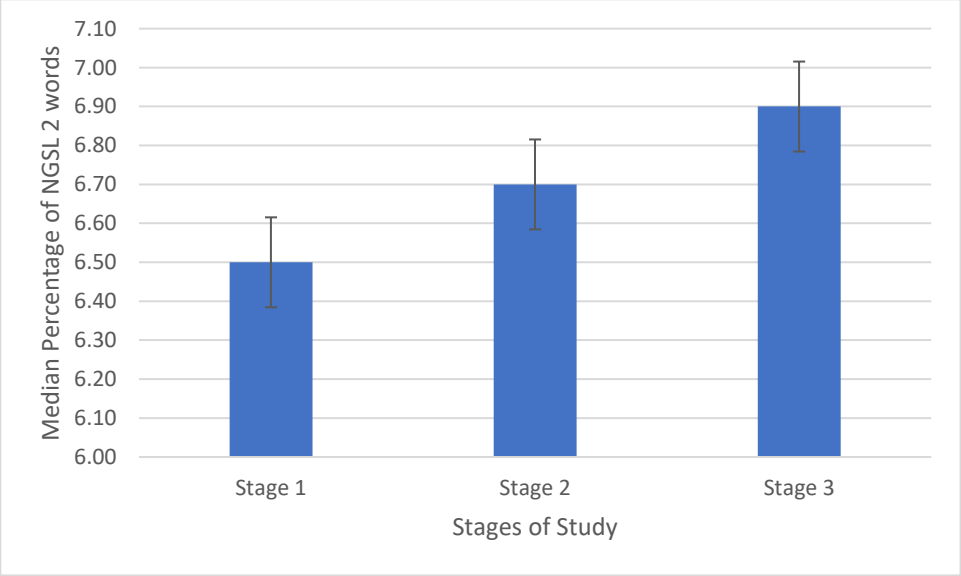


Figure 4.3 Median percentage of NGSL 2 words used in students' assignments compared in stages with standard error bars

In Figure 4.4, there was also a statistically significant effect of the stages of study on the words used from the NGSL 3 list in students' assignments between Stages 1 (*Mdn* = 2.45, *SD* = 0.97) and 3 (*Mdn* = 3.30, *SD* = 0.98), *p* =.02. The size of the effect is medium, *r* = -.400, 95% CI [-.750, -.50]. Between Stages 2 (*Mdn* = 2.35, *SD* = 0.78) and 3 (*Mdn* = 3.30, *SD* = 0.98), *p*<.001, the effect size is large, *r* = -.800, 95% CI [-1.20, -.450].

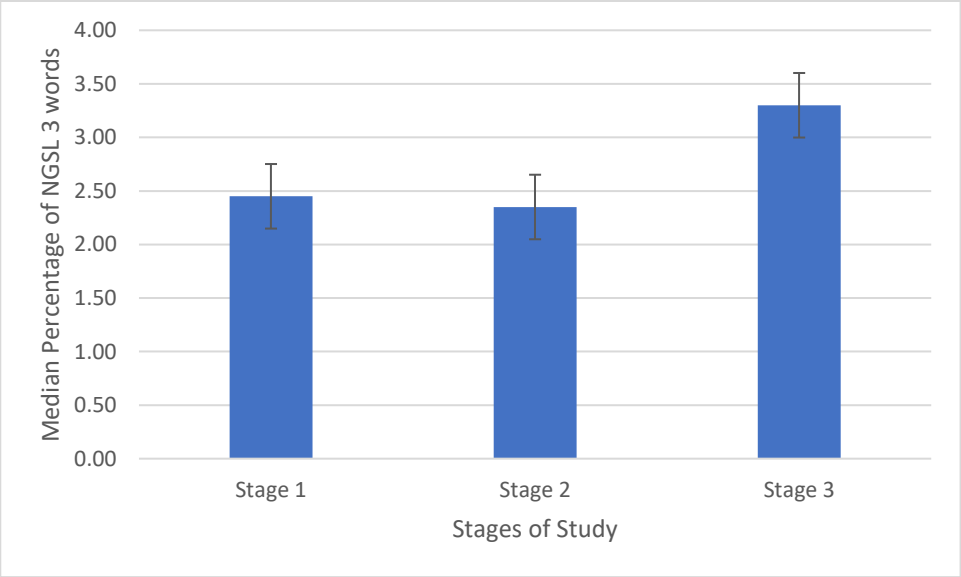


Figure 4.4 Median percentage of NGSL 3 words used in students' assignments compared in stages with standard error bars

In Figure 4.5, there was a statistically significant effect of the stages of study on the words used from the supplementary list in student assignments between Stages 1 (*Mdn* = 0.30, *SD* = 0.28) and 3 (*Mdn* = 0.20 , *SD* = 0.14), *p*=.001.The effect size is small, *r* =.100, 95% CI [.050, .200] but also between Stages 1 and 2 (*Mdn* = 0.20, *SD* = 0.14), *p*<.001. The effect size is small, *r* =.150 [.100, .250].

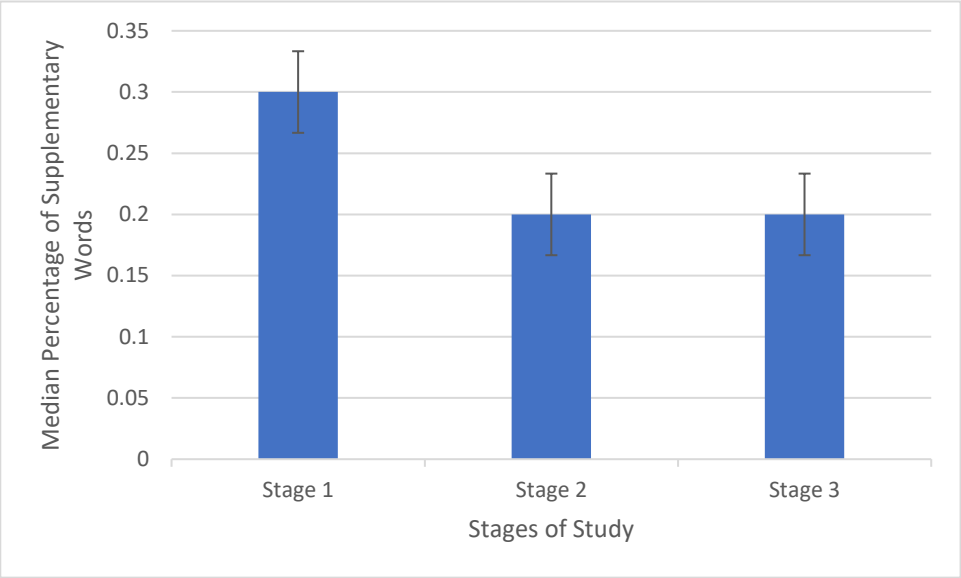


Figure 4.5 Median percentage of words from the Supplementary list used in students' assignments compared in stages with standard error bars

In Figure 4.6 there was a statistically significant effect of the stages of study on the words used from the NAWL in student assignments between stages 1(*Mdn* = 2.00, *SD* = 0.84) and 2 (*Mdn* = 2.50, *SD* = 0.67), *p* =.002. The effect size is medium, *r* =.400, 95% CI [.150, .650]. There was also a statistically significant effect between Stages 1 and 3 (*Mdn* = 2.30, *SD* = 0.75), *p* =.024. The effect size is small to medium, *r* =.300, 95% CI [.050, .600].

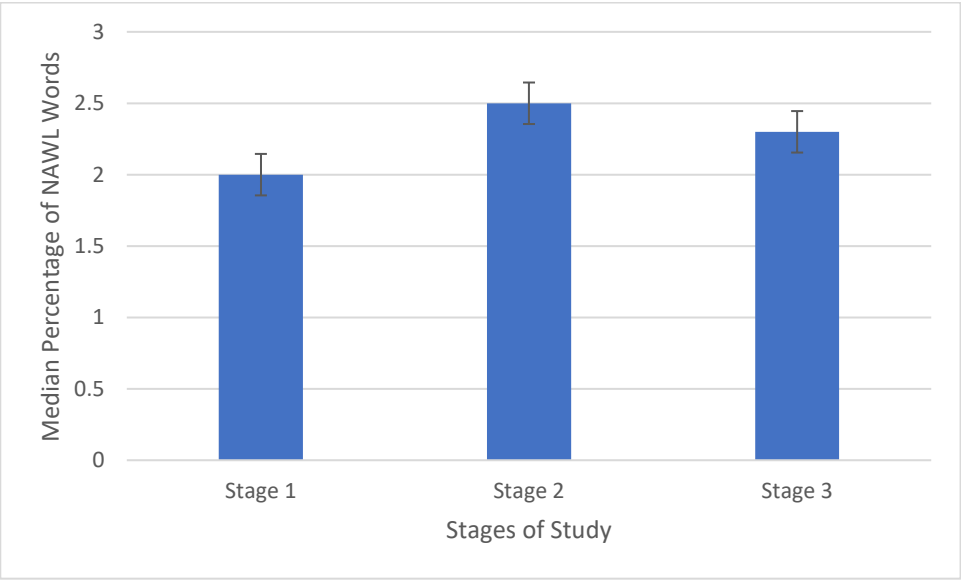


Figure 4.6 Median percentage of words from the NAWL used in students' assignments compared in stages with standard error bars

In Figure 4.7, there is a statistically significant difference effect of the stages of study on 'off-list' words in students' assignments between Stages 1(*Mdn* = 7.10, *SD* = 2.20.) and 3 (*Mdn* = 9.20, *SD* = 2.20), *p*<.001. The effect size is large, *r* = -1.80, 95% CI [-2.60, -.850.

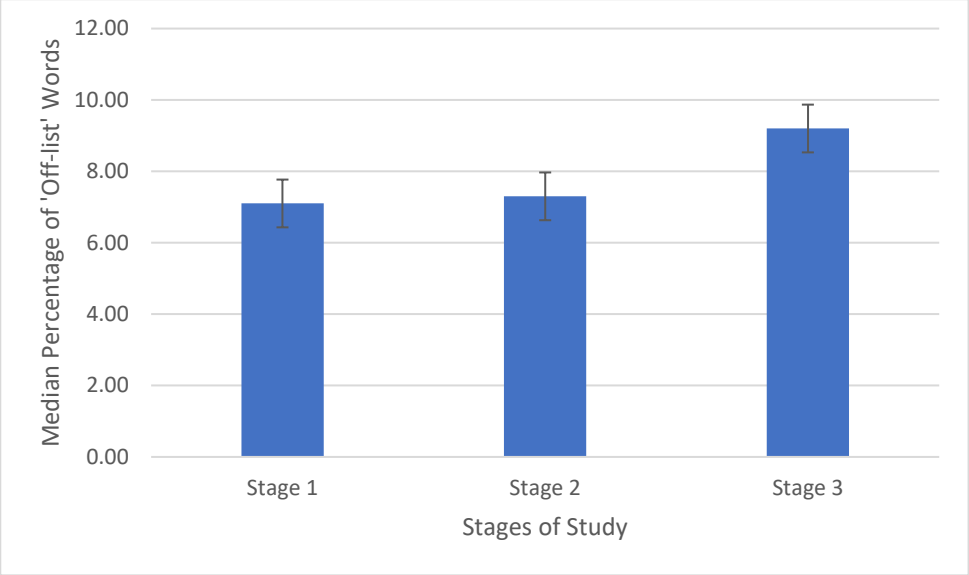


Figure 4.7 Median percentage of words that are 'off-list' in students' assignments compared in stages with standard error bars

In Figure 4.8, there is a statistically significant difference effect of the stages of study on the text coverage all the lists provide for student assignments between Stages 1(*Mdn* = 93.2, *SD* = 2.30) and 3 (*Mdn* = 90.8, *SD* = 2.34), *p*<.001. The effect size is large, *r* = 1.750, 95% CI [.800, 2.55].

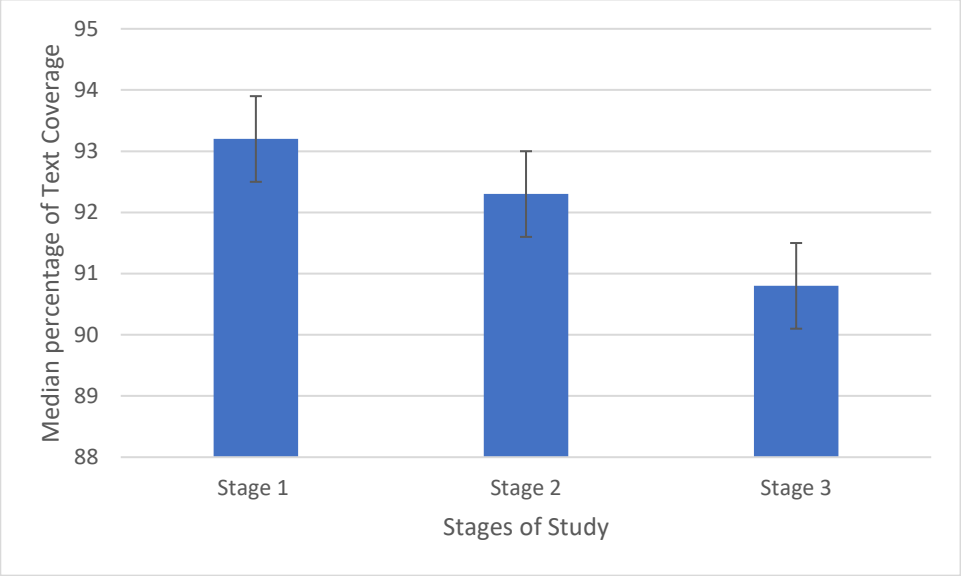


Figure 4.8 Median percentage of words covered by all the lists (TC) with standard error bars

4.3.3 Presentation and analysis of findings for Research question 3

Research question 3 was to investigate whether there is a correlation between the percentage of academic words from the NAWL in students' written work and their marks for each assignment. To do this, Spearman's Rank Correlation (rho) r_2 was calculated with a 99% confidence interval (p<.01) in order to reduce the risk of a Type I error. In addition, for interpreting correlation coefficients, Field (2018) suggests that correlations are small if the values are between 0 and 0.1, small to medium between 0.1 and 0.3, medium to large between 0.3 and 0.5 and large if the value is greater than 0.5. This is key for interpreting the data and the size of the correlation.

Table 4.10 Correlation table (Spearman's Rank Correlation): NGSL 1, NGSL 2, NGSL 3, Supp, NAWL, 'Off-list' and Marks (for each assignment) in Stage 1

	Marks	**NGSL 1 words**	**NGSL 2 words**	**NGSL 3 words**	**Supp words**	**NAWL words**	**'Off-list' words**
Marks	1.00**	-0.45**	0.27**	0.71	0.02	0.15	0.39**
NGSL 1 words	-0.45**	1.00**	-0.49**	-0.37**	-0.18	-0.55**	-0.75**
NGSL 2 words	0.27**	-0.49**	1.00**	0.02	-0.03	0.11	0.09
NGSL 3 words	0.71	0.37**	0.02	1.00**	-0.07	0.34**	0.09
Supp words	0.02	-0.18	-0.03	-0.07	1.00**	0.07	0.22
NAWL words	0.15	-0.55**	0.11	0.34**	0.07	1.00**	0.24
'Off-list' words	0.39**	-0.75**	0.09	0.09	0.22	0.24	1.00**

**p<.01

Table 4.11 Correlation table (Spearman's Rank Correlation): NGSL 1, NGSL 2, NGSL 3, Supp, NAWL, 'Off-list' and Marks (for each assignment) in Stage 2

	Marks	**NGSL 1 words**	**NGSL 2 words**	**NGSL 3 words**	**Supp words**	**NAWL words**	**'Off-list' words**
Marks	1.00**	-0.28**	0.22*	0.25**	-0.66	0.09	0.09
NGSL 1 words	-0.28**	1.00**	-0.74**	-0.45	-0.18	-0.39**	-0.59
NGSL 2 words	0.22*	-0.74**	1.00**	0.26**	0.16	0.20	0.40**
NGSL 3 words	0.25**	-0.45**	0.26**	1.00**	-0.15	0.10	0.35**
Supp words	-0.66	-0.18	0.16	-0.15	1.00**	0.25	0.38**
NAWL words	0.09	-0.39**	0.19	0.10	0.25	1.00**	0.23
'Off-list' words	0.09	-0.59**	0.40**	0.35**	0.38**	0.23	1.00**

**p<.01

Table 4.12 Correlation table (Spearman's Rank Correlation): NGSL 1, NGSL 2, NGSL 3, Supp, NAWL, 'Off-list' and Marks (for each assignment) in Stage 3

	Marks	**NGSL 1 words**	**NGSL 2 words**	**NGSL 3 words**	**Supp words**	**NAWL words**	**'Off-list' words**
Marks	1.00**	-0.21	0.10	0.02	-.00	0.26	0.27**
NGSL 1 words	-0.21	1.00**	-0.19	-0.08	-0.07	-0.36	-0.61
NGSL 2 words	0.10	-0.19	1.00**	0.13	0.19	-0.01	0.39**
NGSL 3 words	0.02	-0.08	0.13	1.00**	0.14	0.01	0.24
Supp words	-.00	-0.07	0.19	0.14	1.00**	-0.15	0.02
NAWL words	0.26	-0.36**	-0.01	0.01	-0.15	1.00**	0.26**
'Off-list' words	0.27**	-0.61**	0.39**	0.24	0.02	0.26**	1.00**

**p<.01

From Tables 4.10, 4.11 and 4.12, it can be seen there is no correlation between marks on students' assignments and their usage of words from the NAWL. However, there is a medium negative correlation (r_2 = -.45, p<.01) between the usage of NGSL 1 words used in student assignments which can be seen in Figure 4.9 and their marks in Stage 1 and a small negative correlation (r_2 = -.28, p<.01) in Stage 2. In other words, as students' marks on their assignments increase, the usage of words on the NGSL 1 decrease, particularly in Stage 1.

Conversely, there is a small positive correlation between the usage of NGSL 2 words used in student assignments and marks in Stage 1 (r_2 = .27, p<.01) and in Stage 2 (r_2 = .27, p<.01) but not in Stage 3. As students' marks on their assignments increase in Stages 1 and 2, so does the usage of words from the NGSL 2. In terms of the

NGSL 3, no correlations were found in Stages 1 and 3 in relation to marks but there is a small positive correlation (r_2 = .25, p<.01) in Stage 2.

No significant correlations were found between the supplementary list (Supp) and marks on students' assignments in all three stages of study as well as the NAWL. Finally, a positive medium/small size correlation (r_2 = .39, p<.01) is seen in Stage 1 between marks and 'off-list words' used and in Stage 3, a small positive correlation can be seen (r_2 = .27, p<.01) between marks and 'off-list' words (see Figure 4.10).

The correlation Tables 4.10, 4.11 and 4.12 also reveal relationships between the word lists used in the research. There is a medium to large, negative correlation (r_2 = -.55, p<.01) between the usage of words in student assignments from the NAWL and the NGSL 1 in Stage 1 and in Stage 2 the correlation is medium (r_2 = -.39, p<.01). As students use more words from the NAWL there is a corresponding decrease in words used from the NGSL 1. There are no correlations between the NAWL and the NGSL 2 as well in the Supp list. However, there is a positive, medium correlation (r_2 = .34, p<.01) between the words used on the NAWL and the NGSL 3 in Stage 1 only.

Finally, as can be seen in Table 4.12, there is a positive, medium correlation (r_2 = .39, p<.01) in Stage 3 between the 'off-list' words and the NGSL 2. There are also positive medium correlations between 'off-list' words and the NGSL 2 (r_2 = .40, p<.01) as well as the NGSL 3 (r_2 = .35, p<.01) and Supp list (r_2 = 38, p<.01) in Stage 2 (see Table 4.12). Conversely, there is a negative, large correlation between 'off-list' words used (r_2 = -.75, p<.01) and usage of NGSL 1 words in student assignments in Stage 1 (see Table 4.11).

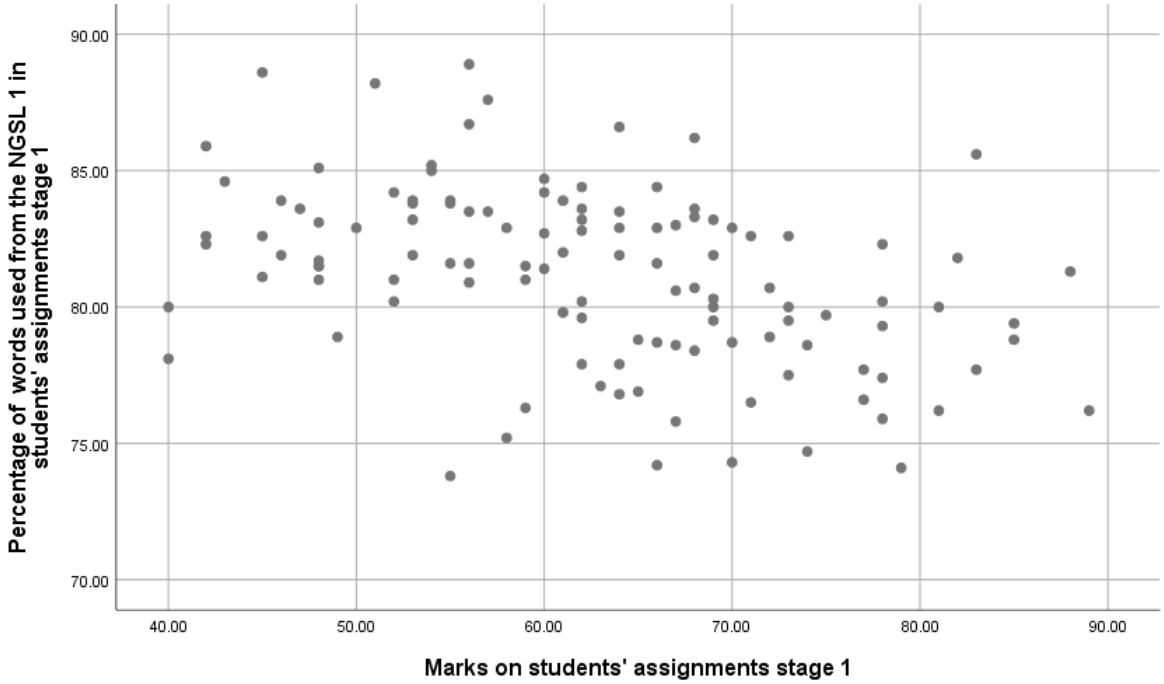


Figure 4.9 Scatter plot of Stage 1 percentage of NGSL1 word tokens used in students' assignments compared to their marks (r_2 = -.45, p<.01)

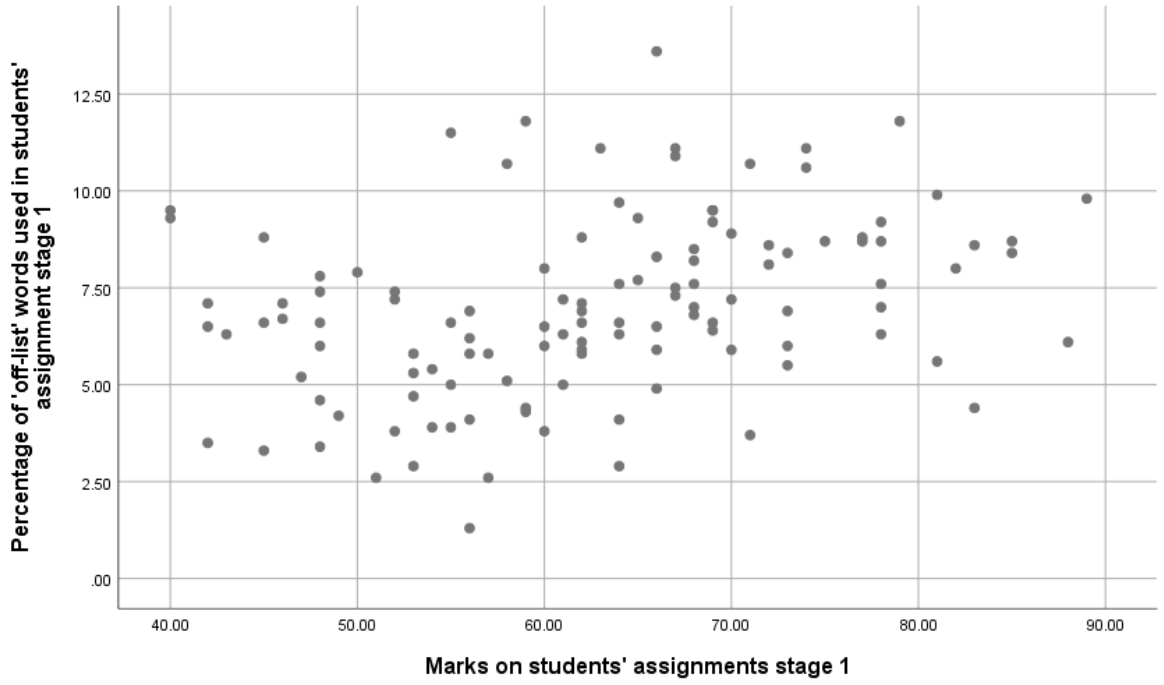


Figure 4.10 Scatter plot of Stage 1 percentage of 'off-list' words used in student's assignments compared to their marks (r_2 = .39, p<.01)

4.3.4 Finer grained analysis of six participants

To place the analysis of six cases in context, it is necessary to establish which words all the students in the corpus have used most frequently from the NAWL. The corpus was divided into three sub-corpora (based on stages) for this purpose to find patterns between the stages of study. In the following word frequency analysis, the absolute and relative frequency were calculated. The absolute frequency (AF) is a count of all the tokens in the corpus that belong to a particular word type (Brezina, 2018a). However, the relative frequency is more useful for comparing two or more corpora (in this research between the stages or sub-corpora) which is the absolute frequency divided by the total number of word tokens, multiplied by the bases of normalisation. In smaller corpora, smaller bases for normalisation are more appropriate (Brezina, 2018a). In this case a normalisation per 1,000 words seems suitable, particularly given the word limits per assignment (3,000, 3,500 and 4,000 for Stages 1,2 and 3 respectively). The range tells the number of texts that the word is found in.

Table 4.13 Top ten word types from the NAWL in the Stage 1 sub-corpus

Rank	***Word***	***Range***	***Absolute frequency***	***Relative frequency***
1	*curriculum*	71	888	3.19
2	*inequality*	29	300	1.08
3	*classroom*	58	247	0.89
4	*aspect*	75	178	0.64
5	*found*	60	170	0.61
6	*impact*	66	162	0.58
7	*assignment*	42	156	0.56
8	*disadvantage*	48	131	0.47
9	*cognitive*	24	119	0.43
10	*diverse*	33	92	0.33

The Stage 1 corpus consists of 278,058 tokens and 6672 types.

The word *curriculum* is the most frequent word type from the NAWL in the Stage 1 corpus occurring 888 times (3.19 times per 1,000 words) but only in 71 texts [*R(w)* = 71] out of 123.

Table 4.14 Top ten word types from the NAWL in the Stage 2 sub-corpus

Rank	***Word***	***Range***	***Absolute frequency***	***Relative frequency***
1	*curriculum*	85	1,643	4.61
2	*classroom*	86	392	1.1
3	*traits*	34	346	0.97
4	*impact*	78	211	0.59
5	*found*	68	182	0.51
6	*aspects*	75	177	0.50
7	*theorists*	50	166	0.47
8	*non*	49	150	0.42
9	*disability*	15	147	0.41
10	*globalisation*	9	135	0.37

In Stage 2, the sub-corpus consists of 356,362 word tokens and 10,545 types. The word *curriculum* is the most frequent word type from the NAWL occurring 1,643 times (4.61 times per 1,000 words) but only in 85 texts out of 123 [*R(w)* = 85].

Table 4.15 Top ten word types from the NAWL in the Stage 3 sub-corpus

Rank	***Word***	***Range***	***Absolute frequency***	***Relative frequency***
1	*classroom*	83	639	1.56
2	*inclusion*	48	597	1.46
3	*curriculum*	74	407	0.99
4	*found*	83	323	0.79
5	*impact*	95	303	0.73
6	*cognitive*	62	256	0.63
7	*socially*	43	247	0.60
8	*disability*	25	245	0.60
9	*aspects*	80	221	0.54
10	*paradigm*	20	178	0.43

In the Stage 3 sub-corpora, there was a total of 409,593 word tokens and 11,385 word types. The word *classroom* is the most frequent word type from the NAWL occurring 639 times (1.56 times per 1,000 words) and in 83 texts out of 123 [*R(w)* = 85].

In all three stage sub-corpora, *curriculum*, *classroom*, *found* and *impact* are in the top ten most frequently used word types. Although different word types, *aspect/aspects* can be found in the top ten most frequently used word types. *Disability* can be found in the top ten most frequently used word types in Stages 2 and 3 but not in Stage 1.

An additional analysis of the output from AntWordProfiler was conducted on six cases, selected at random. This analysis included the number of types and tokens from the NAWL for each assignment, but the focus was concerned with the most frequent types of academic words used from the NAWL across all three sub-corpora (in other words the frequency in the use of *curriculum*, *classroom*, *found* and *impact*). The types and tokens were calculated from the NAWL from each student assignment (three) at each stage of study (see Appendix 8), totalling nine. The types show how many different word forms are used from the NAWL. Topics of the texts were also considered as the general genre as discussed in the methodology, Section 3.16, was stable. Each assignment fell into the category ‘to develop powers of reasoning’ (Nesi & Gardner, 2012). Each of the assignments were from modules which formally sat in a particular strand of study that ran through each stage. The strands were international, psychology, sociology, curriculum, SEND, study skills and lifelong learning.

The word *curriculum* was the most repeated word from the NAWL in all three stages – see Tables 4.16, 4.17 and 4.18 for the AF when greater than 10 for each student.

Table 4.16 Instances per text in Stage 1 where the AF of *curriculum* is greater than ten

Student code	**Text topic**	**AF**
122	international	18
121	international	26
135	international	32

Table 4.17 Instances per text in Stage 2 where the AF of *curriculum* is greater than ten

Student code	Text topic	AF
107	curriculum	61
122	curriculum	29
135	curriculum	55
143	curriculum	51

Example 1

> A contemporary debate surrounding those in control of the ***curriculum*** has always been present in education. Gove's (2011) ***curriculum*** has seen a ***curriculum*** left in the hands of the government and policy makers, in comparison to a ***curriculum*** made up of both government and educational professions in the past. Additionally, Kelly (2009) explains changes to the education system where a pendulum between the importance of knowledge and skills has differed between stakeholders; causing little chance for a collaborative ***curriculum*** to work.

The NAWL item *curriculum* is in bold and italics. The sample text is taken from a text from student 107 in Stage 2; this text had the highest AF of the word *curriculum* in all the texts across all three stages.

Table 4.18 Instances per text in Stage 3 where the AF of *curriculum* is greater than ten

Student code	Text topic	AF
107	psychology	12
124	curriculum	21
143	curriculum	20

In Table 4.16, a closer examination of the topics within the international strand in Stage 1, revealed why *curriculum* was so frequently used. The assignments all compared the English national curriculum to another country. Obviously in Stage 2, the topic curriculum generated the greatest number of the word type *curriculum*. In

Stage 3, the assignment topic for student 107 was well-being in the curriculum, and the topic for both students in 124 and 143 was art in the UK national curriculum.

This data suggests that the largest influence on the word types used from the NAWL stems from the assignment topic rather than the genre at large. Further evidence is taken from a specific look at student 122.

