THESIS

A CORPUS-BASED ANALYSIS OF ENGLISH VOCABULARY INPUT PROVIDED IN K-

12 TEXTBOOKS USED IN SAUDI ARABIA

Submitted by

Elham Alhudithi

Department of English

In partial fulfillment of the requirements

For the Degree of Master of Arts

Colorado State University

Fort Collins, Colorado

Summer 2017

Master's Committee:

    Advisor: Tatiana Nekrasova-Beker

    Anthony Becker
    Mary Vogl

ABSTRACT

A CORPUS-BASED ANALYSIS OF ENGLISH VOCABULARY INPUT PROVIDED IN K-12 TEXTBOOKS USED IN SAUDI ARABIA

The importance of a textbook becomes greater when it is the major source of foreign vocabulary input as is the case in Saudi Arabia with the English language. Given the fact that numerous studies indicated that Saudi EFL learners have a very limited vocabulary size (Albogami, 1995; Alhazemi, 1993; Alsaif, 2011), the present study is concerned with the extent to which input from textbooks contributes to building the learners' lexicon. Through utilizing range and concordance programs to analyze a corpus of 252,517 tokens, the researcher concluded that great opportunities were offered to learn the most frequent words in English, in which they had a relatively low type-token ratio, indicating a great deal of repetition and coverage. The findings obtained from the case study analysis also demonstrated that a sufficient variety of collocations, derivations, and inflections were provided to assist in deepening the learners' vocabulary knowledge. Although the learning of words goes hand-in-hand with the number of occurrences, where the more words that are repeated the greater the learning, the textbooks surprisingly denoted a shortcoming in the repetition of words. More specifically, at least 74% of newly-introduced words appeared four times or less. Another shortcoming was attributable to the distribution of newly-introduced words, where it contradicted with the literature by introducing as high as 30 new words per hour of schooling. Based on what the researcher found, several pedagogical implications were suggested for teaching vocabulary in EFL settings, varying from providing more repetition opportunities for abstract concepts that

carry the central meaning in a given context to following a data-driven approach through the

lexical analysis of concordance lines to promote students' noticing and long-time learning.

ACKNOWLEDGMENTS

I would like to express my profound gratitude to my advisor, Dr. Tatiana Nekrasova-Beker. This thesis would have never been accomplished without her persistent support, guidance and advice, but most of all, her confidence in me through every stage of the process. I am also gratefully indebted to her for the opportunities she has opened to me which have expanded my abilities. Her encouragement, mentorship, and expertise has been instrumental to my overall success at CSU. With my deepest appreciation, thank you for all you have done.

I would like also to acknowledge members of my thesis committee: Dr. Anthony Becker of the Department of English and Dr. Mary Vogl of the Department of Languages, Literatures, and Cultures, whose expertise and recommendations added to the richness of this study. I am greatly thankful and indebted to them for their time, support, and enthusiasm about my research. My appreciation is also extended to the staff and faculty of the Department of English for their commitment and assistance over the course of my MA.

Most importantly, I would like to take this opportunity to express my gratitude to my incredible family for never stop believing in me, and providing me with invaluable support. First and foremost, my eternal gratitude goes to my mother, a wonderful educator who has a tremendous impact on every life she touches. I am greatly appreciative for all of the support, love, and prayers she has been given me. To my father, I would like to thank you very much for your continuous encouragement, support, and prayers. To my sister, who has supported and joined this endeavor since day one, I am especially grateful to her for being a pillar of strength and motivation. This accomplishment would not be possible without their unfailing support, and I am forever indebted to them.