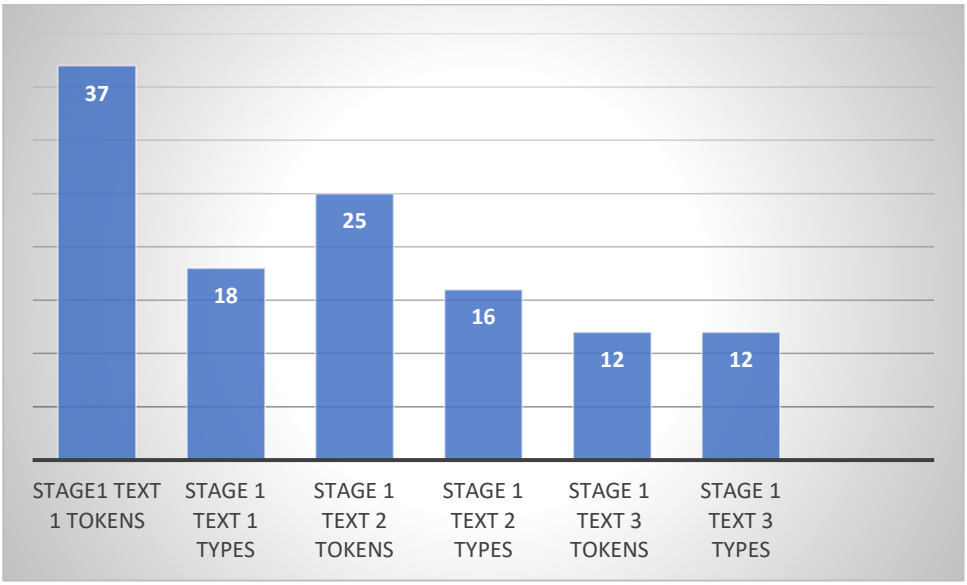


Figure 4.11 The number of tokens and types used from the NAWL in three texts in Stage 1 for student 122

In Figure 4.11, in text 1, there were 37 tokens and 18 types, but the word *curriculum* was used 18 times. The topic of the assignment was a comparison of the curriculum between Luxembourg and England. In text 2, the word *assignment* was used six times, and both *critically* and *thereby* were used three times with no mention of *curriculum*. The topic of this text was concerned with the characteristics of academic writing. It could be argued that the word *curriculum* is not relevant to this topic and perhaps why it was not used in this text. However, the number of tokens from the NAWL in text 3 (topic-vocational courses) was 12 and number of types the same revealing no repetition of words.

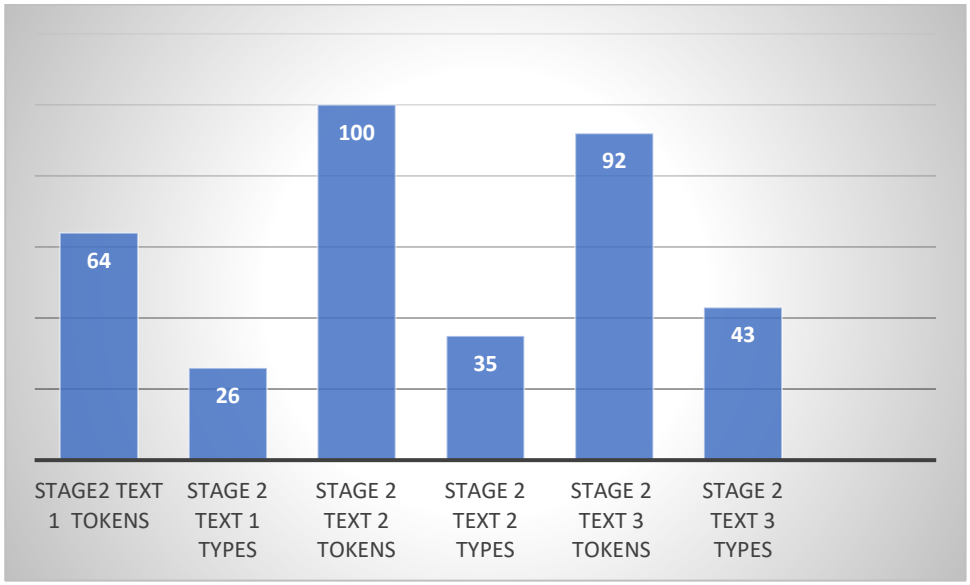


Figure 4.12 The number of tokens and types used from the NAWL in three texts in Stage 2 for student 122

In Stage 2, student 122 used *curriculum* 29 times in text 1 (topic is curriculum) but not in text 2 or 3 at all (Figure 4.12). In comparison, the topic of text 2 is SEND, more specifically the assignment explored definitions of disability. In this text, the word type *disability* from the NAWL appeared 35 times and *curriculum* not once.

Example 2

> In contrast the social model of ***disability*** would view the map as a disabling barrier as it was not laid out clearly. The social model views ***disability*** as a problem created by society. It believes society creates barriers which restrict the individual's life which makes them disabled. Shakespeare (2006) states the social model of ***disability*** 'downplays the role of impairment in the lives of disabled people.'

The NAWL item *disability* is in bold and italics. The sample text is taken from text 2 from student 122 in Stage 2.

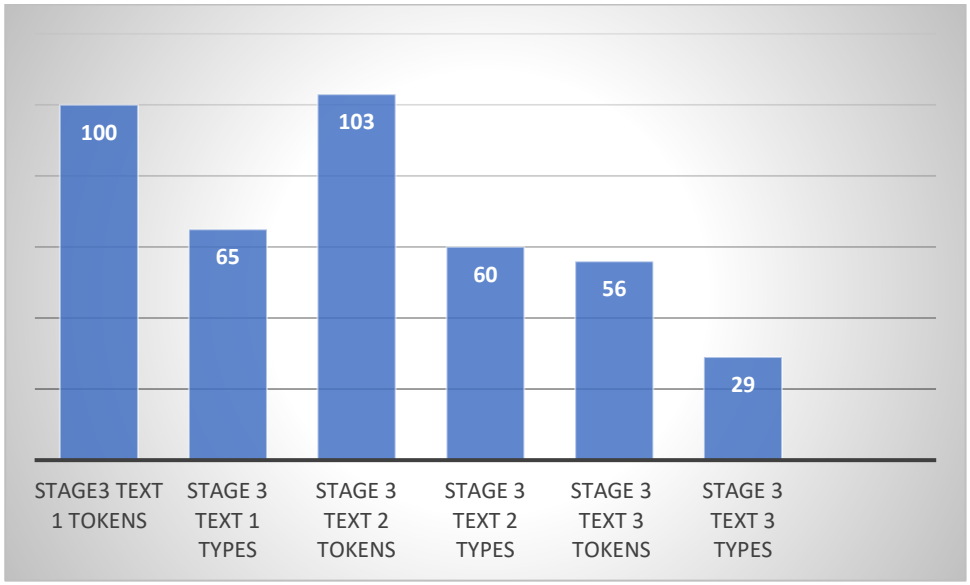


Figure 4.13 The number of tokens and types used from the NAWL in three texts in Stage 3 for student 122

In Stage 3, text 1, student 122 used 100 tokens and 65 types from the NAWL. The most repeated word from the NAWL was *found* with an AF of 9 while *impact* appeared once. The topic for this assignment is social exclusion (sociology). However, the word *exclusion* does not appear on any of the lists, including the NAWL, so is considered 'off-list' and was used 46 times.

Example 3

> There is debate about what social ***exclusion*** signifies and how it is best used in verbal and policy contexts (De Haan, 2001). Notwithstanding the level of discussion, the concept is commonly used and seems to be intensely attractive to the producers of social policy discourse (Blackmore and Warick-Booth, 2013). The idea of social ***exclusion*** endeavours to capture the complexity of helplessness in modern society rather than merely centring on one of its outcomes (Gorard et al, 2003; Levitas, 2005; Tomlinson, 2005). Consequently, in practical terms, the UK government Social ***Exclusion*** Unit (1998) outlines ***exclusion*** in terms of a procedure of linked problems such as unemployment, poor skills, low income, poor housing, high crime environments, bad health and family breakdown (Hills and Stewart, 2005).

The 'off-list' item *exclusion* is in bold and italics. The sample text is taken from a text from student 122 in Stage 3. In this text, the highest AF 'off-list' word was *exclusion.*

In text 2 for student 122, the topic was well-being (psychology). The most repeated words from the NAWL were *found* with an AF of 17 and *classroom*, AF (12). *Curriculum* appeared once. However, the word *well-being* is not on the NAWL or other base word lists so is considered 'off-list' and was used 64 times.

Finally, the topic for text 3 for student 122, was behaviour management (psychology). The most repeated words from the NAWL were *classroom*, AF (7), *found* (2), *curriculum* (1). The most frequent word types being *reinforcement,* AF (9) and *assertive*, AF (8) from the 'off-list'.

Example 4

> Canter and Canter (1976) developed a behaviourist driven approach called 'Assertive Discipline' to deal with behaviour in schools. Assertive discipline is a structured approach which enforces a teacher-controlled ***classroom***. The package enforces pupils should be made fully aware of ***classroom*** rules and what is expected of them. Constant reinforcement is required for both positive and negative behaviour. Unwanted behaviour must be addressed quickly, with the reinforcement of the ***classroom*** rules. Teachers must be assertive instead of being aggressive. Children's behaviour can be put down to the teacher approach. This follows on to Canter and Canter (2005) model in which they categorise teachers into three fields, non-assertive, assertive and hostile teachers.

The NAWL item *classroom* is in bold and italics. The sample text is taken from text 3 from student 122 in Stage 3.

4.3.5 Summary of key findings for Phase II

The median percentage of word tokens from the NGSL 1 used in assignments for Stages 1, 2 and 3 was 81.50%, 80.60% and 79.20% respectively, while the median percentage of word tokens from the NGSL 2 was 6.50%, 6.70% and 6.90% and from the NGSL 3, 2.45%, 2.35% and 3.30% for Stages 1, 2 and 3. From the NAWL, the median percentage of word tokens used in Stage 1 was 2.00%, in Stage 2, 2.50% and in Stage 3, 2.30%. The median percentage of 'off-list' word tokens was 7.10%, 7.30% and 9.20% for Stages 1, 2 and 3. Finally, the median percentage of word tokens covered by all the list (TC) was 93.20%, 92.30% and 90.80% for Stages 1, 2 and 3, respectively.

Key findings for Research questions 1 and 2 indicate that most of the words that students use in their writing are high frequency (Words 1-1,000). However, the proportion of word tokens from the NGSL 1 decreases from Stage 1 to Stage 2 and from Stage 2 to Stage 3. Conversely, students used more words from the NGSL 2 from Stages 1 to 3. Students used more 'off-list' words in Stage 3 compared to Stages 1 and 2. For academic word usage from the NAWL, students use more academic words in Stage 2 as compared to Stages 1 and 3. These findings suggest that students do experience change in their vocabulary while at university.

Key findings for Research question 3 show that there are no correlations between the marks on students' assignments and the percentage of academic words used from the NAWL. The findings also show that there are no correlations between marks and the NGSL 3. However, there are indications that there is a small positive correlation between marks and the usage of NGSL 2 words in student assignments in Stage 1 and 2 but not in Stage 3. Moreover, there is a positive, small correlation in Stage 3 between the 'off-list' words and marks in students' assignments. Finally, positive small size correlations can be seen in Stages 1 and 3 between marks and 'off-list words' used in Stage 3.

A more detailed analysis of the word types students used revealed that in all three stage sub-corpora, *curriculum*, *classroom*, *found*, and *impact* are in the top ten most frequently used word types with the word *curriculum* being the most repeated. It is also surmised that the topic of the assignment plays a key role in which words are used from the NAWL and the usage of 'off-list' word types.

The following chapter will further discuss these findings with reference to the literature that was reviewed in Chapter 2.

Chapter 5: Discussion

5.1 Introduction

The aim of this research is to explore undergraduates' vocabulary size and lexical richness in a UK university. It was conducted in two phases with the first phase centred on measuring students' receptive knowledge of vocabulary through testing vocabulary sizes, formulating an understanding of the words students do not understand and exploring whether there is a link between vocabulary size and academic achievement (measured by expected final degree classification). The second phase focused on measuring students' productive vocabulary knowledge by estimating the level of vocabulary richness in written assignments of English-speaking undergraduate students by using word frequency profiling; exploring whether undergraduate students experience changes in their vocabulary knowledge by measuring their academic word usage from Stage 1 to 3; and investigating whether there is a correlation between the percentage of academic words in students' written work and their assignment marks. The purpose of this chapter is to discuss the research questions with respect to the findings and previous research considered in the literature review. It will establish whether the findings of this research support, contest or add new knowledge to existing research. The chapter is organised by phases and research questions and themes identified in the literature review.

5.2 Phase I

5.2.1 Research question 1: Receptive vocabulary sizes

Before discussing the research findings, it is important to establish what inferences can and cannot be made from examining vocabulary sizes. The aim of vocabulary size testing is to find out how many words someone knows but this can only ever be an estimate as not every single word in the English language is tested. Additionally, receptive vocabulary size tests cannot measure all aspects of knowing a word (see Section 2.8) so do not indicate how well the word is known and whether it can be used in speaking or writing (Nation & Coxhead, 2021). Goulden *et al*'s test (1990) measures the form-meaning link; in other words, what the word looks like and means. It does not test the strength of that link. Additionally, because the data is taken from a cross section of the population, inferences about vocabulary growth need to be taken with caution since the same individuals were not tested at each stage of study.

With these points in mind, this section discusses the findings in relation to previous research and how the data could be useful in a practical sense in the HE context.

Previous research has shown that estimates of receptive vocabulary sizes for university undergraduates vary from 215,000 words (Hartmann, 1946) to 9,800 (Treffers-Daller & Milton, 2013). Findings from this study suggest that a typical undergraduate may have a vocabulary size of around 10,000-12,000-word families. This supports Schmitt and Schmitt's (2020) conclusion that native speakers of English have a vocabulary size of about 10,000 to 13,000-word families. As outlined previously, the test used in this research was created by Goulden *et al* (1990) who found the average vocabulary size of a native English-speaking university graduate to be 17,200 words while Zechmeister *et al* (1995) reported that first-year college students were able to recognize the meanings of about 12,000 words. Finally, Treffers-Daller and Milton (2013) using the same test as this research (i.e., Goulden *et al*'s) found that the average vocabulary size was around 10,000 to11,000-word families.

As argued in Section 3.3.1, Goulden *et al*'s (1990) test was chosen, partly, in order to make valid comparisons between this research and previous research as it was also used in Treffers-Daller and Milton's (2013) UK study. Another reason for choosing the test was Treffers-Daller and Milton (2013) identified that a limitation of their study was their small sample size (n = 161) despite being taken from three universities. A further limitation of their sample was it only consisted of students studying degrees in Humanities and Speech and Language Therapy. The differences in the sample and size of this research as compared to Goulden *et al* (1990) and Treffers-Daller and Milton (2013) could partially account for the different vocabulary size estimates. Goulden *et al*'s sample consisted of 20 native speakers who were university graduates over the age of 22 (in the USA). It could be argued that their sample size was too small to make any generalised conclusions. As discussed in the methodology chapter, a sample should be large enough to represent the population (Salkind, 2014) and have at least 30 participants (Gray, 2018). Even though the sample in this research was taken from only one university, it consisted of 389 participants from 14 degree courses ranging from Biology to Sociology (see Appendix 4) which is arguably more representative of the population and provides confidence in the potential to draw more valid conclusions and generalisations from this study.

Treffers-Daller and Milton (2013, p. 159) stated ‘To our knowledge, vocabulary size estimation on a significantly large sample of British university students has not been reported in the academic literature. Almost all previous studies are based on speakers of American English’. Table 4.3 illustrates this point. However, the sample size of Treffers-Daller and Milton (2013) was again smaller than in this research but larger than that of Goulden *et al* (1990) with a total of 161 participants (see Table 4.5). Even though they used the test devised by Goulden *et al* (1990), their results are the lowest in the literature (see Table 1.1). Most of their data came from first year participants with only 18 third year students taking part in their research. This could suggest age and/or stage is a factor in vocabulary size. Their study was also conducted in the UK so another potential variable influencing the scores could be due to differences between American and British English.

Moreover, another potential issue rests in the age boundary of the population in vocabulary size research. It is clear from the critique of vocabulary tests in Section 2.9 ‘… that one vocabulary measure will not be suitable for native speakers of all ages’ (Nation & Coxhead, 2021, p. 115). The aim of vocabulary size research using dictionary-sampling methodology is, in general, to find out the average vocabulary size of an adult. However, the age of an ‘adult’ is not usually defined, and samples are typically first-year university students (Brysbaert *et al*, 2016) (see also Table 2.1). If the average age of Goulden *et al*’s (1990) sample is over 22, this could have had an impact on their results leading to an overestimation. As discussed in the literature review, ‘vocabulary knowledge seems to peak around age 50 or possibly later, and decline only slowly, if at all, into old age’ (Bowles *et al*, 2005, p. 234). Brysbaert *et al*’s (2016) research supports this notion as they found that the average 60-year-old knows more words than a 20-year-old (13,400 and 11,100 words respectively).

In addition to age, other factors such as low socio-economic status, lack of access to technology and poor reading skills can negatively affect vocabulary sizes (Nation & Coxhead, 2021). The literature review also showed there is a possible vocabulary gap linked to poor reading skills from primary through to secondary school, extending into higher education (Brost & Bradley, 2006; Duff *et al*, 2015). ‘Vocabulary knowledge affects learning to read and reading in its turn affects vocabulary knowledge’ (Nation & Coxhead, 2021, p. 5). Lower vocabulary size scores could be linked to poor reading skills.

Therefore, a key question is what vocabulary size is necessary in order to 'read for a degree' at university. The literature review highlighted variations in the number of word families needed by native speakers to comprehend certain texts. Laufer (1989,1992) reported it is necessary to know around 3,000-word families in order to comprehend authentic texts while Nation (2006) stated around 8,000 to 9,000 families are necessary to read widely. However, Laufer and Ravenhorst-Kalovski (2010) believe that for L2 learners a minimum threshold (the ability to read with some guidance) of 4,000 to 5,000 words is necessary and to read independently around 8,000-word families is necessary. Finally, Hsu's (2011, 2014) research (albeit in a EFL context) suggests that students need knowledge of around 5,000-word families to access discipline specific texts.

In Stage 1, the mean ($\bar{x}$) vocabulary size of undergraduates was 10,070 and in Stages 2 and 3, 11,614 and 11,582, respectively. There was a small, significant difference in the mean ($\bar{x}$) between Stages 1 and 2 as well as 1 and 3. All three mean scores are well above Laufer's (1989,1992) estimate of 3,000 and Hsu's (2011, 2014) 5,000. However, the mean vocabulary size for Stage 1 is closer to Nation's (2006) top figure of 9,000 but, arguably, only just over this. There was a small, significant difference [(*r)* = 0.26] between Stages 1 and 2 and Stages 1 and 3 in the mean ($\bar{x}$) vocabulary sizes so a closer look at the minimum scores is necessary since this difference is not large.

The minimum scores for each stage were 5,500, 5,000 and 4,000 meaning some participants scored well below the average. These scores are lower than Nation's (2006) estimates and just at or below Hsu's (2011,2014), particularly in Stage 3. As discussed in Section 2.3, research on L1 8-year-olds in New Zealand reported they have vocabulary sizes of between 4,540 to 5,056 words (Nation & Coxhead, 2021). The minimum scores for this research are at the minimum threshold for L2 learners (Laufer & Ravenhorst-Kalovski, 2010) and potentially that of an 8-year-old suggesting some native speakers are likely to need some support with their vocabulary and accessing academic texts.

5.2.2 Research question 2: Types of words students do not understand

The findings from the vocabulary sheet (see Table 4.7) cannot be generalised as the sample only consisted of 27 participants (157 words reported in total). The small sample size also made it difficult to locate any trends in the types of words students

did not know. However, an analysis of the words from the vocabulary record sheet does give a snapshot of the types of words students do not understand.

Most words students recorded were low frequency, not on the NAWL nor NGSL lists. Further analysis revealed that many of them were highly technical or subject specific (e.g., *aetiology*, *epidemiology* and *proselytism*) suggesting students are encountering these types of words often. The literature review highlighted that the percentage of subject-specific words in academic texts varies from as little as 5% (Gillet, 2012) up to around 31% (Chung & Nation, 2003). Given the high number of technical words recorded as unknown, a figure of 31% seems more likely. Technical vocabulary as defined by Schmitt and Schmitt (2020, p. 8) is '… the jargon that is specific to particular domains (e.g., business, medicine, chemistry) and that represents concepts and ideas specific to those domains (*ledger, scalpel*, *catalyst*)'. From this definition, words such as *aetiology* and *epidemiology* are related to the field of medicine and related to the concept of disease and therefore can be considered technical. This subject specific vocabulary is crucial for students to understand key concepts, ideas and knowledge in their discipline (Schmitt, 2010; Schmitt & Schmitt 2020). Since most of the unknown words students recorded were technical, this could have repercussions for student learning as these words might be a potential barrier to accessing important subject knowledge.

A few of the words were slang (e.g., *bae*, *cotch*, *lecky*, *mandem*). The NGSL does not contain proper nouns, abbreviations, slang or 'other noise' so it is understandable why these words were not found on any of the lists (Browne, 2014, p. 4). It is also important to note that this task has highlighted a limitation of lexical frequency profiling in that it cannot identify fixed expressions as one unit of meaning – for example, *doing a big un* – so has separated the words out in the analysis.

The NAWL offers 6% coverage as discussed in Section 3.13.2 and with 10% of the total words students recorded being unknown on the NAWL, it could be perceived that the list offers a reasonable level of representation in an academic environment. Conversely, the NGSL has not offered the same level of coverage at 2.4% but given the majority of unknown words were derived from a higher education context this may be the reason (see Table 4.7).

For comparison, 11 words out of the 157 words were found on the AWL (Coxhead, 2000). Both the NAWL and the AWL offer similar coverage which could partially be

explained because academic words account for only around 10% of the total words in academic texts (Coxhead, 2000).

In the literature review (see Section 2.5.1), it was noted that academic vocabulary is not easy to learn due to its Graeco-Latin origins (Nation, 2013). Therefore, it was anticipated that students would report more academic words as unknown. Students might not have reported academic words as unknown because they are not encountering them in their daily academic life of which reading plays a significant part (Bharuthram, 2012). The literature review highlighted the fact that there are numerous challenges around getting students to read (Brost & Bradley, 2006; Hatteberg & Steffi, 2013; Sharma *et al*, 2013; Cressman, 2018). If students are not reading, they are unlikely to come across academic words in this way and therefore are equally unlikely to report them.

5.2.3 Research question 3: Correlation between vocabulary sizes and academic achievement

The third objective of Phase I was to explore whether there is a link between vocabulary size and academic achievement (measured by expected final degree classification). This research found there is no significant correlation between vocabulary sizes and academic achievement. This diverges from the literature highlighted in Chapter 2 which suggested there is much evidence to indicate a link between vocabulary knowledge/size and academic achievement (Smith *et al*, 1991; Treffers-Daller & Milton, 2013; Bleses *et al*, 2016; Schuth *et al*, 2017; Masrai & Milton, 2018). Moreover, Masrai and Milton (2018) found a strong positive correlation between overall vocabulary size (measured using XK-Lex), academic vocabulary (measured using AVST) and learners' academic performance as measured by their average grade.

However, most participants in Masrai and Milton's (2018) study were non-native speakers of English. The sample in Masrai and Milton's (2018) study consisted of 96 native Arabic speaking students taking an English course at a university in Saudi Arabia, 16 native English speakers all at Doctoral level and 120 non-native English speakers who were enrolled at three levels of study (Bachelor, Master and Doctoral) at British universities. To evaluate general vocabulary size, they used XK-Lex which takes a sample from across a 10,000-word range with ten tests from each 1,000-word band (see Masrai & Milton, 2012). Masrai and Milton (2018) used a different test and

a different set of students and context that could account for the differences in correlations between this study and theirs.

Although not an exact replication of Treffers-Daller and Milton's (2013) study, the context and process are somewhat similar as previously mentioned. However, Treffers-Daller and Milton (2013) found a correlation between vocabulary size and academic achievement (measured by mean scores from academic modules for Stages 1 and 2 and degree classifications for Stage 3 students) and this study did not. Treffers-Daller and Milton (2013) used a Pearson correlation (*r*) as they stated that both variables were normally distributed but as argued in the methodology chapter (see Section 3.7.2) Pearson's (*r*) can only be calculated if both variables are continuous (e.g. interval) (Salkind, 2014) while Spearman's (r_s) can be used when one variable is ordinal and the other interval/ratio (Bryman, 2008). This research used Spearman's (r_s) so the difference between the findings could be as a result of not using the same correlation coefficients to analyse the data.

Treffers-Daller and Milton (2013) reported correlation coefficients of 0.387 for Stage 1 (n = 113); 0.477 for Stage 2 (n = 30) and 0.315 for Stage 3 (n = 18). The number of observations (in this case the sample size) is directly related to the statistical significance of a correlation where a larger correlation is needed with fewer cases to reach statistical significance (Brezina, 2018a). Specifically, Brezina (2018a) states that when there are only ten observations a large correlation of 0.63 or greater is needed to reach statistical significance, indicating there is very little fluctuation in the data. Conversely, with 100 observations, the necessary critical value of Pearson's correlations drops to 0.2 (Brezina, 2018a). This is important for interpreting the data and the size of the correlation. Treffers-Daller and Milton (2013, p. 165) reported that 'all correlations are significant at the 0.05 level and suggest a modest connection between vocabulary size and academic performance'. Because the sample size was small and the correlations modest, particularly in Stage 3, their findings need to be interpreted with caution to determine whether there is enough evidence to generalise the correlation to the population.

The findings relating to Research question 3 suggest there are factors that may contribute to academic success other than general vocabulary and one's vocabulary size. As discussed in the introduction, academic literacy, as defined by Weideman and Van Dyk (2014, p. ii), is 'the ability to use language to meet the demands of tertiary education …'. To reiterate, some of the components of academic literacy

include understanding a range of *academic* vocabulary in context; interpreting the use of metaphor and idiom in *academic* usage; perceiving connotation, word play and ambiguity (Weideman, 2007 and Weideman & Van Dyk, 2014). This suggests measuring vocabulary sizes in the context of general English may not be the best method to ascertain links with vocabulary and academic achievement. In fact, Masrai and Milton (2018, p. 46) question '… whether it is general vocabulary size or specialist vocabulary knowledge that is the more important criterion for academic success'.

It might have been better to measure students' academic vocabulary size instead. Goulden *et al* (1990) state that academic vocabulary size is an important indicator of the ability of second language learners (in the context of schooling) to achieve academic success. As mentioned in Section 2.11, Masrai and Milton (2018) investigated whether there is a relationship between academic vocabulary size measured using the AVST and overall vocabulary size measured using the XK-Lex. They found that a correlation between the AVST and XK-Lex is high and is statistically significant. They suggest that the AVST performs as a general vocabulary size measure and is able to discriminate between learners with different levels of general lexical knowledge.

As discussed in the literature review (Section 2.7.3), the AWL (Coxhead, 2000) has major limitations. Perhaps a revision of the AVST using the NAWL (Browne *et al*, 2013) or AVL (Gardner & Davies, 2014) instead of the AWL is possible in order to provide a more updated test. A key question in research around native-speaker vocabulary still remains as Nation and Coxhead (2021, p. 130) ask 'Do vocabulary size measures predict educational performance?'.

Degree classifications might not be the best measure of educational performance or academic achievement. Universities UK (2004) argued that the current degree classification system is no longer fit for purpose because of grade inflation and that the UK needs an alternative. If higher grades are easier to obtain and degrees not cognitively challenging, then vocabulary knowledge and size would matter less. In addition, as pointed out in the methodology chapter, students had to self-report what their expected degree classification would be due to the timing of the test. This leaves room for error as students could under/overestimate their self-predicted grades and their final degree classification in reality could be different.

In Section 2.3 of the literature review, it was also argued that students do not read their assigned texts (Brost & Bradley, 2006; Hoeft, 2012; Hatteberg & Steffi, 2013;

Sharma *et al*, 2013). Perhaps there is no link between grades and vocabulary sizes because students do not read enough academic texts, which are considered challenging (Ernawati & Ardi, 2017). There has been very little research on the reading habits of students in HE (Nadelson *et al*, 2013). There could be a link between reading habits and academic achievement rather than vocabulary. Much incidental vocabulary learning comes from reading (Nagy *et al*, 1985; Brown *et al*, 2008) but students are more likely to read with a purpose and focus to meet assignment deadlines (Fairburn & Winch, 2011). Perhaps vocabulary sizes are small and little gains are made in terms of academic achievement if students are not reading.

5.2.4 Conclusions for Phase I

There are no standard measures of vocabulary size (Read, 2000; Treffers-Daller & Milton, 2013; Schmitt *et al*, 2020; Nation & Coxhead, 2021) and there are still questions around the methodology of vocabulary size research (Schmitt *et al*, 2020; Nation & Coxhead, 2021) despite its long history in educational research (Kirkpatrick, 1891). Any comparisons between research findings need to be interpreted with this in mind. Sample size, geographical location and age all have an impact on the results as well. However, taking these factors into consideration, the findings from this study suggest the average vocabulary size of a UK undergraduate to be around 11,000 words. In comparison to the findings from previous research (Zechmeister *et al*, 1995; Treffers-Daller & Milton, 2013; Brysbaert *et al*, 2016) the average vocabulary size estimate of around 11,000 seems reasonable.

It is also important to consider the finding that UG students experience changes in their vocabulary sizes between the first and second stage of study but plateau in their final year. Although outside the scope of this project, understanding this initial change in vocabulary sizes and subsequent stagnation could have potential implications for teaching and learning in the HE sector. This plateau might indicate poor reading skills as this can negatively affect vocabulary sizes (Nation & Coxhead, 2021). The importance of reading in HE was argued in Section 2.4 (Leamnson, 1999; Bharuthram, 2012; Colombo & Prior, 2016; Gunobgunob-Mirasol, 2019). Measures to help students improve their reading skills could be actioned in order to support students in continually improving their receptive vocabulary knowledge.

Finally, no link was found between vocabulary sizes and academic achievement despite a plethora of literature (Smith *et al*, 1991; Treffers-Daller & Milton, 2013; Bleses *et al*, 2016; Schuth *et al*, 2017; Masrai & Milton, 2018) suggesting otherwise.

Not finding a correlation between these does not mean it does not exist. Masrai and Milton (2018) used a different measure for academic success – the Grade Point Average or GPA – and gauging academic achievement in an alternative way may yield an altered result. A GPA is more precise than a 'First Class' which could range anywhere from a 70% to a 100%. Furthermore, a test measuring just academic vocabulary size using a more up to date academic word list as per the one devised by Masrai and Milton (2018) might be the answer.

While Phase I explored UG student's receptive knowledge of general vocabulary, Phase II of the research aimed to focus on academic vocabulary in particular. It measured students' productive vocabulary knowledge by estimating the level of vocabulary richness in written assignments; explored whether undergraduate students experience changes in their vocabulary use by measuring their academic word usage from their first to third year; and investigated whether there is a correlation between the percentage of academic words in students' written work and the marks received on their assignments.

5.3 Phase II

5.3.1 Research question 1: Vocabulary richness

This research question – What is the level of vocabulary richness in written assignments of monolingual and bilingual (using two or three languages including English routinely in the home) English speaking undergraduate students using lexical frequency profiling? – was met by measuring the level of vocabulary richness in written assignments of undergraduate students by using word frequency profiling software called AntWordProfiler (Anthony, 2021). As previously mentioned, vocabulary profilers are based on Laufer and Nation's Lexical Frequency Profiler (see Laufer & Nation,1995) and measure lexical richness by calculating the percentage of words in a text that fall into the most frequent words in the English language (1-1,000 band or K1), the second most frequent words (1,001-2,000 band or K2), academic words and less frequent words not found in any of those lists (Laufer & Nation, 1995).

As a reminder, the NGSL used on AntWordProfiler contains the first 1,000 most frequent flemmas (1K), the second 1,000 most frequent flemmas (2K) and the third 1,000 most frequent flemmas (3K). Findings showed (see Table 4.9) that the median percentage of word tokens from the NGSL 1 used in assignments for Stages 1, 2 and 3 was 81.50%, 80.60% and 79.20% respectively, while the median percentage of

word tokens from the NGSL 2 was 6.50%, 6.70% and 6.90% and from the NGSL 3, 2.45%, 2.35% and 3.30% for Stages 1, 2 and 3.

Traditionally, high-frequency vocabulary has been operationalised as around the first 2,000 most frequent word families (Schmitt & Schmitt, 2014) probably since around 80% of written English consists of only the first 2,000 most frequent words in English (Laufer & Nation, 1999). These findings indicate that most of the words that students use in their writing are high frequency as the majority of word tokens students use are from the NGSL 1 and NGSL 2 (words 0-1,000 and 1-2,000 or K1 and K2). If we extend our high frequency boundary as Schmitt and Schmitt (2014) argue to the 3,000-word level, then around 90% of all the word tokens students use in their assignments is to this level. This is expected given that higher frequency words are more likely to be needed for communication than lower frequency words (Webb & Nation, 2017).

Word lists are an invaluable resource for both learners and teachers (Webb & Nation, 2017). They help to reduce the learning load of vocabulary by often identifying the most frequent words as in the GSL (West, 1953) or specialised words such as the AWL (Coxhead, 2000). Word lists provide the most useful words to learn in a given context. However, with the abundance of word lists available, students and teachers need to know which lists will provide the greatest return in learning and studies that have addressed this (e.g., Browne, 2014; Brezina & Gablasova, 2015; Dang & Webb, 2016) used lexical coverage as the only criteria to ascertain which list is best (Dang *et al*, 2020).

The NGSL provides 86% coverage of the Academic Corpus of the Cambridge English Corpus (CEC) (Browne, 2016) which contains '400 million words of written and spoken academic language at undergraduate and post-graduate level from a range of US and UK institutions, including lectures, seminars, student presentations, journals, essays and textbooks' (Durkin, 2019). In Stage 1, the NGSL provided 90.45% coverage of students' written work and in Stage 2, 89.65% and 89.40% in Stage 3. This is slightly above the 86% that Browne (2016) found. The corpus used in this research consists of academic assignments of native English speaking UG students from Stages 1-3 which could serve as an appropriate academic writing norm for both L1 and L2 UG students. If lexical coverage of word lists is used as a measure of usefulness for L2 learners, then the NGSL has shown to be of value in this research.

As a reminder, Browne *et al* (2013) found that the NAWL has around a 6% coverage in the source corpus (Academic CEC, MICASE and the BASE) and in combination with the NGSL, both lists have a reported coverage of 92% of the academic corpus used to create the NAWL. A recent study using the NAWL for lexical frequency profiling of L2 university pre-sessional students revealed that the NAWL accounted for 5% of the words in a sample of essays (n = 118) that scored in the upper quartile (Higginbotham & Reid, 2019). In this research, the median percentage of word tokens used from the NAWL in Stage 1 was 2.00%, in Stage 2, 2.50% and in Stage 3, 2.30%. No other studies examining the NAWL in native speaker student writing were found in the literature, so this finding gives insight into how the list functions in an HE context, albeit on an Education Studies degree course in particular.

The percentage of word tokens used from the NAWL in student assignments is lower than the 5% and 6% coverage that both Higginbotham and Reid (2019) and Browne *et al* (2013) found, respectively. It is also well below the coverage of other lists such as the AWL and the AVL. Coxhead (2000) reported 10% of the total words (tokens) in academic texts are from the AWL while the AVL accounts for 14% (Gardner & Davies, 2014). This could suggest that the NAWL may not be as useful as the AVL for learners of English since the lexical coverage is low. This could be helpful for teachers and students when selecting word lists for vocabulary teaching and learning.

There are a couple of reasons which account for the differences in coverage. One of the main differences is the counting unit for each list is different. The AWL uses word families, the NAWL, flemmas and the AVL, lemmas. Because the NAWL consists of inflected forms rather than whole word families it has fewer word forms (2,604) compared to the AWL (3,110) (Browne *et al,* 2013; Coxhead, 2000), potentially providing less coverage. In order to make direct comparisons with the AWL, Gardner and Davies (2014) converted the lemma-based AVL into approximately 2,000-word families. This is much larger than the 570-word families of the AWL which could account for the highest coverage of all the lists (around 14%). The AVL also contains many high frequency words such as *group, use* and *level* (Webb & Nation, 2017) which arguably may be suitable for general purposes rather than formulating part of an academic core vocabulary at university level.

Kyle and Crossley (2016) contend that because academic language is found less frequently in general corpora, the number of academic words used in a text can be used to calculate lexical sophistication. Lexical sophistication (LS) is the ratio of

lexical words above a certain grade level or ‘advanced’ to the total number of words and is one measure of lexical richness (Engber, 1995; Šišková, 2012). Therefore, based on the fact that academic words are less frequent and therefore ‘advanced’, it can be argued that according to this principle, UG students in this research do not have a high level of lexical sophistication demonstrated in their written work since their median usage was below the 5% and 6% reported by Higginbotham and Reid (2019) and Browne *et al* (2013) using the same word list.

Laufer and Nation (1995) conducted a study using LFP analysing two pieces of writing (around 300 words) on 20 Israeli first year first semester UGs and 23 Israeli first year second semester university students who were non-native speakers of English. They found that 79.6% of the first semester students’ work consisted of words in the first 1,000 band and in the second semester this figured dropped to 75.5%. They also found that 8% of the words came from AWL in the first semester work and 9.1% in the second semester work. In this research, the percentage of word tokens used from the NGSL 1, or the first 1,000 band was 81.50%. While the lists used were different, the usage of K1 words in this research is higher even though English was not the first language for Laufer and Nation’s (1995) participants and was reportedly at a level of the Cambridge First Certificate (equivalent to IELTS 6.0). Students who pass The Cambridge First Certificate are considered independent rather than proficient users of English (UCLES, 2019). The data from this research would suggest that UG students rely on high frequency words in their English written work, particularly in Stage 1.

Caution is needed with any comparisons, though; not only is the word list different but the text length is as well. Word lists have also been designed to meet the needs of non-native speakers of English as an efficient way to develop vocabulary knowledge in the target language (Webb & Nation, 2017). Consequently, they may not reflect the vocabulary interests of native speakers. Moreover, Meara (2005) found that LFP is not as sensitive as Laufer and Nation (1995) have claimed and Smith (2005) attributes this partially to text length. Nation (2013) argues that text length can affect the results and in comparison studies, the content should ideally be similar. Laufer and Nation (1995) do not state what the subject of the essays were so it could be completely different to the ones in this study.

Like the AWL, the NAWL and AVL are not based on frequency in LFP, so it is difficult to know whether students are using a number of high or low frequency words

(Schmitt, 2010). Not knowing this information is a major disadvantage of using LFP as using a number of low frequency words rather than high frequency would indicate a larger vocabulary and greater lexical sophistication or richness (Kyle & Crossley, 2016). Some students could be using a large number of high frequency academic words; these individuals might have smaller vocabularies than LFP might suggest.

Having discussed K1 words and academic words, further data output from AntWordProfiler to consider is the percentage of 'off-list' word tokens found in students' work. 'Off-list' words do not appear on the other lists; more specifically, words not on either the NGSL, the supplement list or the NAWL. Off-list words include proper nouns, subject-specific words, general low frequency words (not on the NGSL), acronyms, numbers, and letters of the alphabet. In this research, the median percentage of 'off-list' word tokens for Stage 1 was 7.10 %, Stage 2, 7.30% and Stage 3, 9.20%. As stated in the methodology how proper nouns are dealt with makes a significant difference to an output profile (Cobb, 2010). If proper nouns are treated in the same way as nouns, it could lead to overestimation of the number of high frequency words, so putting them in the off-list category gives a more realistic account of the types of words students use in the K1 band.

Although different lists were used, further comparisons can be made with Laufer and Nation's (1995) study on 20 Israeli first year first semester UGs and 23 Israeli first year second semester university students who were non-native speakers of English. The first semester students' work comprised 6.1% and the second semester work 8.1% of 'off-list' words. The median percentage of 'off-list' words is similar to Stage 1 in this research. This suggests that using word lists up to the 2-3K frequency level in LFP will generate around 7% of the word tokens in this category for first year UGs. It is likely that if the NGSL contained words beyond the 3K level to include mid-frequency words (between the 3,000 and 9,000-word level) and low frequency (above 9,000) as per Schmitt and Schmitt's (2014) categorisation, a finer grained analysis could be achieved in LFP. This would most likely reduce the number of words in the 'off-list' category.

To conclude this section, it is important to acknowledge the limitations of using LFP. There is still some disagreement on the effectiveness of using LPF to measure productive vocabulary knowledge (Meara, 2005; Smith, 2005; Csomay & Prades, 2018; Higginbotham & Reid, 2019). There are limitations to measuring lexical richness that involve quantifying the degree to which a writer is using a varied and

large vocabulary (Laufer & Nation, 1995). This approach does not indicate whether the words are used appropriately or accurately; it counts word usage and cannot measure quality (Schmitt, 2010). In other words, students might be using several key academic words, for example, in their work, but doing so incorrectly and with frequency. Assessing the number of errors found in a text could provide a possible measure of writing quality (Read, 2020) but requires the researcher to analyse and code the data themselves and not solely rely on software applications (for example, see Faisal *et al,* 2017). Despite these issues, a widely accepted alternative to LFP does not exist (Higginbotham & Reid, 2019).

5.3.2 Research question 2: Word usage between stages of study

Research question 2 asks the question, 'What are the changes, if any, in word usage from the NGSL 1, NGSL 2, NGSL 3, NAWL and 'off-list' in student assignments between the stages of study?'.

While the proportion of word tokens from the NGSL 1 decreases from Stage 1 to 2, the Wilcoxon test indicated that the difference in the median was not statistically significant (see Figure 4.2). Conversely, students used more words from the NGSL 2 from Stages 1 to 3 (6.50%, 6.70%, 6.90%, respectively). There was a statistically significant effect of the stages of study on the words used from the NGSL 2 list in students' assignments between Stages 1 and 3 (see Figure 4.3) as well as the NGSL 3.

To summarise, students are using more words from the NGSL 2 and NGSL 3 as they progress through their degree. This could be down to two main factors. As the length of assignments increase (from approximately 3,000 words in Stage 1, 3,500 for Stage 2 and 4,000 for Stage 3), the opportunity for students to demonstrate their lexical knowledge increases. Moreover, research on language and age suggests that individuals will keep learning new vocabulary words throughout their lifetime and one's vocabulary grows throughout adulthood (Jones & Conrad, 1933; Gold *et al*, 1995; Schaie,1996; Bowles *et al*, 2005; Boaz *et al*, 2015). Brysbaert *et al* (2016) reported that between the ages of 20 and 60, the average person learns 6,000 extra base words or about one new base word every two days. They also found that the knowledge of words increases not only with age but education as well. From this, it might be possible that as a student progresses through their degree they will increase their vocabulary knowledge and use more 2K and 3K words the more education they receive.

In terms of academic word usage from the NAWL, there was a statistically significant effect of the stages of study on the words used from the NAWL in student assignments between Stages 1 and 2; students use more academic words in Stage 2 as compared to Stages 1 and 3. In other words, as students progressed from Stage 1 to Stage 2 there was an increase in the number of word tokens used in students' assignments from the NAWL. This suggests students may have developed a greater awareness of academic words and are using them in their writing in this period.

This is a similar trajectory that Durrant and Benchley (2019) observed in their study on children's use of vocabulary in writing. They found that as children progress through compulsory schooling, they made greater use of academic vocabulary. This could be true when students enter university and progress from Stage 1 to 2 in particular; they begin to use more academic vocabulary.

In all three stages' sub-corpora, *curriculum*, *classroom, found* and *impact* are in the top ten most frequently used word types from the NAWL. It is understandable that students studying a degree in education would use such words as *curriculum* and *classroom* in their work. However, word usage from the NAWL did change as revealed in Section 4.3.4. The finer grained analysis of the words used from the NAWL suggests that the topic of the assignment has the greatest influence on the words used from the list rather than the genre. For example, the word type *disability* from the NAWL can be found in the top ten most frequently used word types in Stages 2 and 3 but not in Stage 1. Students do not take modules in SEND in Stage 1 but do in Stages 2 and 3.

This finding is similar to the results of Durrant's (2016) study which examined the usefulness of the AVL in university student writing by exploring its coverage from texts taken from the BAWE. Durrant (2016) concluded that as students progressed through their degree there was a corresponding increase in the word usage from the AVL. However, Durrant (2016) also found significant variation across text types and disciplines and that the AVL may be more relevant to some student writers than others. The percentage of words used from the NAWL in this research combined with Durrant's findings (2016) suggest that a generic productive, core, academic vocabulary exists but it is potentially much smaller in scope than either the NAWL or AVL word lists.

For academic writing to be cohesive, students can include lexical cohesion which involves repeating the exact same words or phrases to link sentences and

paragraphs or using words from the same lexical set (Harmer, 2015; Valenzuela, 2020). Example 1 from student 107 in Stage 2 (see Section 4.3.4) demonstrates this:

Example 1

> A contemporary debate surrounding those in control of the ***curriculum*** has always been present in education. Gove's (2011) ***curriculum*** has seen a ***curriculum*** left in the hands of the government and policy makers, in comparison to a ***curriculum*** made up of both government and educational professions in the past.

The lexical set for the word *curriculum* is very small (e.g., subjects studied at school) which possibly leaves the writer with little choice but to repeat the word many times to maintain cohesion. It may have been possible in this example to replace the third *curriculum* with *it* (anaphoric referencing) but overall, it could be a challenge to substitute different words to replace *curriculum* in order to vary the language which would contribute to the lexical richness of the assignment. This could partially explain why the word *curriculum* was the most repeated word from the NAWL in all three stages. Similarly, it could also explain why the word *classroom* was the second most repeated word in all three stages (see Example 4 from student 122 in Stage 3 – Section 4.3.4) and why there is not an increase in word tokens and types from the NAWL in Stage 3.

Example 4

> Canter and Canter (1976) developed a behaviourist driven approach called 'Assertive Discipline' to deal with behaviour in schools. Assertive discipline is a structured approach which enforces a teacher-controlled ***classroom***. The package enforces pupils should be made fully aware of ***classroom*** rules and what is expected of them.

Additionally, students used more 'off-list' words in Stage 3 compared to Stages 1 and 2. There is a statistically significant difference effect of the stages of study on 'off-list' words in students' assignments between Stages 1 and 3 (see Figure 4.7). This could be due to the fact that there is a greater degree of specialisation across the stages of study with many subject-specific words not appearing on the NGSL or NAWL so are considered 'off-list'. For example, an analysis of an assignment from student 122 in Stage 3 revealed that the word *exclusion* was used 46 times but does not appear on any of the lists, including the NAWL, so is considered 'off-list' (see Example 3 from

student 122 in Stage 3, in Section 4.3.4 and below). The topic of the assignment was social exclusion, part of the sociology strand. Many of the 'off-list' words could be technical words or subject-specific related to particular topics.

Example 3

> There is debate about what social ***exclusion*** signifies and how it is best used in verbal and policy contexts (De Haan, 2001). Notwithstanding the level of discussion, the concept is commonly used and seems to be intensely attractive to the producers of social policy discourse (Blackmore and Warick-Booth, 2013). The idea of social ***exclusion*** endeavours to capture the complexity of helplessness in modern society rather than merely centring on one of its outcomes (Gorard et al, 2003; Levitas, 2005; Tomlinson, 2005).

Subject-specific words are those that are closely related and common to a topic or subject area but not necessarily elsewhere (Nation, 2013). As discussed in the literature review, a knowledge and understanding of this vocabulary is essential to access disciplinary or subject knowledge (Schmitt, 2010) as many of the key concepts are represented in this vocabulary (Schmitt & Schmitt, 2020). It is through language that students demonstrate their understanding of subject knowledge which is then evaluated (Schleppegrell, 2004). As per the marking criteria used in this research – 'For the highest marks (90-100%): an exceptional standard of work illustrating thorough and in-depth understanding, communicated with exceptional authority' (University of Derby, 2020) – students must demonstrate in-depth knowledge of the subject of which vocabulary plays a vital role.

Subject-specific words occur with low frequency over a range of texts which is why they are not included in the NGSL or NAWL; they are particular to specialist domains and differ according to subject area (Nation, 2001). However, they can appear frequently within a specific text or number of texts related to a particular discipline (Nation & Kyongho, 1995; Webb & Nation, 2017). The example above from student 122's assignment with the word *exclusion* is a case in point.

The median percentage of word tokens which are 'off-list' is 9.20% in Stage 3. Sutarsyah *et al* (1994) state subject-specific words make up around 10% of the words in academic texts but Gillet (2012) suggests a figure of about 5%. As previously mentioned, 'off-list' words include proper nouns, general low frequency words (not on the NGSL), acronyms, numbers, and letters of the alphabet in addition to subject-

specific words. Some examples of proper nouns found in the 'off-list' for stage 1 include *Britain, England, Piaget, Skinner, Slovenia, Watson*; in stage 2 *Gardner, Vygotsky, Woolfolk* and in stage 3 *Bandura, India* and *Smith.* The proper nouns are mostly surnames of authors cited in the assignments and geographical name places. Acronyms found on the 'off-list' for stages 1-3 include *ICT, SEN, OECD, PISA, GCSE* and OFSTED. Low frequency words (not on the NGSL 1, 2 or 3) for stages 1-3 include *esteem, optimistic, proximal and nomothetic*. Examples of subject specific words in the 'off-list' pertaining to education in stages 1-3 are *achievement, attainment, behaviourism, efficacy, humanistic, intersectionality* and *scaffolding*.

Although not within the scope of this research, a further systematic analysis of the 'off-list' word types would help to ascertain what proportion of the 9.2% consists of subject-specific vocabulary or general low frequency words. There were too many 'off-list' types to compile in a single list so for presentation purposes, a table of the 140 most frequent types for each stage can be found in appendix 9. While there are no obvious significant trends, it can be observed that from stages 1-3, the number of letter types (e.g., 'p') decrease and the proper nouns increase. There also fewer abbreviations from stages 1-3 and the appearance of more subject specialised words related to education seems to increase (e.g., *formative, efficacy, exclusion, intersectionality,* and *pedagogy*). However, key questions remain regarding the 'off-list' words in term of how specialised they are and whether they are commonly found outside particular topic areas/fields of study. For example, the word *learners* appears on the 'off-list' but *learn* and *learning* are in the NGSL 1 even though they are in the same word family. If word families are used as the unit of counting, the number of 'off-list' words could potentially decrease and words like *learners* may not be considered specialised.

As students progress through their degree, they are using fewer words from all the lists combined which could contribute to the increase in 'off-list' words, particularly in Stage 3 as discussed. This is evident in the data where there is a statistically significant effect of the stages of study on the text coverage all the lists provide for student assignments between Stages 1(Mdn = 93.2, SD = 2.30) and 3 (Mdn = 90.8, SD = 2.34), p<.001 (see Figure 4.8). The effect size is large r = 1.750, 95% CI [.800, 2.55]. The decrease in text coverage is partly attributed to a decrease in the number of word tokens used from the NGSL 1 and supplementary list.

There was a statistically significant effect of the stages of study on the words used from the supplementary list in student assignments between Stages 1 and 2 (see Figure 4.5) but also between Stages 1 and 3. From Stages 1 to 2, the percentage of words used in student assignments from the supplementary list decreased by 0.10% but remained the same from Stages 2 to 3. Since the effect size is small and the word list not particularly large in terms of content, this finding is not significant. The supplementary list includes days of the week, months and numbers (spelled out). The slight decrease in usage could be due to students using numerals instead of writing them in full, for example.

To summarise, as students progress through their degree, they use more words from the NGSL 2 and 3 as well as the NAWL in Stage 2. Additionally, students used more 'off-list' words in Stage 3 compared to Stages 1 and 2. These findings suggest that students do experience change in their vocabulary while at university. A greater specialisation throughout the course of the degree and the opportunity to demonstrate lexical knowledge may contribute to this change. The findings suggest that the topic of the assignment appears to have had the most influence on the words used from the NAWL, the NGSL and those that become 'off-list'. This suggests that the NAWL and the NGSL may be more useful in certain subjects than others. This is useful for those teaching English and creating resources based on these word lists.

5.3.3 Research question 3: Correlation between academic words and assignment scores

Finally, Phase II examined the relationship between the percentage of academic word tokens from the NAWL in students' written work and the marks on each assignment. No correlation between assignment marks and the number of word tokens from the NAWL for all three stages was found. Thus, a key question remains as to why there is no relationship between academic achievement and academic word usage despite other research findings suggesting there is one (Smith *et al*, 1991; Treffers-Daller & Milton, 2013; Bleses *et al*, 2016; Schuth *et al*, 2017; Csomay & Prades, 2018; Masrai & Milton, 2018).

The marking criteria used to derive the scores needs a closer inspection in order to partially answer this. For example, the undergraduate marking descriptors for a pass (40-49%) at the University of Derby (2020) state:

> A sound standard of work; a fair level of critical analysis and evaluation; little evidence of original thinking or originality; adequately researched; a sound standard of presentation; ideas fairly clear and coherent, some significant mis-understandings and errors; some weakness in style or presentation but satisfactory overall.

To present ideas clearly, cohesion and coherence are not only important aspects of this criteria but in academic writing at large as previously discussed. Important cohesive devices or linking words such as *thus, but, therefore* and *though* are not found on the NAWL but on the NGSL 1. Additionally, there is no overt reference to vocabulary in use in the marking criteria. As a result, assignments are not marked according to what kind of vocabulary students employ and at the time of marking the assignments, the lecturer does not focus on the occurrence of academic vocabulary the students used.

However, as previously mentioned, an exceptional standard of work which demonstrates thorough and in-depth understanding is necessary to be awarded the highest marks (90-100%) (University of Derby, 2020). Students must demonstrate extensive knowledge of their subject of which vocabulary plays a vital role. Arguably, subject-specific vocabulary plays a key role in demonstrating in-depth understanding of the topic students are writing about.

Indeed, a positive medium/small size correlation (r_2 = .39, p<.01) is seen in Stage 1 between marks and 'off-list words' used and in Stage 3, a small positive correlation can be seen (r_2 = .27, p<.01) between marks and 'off-list' words (see Figure 4.10). Further analysis of the 'off-list' words is necessary to understand why a relationship, albeit weak, exists. The 'off-list' words contain punctuation, proper nouns, abbreviations, low frequency and subject-specific words. It was beyond the scope of this research to analyse these, so it is unknown how many 'off-list' word tokens fall into each category. Conclusions cannot be made until the exact nature of these words is confirmed. Further research on identifying subject-specific words and the role they play in student achievement is needed.

Furthermore, no correlation exists possibly because of the variation of word usage from the NAWL within assignments but also between assignments. This could possibly render the usage of words from the NAWL as an unstable variable in determining links between academic achievement. Student 122 is a good example of this. The number of tokens and types from the NAWL varies from assignment to

assignment within the first stage of their study but also between subsequent stages (see Tables 4.13, 4.14 and 4.15). Again, the use of the word *curriculum* is a case in point. For example, in Stage 2, student 122 used *curriculum* 29 times in text 1 (topic is curriculum) but not in text 2 or 3 at all. Perhaps word usage from the NAWL has a stronger relationship with the topic of the assignment than the mark given for it.

In terms of the other lists, there is a medium negative correlation (r_2 = -.45, p<.01) between the usage of NGSL 1 words in student assignments and their marks in Stage 1 which can be seen in Figure 4.9 and a small negative correlation (r_2 = -.28, p<.01) in Stage 2. In other words, as students' marks on their assignments increase, the usage of words on the NGSL 1 decreases, particularly in Stage 1. Conversely, there is a small positive correlation between the usage of NGSL 2 words used in student assignments and marks in Stage 1 (r_2 = .27, p<.01) and in stage 2 (r_2 = .27, p<.01) but not in Stage 3. As students' marks on their assignments increase in Stages 1 and 2, so does the usage of words from the NGSL 2. In terms of the NGSL 3, no correlations were found in Stages 1 and 3 in relation to marks but there is a small positive correlation (r_2 = .25, p<.01) in Stage 2.

This suggests that students who are awarded higher marks use fewer K1 words (first 1,000) in their writing and more K2 words in Stages 1 and 2, but the correlations are small (except for the NGSL 1 and marks in Stage 1) indicating weak relationships between the NGSL word lists and marks. There could be other factors involved, topic being a main one as previously mentioned. Interest, motivation, and time management at the time of writing as well as personal circumstances could also play a role in the quality of writing students produce, affecting their scores.

The marking criteria used to assess the student assignments refers to aspects of content (e.g., critical analysis, evaluation, originality, well researched and in-depth understanding) and structure (e.g., a sound standard of presentation). This is not to say that academic word usage is unimportant in student writing. Indeed, in the introductory chapter, it was established that vocabulary plays a key role in academic literacy in meeting the demands of tertiary education (Weideman, 2007; Nagy & Townsend, 2012; Weideman & Van Dyk, 2014). However, the content and structure of assignments are likely to play a more significant role in scoring students' work in terms of the quality of their writing than academic word usage since no correlations were found. It is very difficult to determine how much vocabulary ability plays in the overall quality of student writing (Read, 2000). Therefore, the discussion continues

around the role of quantitative measures of vocabulary use in texts in determining their quality and whether this links to achievement (Csomay & Prades, 2018).

Correlations between the word lists demonstrate slightly stronger relationships as there is a medium to large, negative correlation (r_2 = -.55, p<.01) between the usage of words in student assignments from the NAWL and the NGSL 1 in Stage 1 and in Stage 2 the correlation is medium (r_2 = -.39, p<.01). Even though there is no correlation between the NAWL and marks on student assignments this does indicate that as students use more words from the NAWL there is a corresponding decrease in words used from the NGSL 1 in Stage 2. Furthermore, there is a positive, medium correlation (r_2 = .34, p<.01) between the words used on the NAWL and the NGSL 3 in Stage 1 only. These findings are important in terms of demonstrating the relationship between the NAWL and the NGSL in undergraduate student writing.

To conclude this section, controlling for text topic and text length when using LFP could give a greater insight into the role of academic word usage in academic achievement. Whether or not quantitative measures of academic word usage from word lists equates to quality writing is still very much open to debate. However, this study does contribute to an understanding of how academic words from the NAWL are used on an Education Studies degree in the UK by native speaking UG student writers.

5.3.4 Conclusions for Phase II

In light of the literature, the data findings have revealed that there may not be a proficient level of lexical richness in UG students' written work as measured by lexical sophistication. For all three stages, the median usage of academic word tokens from the NAWL was between 2% and 2.5%. Kyle and Crossley (2016) contend that because academic language is found less frequently in general corpora, the number of academic words used in a text can be used as a measure of lexical sophistication. Because the median percentage of academic words is much lower than the 5% to 6% average that Higginbotham and Reid (2019) and Browne *et al* (2013) found, this study suggests that the level of lexical richness found in UG students' work is lower than expected.

However, student writers are using more academic words from the NAWL in Stage 2 as compared to Stages 1 and 3. This also suggests there is change in the knowledge

of academic vocabulary from Stages 1 to 2. It was also found that as students progress through their degree, they use more words from the NGSL 2 and 3 in addition to the NAWL. Furthermore, students used more 'off-list' words in Stage 3 compared to Stages 1 and 2. These findings suggest that students do experience change in their vocabulary while at university. A greater specialisation throughout the course of the degree and the opportunity to demonstrate lexical knowledge may contribute to this change. The findings suggest that the topic of the assignment appears to have had the most influence on the words chosen from the NAWL, the NGSL and those that become 'off-list'.

No correlations between marks on students' written work and the usage of academic word tokens from the NAWL for all three stages was found. However, findings suggest there is a positive medium/small size correlation between marks and 'off-list words' used in Stage 1 and 3. In terms of the other lists, there is a medium negative correlation between the usage of NGSL 1 words used in student assignments and their marks in Stage 1 and a small negative correlation in Stage 2. In other words, as students' marks on their assignments increase, the usage of words on the NGSL 1 decrease, particularly in Stage 1. Conversely, there is a small positive correlation between the usage of NGSL 2 words in student assignments and marks in Stage 1 but not in Stage 3. As students' marks on their assignments increase in Stages 1 and 2, so does the usage of words from the NGSL 2. In terms of the NGSL 3, no correlations were found in Stages 1 and 3 in relation to marks but there is a small positive correlation in Stage 2.

In Figure 5.1 a funnel shape is used to demonstrate how the component parts of the research (Phase I and Phase II) merge together to form a final outcome for the students – higher marks. The findings from this research indicate that while bigger vocabulary sizes and word usage from the NAWL do not necessarily lead to higher marks, there is the possibility the NGSL 2, NGSL 3 and 'off-list' words may contribute to higher marks and therefore better outcomes. The findings suggest that students who are awarded higher marks use fewer words K1 words (first 1,000) in their writing and more K2 words in Stages 1 and 2, but the correlations are small (except for the NGSL 1 and marks in Stage 1) indicating weak relationships between the NGSL word lists and marks.

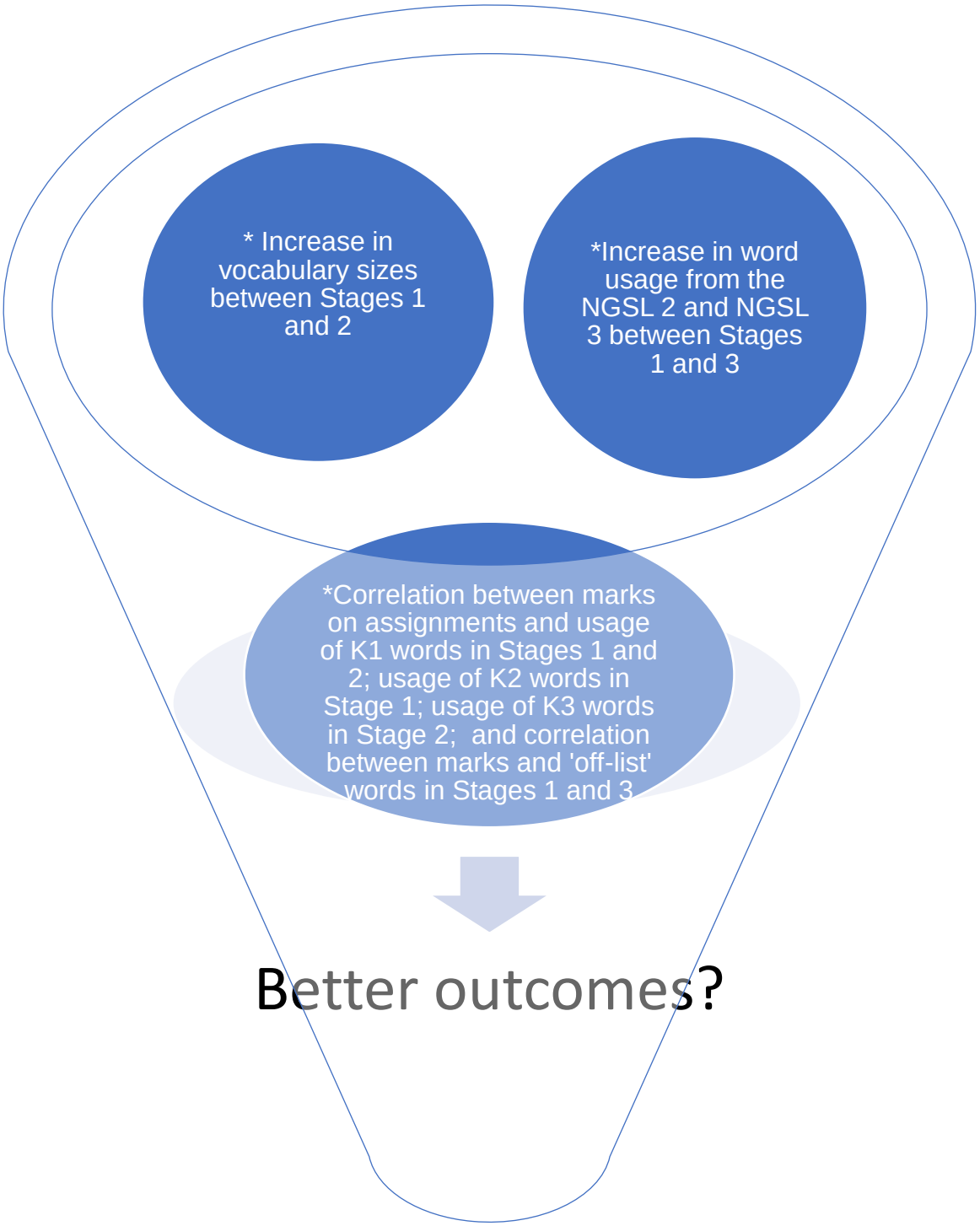


Figure 5.1 Summary of key findings from Phase I and Phase II

Chapter 6: Conclusion and Recommendations

Considering the research aim and objectives, this chapter will present a number of conclusions from the key findings that have emerged from the discussion of the data. How these conclusions contribute to the existing knowledge base is considered. Implications of the findings for practice will also be explored. Areas for further investigation are highlighted followed by an outline of the dissemination strategy. Finally, the chapter will conclude with offering a number of recommendations which arise from the research conclusions.