DEDICATION

Dedicated to my mother, Faitmah, for her unwavering support and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

During the past few decades, second language acquisition researchers have recognized the importance of vocabulary in learning any language in which it becomes "the fastest growing area of second language education in terms of research output and publication" (Maiguashca, 1993). What researchers have come to conclude is that having a deep knowledge of vocabulary would enable language learners to improve their comprehension and reach higher levels of proficiency (Milton, 2009; Nation, 2001). Due to the importance of vocabulary, researchers have tried to estimate the needed vocabulary size that learners of languages need to obtain for comprehension, including learners of English. Studies show that learners of English as a second language (ESL) need to have a vocabulary of 5,000 to 10,000 in order to comprehend authentic spoken and written texts (Laufer, 1989; Milton, 2009; Nation, 2006; Webb & Rodgers, 2009). With regard to an English as a foreign language (EFL) context, it is estimated that a vocabulary knowledge of 3,000 to 5,000 would enable English as a foreign language learners (EFLLs) to be able to comprehend the language to a certain degree effectively (Nation, 1990; Sutarsyah, Nation, & Kennedy, 1994).

Since vocabulary is seen as an essential component of language learning, it is important for English language learners (ELLs) to be exposed to the target language and have many sources to deepen their vocabulary knowledge and improve their comprehension. Although a number of vocabulary sources can be easily accessed today, the only source of vocabulary exposure for EFL learners oftentimes is what is being introduced and taught in classrooms (Milton, 2009). When looking at the Saudi Arabian context, there is a lack of exposure to the English language since it is not the official language of the country (Alsaif & Milton, 2012; Alqahtani, 2011; Khan, 2011). That is, the majority of Saudis use their native language, Arabic,

to interact with friends, workers, and the rest of the members in their communities. Even though the English language plays an important role in the country's economy, where, in particular, millions of expatriates are working in the private sectors and using English as the medium of communication, they tend to learn the Saudis' native language to be able to communicate with them. Due to the absence of using English in Saudis' daily life, textbooks are considered by far the main source of vocabulary knowledge.

Particular attention has been given towards vocabulary in textbooks during the past few decades (Laufer, 1989; Milton, 2009; Nation, 2001; O'Dell, 1997). The most important aspects of vocabulary presentation in textbooks that second language acquisition researchers tend to emphasize are related to vocabulary selection, coverage, and recycling. It is believed that a good language textbook would select appropriate types of vocabulary for the intended audience, include a suitable amount of vocabulary in relation to the total study hours, and repeat new lexical items at least five times. Since textbooks play a key role in foreign vocabulary learning in Saudi Arabia as it contributes to the learners' vocabulary size, it is necessary to investigate the vocabulary presented in Saudi EFL textbooks.

1.1 STATEMENT OF PROBLEM

The Saudi Ministry of Education has been expressing major concerns regarding the outcomes of English language teaching in K-12 Saudi public schools (Alrashidi & Phan, 2015). Despite the ongoing funds and enormous efforts to improve English language learning, the majority of Saudi public school students have a very low level of English proficiency in a way that they could not produce any flawless English sentence (Aljohani, 2009; Alshumaimeri, 2003; Fareh, 2010; Khan, 2011; Rajab, 2013). Although the English language is taught as a compulsory course for nine years in Saudi public schools, researchers have found that the vocabulary size of

2

secondary school graduates is limited to about 1,000 words (Albogami, 1995; Alhazemi, 1993; Alsaif, 2011).

A number of research studies have been conducted to investigate the suitability of teaching methods, the quality of teaching preparation, and the effects of non-linguistic factors including motivation and anxiety as potential factors for the learners' small vocabulary size and low English language achievements (Alakloby, 2001; Alharbi, 2015; Alnasser, 2015; Alshammari, 2011). Yet, only one study has linked the learners' poor vocabulary knowledge to the Saudi EFL textbooks. Alsaif and Milton's study (2012) showed that most new vocabulary items were introduced in intermediate school, from grades 7-9. To elaborate, around 900 lexical items were introduced in grade 9, where the number of new lexical items dropped down sharply to approximately 100 in grade 10. When only introducing one new vocabulary word per hour of schooling, it would not be surprising if students at higher school levels have a very limited vocabulary size and poor English language proficiency.