6.1 Key conclusions and original contribution to knowledge from Phase I and Phase II

Based on the findings from the study a number of conclusions can be drawn which contribute to new knowledge in the field. These conclusions are formulated around three key themes pertaining to: changes in UG students' receptive and productive vocabulary knowledge during UG study; vocabulary and academic achievement; and methods of measuring vocabulary knowledge. The implications for practice in HE are explored within the three themes.

6.1.1 Changes in vocabulary knowledge

The study contributes to the knowledge base around vocabulary in several ways. Firstly, the data has shown that we can be confident that a typical undergraduate may have a vocabulary size of around 11,000 words rather than in the region of 215,000 words as previous research (Hartmann, 1946) suggested. This finding can contribute to the growing consensus around the number of English words native speakers know (Nation & Coxhead, 2020). It can also provide a learning goal for learners of English wishing to attend HE institutions in the UK. However, minimum scores for each stage were 5,500, 5,000 and 4,000 meaning some native English-speaking students are at the minimum threshold for L2 learners (Laufer & Ravenhorst-Kalovski, 2010). This suggests some native speakers may need some language support in order to close this vocabulary gap.

Secondly, the research has captured a snapshot of changes in students' receptive and productive vocabulary knowledge in a higher education context. Emerging from

the data findings a picture can be formulated around these changes. Students' receptive knowledge of vocabulary appears to increase during UG study; there are statistically significant differences in the mean vocabulary size of students between Stage 1 ($\bar{x}$) =10,070 and Stage 2 ($\bar{x}$) = 11,614. In Stage 3, receptive vocabulary sizes did not change significantly.

The second year of UG study appears to be a key stage in terms of changes in vocabulary knowledge. Not only do vocabulary sizes increase but students are using more academic words in their writing from the NAWL in Stage 2 as compared to Stages 1 and 3. It was also found that as students progress from Stage 2 to 3, they use more words from the NGSL 2 and 3 (see Figure 6.1). Furthermore, students used more 'off-list' words in Stage 3 compared to Stages 1 and 2. Finally, students used fewer high frequency, K1 words, as they progressed through their degree.

There are several possible reasons for these changes. They may indicate that students are learning new academic and subject-specific words as they progress through their degree. An increase in assignment length from year to year gives student writers further opportunities to demonstrate lexical knowledge whether new or existing. Greater specialisation throughout the course of an UG degree could also require students to demonstrate subject-specific knowledge of which vocabulary plays a key role. Indeed, the number of 'off-list' words increases in students' work from Stage 1 to 3. However, further examination of the 'off-list' is needed to understand the nature of these words and whether they are indeed specific to a particular subject.

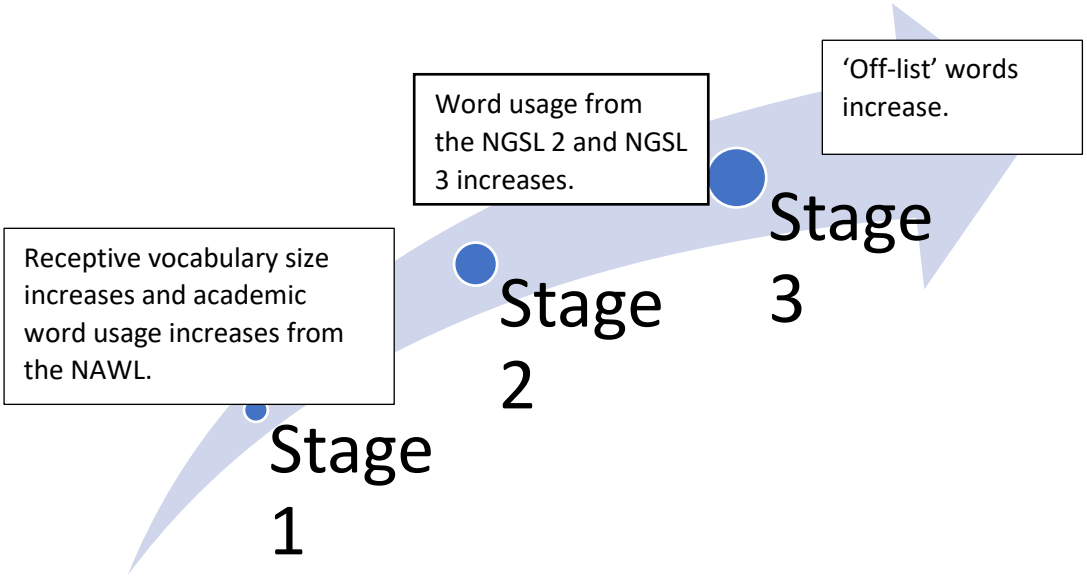


Figure 6.1 Changes in UG students' vocabulary as they progress through their degree

The changes in vocabulary highlighted in this study are very important for students. This research found that vocabulary sizes of UG students and usage of words from the NAWL in student writing plateau after Stage 2. It was argued in Chapter 1 that vocabulary is a significant component of effective communication and as a result enhances employability. It was also reasoned that vocabulary is vital for social mobility. In the words of the German philosopher Wittgenstein (2009, p. 73), 'The limits of my language mean the limits of my world'. By closing any potential vocabulary gaps for those students who begin university with lower vocabulary sizes and lexical sophistication, we can facilitate the expansion of their ability to communicate with confidence in the world outside of educational institutions (Quiqley, 2018). Practitioners can build on students' existing vocabulary knowledge by encouraging them to continually learn unfamiliar words to increase their receptive vocabulary sizes beyond Stage 2. In order to do this, HE practitioners should consider practice that overtly embeds vocabulary awareness and vocabulary instruction to lead to improved receptive and productive vocabulary knowledge. It is suggested that these strategies need to focus and include the three types of vocabulary found in the HE context: general, subject specific and academic (see Figure 6.2).

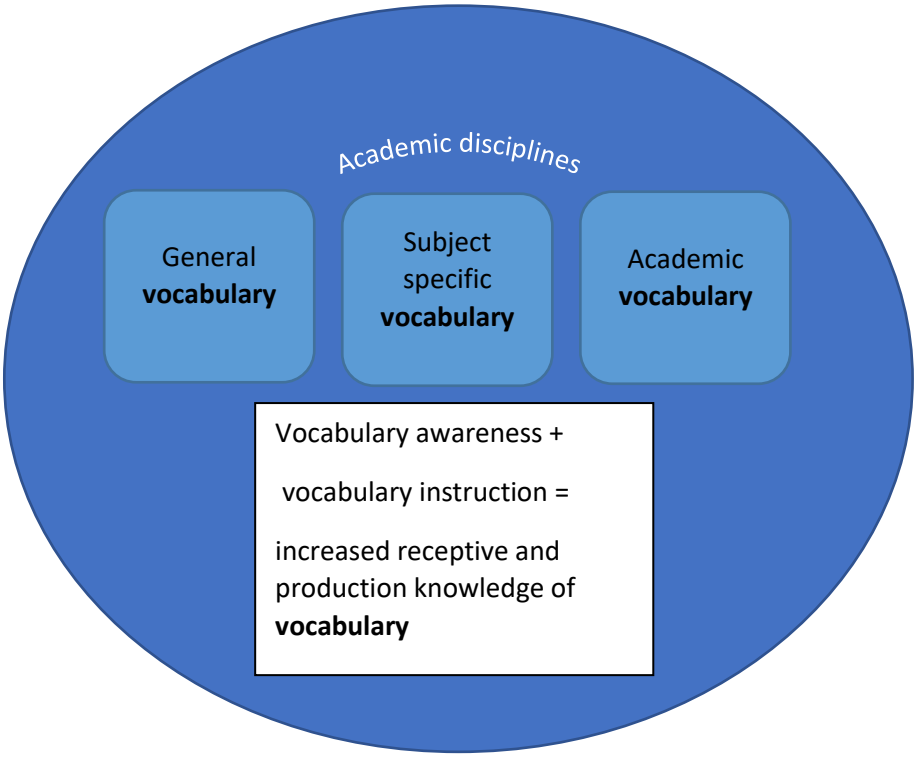


Figure 6.2 Embedding vocabulary within academic discipline contexts

6.1.2 Language acquisition

Before considering strategies, it is necessary to explore how students learn vocabulary. There is no generally accepted theory on how vocabulary is acquired (Tseng & Schmitt, 2008). However, much of the literature suggests that it is acquired through two main modes: implicit and explicit. Implicit learning is incidental in nature and 'is the process of learning something without the intention of doing so' (Brown *et al*, 2008, p. 136). In other words, learning vocabulary is a by-product of another activity such as reading comprehension, and it happens without conscious effort. On the other hand, explicit learning, also known as intentional or deliberate, is defined as any learning activity geared at committing lexical information to memory (Seel, 2012).

Nagy *et al* (1985) argued that through reading, children incidentally learn a large amount of vocabulary from context without much help from teachers. Incidental vocabulary learning is linked to extensive reading (Brown *et al,* 2008). Extensive reading (or listening), according to Harmer (2015) involves reading at length for pleasure. This poses a number of challenges for university undergraduates. Do students read academic texts for the purpose of pleasure? This is possible for a few, but the research on reading compliance previously discussed suggests otherwise and reading academic texts for pleasure is highly unlikely for the majority of undergraduates. What is more likely, though, is that students read with a purpose and with focus in order to meet assignment deadlines (see Fairburn & Winch, 2011). On the other hand, if this is true, how can reading at university be a pleasurable experience? Research on vocabulary acquisition through extensive reading found that large scale vocabulary growth is more likely to occur from expository reading materials (Gardner, 2004). Expository texts or non-fiction are the main kind of literature found at university unless English literature is being studied. A key question remains as to whether students at university acquire new words by reading textbooks.

Learning words implicitly through extensive reading is a slow process (Sökmen, 1997). Sökmen (1997, p. 152) also argues that 'guessing from context does not necessarily result in long-term retention'. Another way of acquiring new vocabulary is to make a deliberate effort to do so. This can often mean learning vocabulary out of context through word lists or exercises but can also occur in context when looking up unfamiliar words from either written texts or spoken discourse (Chinese University

Hong Kong, 2012). In other words, there is a conscious attempt on the part of the learner to learn an unfamiliar word.

Students can encounter new words listening to a lecture, speaking to other students, or through reading. Nation (2001 cited in Lightbown & Spada, 2006) suggests that in order for new words to become committed firmly to memory, learners need to have a number of meaningful encounters with a new word. Exactly how many encounters is meaningful remains less clear (Lightbown & Spada, 2006). In terms of subject-specific vocabulary, how many times a student would need to encounter an unfamiliar word before it is learnt remains unknown. And more importantly, what it means to have a 'meaningful' encounter is equally indeterminate.

A major criticism of explicit or deliberate vocabulary learning is that there are thousands of words in the English language so learners would do better to concentrate on their reading as long term vocabulary growth is greater from incidental learning (Nation, 2013). Nagy *et al* (1985) found that the clear majority of words are learnt implicitly and much research (see Hunt & Beglar, 1998; Shokouhi *et al,* 2009; Alemi & Tayebi, 2011) has been done to support this view. While this may be true, others such as Schmitt (2008) argue that learners will need both in order to learn the large amount of vocabulary needed to master the English language. To support this, in a longitudinal case study of a Chinese MA student, Li and Schmitt (2009) found that she learned 166 new lexical phrases during her studies through both explicit and implicit means, particularly from her academic reading. Furthermore, Ellis (1995) identified four main points on an explicit–implicit continuum whereby words are learnt unconsciously at the one end and at the other end words are learnt explicitly by adopting metacognitive strategies.

A key distinction between implicit and explicit learning has to do with whether the student is conscious or aware that learning is taking place. Curtis (2006) stresses the importance of promoting word consciousness in the role of vocabulary instruction in adult basic education. When students have an awareness of and an interest in words, they make greater gains in growing their vocabulary (Anderson & Nagy, 1996 cited in Curtis, 2006). This suggests that metacognitive strategies are important in vocabulary acquisition. There is much on the effectiveness of these strategies in the literature (see Hedge, 2000; Rasekh & Ranjbary, 2003; Zhao, 2009; Rahimia & Katala, 2012; Amirian *et al*, 2015; Diaz, 2015; Trujillo Becerra *et al*, 2015) which

suggests that educators can and should facilitate students' awareness of their own language learning and needs.

Metacognitive strategies deal with learning to learn or thinking about one's learning and are sometimes referred to as self-regulation. Metacognition is a process where students plan, monitor and evaluate their own learning (EEF, 2018). Metacognitive strategies for language learners include interacting with native speakers as much as possible to maximise their exposure to new language, testing oneself and reviewing new material (Schmitt, 1997). Schmitt (1997) also found the most useful strategies specific to learning vocabulary were using the dictionary, written and verbal repetition, saying an unfamiliar word aloud, studying a word's spelling, and taking notes in class. Some studies (e.g., Diaz, 2015; Trujillo Becerra *et al*, 2015) that have looked at the practice of keeping vocabulary notebooks and journals found them to be very effective strategies for learning new words as well.

Given the complexities of learning and despite the plethora of research in the area of vocabulary acquisition, there is still very little in the way of knowing the best way to achieve it (Schmitt, 2008). However, raising awareness of the importance of vocabulary could help facilitate and may even accelerate the vocabulary growth of our students. Programme teams and individual lecturers/tutors can help to make vocabulary learning explicit. Students make greater progress in their vocabulary learning when they have an awareness and interest in words (Anderson & Nagy, 1996 cited in Curtis, 2006). We need to help raise awareness of key vocabulary as well as motivate students to become interested in learning new words by making vocabulary accessible. Some of the most useful vocabulary learning strategies include written and verbal repetition, saying an unfamiliar word aloud, studying a new word, and keeping a vocabulary notebook (Schmitt, 1997; Diaz, 2015; Trujillo Becerra *et al,* 2015). Highlighting words in sessions, facilitating class discussion on new terminology, using glossaries, and encouraging students to keep a vocabulary notebook will help to raise vocabulary awareness, make it accessible and hopefully motivate students to have an interest in it.

Explicit vocabulary instruction in HE may make journal articles more accessible as well. Proctor (2015) found that one of the main reasons teachers do not engage in research-related activities is due to the inaccessibility of the language used in research articles found in academic journals. Butcher *et al*'s (2017) research supports this notion as well. Students in their research reported they were unfamiliar with the

'type' of language experienced in HE. It was also argued that students do not read their assigned texts (see Section 2.4) and one explanation might be that they find the language hard to understand just as the teachers in Proctor's (2015) research did. Explicit vocabulary instruction could potentially help students find journal articles more readable which in turn may further increase their vocabulary (see Figure 6.3).

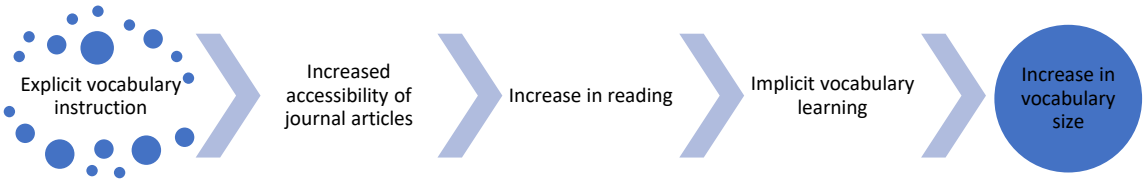


Figure 6.3 Potential implications of explicit vocabulary instruction

Explicit vocabulary instruction can also help students learn academic and subject-specific words in the initial stages of their study. Using word lists such as the NAWL (Browne *et al*, 2013) and subject-specific glossaries with students at the beginning of courses could help students identify possible gaps in their vocabulary. Teaching staff in HE could also conduct an audit or evaluation around vocabulary in their practice. For example,

1. What vocabulary knowledge do my students already have?
2. What vocabulary knowledge do my students need to know?
3. Is there a word list or glossary of key terminology for my subject area/discipline?
4. What examples or evidence can I give where vocabulary is being taught/learnt/discussed or noted?

To summarise, through raising awareness and explicit vocabulary instruction in the context of academic disciplines, HE practitioners can help facilitate an increase in students' general, subject-specific and academic vocabulary knowledge. This would be particularly beneficial for those native-speaking students whose vocabulary sizes were found to be at the minimum threshold for L2 learners (Laufer & Ravenhorst-Kalovski, 2010). This may aid in reducing the vocabulary gap for them by increasing their communication skills, helping to secure employment, and thus reducing economic inequality. As argued in Chapter 1, closing the vocabulary gap may be the

key to improving social mobility (Quigley, 2018). Explicit vocabulary instruction in HE could play a key role for these students.

6.1.3 Vocabulary and academic achievement

The data revealed that having a 'small' vocabulary size does not seem to hinder achievement in terms of gaining higher marks. The data findings suggest there is no link between vocabulary size or the usage of academic words in students' written work and academic achievement as measured by assignment marks despite a large body of research stating otherwise (Smith *et al*, 1991; Treffers-Daller & Milton, 2013; Bleses *et al*, 2016; Schuth *et al*, 2017; Masrai & Milton, 2018). There is no statistically significant difference between the mean vocabulary size of students predicted a first [($\bar{x}$)= 11,521] and those who were predicted a 2:1 [($\bar{x}$)= 11,312]; a 2:2 [($\bar{x}$)= 11,450] and a 3rd [($\bar{x}$)= 9,833]. It was also found that there is no correlation between marks and the percentage of academic word tokens used from the NAWL in assignments. Both of these findings suggest there are other factors that determine degree success.

There are also uncertainties as to whether quantitative measures of academic word usage from word lists equates to quality writing and academic achievement. Added to this is a key question around how useful the NAWL is for students studying education, given the list has shown a coverage of around 2.3%. The median percentage of word tokens used in each stage, around 2.3%, is much lower than the 5% to 6% average that Higginbotham and Reid (2019) and Browne *et al* (2013) found. As previously stated, this could indicate that a generic, productive, academic vocabulary exists but is probably much smaller in scope than originally thought.

The lower coverage may partly be explained by the subject area in which this research is focused, namely education. The finer grained analysis indicated that word usage from the NAWL was greatly influenced by the topic of the assignment rather than the academic genre at large. This has important implications for L2 learners using the NAWL to learn academic vocabulary in HE. There are many word lists available so both teachers and students need to know which lists will give them the greatest return on their learning (Dang *et al*, 2020). The lexical coverage of a word list is an effective way to identify whether word lists are useful, particularly for L2 learners (Dang *et al*, 2020). The findings from this research suggest the NAWL may be more useful in certain subjects as the coverage could potentially vary accordingly. More research on how the NAWL functions in different disciplines would be helpful

for those teaching English for Academic Purposes either in pre-sessional or in-sessional contexts.

There is no correlation between academic word usage and marks on students' assignments; however, correlations were found with the general English word lists. A medium, negative correlation was found between marks on assignments and K1 words in Stage 1 and a small correlation in Stage 2. There is also a small positive correlation between the usage of K2 words used in student assignments and marks in Stage 1 but not in Stage 3. In terms of the NGSL 3, no correlations were found in Stages 1 and 3 in relation to marks but there is a small positive correlation in Stage 2.

This indicates that students achieving higher marks use fewer high frequency and more lower frequency words in their assignments. Although these correlations are not strong enough predictors, they still could have some consequence for student achievement; it implies that students need to write with some degree of lexical sophistication to achieve higher marks. It could also imply that examining the quantity of academic word usage specifically may not be a good measure of lexical sophistication in relation to achievement. This finding also provides further evidence that vocabulary awareness and instruction of more 'advanced' vocabulary needs to be embedded in courses, particularly in the initial stages of study.

6.1.4 Methods of measuring vocabulary size/knowledge

There are still questions around the methodology of vocabulary size research and the validity of testing (Schmitt *et al*, 2020; Nation & Coxhead, 2021) despite its long history in educational research (Kirkpatrick, 1891). While there are a number of receptive vocabulary tests widely available (see Section 2.9), there is much improvement that can be made in terms of format and content (Nation & Coxhead, 2021) and Goulden *et al*'s (1990) test is no exception.

Word lists play a key role in measuring receptive vocabulary sizes (Nation & Coxhead, 2020) and evaluating productive vocabulary knowledge (e.g., Morris & Cobb, 2004; Csomay & Prades, 2018; Higginbotham & Reid, 2019). Additionally, measures of productive vocabulary knowledge and the quality of student writing is partly determined by the measures used (Olinghouse & Wilson, 2013). The word list and marking criteria were key determinants of both in this research. Word lists are designed for learners of English so whether they are useful in measuring the productive output of native speakers remains to be seen. As stated in Section 6.1.2,

the amount of vocabulary used from these word lists using LFP might not serve as a good measure of productive knowledge. Word lists should also be designed to reflect UG native speakers of English writers and their needs in terms of subject content as well as interest in particular topics. Nonetheless, this research contributes to new knowledge in the field by demonstrating how word lists such as the NAWL and NGSL operate in UG student writing in an Education subject degree.

A key finding from this research is the topic of a written piece has a profound influence on the words used from the NAWL. The diversity of topic strongly affects the nature of the output making comparisons using LFP slightly challenging. Controlling for text topic and text length when using LFP could give a greater insight into the role of academic word usage in academic achievement but in naturalistic studies it is very difficult to control for this (Nation & Coxhead, 2021). Nonetheless, despite this, LFP is useful for gauging the types of vocabulary found in a text.

The literature review highlighted that the vocabulary profiling software used in Phase II is typically used in the field of second language acquisition and studies using LFP mainly focus on ESOL students rather than native speakers (e.g., Meara & Fitzpatrick, 2000; Muncie, 2002; Morris & Cobb, 2004; Kojima & Yamashita, 2014; Lutviana *et al,* 2015; Csomay & Prades, 2018; Higginbotham & Reid, 2019). To date, there are no other known studies that have used LFP to analyse the work of UG students in the UK. Laufer and Nation's (1995) study used LFP, but the participants were all foreign learners of English and the written work consisted only of 300 words each. Similarly, Morris and Cobb's (2004) study used LFP on 122 TESL students in a Canadian university by means of an analysis of 300-word samples of their writing. The writing samples consisted of their entrance exam opinion essays which falls into a different genre according to Nesi and Gardner's (2012) categorisation. The word count is very short in comparison to the assignments analysed in this research so students do not necessarily have many opportunities to demonstrate their linguistic knowledge. In other words, this study partly contributes to original knowledge in the field by examining academic written work in the critique and essay genre (over 1,000 words for each piece) of undergraduate native speakers of English.

On the Lextutor website, which features a number of vocabulary profilers, there are also many other word activities (e.g., word games, vocabulary tests) suitable for UG students to help build their vocabulary. Students might need guidance as many of the activities are aimed at non-native speakers so some assistance as to which ones are

most suitable may be necessary. Finally, students could also use vocabulary profilers to create their own word lists and glossaries particularly focused on subject-specific vocabulary. Using this software may have the added benefit of increasing students' digital skill set at the same time as learning unfamiliar words.

6.1.5 Summary of key original contributions to knowledge

- A typical undergraduate may have a vocabulary size of around 11,000 words.
- Vocabulary sizes and word usage from the NAWL in student writing do not change from Stage 2 to 3 which may require additional learning and teaching strategies to be adopted, particularly for those students who start university with low sizes as stated above.
- The topic of student assignments determine which words are used from the NAWL. This may partially explain its low text coverage suggesting the NAWL may be more useful in certain subject areas and not as general in scope.
- Factors other than vocabulary size and academic word usage may play a greater role in student achievement.

6.2 Further research

Borne out of the data findings, discussion and conclusion a number of areas for future research have been identified. This research found that vocabulary size and academic word usage do not correlate to academic achievement but that does not mean a relationship does not exist. It was concluded that there are still questions around testing methodology necessitating calls for improvements in this area. With improved vocabulary size testing, the question of whether vocabulary knowledge is a good predictor of academic achievement needs to be revisited (Nation & Coxhead, 2020).

Phase I explored students' receptive vocabulary sizes while Phase II explored their productive knowledge. By looking at both, it does give a snapshot view of students' vocabulary knowledge, but a future project could use a newly improved, updated test to measure receptive knowledge in combination with a productive measure using the same participants over the stages of study. For example, the NAWL could replace the AWL in an adapted version of the Academic Vocabulary Size Test (AVST) (Masrai & Milton, 2018) to provide a more up to date test. This could capture vocabulary size growth in individuals as well as identifying any needs. Further research is also needed

to see how the NAWL performs in other subject areas to determine its usefulness to students in other disciplines.

Vocabulary testing aside, the literature review highlighted other factors which may affect academic achievement and obtaining higher marks resulting in a 'good' honours degree. The research findings suggest the content of the assignment as per the marking criteria is more likely to influence the score given to a piece of work. According to the criteria, the content must be adequately researched which involves wider reading. There has been very little research on the reading habits of students in HE (Nadelson *et al*, 2013). There could be a link between reading habits and academic achievement rather than vocabulary. Reading is one of the most important academic tasks encountered by students in higher education (Bharuthram, 2012) so further investigation into the role it plays in achievement would have a significant impact on understanding the needs of our students.

Additionally, a greater understanding of 'off-list' words is required in order to ascertain the extent to which these words play a role in academic achievement. A further, detailed analysis of the 'off-list' word types would help to clarify which words are potentially subject-specific. As previously stated, key questions remain regarding the 'off-list' words in term of how specialised they are and whether they are commonly found outside particular topic areas/fields of study. As pointed out in the literature review, Chung and Nation (2003) found in their research that almost one out of every three words in an anatomy text (31.2%) was a technical word. They also reported that 20.6% of the running words in an Applied Linguistics text were classified as technical. This indicates there is potentially a large number of subject-specific words within each discipline. Lastly, a positive medium/small size correlation was found in Stages 1 and 3 between marks and 'off-list words' which indicates more research needs to be conducted to understand the relationship between this and marks awarded on assignments. There is a possibility that these words are important for determining the quality and thus mark of an assignment.

Finally, the creation of a corpus consisting of UG student academic writing is a valuable resource for further language related research. The majority of vocabulary studies have focused on single words until recently (Webb, 2020) and research on formulaic language such as collocations, idioms, set phrases, phrasal expressions, phrasal verbs, lexical bundles, fixed and variable expressions is much needed, particularly in an academic context.

6.3 Recommendations

In light of the literature, findings, discussion and conclusions, a number of recommendations with practitioners in mind has been devised around embedding vocabulary learning within subject areas/disciplines and areas for further research.

1. The findings suggest that some UG students may need some English language support while at university. Practitioners can facilitate this learning in a number of ways and there is much literature on the effectiveness of metacognitive strategies (outlined in Section 6.1.2) in particular (see Hedge, 2000; Rasekh & Ranjbary, 2003; Zhao, 2009; Rahimia & Katala, 2012; Amirian *et al,* 2015; Diaz, 2015; Trujillo Becerra *et al*, 2015). Academics can and should help make vocabulary knowledge accessible by facilitating students' awareness of their own language learning needs using these strategies, particularly in Stages 1 and 2.

2. Consideration also needs to be given to the key words and terminology in a subject area. Are practitioners always aware of this language; how they are using it and how they are enabling students to access it? The way in which lecturers deal with unknown words in sessions could potentially have a significant impact on their students' learning.

3. Further research is needed on the reading habits of UG students as there could be a link between reading habits and academic achievement rather than vocabulary. Reading is one of the most important academic tasks encountered by students in higher education (Bharuthram, 2012). This conclusion has led the researcher to question what they really know about the reading habits of their own students. There are too many assumptions on whether students have done the pre-reading before sessions based on their ability to contribute to group discussions. A key assumption is that when students comprehend the assigned texts, they can then apply what they have read in sessions and assignments. The ability to demonstrate wider reading is a key feature of UG study as reflected in marking criteria and linked to higher marks. Future

research is necessary to determine the nature of this relationship between UG students' reading habits, the quality of their work and gaining higher marks.

4. The findings from this research can serve as a benchmark for learners of English in a HE context. They also give some indication on the usefulness of the NAWL and NGSL in terms of student writing, although further research using these lists is needed in other subject areas.

5. Further to point 4, the research may be of interest to those teaching English for Academic Purposes and providing academic support to students, both native and non-native speakers. This research may be of value to academics/researchers interested in the role of language in higher education or education generally; for example, the Educational Role of Languages (ERL) which is a world-wide network of academics who research on issues at the intersection of pedagogy and language (University of Gdansk, 2019).

References

Alemi, M. & Tayebi, A. (2011) The influence of incidental and intentional vocabulary acquisition and vocabulary strategy use on learning L2 vocabularies, *Journal of Language Teaching and Research*, 2:1, 81-98

Alim, H.S. & Paris, G. (2015) Whose Language Gap? Critical and Culturally Sustaining Pedagogies as Necessary Challenges to Racializing Hegemony, in *Journal of Linguistic Anthropology*, 25:1, 66–86

Alqahtani, M. (2015) The importance of vocabulary in language learning and how to be taught, *International Journal of Teaching and Education*, 3:3, 21-34

Amirian, S., Mallahi, O. & Zaghi, D. (2015) The relationship between Iranian EFL learners' self-regulatory vocabulary strategy use and their vocabulary size, *Iranian Journal of Language Teaching Research*, 3:2, 29-46

Anderson, R. & Freebody, P. (1981) Vocabulary knowledge, in Guthrie, G.T. (Ed.), *Comprehension and Teaching: Research Reviews*, Newark, DE: International Reading Association, pp. 77-117

Anderson, R. & Nagy, W. (1993) *The Vocabulary Conundrum*, Technical Report No 570 from the Center for the Study of Reading, Urbana, IL: University of Illinois.