With new EFL curricula being utilized in Saudi public schools since the 2011-2012 academic year, it is necessary to investigate the vocabulary recycling, selection, and coverage to ensure the appropriateness of lexical items for the intended audience. The present study investigates and measures the vocabulary presented in the most recently developed EFL textbooks being implemented in Saudi public schools from grades 4 to 12 to ensure that the learners are receiving appropriate vocabulary that would lead them to expand their vocabulary size and improve their comprehension.

1.2 SIGNIFICANCE OF THE STUDY

The importance of a textbook becomes greater when it is the major source of foreign vocabulary input as is the case in Saudi Arabia with the English language. That being said, it is of vital importance to investigate the recycling, selection, and coverage of vocabulary items in the most recently developed EFL textbooks being utilized in Saudi public schools. The investigation revolves around measuring the amount of newly introduced words, the frequency of word occurrences, the characteristics of words, and the percentage of different word lists presented in the Saudi EFL textbooks. The findings of this study help to determine whether the lexical material used in the textbooks is adequate to enable Saudi public school EFL learners to succeed in improving their comprehension and expanding their vocabulary size.

1.3 RESEARCH QUESTIONS

RQ 1: To what extent are words from K1, K2, AWL, and additional word lists represented in Saudi EFL textbooks?

RQ 2: How many newly-introduced word types are presented per grade?

RQ 3: How often are newly-introduced words repeated in the textbooks?

RQ 4: What types of learning opportunities do Saudi EFL textbooks present to acquire new vocabulary?

1.4 BACKGROUND INFORMATION ON THE USE OF ENGLISH IN SAUDI ARABIA

Although Saudi Arabia was never under European power, the English language was introduced there in the 1930s as an educational subject. Shortly after the discovery of oil, the Saudi government realized the importance of introducing the English language to the society in order to enable them to communicate with the outside world (Aljohani, 2009; Hayer, 2005). According to Alseghayer (2005), "the early stage of oil production required Saudis who could successfully communicate in English with interests outside the country and with foreign experts" (p. 126). Up until the 1950s, the English language was mainly used within the setting of business in Saudi Arabia due to the great expansion of the oil industry. Since then, the importance of the English language has grown rapidly in the country.

Although the English language is not the official language in Saudi Arabia, it still plays a major role in various disciplines today, including business, media, and medicine; thus, it is considered by far the dominant foreign language (Alfotais, 2012; Rahman & Alhaisoni, 2013). The majority of Saudi job applicants are required today to have some level of English proficiency in order to get hired in both the public and private sectors. Aside from the labor market, the presence of the English language is also shown in the mass media in Saudi Arabia. A considerable number of newspapers are written in English and published daily in the country; the *Saudi Gazette* and the *Daily Riyadh* to name a few. Likewise, the country broadcasts various English and American programs 24 hours a day.

Due to the importance of English, the Saudi Arabian government has been spending billions of dollars in order to improve English education by training teachers of English, developing the English curriculum, offering scholarship programs to English-speaking countries, and preparing language laboratories (Rahman & Alhaisoni, 2013). In Saudi Arabian K-12

schools, the English language is implemented as a compulsory course starting from grade four at public schools and grade one at private schools (Abdan, 1991), where the majority of international schools utilize English as a medium of instruction for all K-12 grades. As for the university level, English is the sole language of instruction in various scientific majors, including medicine and engineering. Additionally, there are two universities that use English as the language of instruction for all courses they provide: King Fahad Petroleum and Mineral University and King Abdullah University of Science and Technology. For the rest of the non-scientific fields including humanities, students are asked to complete courses in English for academic purposes to earn a bachelor degree.

1.4.1 ENGLISH LANGUAGE TEACHING IN K-12 SAUDI PUBLIC SCHOOLS.

The Saudi Arabian government officially introduced English as the only foreign language being taught at Saudi public K-12 schools during the late 1950s. General education is comprised of four stages: kindergarten, primary, intermediate, and secondary (BinObaid, 2016). At the age of six, students are expected to start their general education and spend six years in the primary stage, three years in the intermediate level, and finally three years in the secondary stage. During the 2015-2016 academic year, there were 4,153,949 students attending the 23,130 Saudi K-12 public schools (Ministry of Education, 2016a).

All K-12 public schools across the country are administered and supervised by the Ministry of Education (Alrashidi & Phan, 2015; Badawood, 2003), the largest centralized educational agency in Saudi Arabia. The ministry offers a unified educational system and provides a tuition-free education for all students. Some of the responsibilities that the ministry manages are setting the curricula, hiring school teachers and staff, publishing the textbooks, and developing the educational policies.

In the 1930s, the teaching of English began in a number of Saudi schools with no defined curriculum (Aljohani, 2008). During the 1950s, the Saudi government implemented an English language syllabus which was imported from Egypt in Saudi public schools (Alamri, 2008). Since the implementation of the Egyptian curriculum did not meet the expected results, a new comprehensive English language curriculum emerged in order to meet the needs of Saudi learners in 1960. The developed curriculum, entitled *Living English for the Arab World,* was implemented for 20 years in Saudi public schools. According to Alseghayer (2005), all English language materials were developed by the Department of Curriculum Development in Saudi Arabian Ministry of Education.

The English language branch at the Department of Curriculum Development launched a number of projects throughout the years to further improve the English curriculum (Alseghayer, 2005). In collaboration with Macmillan Press in 1980, public school students were introduced to a new English curriculum called *Saudi Arabian Schools English*. Because it better corresponded to students' needs, this curriculum was utilized for almost 15 years. With the help of English as a foreign language specialists from King Fahad University, another curriculum, entitled *English for Saudi Arabia*, was introduced to public school students during the academic year of 1995-1996. After eight years, another textbook series was introduced to public school students, entitled *Say it in English* in 2004.

As for the current English curriculum, it started to be implemented officially during the academic year of 2011-2012 in K-12 public schools (Ministry of Education, 2016b). This curriculum emerged in collaboration with three global organizations: Macmillan Press, Cengage Learning, and MM publications. Aside from these projects, curriculum developers depend on

supervisors, teachers, and researchers' recommendations and findings to improve the English language curriculum.

In all K-12 public schools, three types of English language materials are being utilized: a student's textbook, a workbook, and a teacher's manual (Alotaibi, 2004). These books integrate the four skills of English language: reading, writing, listening, and speaking, along with the presentation of grammatical and lexical aspects of the language. All students receive the learning materials for free, which include the student's textbook and the workbook. For each grade level, teachers of English are asked to adhere to identical syllabus guidelines. With respect to the minimum qualifications to teach the English language in these schools, teachers are required to have a bachelor's degree in the English language (Alfahadi, 2014). However, English language teachers are not required to have any previous training or teaching experience to start their teaching journey at K-12 public schools.

Initially, the English language was a compulsory course for only intermediate and secondary school students in K-12 Saudi public schools (Alrashidi & Phan, 2015; Ebad, 2014). To benefit from earlier exposure of the language, the Ministry of Education made a decision to teach the English language as a required subject at primary public schools in 2003, starting from grade six. Today, English is being taught from grade four at Saudi public schools. In primary schools, students receive English instruction during two 45-mintue classes each week throughout the semester, whereas students from grade seven and higher are required to attend four 45-mintue English classes weekly.

Although Saudi public school students are learning the English language as a required course starting from grade four, they tend to be less interested in improving their level of English proficiency since English is considered one of the overridden courses (Alharbi, 2015). That is,

8

whenever a student fails to reach the minimum passing score, which is 50%, the student is still able to progress from one grade to another (Ministry of Education, 2016b). However, if students fail to get the minimum score in non-overridden courses such as the Arabic language, they will not be able to take any course from the next level, and they will have to retake all courses from the previous year.