Angoff, W. (1953) Test reliability and effective test length, *Psychometrika*, 18, 1-14

Anthony, L. (2013) A critical look at software tools in corpus linguistics, *Linguistic Research*, 30:2, 141-161

Anthony, L. (2021) AntWordProfiler (Version 1.5.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from: www.laurenceanthony.net/software (accessed 3 November 2020)

Anthony, L. and Nation, I.S.P. (2017). Picture Vocabulary Size Test (Version 1.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from: www.laurenceanthony.net/software/pvst (accessed 25 May 2021)

Augar Report (2019) *Review of Post-18 Education and Funding*. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/805127/Review_of_post_18_education_and_funding.pdf (accessed 12 July 2019)

Bahr, S. Lebby, S. & Wilkinson, L. (2020) Spelling error analysis of written summaries in an academic register by students with specific learning disabilities: phonological, orthographic, and morphological influences, *Reading and Writing*, 33, 121-142

Barclay, S. & Schmitt, N. (2019) Current perspectives on vocabulary teaching and learning, in Gao, X., *Second Handbook of Second Language Teaching and Learning*, Switzerland: Springer

Barnett, V. & Lewis, T. (1994) *Outliers in Statistical Data*, 3rd edn, Chichester: Wiley & Sons

BERA (2011) *Ethical Guidelines for Educational Research*. Available from: www.bera.ac.uk/resources/using-numeric-secondary-data-education-research (accessed 11 November 2012)

BERA (2018) *Ethical Guidelines for Educational Research*, 4th edn. Available from: www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2018 (accessed 20 February 2019)

Berberich, K. & Keiber, K. (2020) *Tools for Corpus Linguistics*. Available from: https://corpus-analysis.com/ (Accessed 23 December 2020)

Bereby-Meyer, Y., Meyer, J. & Flascher, O. (2002) Prospect theory analysis of guessing in multiple choice tests, *Journal of Behavioral Decision Making*,15, 313-327

Berko, J. (1958) The child's learning of English morphology, *Word*, 14, 150-177

Bharuthram, S. (2012) Making a case for the teaching of reading across the curriculum in higher education, *South African Journal of Education*, 32, 205-214

Biemiller, A. (2003) Vocabulary: needed if more children are to read well, *Reading Psychology*, 24, 323-335

Biemiller, A. (2012a) The influence of vocabulary on reading acquisition, *Encyclopedia of Language and Literacy Development*, London, ON: Western University. Available from: www.literacyencyclopedia.ca/pdfs/topic.php?topId=19 (accessed 26 December 2020)

Biemiller, A. (2012b) Teaching vocabulary in the primary grades: vocabulary instruction needed, in Baumann, J. & Kame'enui, E. (Eds), *Reading Vocabulary: Research to Practice*, 2nd edn, New York, NY: Guilford Press

Black, T. (1999) *Doing Quantitative Research in the Social Sciences: An Integrated Approach to Research Design, Measurement and Statistics*, London: Sage

Blake, M. & Dearborn, W. (1935) The improvement of reading habits, *Journal of Higher Education*, 6:2, 83-88

Bleses, D., Makransky, G., Dale, P. & Hojen, A. (2016) Early productive vocabulary predicts academic achievement 10 years later, *Applied Psycholinguistics*, 37:6, 1461-1476

Blum, S. & Riley, K. (2014) Selling the Language Gap, in *Anthropology News*, Available from http://linguisticanthropology.org/wp-content/uploads/2015/02/BlumRiley2014_SellingTheLanguageGap.pdf (accessed 10 June 2022)

Boaz, M., Erel, H., Goy, H. & Schneider, B. (2015) 'Older is always better': age-related differences in vocabulary scores across 16 years, *Psychology and Aging*, 30:4, 856-862

Bodner, C. & Clark, R. (2017) Can game-based learning enhance engineering communication skills?, *IEEE Transactions on Professional Communication*, 6:1, 24-41

Bogaards, P. & Laufer, B. (2004) *Vocabulary in a Second Language*, Amsterdam/Philadelphia: John Benjamin Publishing

Bowey, J. A. (1986) Syntactic awareness in relation to reading skill and ongoing reading comprehension monitoring, *Journal of Experimental Child Psychology*, 41:2, 282-299

Bowles, R., Grimm, K. & McArdle, J. (2005) A structural factor analysis of vocabulary knowledge and relations to age, *Journal of Gerontology: Psychological Sciences*, 60B:5, 234-241

Brezina, V. (2017) *LancsLex: Lancaster Vocab Analysis Tool*. Available from: http://corpora.lancs.ac.uk/vocab/index.php (accessed 28 January 2021)

Brezina, V. (2018a) *Statistics in Corpus Linguistics*, Cambridge: CUP

Brezina, V. (2018b) Collocation graphs and networks: selected applications, in Cantos-Gómez P. & Almela-Sánchez M. (eds), *Lexical Collocation Analysis. Quantitative Methods in the Humanities and Social Sciences*, Cham: Springer

Brezina, V. & Gablasova, D. (2015) Is there a core general vocabulary? Introducing the New General Service List, *Applied Linguistics*, 36:1, 1-22

Brezina, V. & Timperley, M. (2017) How large is the BNC? A proposal for standardized tokenizations and word counting. CL2017 Birmingham. Available from: www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2017/general/paper303.pdf (accessed 15 May 2021)

British Chambers of Commerce (BCC) (2014) *Workforce Survey 2014*. Available from: https://wcnwchamber.org.uk/british-chambers-of-commerce-workforce-survey-2014/ (accessed 10 June 2019)

Brost, B. & Bradley, K. (2006) Student compliance with assigned reading: a case study, *Journal of Scholarship of Teaching and Learning*, 6:2, 101-111

Brown, J. (1998) *Understanding Research in Second Language Learning: A Teacher's Guide to Statistics and Research Design,* Cambridge: Cambridge UP

Brown, J. & Rogers, T. (2002) *Doing Second Language Research*, Oxford: Oxford University Press

Brown, R., Waring, R. & Donkaewbua, S. (2008) Incidental vocabulary acquisition from reading, reading-while listening, and listening to stories, *Reading in a Foreign Language*, 20:2, 136-163

Browne, C. (2014) A new General Service List: the better mousetrap we've been looking for?, *Vocabulary Learning and Instruction*, 3:2, 1-10.

Browne, C. (2016) *New General Service List Project: The Most Important Words for Second Language Learners of English*. Available from: www.newgeneralservicelist.org/nawl-new-academic-word-list/ (accessed 30 May 2021)

Browne, C., Culligan, B., & Phillips, J. (2013) *New Academic Word List*. Available from: www.newgeneralservicelist.org/nawl-new-academic-word-list/ (accessed 10 November 2019)

Bryman, A. (1988) *Quantity and Quality in Social Research*, New York: Routledge

Bryman, A. (2008) *Social Research Methods*, 3rd edn, Oxford: Oxford UP

Brysbaert, M., Stevens, M., Mandera, P. & Keuleers, E. (2016) How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age, *Frontiers in Psychology*, 7:1116, 1-11

Burgess, H., Sieminski, S. & Arthur, L. (2006) *Achieving Your Doctorate in Education*, London: Sage

Burke, P. (2017) Widening access to higher education, in Shin, J.C. & Teixeira, P. (eds), *Encyclopedia of International Higher Education Systems and Institutions*, pp. 1-10

Burns, R. (2000) *Introduction to Research Methods*, London: Sage

Butcher, J., McPherson, E., Shelton, I., Clarke, A., Hills, L. & Hughes, J. (2017) Unfit for purpose? Rethinking the language of assessment for Widening Participation students, *Widening Participation and Lifelong Learning*, 19:2, 27-46

Cambridge Dictionary (2020) Bulk. Available from: https://dictionary.cambridge.org/dictionary/english/bulk (accessed 6 December 2020)

Campion, M. & Elley, W. (1971) *An Academic Vocabulary List*, Wellington: New Zealand Council for Educational Research.

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

Carrell, S., Fullerton, R. & West, J. (2009) Does your cohort matter? Measuring peer effects in college achievement, *Journal of Labor Economics*, 27:3, 439-464

Carroll, J. (2002) *Suggestions for Teaching International Students More Effectively*, Oxford Brookes University, Learning and Teaching Briefing Papers

Carroll, J. & Ryan, J. (2005) *Teaching International Students*, Oxon: Routledge

Carver, R. (1994) Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: implications for instruction, *Journal of Reading Behavior*, 26, 413-437

Castles, A., Rastle, K. & Nation, K. (2018) Ending the reading wars: reading acquisition from novice to expert, *Psychological Science in the Public Interest*, 19:1, 5-51

Catts, H. W. & Kamhi, A. G. (2017) Prologue: Reading comprehension is not a single ability, *Language, Speech, and Hearing Services in Schools*, 48:2, 73-76.

Chall, J., Jacobs, V. & Baldwin, L. (1990) *The Reading Crisis: Why Poor Children Fall Behind*, Cambridge, MA: Harvard University Press

Charles, E. (2019) Decolonizing the curriculum, *Insights,* 32:1. Available from: https://insights.uksg.org/articles/10.1629/uksg.475/ (accessed 1 October 2019)

Chen, M. (2018) Beyond frequencies: investigating the semantic and the stylistic features of phrasal verbs in a three-year longitudinal corpus by Chinese university students, in Brezina, V. & Flowerdew, L., *Learner Corpus Research*, London: Bloomsbury

Chinese University Hong Kong (2012) *What Roles do Explicit and Implicit Learning Play in Vocabulary Acquisition?* Available from: http://engres.ied.edu.hk/vocabulary/vocabulary2-9.html (accessed 11 November 2012)

Chiu, K. (1972) Measuring register characteristics: a prerequisite for preparing advanced level TESOL programs, *TESOL Quarterly*, 6:2, 129-141

Chung, T. & Nation, P. (2003) Technical vocabulary in specialised texts, *Reading in a Foreign Language*, 15:2, 103-116

Chung, T. & Nation, P. (2004) Identifying technical vocabulary, *System*, 32, 251-263

Coady, J. & Huckin, T. (1997) *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*, Cambridge: Cambridge UP

Cobb, T. (1997) From Concord to Lexicon: Development and Test of a Corpus-Based Lexical Tutor, PhD thesis. Available from: www.lextutor.ca/cv/webthesis/Thesis0.html (accessed 23 June 2014)

Cobb, T. (2010) Learning about language and learners from computer programs, *Reading in a Foreign Language*, 22:1, 181-200

Cobb, T. (2017) *VocabProfile*, Available from: www.lextutor.ca/vp/

Cohen, L., Manion, L. & Morrison, K. (2007) *Research Methods in Education*, 6th edn, Abingdon: Routledge

Colombo, L. & Prior, M. (2016) How do faculty conceptions on reading, writing, and their role in the teaching of academic literacies influence their inclusive attitude, *Ilha do Desterro*, 69:3, 115-124

Corino, E. & Onesti, C. (2019) Data-driven learning: a scaffolding methodology for CLIL and LSP teaching and learning, *Frontiers in Education*. Available from: www.frontiersin.org/articles/10.3389/feduc.2019.00007/full (accessed 1 January 2020)

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

Coventry University (2020) *(BAWE) British Academic Written English Corpus*. Available from: www.coventry.ac.uk/research/research-directories/current-projects/2015/british-academic-written-english-corpus-bawe/ (accessed 1 January 2021)

Cowan, J. (1974) Lexical and syntactic research for the design of EFL reading materials, *TESOL Quarterly* , 8:4, 389-399

Coxhead, A. (2000). A new academic word list, *TESOL Quarterly*, 34:2, 213-238

Crème, P. & Lea, M. R. (1997) *Writing at University*, 2nd edn, Maidenhead: OUP

Cressman, J. (2018) 'Doing' the reading: what neuroscience instructors should know about reading compliance and comprehension in college, *Journal of Undergraduate Neuroscience Education*, 16(3), A217-A219

Creswell, J. (2009) *Research Design*, 3rd edn, London: Sage

Creswell, J. (2014) *Research Design*, 4th edn, London: Sage

Crystal, D. (2020) *The Cambridge Encyclopedia of the English Language,* 3rd edn, Cambridge: CUP

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

Curtis, M. (2006) The role of vocabulary instruction in adult basic education, in Comings, J., Garner, B. & Smith, C. (eds), *Annual Review of Adult Learning and Literacy*, 6, 43-68. New Jersey: Lawrence Erlbaum Associates

Dancer, D., Morrison, K. & Tarr, G. (2015) Measuring the effects of peer learning on students' academic achievement in first-year business statistics, *Studies in Higher Education*, 40:10, 1808-1828.

Dang, T. (2020) Corpus-based word lists in second language vocabulary research, learning and teaching, in Webb, S. (ed.), *The Routledge Handbook of Vocabulary Studies*, Abingdon: Routledge

Dang, T. & Webb, S. (2016) Making an essential word list for beginners, in Nation, P. (ed), *Making and Using Word Lists for Language Learning and Testing*, Amsterdam: John Benjamins, pp. 153-167

Dang, T., Webb, S. & Coxhead, A. (2020) Evaluating lists of high frequency words: teachers' and learners' perspectives, *Language Teaching Research*, 1-25

D'Anna, C., Zechmeister, E. & Hall, J. (1991) Toward a meaningful definition of vocabulary size, *Journal of Reading Behaviour,* 23:1, 109-121

Davies, A. (2007) *An Introduction to Applied Linguistics: From Practice to Theory*, 2nd edn, Edinburgh: EUP

Davis, F. (1942) Two new measures of reading ability, *Journal of Educational Psychology,* 33:5, 365-372

Davis, F. (1944) Fundamental factors of comprehension in reading, *Psychometrika*, 9:3, 185-197

Davis, F. (1968) Research in comprehension in reading, *Reading Research Quarterly*, 3, 499-545

De Gruyter (2021) *Applied Linguistics Review*: *about this journal*. Available from: www.degruyter.com/journal/key/ALR/html (accessed 13 February 2021)

Denscombe, M. (2007) *The Good Research Guide*, Maidenhead: Open University Press

DfE (2014a) *The National Curriculum in England: Complete Framework for Key Stages 1 to 4*. Available from: www.gov.uk/government/publications/national-curriculum-in-england-framework-for-key-stages-1-to-4 (accessed 27 December 2020)

DfE (2014b) *Phonics: Choosing a Programme*. Available from: www.gov.uk/government/collections/phonics-choosing-a-programme (accessed 27 December 2020)

DfE (2015) *Reading: The Next Steps, Supporting Higher Standards in Schools*. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/409409/Reading_the_next_steps.pdf (accessed 27 December 2020)

DfE (2017) *Unlocking Talent, Fulfilling Potential: A Plan for Improving Social Mobility Through Education*. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attach

ment_data/file/667690/Social_Mobility_Action_Plan_-_for_printing.pdf (accessed 8 November 2018)

DfE (2018) *Employer Skills Survey*. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/746493/ESS_2017_UK_Report_Controlled_v06.00.pdf (accessed 7 June 2019)

DfE (2019) *National Curriculum Assessments at Key Stage 2 in England, 2019 (Provisional)*. Available from: www.gov.uk/government/statistics/national-curriculum-assessments-key-stage-2-2019-provisional (accessed 21 December 2020)

Diaz, I. (2015) Training in metacognitive strategies for students' vocabulary improvement by using learning journals, *Issues in Teachers' Professional Development*, 17:1, 87-102

Donley, A. (2012) *Student Handbook to Sociology: Research Methods,* New York: Infobase Learning

Dörnyei, Z. (2007) *Research Methods in Applied Linguistics*, Oxford: OUP

Drake, P. & Heath, L. (2011) *Practitioner Research at Doctoral Level: Developing Coherent Research Methodologies*, London: Routledge

Duff, D., Tomblin, J. & Catts, H. (2015) The influence of reading on vocabulary growth: a case for a Matthew effect, *Journal of Speech, Language and Hearing Research*, 58:3, 853-864

Dunn, L. M. (1959) *Peabody Picture Vocabulary Test*, 1st edn, Circle Pines, MN: American Guidance Service

Dunn, L. M. & Dunn, L. M. (1997) *Peabody Picture Vocabulary Test*, 3rd edn, Circle Pines, MN: American Guidance Service

Dunworth, K. (2010) Clothing the emperor: addressing the issue of English language proficiency in Australian universities, *Australian Universities' Review*, 52:2, 5-9

Dupuy, H. (1974) *The Rationale, Development and Standardization of a Basic Word Vocabulary Test,* Washington, DC: National Center for Health Statistics

Durkin, J. (2019) *Cambridge Language Sciences Interdisciplinary Research Centre*. Available from: www.languagesciences.cam.ac.uk/news/cambridge-english-corpus-available-academic-use (accessed 30 May 2021)

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

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

Durrant, P. & Brenchley, M. (2019) Development of vocabulary sophistication across genres in English children's writing, *Reading and Writing*, 32, 1927-1953

ECU (2018) *Degree Attainment Gaps*. Available from: www.ecu.ac.uk/guidance-resources/student-recruitment-retention-attainment/student-attainment/degree-attainment-gaps/ (accessed 6 June 2019)

EEF (2018) *Metacognition and Self-Regulated Learning*. Available from: https://educationendowmentfoundation.org.uk/public/files/Publications/Campaigns/Metacognition/EEF_Metacognition_and_self-regulated_learning.pdf (accessed 14 March 2019)

EEF (2019) *Improving Literacy in Secondary Schools*. Available from: https://educationendowmentfoundation.org.uk/tools/guidance-reports/ (accessed 27 December 2020)

Ehren, B., Lenz, B.K. & Deshler, D. (2004) Enhancing literacy proficiency with adolescents and young adults, in Stone, C., Silliman, E., Ehren, B. & Apel, K., (eds), *Handbook of Language &* Literacy, New York: Guilford Press, pp. 681-701

Elliott, H. (1997) The use of diaries in sociological research on health experience, *Sociological Research Online*, 2:2. Available from: www.socresonline.org.uk/2/2/7.html (accessed 8 November 2018)

Ellis, R. (1995) Modified oral input and the acquisition word meanings, *Applied Linguistics*, 16:4, 401-441

Elsevier (2021) *What is Peer Review?* Available from: www.elsevier.com/reviewers/what-is-peer-review (accessed 13 February 2021)

Engber, C. (1995) The relationship of lexical proficiency to the quality of ESL compositions, *Journal of Second Language Writing*, 4:2, 139-155

Ernawati, I. & Ardi, P. (2017) Exploring students difficulties in reading academic tests, *UNNES International Conference on ELTLT*, 0:0, 428-444

Fair Education Alliance (FEA) (2018) *Report Card 2018*. Available from: https://static1.squarespace.com/static/543e665de4b0fbb2b140b291/t/5b8928a10e2e728743b52f37/1535715549432/report_card_2018_FINAL_Aug-31b.pdf (accessed 17 June 2019)

Fairburn, G. & Winch, C. (2011) *Reading, Writing and Reasoning*, 3rd edn, Maidenhead: OUP

Faisal, M. Mulaya, K. & Syamsul, B. (2017) Errors in EFL writing by junior high students in Indonesia, *International Journal of Research Studies in Language Learning*, 6:1, 38-52

Fang, Z., Caoa, P. & Murray, N. (2020) Language and meaning making: register choices in seventh- and ninth-grade students' factual writing, *Linguistics and Education,* 56, 1-12

Farrell, P. (1990) *A Lexical Analysis of the English of Electronics and a Study of Semi-technical Vocabulary* (CLCS Occasional Paper No.25), Dublin: Trinity College

Fasih, P., Izadpanah, S. & Shahnavaz, A. (2018) The effect of mnemonic vocabulary instruction on reading comprehension of students, *International Journal of Applied Linguistics & English Literature*, 7:3, 49-59

Field, A. (2013) *Discovering Statistics Using IBM SPSS*, 4th edn, London: Sage

Field, A. (2016) *An Adventure in Statistics*, London: Sage

Foster, L., Diamond, I. & Jeffries, J. (2015) *Beginning Statistics*, 2nd edn, London: Sage

Frankfort-Nachmias, C. & Leon-Guerrero, A. (2015) *Social Statistics for a Diverse Society*, 7th edn, London: Sage

French, A. (2013) 'Let the Right Ones In!': widening participation, academic writing and the standards debate in higher education, *Power and Education*, 5:3, 236-247

Fry, J. (1988) Diary studies in classroom SLA research: problems and prospects, *Jalt Journal*, 9:2, 158-167

Gablasova, D. (2014) Learning and retaining specialized vocabulary from textbook reading: comparison of learning outcomes through L1 and L2, *The Modern Language Journal*, 98:4, 976-991

Gablasova, D. (2015) Learning technical words through L1 and L2: completeness and accuracy of word meanings, *English for Specific Purposes*, 39, 62-74

Gablasova, D. & Brezina, V. (2018) Disagreement in L2 spoken English: from learner corpus research to corpus-based teaching materials, in Brezina, V. & Flowerdew, L., *Learner Corpus Research*, London: Bloomsbury

Gablasova, D., Brezina, V. & McEnery, T. (2017) Exploring learner language through corpora: comparing and interpreting corpus frequency information, *Language Learning*, 67: S1, pp. 130-154

Galal, Z. (2020) Limitations and solutions of the Academic Word List, in Scatolini, S. & George, M. (Eds), *Education and Society in the Middle East and North Africa*, Newcastle: Cambridge Scholars Publishing

Galloway, E., Stude, J. & Uccelli, P. (2015) Adolescents' metalinguistic reflections on the academic register in speech and writing, *Linguistics and Education*, 31, 221-237

Gardner, D. (2004) Vocabulary input through extensive reading: a comparison of words found in children's narrative and expository reading materials, *Applied Linguistics*, 25:1, 1-37

Gardner, D. (2007) Validating the construct of word in applied corpus-based vocabulary research: a critical survey, *Applied Linguistics*, 28, 241-265

Gardner, D. (2013) *Exploring Vocabulary*, Abingdon: Routledge

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

GDPR (2018) *General Data Protection Regulation*. Available from: www.gov.uk/government/publications/guide-to-the-general-data-protection-regulation (accessed 28 February 2019)

Gilbert, N. & Stoneman, P. (2016) *Researching Social Life*, 4th edn, London: Sage

Gillet, A. (2012) *Using English for Academic Purposes: A Guide for Students in Higher Education.* Available from: www.uefap.com/vocab/select/selframs.htm (accessed 7 May 2014)

Gilner, L. (2011) A primer on the General Service List, *Reading in a Foreign Language*, 23:1, 65-83

Gold, D., Andres, D., Etezadi, J., Arbuckle, T., Schwartzman, A. & Chaikelson, J. (1995) Structural equation model of intellectual change and continuity and predictors of intelligence in older men, *Psychology and Aging*, 10, 294-303

Golinkoff, R., Hoff, E., Rowe, M., Tamis-LeMonda, C., Hirsh-Pasek, K. (2019) Language Matters: Denying the Existence of the 30-Million-Word Gap Has Serious Consequences, in *Child Development*, 90:3, 985–992

González Fernández, B. & Schmitt, N. (2020). Word knowledge: exploring the relationships and order of acquisition of vocabulary knowledge components, *Applied Linguistics*, 41:4, 481-505

Goodson, P. (2017) *Becoming an Academic Writer*, 2nd edn, London: Sage

Gorard, S. (2013) *Research Design*, London: Sage

Gorski, P. (2008) The myth of the 'Culture of Poverty' in, *Educational Leadership*, 65:7, 32-36

Goulden, R., Nation, P. & Read, J. (1990) How large can a receptive vocabulary be? *Applied Linguistics*, 11:4, 341-363

GovUK (2020) *GCSE English and Maths Results*. Available from: www.ethnicity-facts-figures.service.gov.uk/education-skills-and-training/11-to-16-years-old/a-to-c-in-english-and-maths-gcse-attainment-for-children-aged-14-to-16-key-stage-4/latest (accessed 28 December 2020)

Gray, D. (2018) *Doing Research in the Real World*, 4th edn, London: Sage

Green, C. & Lambert, J. (2018) Advancing disciplinary literacy through English for academic purposes: discipline-specific wordlists, collocations and word families for eight secondary subjects, *Journal of English for Academic Purposes*, 35, 105-111

Gunobgunob-Mirasol, R. (2019) Vocabulary size, reading motivation, reading attitudes and reading comprehension performance among Filipino college learners of English, *International Journal of Evaluation and Research in Education* (IJERE), 8:1, 64-70

Guo, Y., Roehrig, A. & Williams, R. (2011) The relation of morphological awareness and syntactic awareness to adult's reading comprehension: is vocabulary knowledge a mediating variable? *Journal of Literacy Research*, 43:2, 159-183

Ha, A. & Hyland, K. (2017) What is technicality? A Technicality Analysis Model for EAP vocabulary, *Journal of English for Academic Purposes*, 28, 35-49

Hammersley, M. (2008) Assessing validity in social research, in Alasuutari, P., Bickman, L. & Brannen, J. (eds), *The Sage Handbook of Social Research Methods*, London: Sage

Harmer, J. (2015) *The Practice of English Language Teaching*, 5th edn, Harlow: Pearson

Hart, B., & Risley, T. R. (1995) *Meaningful differences in the everyday experience of young American children*, Baltimore, MD: Paul H Brooke

Hartmann, G. (1941) A critique of the common method of estimating vocabulary size, together with some data on the absolute word knowledge of educated adults, *Journal of Educational Psychology*, 32:5, 351-358

Hartmann, G. (1946) Further evidence on the unexpected large size of recognition vocabularies among college students, *Journal of Educational Psychology,* 37:7, 436-439

Hatami, S. (2014) Teaching formulaic sequences in the ESL classroom, *TESOL Journal*, 6:1, 112-129

Hatteberg, S. & Steffi, K. (2013) Increasing reading compliance of undergraduates: an evaluation of compliance methods, *Teaching Sociology*, 41:4, 346-352

Haveman, R. & Smeeding, T. (2006) The role of higher education in social mobility, *The Future of Children*, 16:2, 125-150

He, C. (2019) A review on diagnostic vocabulary tests, *International Journal of Higher Education,* 8:3, 126-133

Heatley, A., Nation, I.S.P. & Coxhead, A. (2002) *RANGE and FREQUENCY programs*. Available from: www.victoria.ac.nz/lals/staff/paul-nation.aspx (accessed 14 March 2019)

Hedge, T. (2000) *Teaching and Learning in the Language Classroom*, Oxford: OUP

Helberg, A., Poznahovska, M., Ishizaki, S., Kaufer, D., Werner, N. & Wetzel, D. (2018) Teaching textual awareness with DocuScope: using corpus-driven tools and reflection to support students' written decision-making, *Assessing Writing*, 38, 40-45

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

Hewitt, R. (2020) *Getting On: Graduate Employment and its Influence on UK Higher Education*. Available from: www.hepi.ac.uk/wp-content/uploads/2020/03/Getting-on.pdf (accessed 19 September 2021)

Higginbotham, G. & Reid, J. (2019) The lexical sophistication of second language learners' academic essays, *Journal of English for Academic Purposes*, 37,127-140

Hillege, S., Catterall, J. Beale, B. & Stewart, L. (2014) Discipline matters: embedding academic literacies into an undergraduate nursing program, *Nurse Education in Practice,*14:6, 686-691

Hirsch, E.D. (2013) A wealth of words. The key to increasing upward mobility is expanding vocabulary, *City Journal*, 23:1. Available from: www.city-journal.org/html/wealth-words-13523.html (accessed 3 August 2018)

Hirsh-Pasek, K., Adamson, L., Bakeman, R., Owen, M., Golinkoff, R., Pace, A., Yust, P., Suma, K. (2015) The Contribution of Early Communication Quality to Low-Income Children's Language Success in Psychological Science, 1–13

Hoeft, M. (2012) Why university students don't read: what professors can do to increase compliance, *International Journal for the Scholarship of Teaching and Learning*, 6:2, 1-19

Hsu, W. (2011) The vocabulary thresholds of business textbooks and business research articles for EFL learners, *English for Specific Purposes*, 30:4, 247-257

Hsu, W. (2014) Measuring the vocabulary load of engineering textbooks for EFL undergraduates, *English for Specific Purposes*, 33, 54-65

Hu, M. & Nation, I. (2000) Vocabulary density and reading comprehension, *Reading in a Foreign Language*, 23, 403-430

Hudson, R. (2009) *What is Applied Linguistics*. Available from: cambridge.org › reading › Applied_ART_WhatIsAL (accessed 2 January 2020)

Hunt, A. & Beglar, D. (1998) Current research and practice in teaching vocabulary, *The Language Teacher*, 22,1

Hyland, K. (2006) *English for Academic Purposes*, Abingdon: Routledge

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

Inside Careers (2018) *Getting Started: What Skills do Graduate Employers Look for?* Available from: www.insidecareers.co.uk/career-advice/what-skills-do-graduate-employers-look-for/ (accessed 9 May 2019)

Jameson, D. (2007) Literacy in decline: untangling the evidence, *Business and Professional Communication Quarterly*, 70:1, 16-33

Jones, E. & Brown, S. (eds) (2007) *Internationalising Higher Education*, Abingdon: Routledge.

Jones, J. C. & Conrad, H. S. (1933) The growth and decline of intelligence: a study of a homogeneous group between the ages of ten and sixty, *Genetic Psychology Monographs*, 13, 223-275

Keele University (2019) *Keele Decolonising the Curriculum Network*. Available from: www.keele.ac.uk/equalitydiversity/equalityawards/raceequalitycharter/keeledecolonisingthecurriculumnetwork/#relevant-material (accessed 1 October 2019)

Kettley, N. (2007) The past, present and future of widening participation research, *British Journal of Sociology of Education*, 28:3, 333-347

Kirkpatrick, E.A. (1891) Number of words in an ordinary vocabulary, *Science*, 18:446, 107-108

Klugh, H.E. (1986) *Statistics: The Essentials for Research*, Hillsdale: Lawrence Erlbaum Associates

Koester, A. (2010) Building small specialised corpora, in O'Keefe, A. & McCarty, M. (Eds), *The Routledge Handbook of Corpus Linguistics,* Abingdon: Routledge

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

Kremmel, B. & Schmitt, N. (2016) Interpreting vocabulary test scores: what do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly*, 13:4, 377-392

Kremmel, B. & Schmitt, N. (2018) Vocabulary levels test, in Lionta, J., *The TESOL Encyclopedia of English Language Teaching,* New Jersey: Wiley Blackwell

Kurniawan, I. (2017) Assessing English students` vocabulary size of Lampung State Islamic University, *Humaniora*, 8:4, 381-390

Kyle, K. & Crossley, S. (2016) The relationship between lexical sophistication and independent and source-based writing, *The Journal of Second Language Writing*, 34, 12-24

Kyle, K., Crossley, S. & Berger, C. (2018) The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0, *Behavior Research Methods*, 50, 1030-1046

Laflamme, G. (1997) The effect of the multiple exposure vocabulary method and the target reading/writing strategy on test scores, *Adolescent & Adult Literacy* , 40:5, 372-381

Lan, L. & MacGregor, L. (2010) Investigating the receptive vocabulary size of university-level Chinese learners of English: how suitable is the Vocabulary Levels Test? *Language and Education*, 24:3, 239-249.

Larsson, T. & Kaatari, H. (2020) Syntactic complexity across registers: investigating formality in second-language writing, *Journal of English for Academic Purposes,* 45, 1-16

Laufer, B. (1989) What percentage of text-lexis is essential for comprehension? in Lauren, C. & Nordman, M. (eds), *Special Language: From Humans to Thinking Machines*, Clevedon, England: Multilingual Matters, pp. 316-323

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

Laufer, B. (2013) Vocabulary and writing, in Chapelle, C., *The Encyclopaedia of Applied Linguistics*, 1-5, Boston, MA: Wiley-Blackwell.

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

Laufer, B. & Nation, P. (1995) Vocabulary size & use: lexical richness in L2 written productions, *Applied Linguistics*, 16:3, 307-322

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

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

Layton, A., Robinson, J. & Lawson, M. (1998) The relationship between syntactic awareness and reading performance, *Journal of Research in Reading*, 21:1, 5-23

Leamnson, R. (1999) *Thinking about Teaching: Developing Habits of Learning with First Year College and University Students,* Sterling, VA: Stylus.