CHAPTER 2: LITERATURE REVIEW

The importance of vocabulary learning started to be recognized during the late 1980s. O'Dell (1997) noted that "the words *lexis* and *vocabulary* are remarkable by their absence from either chapter headings or indexes in the major writers on syllabus of the 1970s and 1980s" (p. 258). Since then, it has been considered a necessary component of learning any language due to the growing understanding of the fact that without gaining a lexical knowledge, very little can be achieved (Laufer & Nation, 1999; Milton, 2009; Nation, 2001; Schmitt, 2010). This recognition has been leading a number of language acquisition researchers, textbook designers, and language teachers to conduct studies regarding the needed vocabulary size as well as amounts and types of vocabulary for second/foreign language learners to comprehend the target language and improve their proficiency (Milton, 2009; Nation, 2001; O'Dell, 1997). A discussion regarding vocabulary size, selection, coverage, and repetition is outlined after clarifying three important questions: *What types of vocabulary do learners need to learn? What does knowing a word mean? How are words being counted?*

2.1 TYPES OF WORDS

Vocabulary acquisition researchers have developed a number of word lists that would enable language teachers and material writers to know the most beneficial vocabulary types for learners of languages. Nation (2001) illustrated the differences between four different types of vocabulary: high-frequency, low-frequency, technical/specialized, and academic words. An estimate of the lexical coverage of those four types of vocabulary in a typical textbook is presented in Table 1.

Table 1
Lexical Coverage of Common Academic Texts

Type of Vocabulary	Lexical Coverage
High-frequency Words (K1 & K2)	82%
Low-frequency Words	5%
Technical/specialized Words	5%
Academic Words	9%

As it is illustrated in Table 1, words in the first 1,000 (K1) and second 1,000 (K2) levels are generally thought to be the high-frequency words. These high-frequency words are included in almost all contexts and cover a very large proportion of running words in both written and spoken texts. In respect to the estimated percentage of lexical coverage for these two frequency levels, researchers generally agreed that the K1 word list offers around 77% coverage of running words in academic texts, whereas the second K2 word list offers about 5% coverage.

When examining the conversation genre, the K1 word list is believed to cover about 85% and the K2 word list provides about 6% coverage of the running words (Nation, 2001). Since these types of words make up most of the words in texts (around 80% of coverage) and provide the needed lexical foundation, it is very beneficial for learners of English to learn them first. A productive knowledge of these words will enhance the learners' abilities to comprehend the content independently. According to Milton (2009), "most frequent words will be learned earliest" and they "are also the most useful to the learner; they are the words that will enable the learner to understand and express himself/herself most efficiently" (p. 44).

Low-frequency words, on the other hand, include adverbs, nouns, adjectives, and verbs. All these major word classes are used very frequently in English. Because of that, these words tend to cover a small proportion of texts. According to Nation (2001), words that are not included in the first 1,000 and second 1,000 most frequent word lists are generally referred to as

low-frequency words. However, it is imperative to note that these words do not occur very rarely in all texts. Nation (2001) noted that the range of low-frequency words in texts differs as they tend to occur more in some texts but not the others. That is, these types of words would occur very often once the message of the text is closely related. In terms of coverage for these type of words, they tend to make up about 5% of running words in texts, as seen in Table 1.

Technical words differ based on the subject area (Nation, 2001). Accordingly, these words "are very closely related to the topic and subject area of the text" (p. 12). That is, these types of words would be very frequent in a particular topic but not the other. Nation (2001) noted that once the topic of any given text is selected, it is estimated that technical words would probably make up about 5% of the running words.

The last group is academic words which are commonly included in different types of academic texts. An academic word like *illustrate* can be frequently used in journal articles, lectures, and essays. These types of words typically cover around 9% of the tokens in any academic text. Because of this, it is of a vital importance for learners of English who are interested in using the language for academic purposes to improve their knowledge of those words.

2.2 KNOWLEDGE OF WORDS

One of the widely used concepts to determine the knowledge of words is the distinction between perception and production, also known as passive and active, respectively (Melka, 1997; Milton, 2009). The receptive knowledge of a word is associated with recognizing words in reading and listening whereas the productive knowledge is associated with using words in speech and writing. It has been thought that learners of languages develop their vocabulary knowledge moving from the receptive to the productive stage. Since it has been largely accepted that the

receptive lexical knowledge that language learners own is larger than the productive one, many textbook developers have been dividing wordlists that are expected to be learned either passively or actively. For instance, when looking at the Hungarian context, the curriculum designers suggest that "learners should learn some 1,600 words by the 8th grade, 1,200 of which should be known actively and a further 400 passively" (Milton, 2009, p. 13).