Lee, L., Cheng, S., Lin, C., Chi, M., Lee, I. & Yang, Y. (2018) Communication skills scores of medical students and interview performance 3 years later, *Health Education Journal*, 77:7, 849-854

Lee, N.J. (2009) *Achieving your Professional Doctorate*, Maidenhead: Open University Press

Leedy, P. & Ormrod, J. (2015) *Practical Research* (11th edn), Harlow: Pearson

Lesaux, N. & Kieffer, M. (2010) Exploring sources of reading comprehension difficulties among language minority learners and their classmates in early adolescence, *American Educational Research Journal*, 47:3, 596-632

Lewis, M. (1993) *The Lexical Approach*, Hove: LTP

Li, J. & Schmitt, N. (2009) The acquisition of lexical phrases in academic writing: A longitudinal case study, *Journal of Second Language Writing,* 18:2, 85-102

Lightbown, P. & Spada, N. (2006) *How Languages are Learned* (3rd edn), Oxford: Oxford UP

LinkedIn (2019) *The Skills Companies Need Most in 2018 – and the Courses to Get Them*. Available from: https://learning.linkedin.com/blog/top-skills/the-skills-companies-need-most-in-2018--and-the-courses-to-get-t (accessed 27 May 2019)

Liu, J. & Han, L. (2015) A corpus-based environmental academic word list building and its validity test, *English for Specific Purposes*, 39, 1-11

Lorge, I. & Chall, J. (1963) Estimating the size of vocabularies of children and adults: an analysis of methodological issues, *Journal of Experimental Education*, 32: 2, 147-57

Lutviana, R., Kadarisman, E. & Laksmi, E. (2015) Correlation between lexical richness and overall quality of argumentative essays written by English department students, *Pendidikan Humaniora*, 3:1, 41-51

Malmström, H., Pecorari, D. & Shaw, P. (2018) Words for what? Contrasting university students' receptive and productive academic vocabulary needs, *English for Specific Purposes*, 50, 28-39

Mancilla-Martinez, J. & Lesaux, N. (2010) Predictors of reading comprehension for struggling readers: the case of Spanish-speaking language minority learners, *Journal of Educational Psychology*, 102: 3, 701-711

Marsden, E., Morgan-Short, K., Thompson, S. & Abugaber, D. (2018) Replication in second language research: narrative and systematic reviews and recommendations for the field, *Language Learning*, 68:2, 321-391

Marshall, J. (2009) An examination of the attitudes and perceptions of ESOL students towards their English language learning needs at the University of Derby, MA Ed submission, University of Derby

Masini, F. (2019) Multi-word expressions and morphology, *Oxford Research Encyclopedia*, Oxford: OUP

Maskor, Z. & Bharudin, H. (2016) Receptive vocabulary knowledge or productive vocabulary knowledge in writing skill, which one important? *International Journal of Academic Research in Business and Social Sciences*, 6:11, 261-271

Masrai, A. & Milton, J. (2012) The vocabulary knowledge of university students in Saudi Arabia, *TESOL Arabia Perspectives*, 19:3, 13-19

Masrai, A. & Milton, J. (2017) Recognition vocabulary knowledge and intelligence as predictors of academic achievement in EFL context, *TESOL International Journal,* 12:1 128-142

Masrai, A. & Milton, J. (2018) Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement, *Journal of English for Academic Purposes*, 31, 44-57

Matthews, N., Simon, J. & Kelly, E. (2016) 'Diagnosing' academic literacy needs in higher education: deploying the ethical resources of disability studies to interrogate information sharing in educational transitions, *Pedagogy, Culture & Society*, 24:3, 445-457

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

McEnery, T. & Hardie, A. (2012) *Corpus Linguistics*, Cambridge: CUP

McEnery, T., Xiao, R. & Tono, Y. (2006) *Corpus-based Language Studies: An Advanced Resource Book.* London: Routledge

McGrath, J. & Coles, A. (2013) *Your Education Research Project Companion* (2nd edn), Harlow: Pearson

McLean, S. (2017) Evidence for the adoption of the flemma as an appropriate word counting unit, *Applied Linguistics*, 39:6, 823-845

McLean, S. & Kramer, B. (2015) The creation of a new Vocabulary Levels Test, *Shiken*, 19:2, 1-11

McMillan, K. & Weyers, J. (2009) *The Smarter Study Skills Companion* (2nd edn), Harlow: Prentice Hall

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

Meara, P. (1996) *The Vocabulary Knowledge Framework*. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.716.7973&rep=rep1&type=pdf (accessed 11 December 2020)

Meara, P. (2005) Lexical frequency profiles: a Monte Carlo analysis, *Applied Linguistics*, 26:1, 32–47

Meara, P. & Buxton, B. (1987) An alternative to multiple choice vocabulary tests, *Language Testing*, 4:2, 142-154

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

Melka Teichroew, F. (1982) Receptive vs. productive vocabulary: a survey, *Interlanguage Studies Bulletin*, 6:2, 5-33.

Miller, R. & Pessoa, S. (2018) Corpus-driven study of information systems: project reports, in Brezina, V. & Flowerdew, L., *Learner Corpus Research,* London: Bloomsbury

Mitkov, R. (2016) Computational phraseology light: automatic translation of multiword expressions without translation resources, *Yearbook of Phraseology*, 7:1, 149-166

Mole, J. and Peacock, D. (2005) *Learning, Teaching and Assessment: A Guide to Good Practice for Lecturers Teaching Art, Design and Communication to D/Deaf Students*, Wolverhampton: University of Wolverhampton

Moore, P. & Harrington, M. (2016) Fractionating English language proficiency: policy and practice in Australian higher education, *Current Issues in Language Planning*, 17:3-4, 385-404

Morni, A. & Sahari, S. (2013) The impact of the living environment on reading attitudes, *Procedia - Social and Behavioral Sciences* ,101, 415-425

Morris, L. & Cobb, T. (2004) Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees, *System-Oxford,* 32:1, 75-87

Muncie, J. (2002) Process writing and vocabulary development: comparing Lexical Frequency Profiles across drafts, *System*, 30:2, 225-235

Murray, N. (2012) Ten 'Good Practice Principles' … ten key questions: considerations in addressing the English language needs of higher education students, *Higher Education Research & Development*, 31:2, 233-246

Murray, N. (2013) Widening participation and English language proficiency: a convergence with implications for assessment practices in higher education, *Studies in Higher Education*, 38:2, 299-311

Mykhailiuk, L. (2016) Efficient vocabulary testing techniques, *Vasyl Stefanyk Precarpathian National University*, 3:4, 54-57

Nadelson, L., Villagómez, A., Konkol, D., Haskell, C., McCulley, M. & Campbell, D. (2013) Messages are everywhere: reading perceptions, habits, and preferences of undergraduates, *Journal of College Reading and Learning*, 43:2, 70-90

Nagy, W. & Townsend, D. (2012) Words as tools: learning academic vocabulary as language acquisition, *Reading Research Quarterly*, 47:1, 91-108

Nagy, W., Herman, P. & Anderson, R. (1985) Learning words from context, *Reading Research Quarterly*, 20:2, 233-253

Nagy, W., Berninger, V. W. & Abbott, R. D. (2006) Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students, *Journal of Educational Psychology*, 98:1, 134-147

Nathan, M. (2013) *How can we Equip Graduates with the Workplace Skills They Need?* Available from: www.theguardian.com/careers/careers-blog/equip-graduates-workplace-skills (accessed 8 June 2019)

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

Nation, P. (1990) *Teaching and Learning Vocabulary*, New York: Newbury House

Nation, P. (1993) Using dictionaries to estimate vocabulary size: essential, but rarely followed procedures, *Language Testing*, 10:1, 27-40

Nation, P. (2001) *Learning Vocabulary in Another Language,* Cambridge: CUP

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

Nation, P. (2013) *Learning Vocabulary in Another Language* (2nd edn), Cambridge: Cambridge UP

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

Nation, P. & Coxhead, A. (2021) *Measuring Native-Speaker Vocabulary Size*, Amsterdam: John Benjamins

Nation, P. & Heatley, A. (1994) *Range: A Program for the Analysis of Vocabulary in Texts [Software*]. Available from: www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx (accessed 4 December 2018)

Nation, P. & Kyongho, H. (1995) Where would general service vocabulary stop and special purposes vocabulary begin? *Systems*, 23:1, 35-41

National Literacy Trust (2017) *What is Phonics*? Available from: https://literacytrust.org.uk/information/what-is-literacy/what-phonics/ (accessed 28 December 2020)

Nelson, M. (2010) Building a written corpus, in O'Keefe, A. & McCarthy, M. (eds), *The Routledge Handbook of Corpus Linguistics*, Abingdon: Routledge

Nergis, A. (2013) Exploring the factors that affect reading comprehension of EAP learners, *Journal of English for Academic Purposes*, 12, 1-9

Nesi, H. & Gardner, S. (2012) Families of genres of assessed writing, in Nesi, H. & Gardner, S. (eds), *Genres across the Disciplines: Student Writing in Higher Education*, Cambridge: Cambridge University Press, pp. 21-56

Newby, P. (2010) *Research Methods for Education,* Harlow: Pearson

Nguyen, T., Williams, J. & Trimarchi, A. (2015) Discipline-specific language instruction for international students in introductory economics, *Canadian Journal for the Scholarship of Teaching and Learning*, 6:3,1-21

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

Nusbaum, H., Pisoni, D. & Davis, C.(1984) *Sizing up the Hoosier Mental Lexicon: Measuring the Familiarity of 20,000 Words*, Speech Research Laboratory Progress Report, Bloomington, IN: Indiana University

Oakley, G. (2011) The assessment of reading comprehension cognitive strategies: practices and perceptions of Western Australian teachers, *Australian Journal of Language and Literacy*, 34:3, 279-293

O'Connor, K. & Solity, J. (2020) Assessing the impact of a well-founded intervention with lower-achieving students in Key Stage 2, *Educational Psychology in Practice*, 36:3, 257-280

Ofsted (2019) *School Inspection Handbook*. Available from: www.gov.uk/government/publications/school-inspection-handbook-eif (accessed 27 December 2020)

Olinghouse, N. & Leaird, J. (2009) The relationship between measures of vocabulary and narrative writing quality in second and fourth-grade students, *Read Writ*, 22, 545-565

Olinghouse, N. & Wilson, J. (2013) The relationship between vocabulary and writing quality in three genres, *Read Writ*, 26, 45-65

Oliver, P. (2013) *Writing your Thesis*, London: Sage

Oslund, E., Clemens, N., Simmons, D. & Simmons, L. (2018) The direct and indirect effects of word reading and vocabulary on adolescents' reading comprehension: comparing struggling and adequate comprehenders, *Reading and Writing*, 31:2, 355-379

Oxford University Press (2021) *Oxford Languages*. Available from: https://languages.oup.com/google-dictionary-en/ (accessed 22 March 2021)

Paige, D., Smith, G., Rasinski, T., Rupley, W., Magpuri-Lavell, T. & Nichols, W. (2019) A path analytic model linking foundational skills to Grade 3 state reading achievement, *The Journal of Educational Research*, 112:1, 110-120

Paltridge, B. & Phakiti, A. (eds) (2015) *Research Methods in Applied Linguistics*, London: Bloomsbury

Paquot, M. (2005) Towards a productively-oriented academic word list, in Walinski, J., Kredens, K. & Gozdz-Roszkowski, S., *Practical Applications in Language and Computers*, Peter Lang: Frankfurt and Main, pp.127-140

Paquot, M. (2010) *Academic Vocabulary in Learner Writing: From Extraction to Analysis*, London: Bloomsbury

Pignot-Shahov, V. (2012) Measuring L2 receptive and productive vocabulary knowledge, *University of Reading: Language Studies Working Papers,* 4, 37-45

Porte, G. & McManus, K. (2019) *Doing Replication Research in Applied Linguistics*, Abingdon: Routledge

Preece, S. & Godfrey, J. (2004) Academic literacy practices and widening participation: first year undergraduates on an academic writing programme, *Widening Participation and Lifelong Learning*, 6:1, 6-14

Proctor, R. (2015) Teachers and school research practices: the gaps between the values and practices of teachers, *Journal of Education for Teaching*, 41:5, 464-477

Prospects (2017) *What Skills Do Employers Want?* Available from: www.prospects.ac.uk/careers-advice/applying-for-jobs/what-skills-do-employers-want (accessed 8 June 2019)

Prowse, A. & Delbridge, R. (2013) 'I can't be arsed': A small-scale exploration of students' self-reported motivation on entering a course of study and eventual 'success', *Education + Training*, 55:7, 654-664

Punch, K. (2005) *Introduction to Social Research: Quantitative and Qualitative Approaches*, London: Sage

Punch, K. & Oancea, A. (2014) *Introduction to Research Methods in Education* (2nd edn), London: Sage

Quigley, A. (2018) *Closing the Vocabulary Gap*, Abingdon: Routledge

Rahimia, M. & Katala, M. (2012) Metacognitive strategies awareness and success in learning English as a foreign language: an overview, *Procedia – Social and Behavioral Sciences*, 31, 73-81

Rasekh, Z.E. & Ranjbary, R. (2003) Metacognitive strategy training for vocabulary learning, *TESL-EJ,* 7:2, Available from: www.tesl-ej.org/wordpress/issues/volume7/ej26/ej26a5/ (accessed 23 November 2018)

Read, J. (2000) *Assessing Vocabulary*, Cambridge: CUP

Richards, J. (1976) The role of vocabulary teaching, *TESOL Quarterly,* 10:1, 77-89

Rose, H., McKinley, J. & Briggs Baffoe-Djan, J. (2020) *Data Collection Research Methods in Applied Linguistics*, London: Bloomsbury

Rust, J. & Golombok, S. (2014) *Modern Psychometrics: The Science of Psychological Assessment* (3rd edn), Abingdon: Routledge

Salamonson, Y., Glew, P., Everett, B., Woodmass, J., Lynch, J. & Ramjan, L. (2019) Language support improves oral communication skills of undergraduate nursing students: a 6-month follow-up survey, *Nurse Education Today*, 72, 54-60

Salkind, N. (2014) *Statistics for People Who (Think They) Hate Statistics*, London: Sage

Sanson-Fisher, R., Hobden, B., Waller, A., Dodd, N. & Boyd, L. (2018) Methodological quality of teaching communication skills to undergraduate medical students: a mapping review, *BMC Medical Education*, 18, 1-7

Saunders, M. & Brown, B. (2008) *Dealing with Statistics*, Maidenhead: Open University Press

Schaie, K. W. (1996) *Intellectual Development in Adulthood: The Seattle Longitudinal Study*, New York: Cambridge University Press

Schleppegrell, M. (2004) *The Language of Schooling: A Functional Linguistics Perspective*, New Jersey: Lawrence Erlbaum Associates

Schleppegrell, M. & O'Hallaron, C. (2011) Teaching academic language in L2 secondary settings, *Annual Review of Applied Linguistics*, 31, 3-18

Schmitt, N. (1997) Vocabulary learning strategies. In Schmitt, N. & McCarthy, M. (eds), *Vocabulary: Description, Acquisition, and Pedagogy*, Cambridge: Cambridge University Press

Schmitt, N. (1998) Tracking the incremental acquisition of second language vocabulary: a longitudinal study, *Language Learning*, 48, 281-317

Schmitt, N. (2008) Instructed second language vocabulary learning, *Language Teaching Research*, 12:3, 329-363

Schmitt, N. (2010) *Researching Vocabulary*, Basingstoke: Palgrave Macmillan

Schmitt, N. (2014) Size and depth of vocabulary knowledge: what the research shows, *Language Learning*, 64:4, 913-951

Schmitt, N. (2017) Vocabulary: principles and practice. *English Teaching Professional*, 109, 4-6

Schmitt, N. (2019) Understanding vocabulary acquisition, instruction, and assessment: a research agenda, *Language Teaching*, 52:2 , 261-274

Schmitt, N. & Schmitt, D. (2014) A reassessment of frequency and vocabulary size in L2 vocabulary teaching, *Language Teaching*, 47:4, 484-503

Schmitt, N. & Schmitt, D. (2020) *Vocabulary in Language Teaching* (2nd edn), Cambridge: Cambridge UP

Schmitt, N., Schmitt, D. & 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., Jiang, X. & Grabe, W. (2011) The percentage of words known in a text and reading comprehension, *Modern Language Journal*, 95:1, 26-43

Schmitt, N., Cobb, T., Horst, M. & Schmitt, D. (2017) How much vocabulary is needed to use English? Replication of Van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007), *Language Teaching*, 50:2, 212-226

Schmitt, N., Nation, P. & Kremmel, B. (2020) Moving the field of vocabulary assessment forward: the need for more rigorous test development and validation, *Language Teaching*, 53, 109–120

Schuth, E., Kohne, J. & Weinert, S. (2017) The influence of academic vocabulary knowledge on school performance, *Learning and Instruction*, 49, 57-165

Scouller, K., Bonanno, H., Smith, L. & Krass, I. (2008) Student experience and tertiary expectations: factors predicting academic literacy amongst first-year pharmacy students, *Studies in Higher Education*, 33:2, 167-178

SCI Journal (2021) *What's a Good Impact Factor & Why it Matters (with 2018/2019 Impact Factor Ranking in 27 Categories)*. Available from: www.scijournal.org/articles/good-impact-factor (accessed 13 February 2021)

Scott, J. & Watt, S. (eds) (2010) *Ethnography in Social Science Practice*, London: Routledge

Seel, N.M. (2012) Intentional vocabulary learning. In Seel, N.M. (ed.), *Encyclopedia of the Sciences of Learning*, Boston, MA: Springer

Seliger, H. & Shohamy, E. (1989) *Second Language Research Methods,* Oxford: Oxford UP

Shanahan, T. & Lonigan, C. (2010) The National Early Literacy Panel: a summary of the process and the report, *Educational Researcher*, 39:4, 279-285

Sharma, A., Van Hoof, B. & Pursel, B. (2013) An assessment of reading compliance decisions among undergraduate students, *Journal of the Scholarship of Teaching and Learning,* 13:4,103-123

Shi, J. (2018) Pedagogical application of corpus on EFL of children with Autism Spectrum Disorders (ASD): the function of visualization on attention deficit problem, *Journal of Educational and Developmental Psychology,* 8:2, 120-131

Shokouhi, H., Goosh, S. & Maniati, M. (2009) Learners' incidental vocabulary acquisition: a case on narrative and expository texts, *English Language Teaching*, 2:1, 13-23

Šišková, Z. (2012) Lexical richness in EFL students' narratives, *University of Reading Language Studies Working Papers*, 4, 26-36

Skatova, A. & Ferguson, E. (2014) Why do different people choose different university degrees? Motivation and the choice of degree, *Frontiers in Psychology, 5,* 1244.

Smith, N. & Zinc, A. (1977) A cloze-based investigation of reading comprehension as a composite of subskills, *Journal of Literacy Research,* 9:4, 395-398

Smith, R. (2005) The lexical frequency profile: problems and uses, in Bradford-Watts, K., Ikeguchi, C. & Swanson, M. (eds), *JALT2004 Conference Proceedings*, Tokyo: JALT

Smith, T., Smith, B. & Dobbs, K. (1991) Relationship between the Peabody Picture Vocabulary Test-Revised, Wide Range Achievement Test-Revised, and Weschler Intelligence Scale for Children-Revised, *Journal of School Psychology*, 29:1, 53-56

Snow, C. E. & Uccelli, P. (2009) The challenge of academic language, in Olson, D. & Torrance, N. (eds), *The Cambridge Handbook of Literacy*, Cambridge: Cambridge University Press

Soil Quality Organisation (2020) *Fact Sheet: Bulk Density Measurement*. Available from: http://soilquality.org.au/factsheets/bulk-density-measurement (accessed 6 December 2020)

Sökmen, A. (1997) Current trends in teaching second language vocabulary. Vocabulary: description, acquisition and pedagogy, in Schmitt, N. & McCarthy, M. (eds), *Vocabulary: Description, Acquisition and Pedagogy*, Cambridge: Cambridge University Press (pp. 237-257)

Solity, J. & Vousden, J. (2009) Real books vs reading schemes: a new perspective from instructional psychology, *Educational Psychology*, 29:4, 469-511

Spearritt, D. (1972) Identification of sub-skills of reading comprehension by maximum likelihood factor analysis, *ETS Research Report Series*, 1, i-24

Sperry, D., Miller, P., Sperry, L. (2019) Hazardous intersections: Crossing disciplinary lines in developmental psychology in, *European Journal of Social Theory*, 23:1, 93-112

Sperry, D., Sperry, L. & Miller, P. (2018) Re-examining the Verbal Environments of Children From Different Socioeconomic Backgrounds in, *Child Development,* 90:1, 1303-1318

Stacey, J. & Granville, S. (2009) Entering the conversation: reaction papers in advanced academic literacy, *Teaching in Higher Education*,14:3, 327-339

Stahl, S. & Jacobson, M. (1986) Vocabulary difficulty, prior knowledge, and text comprehension, *Journal of Reading Behavior*, 13:4, 309-323

Stanovich, K. (1986) Matthew effects in reading: some consequences of individual differences in the acquisition of literacy, *Reading Research Quarterly*, 21:4, 360–407

Sudarman, S. & Chinokul, S. (2018) The English vocabulary size and level of English Department students at Kutai Kartanegara University, *Eternal (English, Teaching, Learning & Research Journal)*, 4:1, 1-15

Sunderland, H. & Spiegel, M. (2009) The written word, in Paton, A. & Wilkins, M. (eds), *Teaching Adult ESOL*, Maidenhead: OUP

Sutarsyah, C., Nation, P. & Kennedy, G. (1994) How useful is EAP vocabulary for ESP? A corpus-based study, *RELC Journal*, 25:2, 34-50

Sutton Trust (2014) *Improving Social Mobility Through Education*. Available from: www.suttontrust.com/wp-content/uploads/2014/06/the-sutton-trust-brochure-7.pdf (accessed 28 January 2019)

Tan, A. & Goh, L. (2018) Relationship between vocabulary size and reading comprehension levels of Malaysian tertiary students, *International Journal of English Language and Translation Studies*, 5:4, 149-155

Tanabe, T., Takahashi, M. & Shudo, K. (2014) A lexicon of multiword expressions for linguistically precise, wide-coverage natural language processing, *Computer Speech & Language,* 28:6, 1317-1339

Tardy, C. (2005) 'It's like a story': rhetorical knowledge development in advanced academic literacy, *Journal of English for Academic Purposes*, 4:4, 325-338

Teddle, C. & Tashakkori, A. (2009) *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioural Sciences*, London: Sage

Therova, D. (2020) Review of academic word lists, *TESL-EJ*, 24:1. Available from: www.tesl-ej.org/wordpress/issues/volume24/ej93/ej93a5/ (accessed 1 November 2021)

The Reading Chest (2020) *Synthetic Phonics*. Available from: www.readingchest.co.uk/synthetic-phonics (accessed 27 December 2020)

The Student Room (2017) *Accommodation*. Available from: www.thestudentroom.co.uk/showthread.php?p=83139074&highlight=noise (accessed 3 March 2019)

Thorndike, E.L. (1924) The vocabularies of school pupils, in Carelton, J. (ed.), *Contributions to Education*, New York: World Book Co, pp. 69-76

Thorndike, E.L. & Lorge, I. (1944) *The Teacher's Word Book of 30,000 Words.* New York: World Book Co

Thurstone, L.L. (1946) Note on a reanalysis of Davis' reading tests, *Psychometrika*, 11:3, 185-188

Tolmie, A., Muijs, D. & McAteer, E. (2011) *Quantitative Methods in Educational and Social Researching*, Berkshire: Open University Press

Trafford, V. & Leshem, S. (2008) *Stepping Stones to Achieving your Doctorate*, Maidenhead: Open University Press

Treffers-Daller, J. & Milton, J. (2013) Vocabulary size revisited: the link between vocabulary size and academic achievement, *Applied Linguistics Review*, 4:1, 151-172

Trujillo Becerra, C. L., Alvarez Ayure, C. P., Zamudio Ordoñez, M. N. & Morales Bohórquez, G. (2015) Facilitating vocabulary learning through metacognitive strategy training and learning journals, *Colombian Applied Linguistics Journal*, 17:2, 246-259

Tseng, W. & Schmitt, N. (2008) Toward a model of motivated vocabulary learning: a structural equation modeling approach, *Language Learning*, 58:2, 357-400

UCL (2021) *Longitudinal vs Cross-Sectional Studies*. Available from: https://learning.closer.ac.uk/learning-modules/introduction/types-of-longitudinal-research/longitudinal-versus-cross-sectional-studies/# (accessed 14 February 2021)

UCLES (2019) *The Cambridge English Scale*. Available from: www.cambridgeenglish.org/exams-and-tests/cambridge-english-scale/ (accessed 6 June 2019)

UEfAP (2021) *Grammar in EAP*. Available from: www.uefap.net/grammar/grammar-in-eap-cohesion/grammar-in-eap-cohesion (accessed 12 July 2021)

Universities UK (2004) *Measuring and recording student achievement*. Available from: www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2005/measuring-and-recording-student-achievement.pdf (accessed 25 April 2019)

Universities UK (2016) *Working in Partnership: Enabling Social Mobility in Higher Education – The Final Report of the Social Mobility Advisory Group*. Available from: www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2016/working-in-partnership-final.pdf (accessed 25 May 2019)

Universities UK (2018a) *Tackling Gaps in BME Students' Achievements at University*. Available from: www.universitiesuk.ac.uk/news/Pages/tackling-gaps-in-bme-students-achievements.aspx (accessed 6 June 2019)

Universities UK (2018b) *Higher Education in Numbers*. Available from: www.universitiesuk.ac.uk/facts-and-stats/Pages/higher-education-data.aspx (accessed 6 June 2019)

Universities UK (2019) *Universities Acting to Close BAME Attainment Gap*. Available from: www.universitiesuk.ac.uk/news/Pages/Universities-acting-to-close-BAME-student-attainment-gap.aspx (accessed 6 June 2019)

University of Bradford (2018) *Graduate Attributes and Employability*. Available from: www.bradford.ac.uk/educational-development/curriculum-framework/graduate-attributes-and-employability/ (accessed 4 January 2019)

University of Bradford (2020) *Using the Right Register*. Available from: www.bradford.ac.uk/t4-ssis/ruski-files/academic-voice/page_02.htm (accessed 31 December 2020)

University of Derby (2020) *Academic Regulations*. Available from: www.derby.ac.uk/about/academic-regulations/assessment-regulations-for-undergraduate-programmes/ (accessed 12 June 2020)

University of Essex (2008) *How to Improve your Academic Writing*. Available from: https://www.york.ac.uk/media/study/schoolsandcolleges/sixth-form-resources/how-to-improve-your-academic-writing.pdf (accessed 25 April 2019)

University of Gdansk (2019) *Educational Role of Languages*. Available from: ERL CONFERENCES | Edukacyjna Rola Języka (ug.edu.pl) (accessed 28 October 2021)

University of Hertfordshire (2018) *Graduate Attributes*. Available from: www.herts.ac.uk/about-us/student-charter/graduate-attributes (accessed 4 January 2019)

University of Plymouth (2008) *Widening Participation into Higher Education – A Resource Handbook for Working on the Agenda in Schools and Colleges*. Available from: www.heacademy.ac.uk/system/files/ahpp_wp_he_handbook_schools_colleges.pdf (accessed 4 January 2019)

University of Sheffield (2018) *The Sheffield Graduate*. Available from: www.sheffield.ac.uk/sheffieldgraduate/studentattributes (accessed 4 January 2019)

University of Sheffield (2020) *Style*. Available from: www.sheffield.ac.uk/ssid/301/study-skills/writing/academic-writing (accessed 31 December 2020)

University of Sydney (2020) *Types of Academic Writing*. Available from: www.sydney.edu.au/students/writing/types-of-academic-writing.html (accessed 31 December 2020)

University of York (2020) *Introduction to Academic Register*. Available from: www.york.ac.uk/res/elanguages/index/Modulecd/cu4s4/cu4s40501.htm (accessed 31 December 2020)

Valenzuela, H. (2020) *Linguistics for TESOL*, Cham: Palgrave Macmillan

Van der Slik, F. & Weideman, A. (2005) The refinement of a test of academic literacy, *Per Linguam*, 21:1, 23-35

Van der Slik, F. & Weideman, A. (2007) Testing academic literacy over time: is the academic literacy of first year students deteriorating?, *Ensovoort*, 11:2 ,126-137

Van der Slik, F. & Weideman, A. (2008) Measures of improvement in academic literacy, *Southern African Linguistics and Applied Language Studies*, 26:2, 363-378

Van Dyk, T. & Weideman, A. (2004) Switching constructs: on the selection of an appropriate blueprint for academic literacy assessment, *Journal for Language Teaching*, 38:1,1-13

Veas, A., Castejón, J., Gilar, R. & Miñano, P. (2015) Academic achievement in early adolescence: the influence of cognitive and non-cognitive variables, *The Journal of General Psychology*, 142:4, 273-294

Vilkaitė-Lozdienė, L. & Schmitt, N. (2020). Frequency as a guide for vocabulary usefulness: hi-, mid-, and low-frequency words, in Webb, S. (ed.), *The Routledge Handbook of Vocabulary Studies*, Abingdon: Routledge

Vulperhorst, J., Lutz, C., de Kleijn, R. & van Tartwijk, J. (2018) Disentangling the predictive validity of high school grades for academic success in university, *Assessment & Evaluation in Higher Education*, 43:3, 399-414

Walkup, V. (ed.) (2011) *Exploring Education Studies*, Harlow: Pearson

Walsh, S. (2010) Creating language teaching materials, in O'Keefe, A. & McCarthy, M. (eds), *The Routledge Handbook of Corpus Linguistics*, Abingdon: Routledge

Wang, Z., Sabatini, J., O'Reilly, T. & Weeks, J. (2018) Decoding and reading comprehension: A test of the decoding threshold hypothesis, *Journal of Educational Psychology*, 111:3, 387-401

Webb, S. (2005) Receptive and productive vocabulary learning: the effects of reading and writing on word knowledge, *Studies in Second Language Acquisition*, 27, 33-52

Webb, S. (ed.) (2020) *The Routledge Handbook of Vocabulary Studies*, Abingdon: Routledge

Webb, S. & Chang, A. (2015) How does prior word knowledge affect vocabulary learning progress in an extensive reading programme?, *Studies in Second Language Acquisition,* 37:4, 651-675

Webb, S. & Nation, P. (2017) *How Vocabulary is Learned*, Oxford: OUP

Webb, S. & Sasao, Y. (2013) New directions in vocabulary testing, *RELC Journal*, 44:3, 263-277

Webb, S., Sasao, Y. & Balance, O. (2017) The Updated Vocabulary Levels Test: developing and validating two new forms of the VLT, *International Journal of Applied Linguistics,* 168:1, 34-70

Weideman, A. (2003) Assessing and developing academic literacy, *Per Linguam*, 19:1 & 2, 55-65

Weideman, A. (2006) Assessing academic literacy: a task-based approach, *Language Matters*, 37:1, 81-101

Weideman, A. (2007) *Academic Literacy: Prepare to Learn* (2nd edn), Pretoria: Van Schaik

Weideman, A. (2011) Academic literacy tests: design, development, piloting and refinement, *Language Teaching*, 45:2,100-113

Weideman, A. & Van Dyk, T. (eds) (2014) *Academic Literacy: Test your Competence*, Bloemfontein: Geronimo Distribution

Weideman, A., Patterson, R. & Pot, A. (2016) Construct refinement in tests of academic literacy, in Read, J. (ed.), *Post-admission Language Assessment of University Students.* Cham: Springer, pp. 179-196

Weisser, M. (2016) *Concordancers*. Available from: http://martinweisser.org/corpora_site/concordancers.html (accessed 1 January 2020)

West, M. (1953) *A General Service List of English Words,* London: Longman, Green

Wilkins, D. (1972) *Linguistics in Language Teaching*, London: Arnold

Wittgenstein, L. (2009) *Tractatus Logico-Philosophicus*, Virginia: Wilder Publications

Workman, B. (2007) Casing the joint: explorations by the insider-researcher preparing for work-based projects, *Journal of Workplace Learning*, 19:3, 146-160

Worth, R. (2009) *Communication Skills* (3rd edn), New York: Ferguson

Wray, A. (2002) *Formulaic Language and the Lexicon*, Cambridge: CUP

Wu, K. & Carmichael, S. (2018) Impacts of discipline-specific language instruction on the academic writing of civil engineering students, *Asian Journal of Applied Linguistics*, 5:1, 9-35

Xue, G. & Nation, I.S.P. (1984) A university word list, *Language Learning and Communication*, 3:2, 215-229

Yeari, M. (2017) The role of working memory in inference generation during reading comprehension: retention, (re)activation, or suppression of verbal information? *Learning and Individual Differences*, 56, 1-12

YouGov (2013) *YouGov/News UK Survey Results Good University*. Available from: https://d25d2506sfb94s.cloudfront.net/cumulus_uploads/document/w4o1exrxom/YG-Archive-News-UK-results-130813-Good-University-Guide-B2B.pdf (accessed 14 November 2014)

Zechmeister, E., Chronis, A., Cull, W., D'Anna, C. & Healy, N. (1995) Growth of a functionally important lexicon, *Literature Research,* 27, 201-212.

Zhai, L. (2016) A study on Chinese EFL learners' vocabulary usage in writing, *Journal of Language Teaching and Research*, 7:4, 752-759

Zhang, X. (2013) The 'I don't know' option in the vocabulary size test, *TESOL Quarterly*, 47:4, 790-811

Zhao, N. (2009) Metacognitive strategy training and vocabulary learning of Chinese college students, *English Language Teaching*, 4, 123-129

Zhu, W. (2004) Faculty views on the importance of writing, the nature of academic writing, and teaching and responding to writing in the disciplines, *Journal of Second Language Writing,* 13, 29-48

Appendix 1: An example vocabulary test with definitions

Vocabulary Test 5

These are tests to estimate how many words you know. You will find below a list of 50 words which is part of a sample of all the words in the language. The words are arranged more or less in order of frequency, starting with common words and going down to some very unusual ones.

Procedure

1. Read through the whole list. Put a tick next to each word you know, i.e. you have seen the word before and can express at least one meaning of it.

Put a question mark next to each word that you think you know but are not sure about. (Do not mark the words you do not know.)

2. When you have been through the whole list of 50 words, go back and check the words with question marks to see whether you can change the question mark to a tick.

3. Then find the last five words you ticked (i.e. the ones that are furthest down the list). Show you know the meaning of each one by giving a synonym or definition or by using it in a sentence or drawing a diagram, if appropriate.

4. Check your explanations of the last five words in a dictionary. If more than one of the explanations is not correct, you need to work back through the list, beginning with the sixth to last word you ticked. Write the meaning of this word and check it in the dictionary. Continue this process until you have a sequence of four ticked words (which may include some of the original five you checked) that you have explained correctly.

5. Calculate your score for that 50-item test by multiplying the total number of known words by 500. Do not include the words with a question mark in your scoring. Put your score here:_____. If your score is above 15,000 words, check the list of words not likely to be known. For the words you know on this list, please show you know the meaning by giving a synonym or definition as above (item 3). You do not need to check the definition for this list. Each item represents 100 words. Add your score from the list of 'likely to be unknown' words here:_____. To get your total estimated vocabulary size, add the two scores together.

1. cotton
2. block
3. precious
4. dig

5. hostile
6. accurate
7. inhabit
8. crook
9. blockade
10. microscope
11. deign
12. marrow
13. armada
14. boomerang
15. chowder
16. earring
17. linguistics
18. radium
19. ventilate
20. asperity
21. centripetal
22. dromedary
23. ideograph
24. nuzzle
25. planking
26. welladay
27. brassie
28. huia
29. baobab
30. chomp
31. doubleheader

32. fusilier
33. interplay
34. nubile
35. repartition
36. cockup
37. saddleback
38. hairspring
39. audivision
40. dactylology
41. isomorphy
42. gaper
43. sextodecimo
44. redact
45. capsulectomy
46. volvulus
47. mancipation
48. exceptionalism
49. parasternum
50. sparrowbill

1. cotton 1. a soft, white material that grows on the seeds of a tall plant and that is used to make cloth; also, the plants on which this material grows 2. cloth that is made of cotton; also, clothing that is made of this cloth 3. yarn that is made of cotton
2. block 1. a solid piece of material (such as rock or wood) that has flat sides and is usually square or rectangular in shape 2. an area of land surrounded by four streets in a city 3. the length of one city block

3. precious 1. of great value or high price <*precious* jewels> 2. highly esteemed or cherished <a *precious* friend> 3. excessively refined, affected <*precious* manners> 4. great, thoroughgoing <a *precious* scoundrel>
4. dig 1. to move soil, sand, snow, etc., in order to create a hole 2. to form (a hole, tunnel, etc.) by removing soil, sand, snow, etc. 3. to uncover (something that is underground) by moving earth, soil, sand, etc.
5. hostile 1. of or relating to an enemy <*hostile* fire> 2. marked by malevolence, having or showing unfriendly feelings <a *hostile* act> 3. openly opposed or resisting <a *hostile* critic> <*hostile* to new ideas> 4. not hospitable <plants growing in a *hostile* environment> (2) **:** having an intimidating, antagonistic, or offensive nature <a *hostile* workplace> 5. of or relating to the opposing party in a legal controversy <a *hostile* witness> 6. adverse to the interests of a property owner or corporation management <a *hostile* takeover>
6. accurate 1. free from error especially as the result of care <an *accurate* diagnosis> 2. conforming exactly to truth or to a standard **:** EXACT <providing *accurate* colour> 3. able to give an accurate result <an *accurate* gauge>
7. inhabit 1. to occupy as a place of settled residence or habitat **:** live in <*inhabit* a small house> 2. to be present in or occupy in any manner or form <the human beings who *inhabit* this tale – Al Newman>
8. crook 1. an implement having a bent or hooked form 2. a part of something that is hook-shaped, curved, or bent <the *crook* of an umbrella handle> 3. a person who engages in fraudulent or criminal practices
9. blockade 1. the isolation by a warring nation of an enemy area (as a harbour) by troops or warships to prevent passage of persons or supplies; *broadly* **:** a restrictive measure designed to obstruct the commerce

and communications of an unfriendly nation 2. something that blocks 3. interruption of normal physiological function (as transmission of nerve impulses) of a cellular receptor, tissue, or organ; *also* **:** inhibition of a physiologically active substance (as a hormone)

10. microscope 1. an optical instrument consisting of a lens or combination of lenses for making enlarged images of minute objects; especially : compound microscope 2. a non-optical instrument (as one using radiations other than light or using vibrations) for making enlarged images of minute objects <an acoustic *microscope*>
11. deign 1. to condescend reluctantly and with a strong sense of the affront to one's superiority that is involved 2. stoop <would not even deign to talk to him> 3. to condescend to give or offer
12. marrow 1. BONE MARROW 2. the substance of the spinal cord 3. the choicest of food 4. the seat of animal vigour 5. the inmost, best, or essential part **:** CORE <personal liberty is the *marrow* of the American tradition – Clinton Rossiter> 6. A vegetable
13. armada 1. a fleet of warships 2. a large force or group usually of moving things <an *armada* of fishing boats>
14. boomerang 1. a bent or angular throwing club typically flat on one side and rounded on the other so that it soars or curves in flight; *especially* **:** one designed to return near the thrower 2. an act or utterance that backfires on its originator
15. chowder a soup or stew of seafood (as clams or fish) usually made with milk or tomatoes, salt pork, onions, and other vegetables (as potatoes); *also* **:** a soup resembling chowder <corn *chowder*>
16. earring 1. a piece of jewellery that is worn on the ear and especially on the earlobe 2. an ornament for the ear and especially the earlobe

17. linguistics the study of human speech including the units, nature, structure, and modification of language
18. radium an intensely radioactive brilliant white metallic element that resembles barium chemically, occurs in combination in minute quantities in minerals (as pitchblende or carnotite), emits alpha particles and gamma rays to form radon, and is used chiefly in luminous materials and in the treatment of cancer
19. ventilate 1. to examine, discuss, or investigate freely and openly : expose <*ventilating* family quarrels in public> 2. to make public : utter <*ventilated* their objections at length> 3. *archaic* **:** to free from chaff by winnowing 4. to expose to air and especially to a current of fresh air for purifying, curing, or refreshing <*ventilate* stored grain>; *also* : oxygenate, aerate <*ventilate* blood in the lungs> 5. to subject the lungs to ventilation <artificially *ventilate* a patient in respiratory distress> 6. *of a current of air* **:** to pass or circulate through so as to freshen 7. to cause fresh air to circulate through (as a room or mine) 8. to provide an opening in (a burning structure) to permit escape of smoke and heat
20. asperity 1. rigour, severity 2. roughness of surface : unevenness; *also* : a tiny projection from a surface or roughness of sound 3. roughness of manner or of temper: harshness
21. centripetal 1. moving toward a centre : acting in a direction toward a centre 2. afferent 3. tending toward centralisation : unifying
22. dromedary a camel of western Asia and northern Africa that has one hump on its back
23. ideograph 1. An ideogram – a picture or symbol used in a system of writing to represent a thing or an idea but not a particular word or phrase for

it; *especially* **:** one that represents not the object pictured but some thing or idea that the object pictured is supposed to suggest

24. nuzzle 1. to work with or as if with the nose; *especially* **:** to root, rub, or snuff something 2. to lie close or snug 3. to root, rub, or touch with or as if with the nose 4. to rub or push gently (as one's face) against something
25. planking 1. planks collectively, as in a floor 2. the act of laying or covering with planks. 3. the art of planking is to lay horizontally across any object or the ground with their arms by their sides, aiming to occur in daring situations or a brotherly display of core-strength
26. welladay used to express sorrow or lamentation (archaic)
27. brassie a wooden golf club soled with brass or other metal and used esp. for long low shots from a favourable lie on the fairway
28. huia an apparently extinct, crowlike bird, *Heteralocha acutirostris*, of New Zealand, noted for the completely different bill shapes of the male and female
29. baobab a broad-trunked tropical tree (*Adansonia digitata*) of the silk-cotton family that is native to Africa and has an edible acidic fruit resembling a gourd and bark used in making paper, cloth, and rope; *also* **:** any of several related trees chiefly of Madagascar and Australia
30. chomp to chew or bite on something
31. doubleheader 1. two games (especially baseball games) that are played one after the other on the same day 2. a train pulled by two locomotives
32. fusilier 1. a soldier armed with a fusil 2. a member of a British regiment formerly armed with fusils
33. interplay 1. reciprocal relationship, action, or influence: the *interplay* of plot and character 2. to exert influence on each other

34. nubile 1. of marriageable condition or age 2. sexually attractive – used of a young woman 3. sexually mature; *especially* **:** of marriageable condition or age – used of young women (medical definition)
35. repartition a second or additional dividing or distribution
36. cockup a situation that is complicated, unpleasant, or difficult to deal with because of someone's mistake
37. saddleback 1. Any of various birds, fishes, and other animals having saddle-shaped markings on the back 2. two supporting members (as of timber) placed in the form of an inverted V 3. a hill or ridge having a concave outline at the top 4. a size of wrapping paper measuring 45 by 36 inches
38. hairspring a slender spiralled recoil spring that regulates the motion of the balance wheel of a timepiece
39. audivision the transmission or reception of a succession of images with accompanying sounds over wire or wireless circuits by electrical means
40. dactylology the technique of communicating by signs made with the fingers, especially in the manual alphabets used by the deaf.
41. isomorphy (biology) similarity or identity of form or shape or structure
42. gaper 1. one that gapes 2. any of several large sluggish burrowing clams (families Myacidae and Mactridae) including several used for food
43. sextodecimo 1. the page size of a book composed of printer's sheets folded into 16 leaves or 32 pages 2. a book composed of sextodecimo pages
44. redact 1. to put in writing: FRAME 2. to select or adapt (as by obscuring or removing sensitive information) for publication or release; *broadly***:** EDIT 3. to obscure or remove (text) from a document prior to publication or release
45. capsulectomy 1. excision of a capsule, especially a joint capsule or lens capsule 2. the surgical excision of a capsule, usually the capsule of a joint or of the lens of the eye 3. As commonly used, the surgical removal of a

breast implant (or any capsular implant) and the reactive fibrous tissue that develops around it, which may contract and put pressure on the implant

46. Volvulus twisting of the intestine
47. mancipation 1. the act of enslaving; involuntary servitude **:** SLAVERY 2. an early form of ceremonial conveyance under the jus civile involving the balance scales, bronze money, a balance holder, and five citizens as witnesses in which persons and property (as Italic lands, slaves, beasts of burden, rural praedial servitudes, children under potestas, and various women) subject to the ceremony were transferred by one Roman citizen into the power and control of another
48. exceptionalism 1. the condition of being exceptional or unique 2. the theory or belief that something, especially a nation, does not conform to a pattern or norm 3. (Government, Politics & Diplomacy) an attitude to other countries, cultures, etc. based on the idea of being quite distinct from, and often superior to, them in vital ways
49. parasternum a bony framework formed by the abdominal ribs in various reptiles
50. sparrowbill 1. a small nail 2. a castiron shoe nail 3. a sparable

Words in the *Webster's Third Dictionary* sample that are not likely to be known

anamnestic	cerveliere	ominate
banausic	challoth	oxpecker
beechdrops	chinaball	palar
brachypterous	circumforaneous	panela
clinicopathologic	cladocera	parageosyncline
corium	clowder	patelline
cosmolline	coblenzion	pedalium
decoupage	compacta	perimetrium
didapper	contrist	phaeomelanin

disseminule
enatiomorphism
erythrism
fluerics
greylag
hypermorph
jerkwater
kef
mitogenic
myrmecophagous
octroy
ouabain
pes
pikake
psychrophile
rhus
rubeola
sesamoiditis
smectic
supersedeas
tepa
tragacanthin
tuchahoe
yogh
advertonal
agoura
allopelagic
amphicyrtic
analogon
anthocoridae
arguendo
atacamite
authigenic
baikerinite
barsom
corynomorpha
cuticula
cytozyme
delomorphous
desmolysis
dissilent
elasmosaur
epornitic
espinillo
eumeces
extramitochondrial
fassaile
fluate
follyer
forepost
frontad
garefowl
genistin
glia
goldtit
guignolet
halse
hominal
hursinghar
hydrocalumite
hydrorrhea
hypostase
incus
intocostrin
jones
kantiara
kiaat
koombar
leonite
lychnoscope
phenoxybenzamine
phoniatric
pilum
pisote
poachwood
polyacrylate
poppywort
porta
propria
proxenus
punctum
rackabones
rohu
salinella
saprolegniales
scholarch
sciurus
scudder
sella
shallon
skinball
slipe
snakefish
songman
spikebill
stremmatograph
stylogonidium
symballophone
tanonovicular
theow
tholos
tinsey
tournette
troutbird
twinspur

baya
belcher
bigarade
bostryx
boxwork
breastbeam
bungersome
burdenman
cannilan
castorile
catfit
brachium
cascabel
chicalote
crampfish
deepgoing
dioptrics
dispersoid
einkorn
endexine
erythropoiesis
fibroma
glutarladehyde
hemal
hexamethonium
karyolymph
ketoglutarate
lamin
linoleate
lummox
mooneye
notchback
osmol
playa
rachis
malma
manroot
melilite
meridienne
metarhodopsin
monochloramine
motacilla
nagaika
neral
nosean
okenite
cheilion
chrysogen
circannual
civilite
cloop
coaxation
cohitre
conducta
copellidine
cotyloid
curuba
cynegetic
deadheart
dernivol
diphenoxylate
domnei
dungan
eminento
ervil
etchant
eusynchite
extrophy
fallowchat
ferritungstite
typicon
vashegyite
vervelle
visceripericardial
whiggamore
wobbulator
xurel
yeara
animalist
baldashin
bargeboard
palilolgy
pectoralis
pepperbush
petrolene
phano
pheoporphyrin
phyllade
pinnaglobin
pleonaste
poikilosmotic
proeutectoid
protenoid
pseudowavellite
quiebracha
roding
rotenoid
sampleite
scallom
sciara
scoldenore
sealflower
senam
slipband
slurb

rhinencephalon
ruddle
scouse
snowberry
succinate
synaptosome
theca
triazine
uranic
windage
acidophilia
agammaglobulinemea
aliesterase
amgarn
amphiplatyan
anatta
ascidiozoa
aumakua
avahi
axodendrite
banstickle
bauno
beata
biunial
boughpot
breakax
brookweed
buplever
cannabinol
casualism
centumvir
cessionaire
foedaratus
fonio
frogbit
galenobismutile
gemauve
geta
groundplot
helmetpod
histocyte
howardile
huskanaw
hydrogarnet
insilicate
irreflection
jumpseed
kharmadharaya
koali
kylix
linaloe
lowa
madge
medino
merosome
metasilicate
molave
monorhinic
mouthpipe
naze
norleucine
octanoyl
orseille
ozokerite
softa
sqauloidea
strongback
sucupira
synusia
terna
thermopolymization
tinaja
tiqueur
trochantin
twatchel
tylostyle
vanaprastha
velutino
violescent
weightage
whipster
xanthydrol
xyloketose
yeatmanite

Appendix 2: Personal information questions (pre-testing)

1. Which faculty do you study in? If you are a Joint Honours student you may tick more than one.

Arts, Design and Technology	
Business, Computing and Law	
Education, Health and Sciences	
University of Derby Buxton	

2. What is the subject you are studying? Please write in the box below.

3. Which stage are you at? Please tick one.

Stage 1 (level 4)	
Stage 2 (level 5)	
Stage 3 (level 6)	

4. What is your expected degree classification? Please tick only one.

70-100% First Class	
60-69% 2:1	
50-59% 2:2	
40-49% Third	

5. What is your sex?

Male	
Female	

6. Is English your native (mother tongue) or first language?

Yes	
No	

7. If no to Q6, what is your first language? Please write in the box below.

Language questions

Below are a few questions regarding your vocabulary.

8. Do you find understanding words difficult at University?

Yes	
No	

9. How do you deal with unknown words at university?

Tick as many that apply.

My lecturers explain difficult words in the sessions.	
My lecturers use the module glossary provided on blackboard.	
I ask my lecturers to explain the words I do not understand.	
When I encounter a word I do not know, I ask a friend what it means.	

When I encounter a word I do not know, I look it up in a dictionary or online.	
I keep a list of all the unknown words I encounter.	
I do not do anything when I encounter words I do not know.	

10. Where do you encounter unknown words the most?

Tick as many that apply.

Lectures	
Seminars	
Tutorials	
Journal articles	
Textbooks	
Newspapers	

Other (Please provide an example in the space below.)	

Appendix 3: Vocabulary record sheet

Vocabulary record sheet

Name (Initals): CW

The purpose of this sheet is to identify any unknown words that you encounter. Can you please record the following every time you come across a word you do not know?

After one week, please return this to Jennifer Marshall in E112. You can contact me at J.Marshall@derby.ac.uk if you have any questions or wish to withdraw.

Word: prevaricate	Where: Journal article
Epidemiology	~~text~~ Book
unambiguous	lecture
Grita	lecture
Laissez Faire	leiture
subesequent	Internet research
aspersion	Youtube
Philological	Micro soft word spelling correction
Philology	Google, looking up what the Philological was.

Appendix 4: Single Honours programmes at the University of Derby 2014-2015

Accounting and Finance BA (Hons)
American Studies BA (Hons)
Animation BA (Hons)
Animation MDes
Applied Criminology BSc (Hons)
Applied Social Work BA (Hons)
Architectural Technology and Practice BSc (Hons)
Architectural Venue Design BA (Hons)
Biology BSc (Hons)
Business Management (CMI accredited) BA (Hons)
Business Studies BA (Hons)
Child and Youth Studies BA (Hons)
Civil Engineering BSc (Hons)
Commercial Photography BA (Hons)
Computer Forensic Investigation BSc (Hons)
Computer Games Modelling and Animation BA (Hons)
Computer Games Programming BSc (Hons)

Construction Management and Property Development BSc (Hons)
Computer Networks and Security BSc (Hons)
Computer Science BSc (Hons)
Creative Expressive Therapies BA (Hons)
Creative Writing BA (Hons)
Dance BA (Hons)
Diagnostic Radiography BSc (Hons)
Early Childhood Studies BA (Hons)
Education BEd (Hons)
Education Studies, BA (Hons)
Electrical and Electronic Engineering BEng (Hons)
English BA (Hons)
Fashion Studies BA (Hons)
Film Production BA (Hons)
Fine Art BA (Hons)
Forensic Science BSc (Hons)
Forensic Science with Criminology BSc (Hons)
Geography BSc (Hons)
Geology BSc (Hons)
Graphic Design BA (Hons)

Graphic Design MDes
Health and Social Care BA (Hons)
History BA (Hons)
Hospitality Management (International) BA (Hons)
Human Biology BSc (Hons)
Illustration BA (Hons)
Information Technology BSc (Hons)
Information Technology Management for Business BSc (Hons)
Journalism BA (Hons)
Law - LLB (Hons)
Law - LLB (Hons) Corporate and Commercial Law
Law - LLB (Hons) Employment Law
Law - LLB (Hons) Family Law
Law - LLB (Hons) International and Comparative Law
Law - LLB (Hons) Medical Law
Law - LLB (Hons) Social and Public Law
Law - LLB (Hons) with Criminology
Nursing (Adult) BSc (Hons)
Nursing (Mental Health) BSc (Hons)
Occupational Therapy BSc (Hons)

Photography BA (Hons)
Popular Music with Music Technology BA (Hons)
Product Design BA (Hons)
Product Design BSc (Hons)
Psychology BSc (Hons)
Sociology, BA (Hons)
Specialist Community Public Health Nursing BSc (Hons)
Sport and Exercise Science BSc (Hons)
Sport and Exercise Studies BA (Hons)
Sport Management BA (Hons)
Technical Theatre BA (Hons)
Textile Design BA (Hons)
Theatre Arts BA (Hons)
Visual Effects and Post-Production BA (Hons)
Working with Young People and Communities (Youth Work or Community Development) BA (Hons)
Zoology BSc (Hons)
73 Programmes

Appendix 5: University marking criteria

	Mark descriptors	Class
70-100%	**Excellent** Outstanding; high to very high standard; a high level of critical analysis and evaluation, incisive original thinking; commendable originality; exceptionally well researched; high quality presentation; exceptional clarity of ideas; excellent coherence and logic. Trivial or very minor errors. For the highest marks (90 - 100%): an exceptional standard of work illustrating thorough and in-depth understanding, communicated with exceptional authority.	First
60-69%	**Very good** A very good standard; a very good level of critical analysis and evaluation; significant originality; well researched; a very good standard of presentation; commendable clarity of ideas; thoughtful and effective presentation; very good sense of coherence and logic; minor errors only.	Second division 1
50-59%	**Good** A good standard; a fairly good level of critical analysis and evaluation; some evidence of original thinking or originality; quite well researched; a good standard of presentation; ideas generally clear and coherent, some evidence of misunderstandings; some deficiencies in presentation.	Second division 2
40-49%	**Satisfactory** A sound standard of work; a fair level of critical analysis and evaluation; little evidence of original thinking or originality; adequately researched; a sound standard of presentation; ideas fairly clear and coherent, some significant misunderstandings and errors; some weakness in style or presentation but satisfactory overall.	Third

	Mark descriptors	Class
35-39%	**Unsatisfactory** Overall marginally unsatisfactory; some sound aspects but some of the following weaknesses are evident; inadequate critical analysis and evaluation; little evidence of originality; not well researched; standard of presentation unacceptable; ideas unclear and incoherent; some significant errors and mis-understandings. Marginal fail.	Marginal fail
21-34%	**Poor** Below the pass standard; a poor critical analysis and evaluation; virtually no evidence of originality; poorly researched; presentation unacceptable and not up to graduate standard; ideas confused and incoherent, some serious misunderstandings and errors. A clear fail, short of pass standard.	Fail
1-20%	**Very poor** Well below the pass standard, with many serious errors. Standard of presentation totally unacceptable, incoherent and may be severely under-length. No evidence of evaluation or application. A very clear fail, well short of the pass standard.	Fail
NS	**Non-submission** No work has been submitted.	Fail
Z	**Academic offence notation** Applies to proven instances of academic offence.	Fail

Appendix 6: Ethical approval

Approval Letter

Date: 18th September 2013

Name: Jennifer Marshall

Dear Jen,

Re: Request for ethical approval for study entitled

'An exploration of undergraduates' vocabulary size and academic achievement'

Thank you for submitting your application for the above mentioned study which was considered by 3 reviewers and ratified by Chairs' Action on behalf of the Social Sciences and Post Graduate Research Ethics Committee (SSPG REC) on 18th September 2013.

Your study has been **approved with recommendations**; please see the comments section of the ethics form attached for the detail of these recommendations. No additional submission will be required for this project unless you add to your methods or change them significantly.

I wish you every success with your research.

Yours Sincerely

Dr Neil Radford

Chair of the Social Studies and Post Graduate Research Ethics Committee

Phase I

Request for Ethical Approval for Individual Study / Programme of Research by University Students

Please complete this form and return it to your Independent Studies Supervisor or Co-ordinator as advised by local guidance. Feedback on your application will be via your Independent Studies Supervisor or Co-ordinator

1. Your Name:	Jennifer Marshall	**2. Programme name and code** PX3AA Professional Doctorate in Education (EdD)
3. Contact Info	**Email:** **J.Marshall@derby.ac.uk**	
4. Module name and code	8EU008 PHILOSOPHY AND PRACTICE OF RESEARCH	
5. Name of project supervisor (Director of Studies)		
6. Title or topic area of proposed study		
An exploration of undergraduates' vocabulary size and academic achievement		
7. What is the aim and objectives of your study?		

Aim: To explore the relationship between undergraduates' vocabulary and their academic achievement

Objectives:

1) To estimate the vocabulary size of monolingual and bilingual English speaking undergraduate students

2) To form an understanding of the specific words students do not understand and whether these words are general (everyday English), academic or words found in more than 2 disciplines (e.g. ambiguous, see Academic Word List Coxhead (2000)) or subject

specific (e.g. Legal English, Business English, English for Engineering or words found mainly in one discipline)

3) To investigate whether there is a correlation between the level of unknown words and a student's average marks on their modules and final degree classification

8. Brief review of relevant literature and rationale for study (attach on a separate sheet references of approximately 6 key publications, it is not necessary to attach copies of the publications)

Vocabulary knowledge plays an important part of academic literacy. Academic literacy, defined by Stacey & Granville (2009, p. 327) is 'being able to examine, refute, agree with, unpick or apply - to interact with the ideas of others and to do so in disciplinarily approved ways'. Despite having the required grades to enter higher education, do students have a sufficient vocabulary size to be academically literate?

What size of vocabulary do undergraduates have and need in order to access their discipline? According to Schmitt (2005), a native English speaker with a relatively large vocabulary has around 40,000 words in his or her repertoire while 10,000 words is considered large for a non-native speaker entering university. Nation (2006) uncovered that native speakers need a vocabulary size of around 8,000 to 9,000 words (base word plus its derivatives, e.g. wide, widen, widely) to read the newspaper. A university graduate will have a vocabulary size of around 20,000 words (Goulden, Nation and Read,1990).

However, there is very little research on the vocabulary size of British university students and almost all previous studies have been carried out on American students. In fact, the academic literature is so scarce that to my knowledge there has only been one other study conducted in this area (see Treffers-Daller & Milton, 2013) in the UK. This study examined the vocabulary size of students at three British Universities; City University, Swansea University and the University of the West of England, Bristol.

Lastly, it has been reported by the Complete University Guide (2013) that the percentage of graduates achieving a first or upper second class honours degree at Derby is 54.7% compared with 75.5% at Leicester; 63.0% at Nottingham Trent; 77.5% at Sheffield; and 56.3% at Staffordshire. Is there a link with vocabulary size and academic achievement? If vocabulary plays an important role in students' learning and subsequent achievement, then strategies need to be developed in order to facilitate its acquisition.

9. Outline of study design and methods

Both quantitative and qualitative methods will be employed, although this research will primarily be quantitative in nature as most of the data will measurable.

1) Vocabulary sizes will be measured using a test devised by Goulden, Nation and Read (1990). The test is based on Webster's Third New International Dictionary (1961) + updates and includes a representative sample of the 25,000 most frequent words in Thorndike and Lorge's (1944) frequency lists. The test contains 250 words including five sub-tests where10 words are selected from each of the first 5,000 word bands in this list. It is estimated that it will take students around 25 minutes to complete a sub-test. Students will also be asked about the grades they have achieved and their anticipated final degree classification.

The same test was used by Treffers-Daller & Milton (2013) in their research at City University, Swansea University and the University of the West of England, Bristol. Their published data will serve as a benchmark for interpreting results at Derby. The sample will be taken across EHS, ADT and BCL including each school within the faculty thus ensuring the biggest and best possible representative sample. Random sampling techniques will be employed to ensure that the research is free from bias. The sample will consist of first, second and third year students.

2) Diaries – students will also be asked to keep a diary for a semester to record any unknown words that they encounter. Students will be asked to record: the name of the word; where they heard or read it; what they think it might mean. This will allow the researcher to determine the nature of unknown vocabulary words (e.g. academic, general or subject specific) and where they occur the most (for example, in lectures). By asking students to write down a definition of the word will allow the researcher to distinguish between active and passive vocabulary knowledge.

Rationale

A vocabulary test was chosen as this is the most widely accepted way of measuring whether an individual knows a word. A dictionary sampling method is being used as this is the most common way of estimating vocabulary sizes (D'Anna, Zechmeister & Hall, 1991). This, however, is not without its limitations. Limitations include what does it mean to 'know a word' and differences between being able to recall versus recognise vocabulary. Also there are some issues in vocabulary research as what is counted as a word.

Student diaries were chosen in order to gain a more in-depth insight into the type of words students don't know. There is a lack of data on vocabulary from this methodological approach and the field would benefit from more qualitative research in this area.

Sample

The vocabulary testing will be undertaken with students from the faculties of BCL, ADT and EHS and at each stage of study. Because of issues of access and the complexity of the student body due to the recent merger between Leek College and Buxton, Buxton will not be considered in this project. A convenience sampling technique will be used. According to Bryman (2008, p. 183), 'A convenience sample is one that is simply available to the researcher by virtue of accessibility'. The researcher will purposively choose the students

from certain degree programmes in order to maximize the variety of the sample in an attempt to make the overall sample as representative as possible.

The largest sample size that is feasible will be sought in order to generalise the findings. I am currently waiting for statistics on the number of students in each faculty so I can ensure that the sample is representative of each faculty and that the sampling error is as small as possible.

Analysis of test

To calculate vocabulary size, the total number of words correctly explained in the first, frequency-based, part of the test has been multiplied by 500, and the number of words correctly explained in the second part of the test is multiplied by 100. The sum of these two figures provides the estimate of size.

Analysis of Diary

Words will categorise on the basis of Coxhead's (2000) word list into academic, subject-specific or everyday language. Frequency will be calculated on the basis of the number of word types in each category to determine the nature of the unknown words.

Pilot Study

A pilot study will be undertaken to evaluate the amount of time students will need to take the test . Any problems with the tests will be identified. I will also pilot the instructions and diary on a small number of students to again detect any issues that may arise which can then be rectified before the study is launched.

Governance

This research is part of a recent RTLF award 2013/2014 so has the approval of LEI. Permission will be sought by senior management in order to invite students from other faculties to participate in the study.

Time Scale

Piloting of test and diary

September 2013

Students record diaries (to be given to students)

September to December 2013

January to May 2014

Testing of students vocabulary size

June to August 2014

Analyse Data

September onwards

Write up findings

10. Research Ethics

PROPOSALS INVOLVING HUMAN PARTICIPANTS MUST ADDRESS QUESTIONS 10 - 14.

Does the proposed study entail ethical considerations Yes **/** ~~No~~ **(please circle as appropriate)**

If 'No' provide a statement below to support this position.

If 'Yes' move on to Question 11.

11. Ethical Considerations: Please indicate how you intend to address each of the following in your study. Points a - i relate particularly to projects involving human participants.

Guidance to completing this section of the form is provided at the end of the document.

Consent

Informed consent will be sought for both the vocabulary test and diaries. Please see attached. Background information will be given to participants to the nature of the research to help inform their consent. This research will not employ any covert data collection methods which could affect consent.

a. Deception

This research will not involve deception in any form.

b. Debriefing

Debriefing letters (see attached) will be given to participants taking the vocabulary test in order to thank them and give them contact information and dates if they wish to withdraw from the study.

c. Withdrawal from the investigation

Participation is voluntary and free from coercion. Participants can withdraw from the study at any time before the data is analysed (end of August 2014). They will be given written notice of this in the consent forms and debriefing letters. Contact details will be given so they can notify the researcher if they wish to withdraw after the data has been given.

d. Confidentiality

Participants' anonymity will be protected. The research will comply with the Data Protection Act (1998). The raw data from the vocabulary tests will be securely stored where only the researcher has access to it. The participants' data will not be kept longer than necessary and on completion of the project, be destroyed. The consent forms will be stored separately from the vocabulary tests. The data will only be used for this project.

e. Protection of participants

No physical or psychological harm will result from the research and focus group participants will be handled with respect and dignity.

f. Observation research [complete if applicable] N/A

g. Giving advice

Advice will not be given.

h. Research undertaken in public places [complete if applicable]

Research will not take place in public places.

i. Data protection

Please see above.

j. Animal Rights [complete if applicable] N/A

k. Environmental protection [complete if applicable] N/A

13. Are payments or rewards/incentives going to be made to the participants? If so, please give details below.

No payments or rewards/incentives are going to be made to participants.