Anderson and Freebody (1981) suggested another convention to distinguish the various types of lexical knowledge. According to them, the knowledge of vocabulary differs based on its breadth and depth. The total number of words that a learner knows is referred to as the breadth. That is, breadth includes recognizing word forms passively without necessarily knowing possible meanings that these word forms can carry. The depth, on the other hand, refers to the knowledge that a learner gained about words. The depth of vocabulary knowledge, therefore, is not limited to the meaning of words; it can also include the knowledge of word functions, collocations, and others.

Although the conventions of reception and production, as well as breadth and depth, have been found to be beneficial by a number of material designers and test developers, it is still crucial to determine the complexity of word knowledge. Nation (2001) proposed a more systematic method of analyzing the different types of lexical knowledge and its characteristics. Nation (2001) listed three different ways to know a word: knowledge of form, knowledge of meaning, and knowledge of use, which were further divided into three sub-sections. In each sub-section, Nation distinguished between the receptive (R) and productive (P) knowledge of a word, as presented in Table 2.

Table 2
Knowledge of Form, Meaning, and Use

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

As seen in Table 2, the knowledge of a word form includes the recognition of the written

form of a word as well as the phonological form. It also involves the receptive knowledge of the

word form in which learners know what a word sounds like and what a word looks like. Nation

(2001) added a third part to the knowledge of the word form which is word parts. This type of

knowledge includes knowing the prefixes and suffixes that can be added to words to change the

meaning as in 'happi*ness*' and '*un*happy'.

As for the knowledge of word meaning, it is divided into three different parts: form and

meaning, concepts and referents, and associations. The first part involves the ability to link a

word form to a meaning which, in the case of foreign language learners, revolves oftentimes

around finding a link between a foreign word and its translation. The concepts and referents

involve knowing the possible translations and interpretations of words. In a similar manner, the

word associations show relevant meanings and associations of words in a particular language. For instance, the word 'fat' has a negative meaning in English and a more positive meaning in Nigerian (Milton, 2009). Second language learners need to recognize this aspect of word knowledge in order to avoid offending speakers of the target language without being aware of their wrong word choice.

The last type of word knowledge, use, is likewise divided into three parts: grammatical functions, collocations, and constraints on use. The grammatical functions include knowing the part of speech of a word along with recognizing its use with other words. That is, the English word '*green*', for example, functions as an adjective and is supposed to be used before the noun it qualifies as in '*a green dress*' but not '*a dress green*'. The second part, collocations, consists of a number of words that commonly occur together, such as the word '*work*' with the following combinations: '*work hard*', '*find full-time work*', and '*work together*'. The last part, constraints on use, involves the frequency of word occurrences as well as the context that a word could be used in. *Identify* is an example of a word that is commonly found in academic contexts.

2.3 COUNTING WORDS

Particular attention has been given toward estimating and measuring the amounts of lexical coverage and loading of language textbooks and the vocabulary size of learners during the past few decades. In this respect, it is important to recognize different ways in which words can be counted. Nation (2001) proposed four different ways in which the number of words could be counted in a text, written or spoken, which include token, type, lemma, and word family.

Tokens, sometimes known as running words, represent the total number of words in a spoken or written text in which every reoccurrence of the same word is counted. A sentence like "It is not easy to say it correctly" consists of eight tokens despite the fact that one word, *it*, is

repeated twice (Nation, 2001, p. 7). This unit of measurement assists material developers and researchers to have accurate quantitative information regarding lexical coverage of any text. Nation (2006) defined lexical coverage as the following: "the percentage of running words in the text known by the readers" (p. 61). Counting the tokens would be also beneficial in terms of measuring the total number of words that are being said per minute or written per page, for example.

A type is defined as counting the number of different words only. That is, the same form of a word will be counted only once regardless of how many times it reoccurs in a particular text. Unlike tokens, the sentence "*It is not easy to say it correctly*" contains seven types since the word, *it*, occurs more than once (Nation, 2001, p. 7). Counting the number of words based on their types would be useful to answer a question like "How many words do you need to know to read this book?" (Nation, 2001, p. 7).

A lemma is a unit of counting words which "consists of a headword and some of its inflected and reduced (n't) forms" (Nation, 2001, p. 7). The English language has eight different inflectional categories: plural, past tense, past participle, present participle, third person singular present tense, possessive, comparative, and superlative. It is important to mention that words that have different parts of speech are not counted as one lemma even if they share the same form. To give an example, a word like '*think*' will be counted as two different lemmas in case they are being used as a noun and a verb.

Lastly, a word family refers to "a base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately" (Bauer & Nation, 1993, p. 253). Words like '*develop*', '*develops*', '*development*', and '*developer*' belong to the same word family. It has been believed that the recognition of words included in the same family

16

would be easily achieved once a base form or a closely derived word is known. Given that, the size of each word family increases whenever a learner is exposed to numerous inflectional forms.

2.4 VOCABULARY SIZE

Global Language Monitor reported that the number of words that are included in the English language is 1,025,109 (Global Language Monitor, 2014). When measuring the actual vocabulary size of native speakers of English, researchers have estimated that the lexical knowledge of native speakers varies from 12,000 to 20,000 word families as a result of having different levels of education (McCarten, 2007; Nation & Waring, 1997). It is estimated that native five-year-olds have a lexicon of about 4,000 to 5,000 word families (Schmitt, 2000). Each year, from childhood to adolescence, around 1,000 word families are expected to be added to their lexicons. As for the educated native speakers of English who received higher education, it is estimated that they have a knowledge of 20,000 word families (Nation & Waring, 1997). It has been believed that once English language learners know approximately 10 - 15% of the vocabulary size that educated natives have, they would be able to use it to communicate effectively.

A number of studies have been conducted to measure the vocabulary size of non-English native speakers (AlBogami, 1995; AlHazemi, 1993; Laufer, 2000; Nation, 1990). After receiving English instruction for a period of five years, it is estimated that English as a foreign language learners would be able to know up to 2,000 word families. Laufer (2000) investigated the vocabulary size of learners of English as a foreign language from eight different countries and concluded that their lexical knowledge ranged from 1,000 to 4,000 words. When looking to the Saudi Arabian context, researchers have tried to estimate the size of vocabulary knowledge of

Saudi learners and concluded that they know approximately 1,000 words (AlBogami, 1995; AlHazemi, 1993).

With respect to learning English as a second language, it is estimated that young ESL learners would acquire 1,000 word families each year (Konstantakis & Alexiou, 2012). Despite the great learning opportunities that ESL learners receive, it would be crucial for them to reach a native-like accuracy. Fortunately, words differ not only in their sounds, spellings, and meanings, but also in the level of their occurrences since some words are far more frequent than others in the spoken or written use of the language. According to Nation (2001), some words tend to be more useful than others for English language learners, such as the most frequent words. If learners have a knowledge of words that native speakers of English mostly produce, they will be able to comprehend the English language. That being said, it is imperative for material developers and language teachers to know about word frequency and word coverage and their effects on comprehension. The relationship between word coverage and comprehension is discussed in the following section.

2.5 COVERAGE AND COMPREHENSION

The relationship between coverage and comprehension is incredibly strong since the knowledge of more words increases the level of understanding. That is, the more words that are known, the better the understanding of spoken and written use of the language. This raises a question of how much lexical coverage is needed to aid comprehension. There are several studies that investigated the amount of vocabulary needed for comprehension of different spoken and written genres (Adolphs & Schmitt, 2003; Laufer, 1989; McCarthy, 2004; Milton, 2009; Nation, 2006; Webb & Rodgers, 2009).