14. What study materials will you use? (Please give full details here of validated scales, bespoke questionnaires, interview schedules, focus group schedules etc and attach all materials to the application)

The vocabulary test

The students will take a fifty item test taken from Goulden *et al* (1990) drawn from Webster's Third New International Dictionary (1961) and selected to be a representative sample of the 25,000 most frequent words on Thorndike and Lorge's (1944) frequency lists. There are five sub-tests to this test and 10 words are selected from each of the first 5,000 word bands in this list.

This test is a self-reporting checklist test where, in Goulden et al's original version, the testees are asked to read through the lists and mark the words they know the meaning of. Once they have done this they are asked to go back through the list in frequency order and for the last five words they tick (the least frequent therefore) they are asked to provide a synonym or explanation which should be checked in a dictionary. If they find an explanation is not correct then this word is discounted from the size calculation and the learners check the next least frequent word they marked. The correctness of every word marked is not checked therefore and self-reporting of this kind, as noted by D'Anna et al (1991) can result in overestimation of knowledge. To try to avoid the overestimation in this study, the testees will be asked to provide a synonym, explanation or illustration of use for every word they identify in the test. These responses will then be checked by the researcher for correctness. The test used is provided in the Appendix.

15. What resources will you require? (e.g. psychometric scales, equipment, such as video camera, specialised software, access to specialist facilities, such as microbiological containment laboratories).

Non electronic diaries. Paper-based tests. No specialised equipment will be needed.

16. Have / Do you intend to request ethical approval from any other body/organisation ? ~~Yes~~ / No (please circle as appropriate)

If 'Yes' – please give details below.

17. The information supplied is, to the best of my knowledge and belief, accurate. I clearly understand my obligations and the rights of the participants. I agree to act at all times in accordance with University of Derby Code of Practice on Research Ethics
http://www.derby.ac.uk/research/ethics/policy-document

Date of submission……13/08/2013……..

Signature of applicant……Jennifer Marshall……

Signature of project supervisor (Director of Studies)
……

For Committee Use *Reference Number (Subject area initials/year/ID number)……..*

Date received 03/09/13 Date approved 18/09/13 Signed Dr. Neil Radford

Comments

Approved with recommendations: Approval is given to begin the data collection phrase of the work but the candidate is asked to consider the following recommendations (no second submission is needed for the work detailed in this application, but if additional elements or research tools are to be used at a later date a new ethics application will be required):

- Consent letter – give more detail to participants about the potential benefits of taking part in the study.
- Consent letter – give more detail to participants about what they will be asked to do, e.g. how much time will the test and the diary take?
- Review sample to ensure ethicality in relation to sourcing undergraduate grades. Also, sampling year 1,2,3 students collectively may be seen as ethically challenging and it may not recognise growth in vocabulary over the time of the students' programme of study.

- The administration of the test needs to be explained to ensure ethical adherence – location, time, etc

- Selection of research participants for completion of diaries is not explained for ethical scrutiny

- Check regulations on time periods of holding research data before it can be destroyed

- There is no discussion of focus group activity in your research methods section which describes research participants undertaken vocabulary tests, and completion of diaries. Ethically, you need to explain how and why focus groups form a part of your research design.

- Ethically, you will need to ensure that research participants have enough space on the form to complete each of instructions you are asking of them.

- Discuss the potential of this project to supply original contribution (needed for the doctoral award) with your supervisor – the current project seems appropriate for an RTLF but seems limited in scope for a doctoral study.

Date: 11 September 2015

Name: Jennifer Marshall

Dear Jennifer,

Re: Request for ethical approval for study entitled 'An exploration of undergraduates' vocabulary size and academic achievement'

Thank you for submitting your application for the above study which was considered by 3 reviewers on behalf of the College of Education Research Ethics Committee (CEREC) by Chair's Action on 11 September 2015.

The reviewers commented that the application was very well-written and is conceptually well positioned as part of a sustained ethical and valid approach. I am pleased to inform you that your study has been **approved with recommendations**; please see below. No additional submission will be required for this project, unless you change the methods detailed in this submission significantly. Additional phases of your research will require further ethical applications.

Recommendation:

- Be sure to use your work address for correspondence with participants (Section 3).
- Looking at Appendix 4: There was a concern that students may think that this research may influence their grades. Perhaps including a line in this announcement making clear that the research project will not in any way effect the grades of the assignments used may clarify that point with participants.
- You may want to consider accessing work which has been previously turned in through the software 'TurnItIn' to ensure that all work which has been analysed has not been plagiarised.

I wish you every success with your study.

Yours Sincerely,

Michelle Appleby

(Acting) Vice-Chair of the College of Education Research Ethics Committee

Phase II

Ethics approval form

Request for Ethical Approval for Individual Study / Programme of Research by University Students

Please complete this form and return it to your Independent Studies Supervisor or Co-ordinator as advised by local guidance. Feedback on your application will be via your Independent Studies Supervisor or Co-ordinator

1. Your Name:	Jennifer Marshall	**2. Programme name and code** PX3AA Professional Doctorate in Education (EdD)
3. Contact Info	**Email: J.Marshall@derby.ac.uk**	
4. Module name and code	8EU008 PHILOSOPHY AND PRACTICE OF RESEARCH	
5. Name of project supervisor (Director of Studies)		
6. Title or topic area of proposed study		
An exploration of undergraduates' vocabulary size and academic achievement		
7. What is the aim and objectives of your study?		
Aim: To explore the relationship between undergraduates' vocabulary and their academic achievement Objectives: 1) To estimate the level of vocabulary richness (productive knowledge of words) in written assignments of monolingual and bilingual English speaking undergraduate students by using word frequency profiling 2) To explore whether undergraduate students experience vocabulary growth by measuring their academic word usage from year one to year 3		

3) To investigate whether there is a correlation between the percentage of academic words in students' written work and the mark on the assignment and final degree classification

8. Brief review of relevant literature and rationale for study (attach on a separate sheet references of approximately 6 key publications, it is not necessary to attach copies of the publications)

This is the second phase of data collection for a doctoral research project investigating the vocabulary of undergraduates at the University of Derby. In the first phase, undergraduates' receptive vocabulary size was investigated and an outline of the procedure and findings is below.

Sample

- Undergraduate programmes (excluding Buxton campus) were randomly selected
- 11 single honours and 19 joint honours programmes were represented in the sample
- Test sample size (N) = 389 where the total undergraduate population is 13,595 (full and part-time)
- 27 participants recorded unknown words

Method

- Vocabulary sizes were estimated using a 50 item test devised by Goulden, Nation and Read (1990) based on a combination of dictionary sampling methods, word frequency counts and self-assessment
- Students kept a vocabulary record sheet for one week and recorded unknown words they encountered

Initial Findings

- Undergraduates have a much smaller vocabulary size than one would expect compared to previous research
- There is no correlation between vocabulary size and academic achievement.
- The differences in the means are significant between stages 1 and 2 but not between 2 and 3.
- Unknown words were a mixture of academic and subject-specific

Stage	Sample size	Mean	Standard Deviation
One	**130**	**10070**	**2674**
Two	**139**	**11614**	**3016**
Three	**120**	**11582**	**2944**
Overall test score	**389**	**11088**	**2965**

Degree classification	N	Mean	Standard Deviation
Other	**308**	**11047**	**3007**
1st	**71**	**11521**	**2861**

The initial research has given a snapshot of student's receptive vocabulary size but is inconclusive in terms of determining vocabulary growth, particularly from Stage 2 to 3. Similarly, Cobb & Horst (1999) found by testing the receptive vocabulary of first and second year University students (whose English was not their native language) in Hong Kong over a 6 month period, that there was no vocabulary growth in the first year group. The test used was also designed to measure general vocabulary rather than academic. Initial findings also suggest there is no link between academic achievement and the size of a students' vocabulary. One reason for both of these findings is that the tool (vocabulary test) is not accurate enough. For this reason, the second phase will be exploring productive rather than receptive vocabulary while also ascertaining the level of academic and subject-specific words that students use in their writing. It is important to define here that receptive vocabulary knowledge is being

able to understand a word in its spoken or written form and productive knowledge means to be able to use a word correctly in a written work or a speech (Pignot-Shahov, 2012).

Previous research in this area is scarce, particularly on native speakers of English, as the majority of scholarship centres on second language learners. For example, Ozturk (2015) investigated both the receptive and productive vocabulary sizes of advanced learners of English as a foreign language in an English-medium degree programme. His findings suggest that learners' receptive vocabulary did not grow significantly which corresponds to the findings in the table above. However, his longitudinal data indicated an increase by about 10% in students' productive knowledge of academic vocabulary over a three year period. This does have limitations, though, as this was measured by using Vocabulary Levels Tests (VLT) created by Nation (2001). Ozturk (2015) acknowledges that the validity of these tests have been questioned by Read (2000) and Schmitt (2010) as they do not measure whether learners can use a word in either written or spoken production but rather measures the availability for productive use (see Laufer & Nation, 1999).

Another way of ascertaining vocabulary size is to analyse students' written work. Laufer & Nation (1995, p. 399) believe that 'One of the major determinants of vocabulary used in written production is the vocabulary size of the writer…' and 'measures of lexical richness attempt to quantify the degree to which a writer is using a varied & large vocabulary'. There are several ways to measure lexical richness: lexical diversity (saying how many different words are used e.g. Malvern and Richards, 1997, 2002), lexical sophistication (saying how many advanced words are used) and lexical density (saying what the proportion of content words in the text is) (Šišková, 2012). Each one of these has limitations and it is not within the scope of this proposal to delve into each one. Nevertheless, it is now widely accepted that lexical frequency profiling (LFP) as developed by Laufer & Nation (1995) and since adapted by researchers is an effective way of measuring vocabulary size (Morris & Cobb, 2004).

Lexical frequency profiling breaks a text down into word frequency bands. The software that will be used in this research provides a profile of a text's lexical content by grouping the vocabulary items into four categories:

(1) words from the first 1000 frequency level

(2) words from the second 1000 frequency level

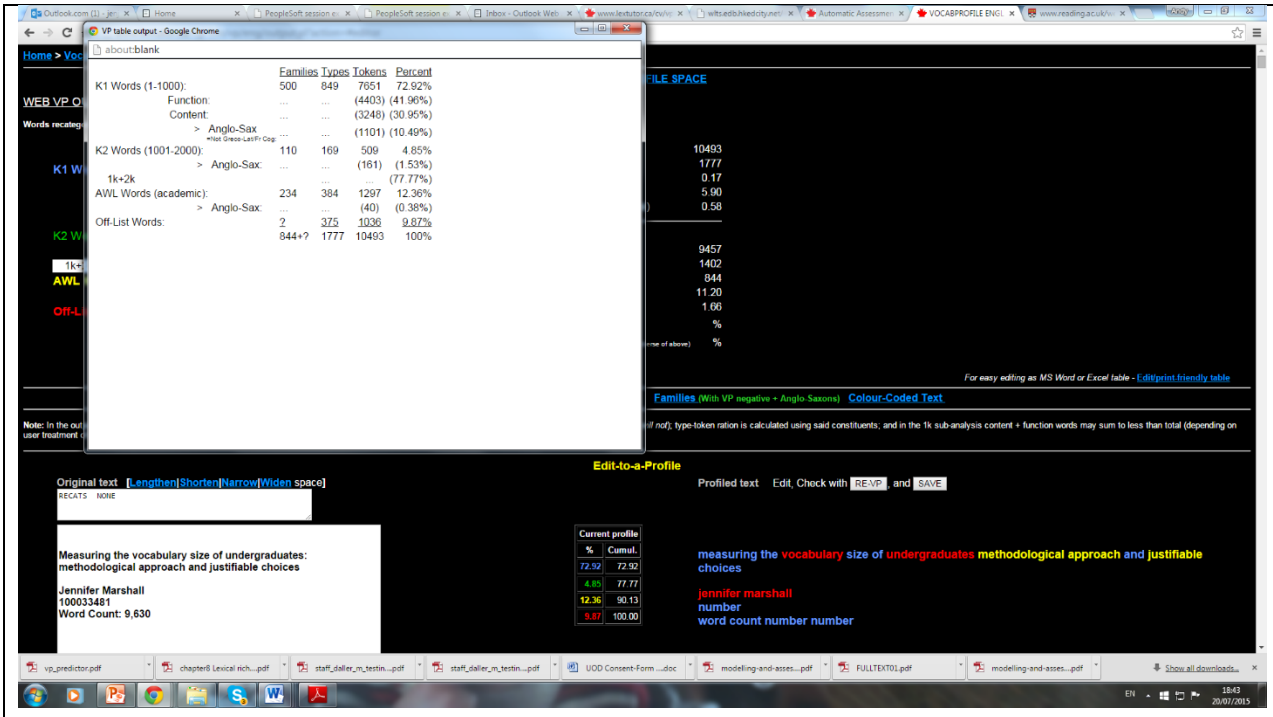
(3) words included in the Academic Word List (AWL) (Coxhead, 2000)

(4) words not included in the previous three categories, i.e. "off-list".

This will provide a picture of the percentage of words

9. Outline of study design and methods

Quantitative methods will be employed by using free software available from http://www.lextutor.ca/vp/eng/ (Cobb, 2002). Student's work (which will be anonymised) is uploaded onto the sight by cutting and pasting into a text box. The results are given in a table and look like:



The example analysis above was taken from my EdD assignment Module Code: 8EU500Philosophy and Practice of Research Methodology. Texts will be analysed by looking at:

1) what % of the vocabulary consists of academic words

2) what % of the vocabulary is in the K1 word band

3) what % of the vocabulary is in the K2 word band

4) the mark for that particular piece of work

I would also like to see what % of students' work contains subject specific vocabulary. An Education Studies glossary was compiled two years ago and this will be used like the AWL. However, I cannot upload the list in order to compare each individual text in the vocabprofiler. Instead, I need to use text_lex_compare for this. I will need to assemble the students' work as a corpus then upload it as file 1 then my glossary as file 2. This will give me collective percentage rather than individual.

Sample

Purposive sampling methods will be employed. The sample will consist of third year BA (Hons) Education Studies students as I have access to their work. I aim to have around 70 participants and analyse one of their assignments from the first, second and third year giving a total of 210 pieces of written work. The modules selected will be from the psychology and sociology strand of the programme to minimise variety in terms of lexical content and structure. Therefore the following modules have been identified from the psychology strand:

Stage 1: Psychology and Education Understanding Teaching and Learning

Stage 2: Individual Differences and Learning

Stage 3: Learning and Motivation

The following modules have been identified from the sociology strand:

Stage 1: Education for All? Exploring Inequality

Stage 2: The Sociology of Education

Stage 3: Contemporary Debates

Individual students' work will be analysed in either the sociology or psychology strand but not both. These two strands were chosen because they are core in the first and second years and attract good numbers of students in the third year. Moreover, there is some benchmarking data on lexical frequency across disciplines (please see table 2 below). The data below suggests that around 14% of words in texts found in the disciplines of sociology and social psychology are from the academic word list (AWL) while texts in medicine and zoology only contain around 7% due to proportion of more specialised terminology (Cobb & Horst, 2004).

Table 2: Lexical frequency profiles across disciplines (coverage percentages).

Brown segment	Discipline	No. of words	1000	2000	1000 + 2000	AWL	1K + 2K + AWL
J32	Linguistics	2031	73.51	8.37	81.88	12.60	94.48
J29	Sociology	2084	74.23	4.75	78.98	13.44	92.42
J26	History	2036	69.3	5.7	75.00	14.49	89.49
J25	Social Psychology	2059	73.63	3.11	76.74	14.38	91.12
J22	Development	2023	76.42	4.55	80.97	12.26	93.23
J12	Medicine (anatomy)	2024	71.05	3.80	74.85	6.72	81.57
J11	Zoology	2026	75.12	6.17	81.29	7.31	88.60
M			73.32	5.21	78.53	11.60	90.13
SD			2.42	1.74	3.01	3.24	4.30

From 'Is there room for an AWL in French?' <HERE> http://www.lextutor.ca/cv/awl_F.htm

The data will be catalogued using an excel spreadsheet which can then be uploaded on to SPSS software for further analysis.

It is important to note for ethical purposes that the data I will upload onto the software is not stored (see appendix 1).

10. Research Ethics

PROPOSALS INVOLVING HUMAN PARTICIPANTS MUST ADDRESS QUESTIONS 10 - 14.

Does the proposed study entail ethical considerations **Yes** **/** ~~**No**~~ **(please circle as appropriate)**

If 'No' provide a statement below to support this position.

If 'Yes' move on to Question 11.

11. Ethical Considerations: Please indicate how you intend to address each of the following in your study. Points a - i relate particularly to projects involving human participants.

Guidance to completing this section of the form is provided at the end of the document.

Consent

Informed consent has been sought and granted from the Deputy Head of Department. Please see appendix 2 and 3. Background information will be given to participants about the research to help inform their consent (see appendix 4). Students will be notified about the research via Blackboard and be given the option NOT to participate (see draft announcement in appendix 4). I will also speak to the stage 3 students at induction about the project. If therefore, I am not contacted by the student that they do not wish to participate, then it is assumed they have given consent to take part. The 'opt-out' approach has been taken in order to maximise the sample size as this is a quantitative study as 'opt-in' approaches tend to have lower response rates.

This research will not employ any covert data collection methods which could affect consent.

a. Deception

This research will not involve deception in any form.

b. Debriefing

Debriefing is not necessary as there is no direct involvement on the part of the participants. The data will be analysed using software and therefore no face to face contact is required between the participant and researcher.

c. Withdrawal from the investigation

Participation is voluntary and free from coercion. Participants can withdraw from the study at any time before the data is analysed (end of June 2016). They will be given written notice of this in the consent announcement via blackboard. Contact details will be given so they can notify the researcher if they wish to withdraw after the data has been given.

d. Confidentiality

Participants' anonymity will be protected in the following way: students' names will be removed from their work before being uploaded into the software. Each student will be given a number which will go into the title section of the software so the data produced can be traced back to the student in the event they want to withdraw. It will look like: WEB VP OUTPUT FOR FILE: Student 1. The software produces a vocabulary profile table (previously shown on page 3) which will be stored electronically along with the mark for that piece of work. This table is referred to as 'data' so is a by-product of their work and not the student's work itself. No student work will be stored and the data will be stored separately from the list of student names with their number.

e. Protection of participants

No physical or psychological harm will result from the research and focus group participants will be handled with respect and dignity.

f. Observation research [complete if applicable] N/A

g. Giving advice

Advice will not be given.

h. Research undertaken in public places [complete if applicable]

Research will not take place in public places.

i. Data protection

The research will comply with the Data Protection Act (1998). The data and list of student names will be securely stored by password protection and only the researcher has access to it. The data will only be used for this project and will be destroyed on completion of the dissertation. No personal details such as address, age, or sex will be stored at all. The researcher has taken steps (see section d) to anonymise the data so individuals cannot be identifiable.

j. Animal Rights [complete if applicable] N/A

k. Environmental protection [complete if applicable] N/A

13. Are payments or rewards/incentives going to be made to the participants? If so, please give details below.

None.

14. What study materials will you use? (Please give full details here of validated scales, bespoke questionnaires, interview schedules, focus group schedules etc and attach all materials to the application)

None.

15. What resources will you require? (e.g. psychometric scales, equipment, such as video camera, specialised software, access to specialist facilities, such as microbiological containment laboratories).

I will use free software which can be found at: http://www.lextutor.ca/

16. Have / Do you intend to request ethical approval from any other body/organisation ? No (please circle as appropriate)

If 'Yes' – please give details below.

17. The information supplied is, to the best of my knowledge and belief, accurate. I clearly understand my obligations and the rights of the participants. I agree to act at all times in accordance with University of Derby Code of Practice on Research Ethics
http://www.derby.ac.uk/research/ethics/policy-document

Date of submission: 21/07/015

Signature of applicant: Jennifer Marshall

Signature of project supervisor (Director of Studies): Neil Radford

For Committee Use *Reference Number (Subject area initials/year/ID number)……..*

Date received…… Date approved ……. Signed……

Comments

Appendix 7: Letter of gatekeeper consent

Dear Deputy Head of Department (Head of Education Studies and SEND),

You are being asked for gatekeeper consent in a research project to investigate the vocabulary size of undergraduate students at the University of Derby. I would like to analyse students' written work from all stages of the BA (Hons) Education Studies programme. I am seeking permission to approach 3rd year students in September 2015 to participate in the project.

This project will be analysing students' productive vocabulary knowledge (being able to use a word correctly in writing or speaking) and usage of academic words. I will do this by looking at word frequency in their written assignments using software (computational text analysis).

Potential risks:

There are no risks believed to be associated with this project as participation is voluntary. No individuals will be identified and data will be anonymised. Students will be given a code and data (in the form of a table) will be kept separately from the students' names, all of which will be stored electronically and password protected. Students' work will not be stored at any point during the research.

Potential benefits:

It is hoped that as a result of this study an understanding of the vocabulary needs of undergraduate students in particular will be gained. The study will also explore the links between vocabulary size and academic achievement. Understanding the role language plays in achievement is important in ensuring the academic success of all students.

Confidentiality:

No individuals are identified in this study and no identifying personal information will be published. Only the researcher will have access to the data. Any data will be stored electronically by password protection.

Consent:

Informed consent will be sought via an 'opt out' approach to ensure a large sample size which is needed for a quantitative approach. This will require minimal effort for the student as well. An announcement will put on Blackboard telling students about the research and they have the option not to participate if they wish. Students' participation in the project is completely voluntary and confidential, and they can withdraw at any time up until the data is analysed. Participation or non-participation will have no effect on their access to potential services that may result from this study.

Please sign here if you consent to granting me access to the participants on the BA (Hons) Education Studies programme.

.

_____ Date:_____

May I thank you very much for your cooperation.

Yours sincerely,

Jennifer Marshall

Reply from gatekeeper
Hi Jen

I am happy to give gatekeeper consent subject to your proposal being agreed by the Ethics Committee.

Best wishes

Ang
Ang Davey LLB MA Ed Cert Ed FHEA
Head of Education Studies & SEND
Deputy Head of Education Studies & Childhood

Appendix 8: Tokens and types from the NAWL and the frequency (*f*) of *curriculum*, *classroom*, *found* and *impact* for all stages

Table X: Six selected cases, three assignments chosen for each, the education topics, tokens and types from the NAWL and the frequency (*f*) of *curriculum*, *classroom*, *found* and *impact* for Stage 1

Student code	Text 1	Text 2	Text 3
122 Topic Tokens/Types (*f*)	international 37/18 *curriculum* 18, *classroom* 1	study skills 25/16 *classroom* 1	curriculum 12/12 *found* 1, *impact* 1
124 Topic Tokens/Types (*f*)	study skills 13/10	psychology 111/60 *classroom* 8	curriculum 27/17 *curriculum* 5, *impact* 1
143 Topic Tokens/Types (*f*)	sociology 56/34 *curriculum* 5, *classroom* 1, *found* 1	international 41/22 *curriculum* 2, *impact* 1	lifelong learning 34/22 *impact* 1

107 Topic Tokens/Types (*f*)	sociology 75/42 *classroom* 2, *curriculum* 2, *impact* 2	international 62/38 *classroom* 2, *curriculum* 4, *found* 2, *impact* 1	lifelong learning 65/36 *curriculum* 4, *impact* 2
121 Topic Tokens/Types (*f*)	sociology 48/28 *found* 9	international 67/29 *curriculum* 26, *impact* 3, *classroom* 1, *found* 1	lifelong learning 53/31 *impact* 4, *curriculum* 2,
135 Topic Tokens/Types (*f*)	sociology 95/46 *classroom* 7, *found* 4, *curriculum* 3, *impact* 2	international 69/22 *curriculum* 32, *aspect* 1	lifelong learning 47/30 *found* 2, *impact* 2

Table Y: Six selected cases, three assignments chosen for each, the education topics, tokens and types from the NAWL and the frequency (*f*) of *curriculum*, *classroom*, *found* and *impact* for Stage 2

Student code	Text 1	Text 2	Text 3
122 Topic Tokens/Types (*f*)	curriculum 64/26 *curriculum* 29, *classroom*, *found* 1	SEND 100/35 *found* 7, *classroom* 1, *curriculum* 1	lifelong learning 92/43 *found* 10, *classroom* 5
124 Topic Tokens/Types (*f*)	SEND 46/23 *found* 2	lifelong learning 63/46 *found* 1	sociology 30/20 *classroom* 2, *curriculum* 2, *found* 1,
143 Topic Tokens/Types (*f*)	curriculum 79/23 *curriculum* 51, *impact* 4, *classroom* 2	lifelong learning 124/59 classroom 5, found 5, *curriculum* 2, *impact* 2	psychology 74/32 *classroom* 9, *impact* 4
107 Topic Tokens/Types (*f*)	curriculum 111/ 33 *curriculum* 61, *found* 1	lifelong learning 142/56	psychology 62/28 *classroom* 13, *impact* 1

		found 5, *classroom* 3, *curriculum* 3, *impact* 2	
121 Topic Tokens/Types (*f*)	international 82/45 *curriculum* 8, *impact* 3	lifelong learning 111/65 *found* 9, *curriculum* 3, *classroom* 2	sociology 60/33 *curriculum* 6, *found* 4
135 Topic Tokens/Types (*f*)	curriculum 90/29 *curriculum* 55, *classroom* 1, *found* 2	international 60/16 *impact* 9, *curriculum* 5, *found* 2	lifelong learning 104/43 *classroom* 5, *impact* 4, *found* 3, *curriculum* 2

Table Z: Six selected cases, three assignments chosen for each, the education topics, tokens and types from the NAWL and the frequency (*f*) of *curriculum*, *classroom*, *found* and *impact* for Stage 3

Student code	Text 1	Text 2	Text 3
122 Topic Tokens/Types (*f*)	sociology 100/65 *found* 9, *impact* 1	psychology 103/60 *found* 17, *classroom* 12, *curriculum* 1	psychology 56/29 *classroom* 7, *found* 2, *curriculum* 1
124 Topic Tokens/Types (*f*)	International 62/45 *impact* 3, *found* 1	psychology 69/39 *found* 13, *curriculum* 3, *classroom* 2	psychology 48/19 *curriculum* 21, *found* 2, *impact* 1
143 Topic Tokens/Types (*f*)	psychology 66/32 *impact* 5, *found* 4	psychology 59/27 *classroom* 9, *found* 4	curriculum 36/15 *curriculum* 20, *impact* 3, *found* 1
107 Topic	psychology	psychology	SEND

Tokens/Types (*f*)	92/53 *curriculum* 12, *classroom* 7, *impact* 2	90/51 *curriculum* 8, *classroom* 1	182/ 60 *impact* 4
121 Topic Tokens/Types (*f*)	international 77/40 *equality* 13	psychology 64/34 *found* 8, *classroom* 4	SEND 159/51 c *classroom* 16, *curriculum* 8, *found* 3
135 Topic Tokens/Types (*f*)	sociology 121/41 *found* 3, *impact* 3	International 78/36 *found* 7	psychology 32/17 *classroom* 6, *curriculum* 2, *found* 2

Appendix 9: List of the 140 most frequent 'off-list' types by stages

List of the top 140 most frequent 'off-list' types for stage 1

england	lifelong	uk
piaget	al	et
vocational	english	reay
skinner	better	p
reinforcement	vygotsky	learners
pisa	kingdom	compulsory
oecd	portugal	best
bandura	malta	gcse
luxembourg	bibliography	website
masculinity	learner	sats
walkup	underachievement	ireland
sen	skelton	british
attainment	ofsted	proofreading
slovenia	e	widening
britain	caribbean	latvia
belgium	wolf	employability
literacy	c	o
apprenticeships	bruner	european
efficacy	sociological	carrington
spain	cottrell	gillborn
operant	kennedy	operational
academies	exclusion	stakeholders
guidance	behaviourist	plagiarism
mainstream	davies	parsons
skinners	warrington	academically
cyprus	doll	mcmillan
behaviourism	hofstede	rogers
weyers	deemed	educators
ethnicity	masculine	sre
v	afro	g
offenders	foster	scaffolding
schemas	co	numeracy
schema	academy	albert
annotated	burton	gcses
b	crozier	grady
ict	imitate	mcleod
racism	underachieving	watson
certificate	derby	constructivist
pritchard	proximal	smith
x	nottingham	portuguese
bbc	gov	j
maltese	slovakia	diane
latvian	netherlands	aspirations
born	btec	esteem
mufti	socio	wright
additionally	assimilation	creativity
feminine	haywood	

List of the 140 most frequent ‘off-list’ types for stage 2

et	al	psychometric
gardner	learners	spearman
eysenck	adler	iq
vocational	g	dyslexia
gove	learner	uk
nomothetic	idiographic	english
wellbeing	best	better
bowlby	kelly	costa
feist	mccrae	freud
stakeholders	attainment	maltby
geddes	benn	vygotsky
pisa	gillard	esteem
educators	psychometrics	cobo
p	woolfolk	dfe
kingdom	additionally	maslow
allport	born	neuroticism
reid	visser	beneficial
mi	cattell	hancock
gardener	erikson	literacy
wolf	autism	sternberg
england	marx	creativity
ego	wilson	e
deemed	gcse	scaffolding
hirsch	feminism	id
michael	ofsted	siblings
callaghan	emotionally	horn
bandura	extraversion	british
sen	burke	efficacy
mcdonaldization	truss	asd
c	disabled	grosvenor
nettelbeck	carroll	dyslexic
fox	pring	leat
adhd	compulsory	guidance
negatively	oecd	optimistic
feminists	inferiority	insecure
superego	howarth	inventory
waterhouse	haslam	psychoticism
strive	summerhill	chc
interpersonal	openness	kinaesthetic
academically	devised	disruptive
engagement	j	proximal
tcks	bolt	educationalists
nettlebeck	ritzer	bbc
caregiver	chitty	durkheim
excel	globally	impairment
prophecy	sibling	solely
kane	mainstream	

List of the most frequent ‘off-list’ types for stage 3

wellbeing	learners	al
et	exclusion	inclusive
learner	efficacy	bullying
mainstream	intrinsic	sen
assertive	levitas	walkup
better	dfe	nurture
underachievement	extrinsic	best
expectancy	esteem	uk
india	robinson	schunk
attainment	lgbtiqa	attribution
bandura	gangs	ofsted
smith	gang	socio
intersectionality	mainstreaming	behaviourist
maslow	p	beneficial
ryan	reinforcement	warnock
watson	weiner	canter
deci	china	hedonistic
additionally	corbett	neets
engagement	henley	intrinsically
educators	impairment	deemed
pedagogy	atkinson	woolfolk
eudaimonic	skinner	norwich
england	hughes	interventionist
solidarity	inappropriate	english
piaget	unesco	unicef
asia	brazil	disruptive
negatively	cognition	expectancies
jordan	afghanistan	africa
gcse	motivational	kingdom
seligman	attributions	teenage
humanistic	sid	disabled
saharan	e	mud
bullied	de	inclusivity
shakespeare	syria	bronfenbrenner
collins	panju	swinson
byrne	cyber	humphrey
operant	russia	segregated
behaviourism	empathy	humanist
segregation	trussler	underachieving
hawkins	holistic	relevance
solely	villa	byman
canney	continuum	neet
formative	pakistan	compulsory
deprived	eccles	rogers
mathieson	modernisation	o
underclass	restorative	academically
bourdieu	carr	