In respect to spoken discourse, vocabulary acquisition researchers have made several attempts to calculate the required lexical coverage for understanding (Adolphs & Schmitt, 2003; Nation, 2006; McCarthy, 2004; Web & Rodgers, 2009). It is believed that a lexical coverage of 98% would contribute to better understanding of different types of spoken genres, including movies and lectures. A number of researchers argued that a vocabulary size of 5,000 to 9,000 word families make up 98% of running words in television programs. Thus, learners of English need to have a knowledge of at least 5,000 word families in order to be able to comprehend the content of these programs independently. In terms of the required vocabulary size required to comprehend the news, learners of the English language need to know around 1,800 word families, which will help to cover over 80% of tokens. Lastly, the minimal vocabulary size that is required to participate in everyday conversation is thought to be around 2,000 to 3,000 word families that occur very frequently.

For written texts, Laufer (1989) reported that good reading comprehension would be achieved with a text coverage of 95%; that is, one word in 20 is unknown. Laufer came to this conclusion after investigating numerous texts, including English for academic purposes, to determine the amount of coverage needed for learners to have minimal reading comprehension. In order for learners of English to reach this level of coverage, they need to have a vocabulary size of at least 5,000 word families. Until recently, it was generally thought that once learners of English know around 95% of lexical items, they would be able to read and understand written texts independently. Nation (2006) stated that learners of English need to be familiar with 98% coverage of running words in order to comprehend reading texts successfully. That is, if one word among 50 words is unknown, it will not affect reading comprehension of English language learners. Because of that, learners of English need to have a vocabulary size of at least 8,000

word families to be able to read any written text independently. Table 3 shows the common lexical coverage of different types of written texts, as identified by Nation (2001).

Table 3
Vocabulary Type and Text Coverage

Levels	Fiction	Newspapers	Academic text
K1	82.3%	75.6%	73.5%
K2	5.1%	4.7%	4.6%
Academic	1.7%	3.9%	8.5%
Others	10.9%	15.7%	13.3%

As can be seen in Table 3, words in the K1 level contribute significantly to the understanding of different written texts. It provides a lexical coverage of 82% fiction, 76% newspapers, and 74% academic texts. Thus, it is imperative for learners of English to acquire these types of words first since they provide the needed lexical foundation. In other words, English language learners would benefit greatly from knowing these high-frequency words since they make up most of the running words in any text. With this in mind, a coverage of 95 - 98% of the running words is required to comprehend written or spoken texts successfully.

2.6 VOCABULARY INPUT IN TEXTBOOKS

It is imperative to choose carefully the type and amount of vocabulary that would be presented in textbooks since an inappropriate selection would hinder the learning. Historically, the selection of vocabulary in foreign language textbooks had few similarities. Milton and Benn (1933) examined the vocabulary input of 29 beginner French textbooks and found that only 19 word types were common among those books out of 6,000 word types. Due to the huge variety, they were concerned with having very little vocabulary in common in order to build on their vocabulary knowledge. They were also concerned about the challenge of creating future learning materials once these learners progressed to the next level.

Fortunately, the majority of publishers in recent decades have been sharing their vocabulary selection criteria along with creating their own corpora, which the words in their developed materials are derived from. Gairns and Redman (1986) listed frequency as the top criteria for vocabulary selection. Two years later, White (1988) suggested the selection of vocabulary be based on frequency, coverage, range, learnability, availability, interest, and learning opportunity. O'Dell (1997) proposed similar criteria for the selection of vocabulary in textbooks which include the following: *frequency, coverage, range, availability, interest, learning opportunity, and learnability*. Recently, Nation (2001) also recommended the word frequency to be the most important criterion for selecting vocabulary.

What influenced most vocabulary acquisition researchers to list frequency as their top criterion for selection of vocabulary is the realization of the importance for learners of English to have a knowledge of words that native speakers use most, which is derived from frequency data. This is thought to help learners of English avoid difficulties in understanding any text they encounter. For this reason, curriculum developers should pay attention to the selection of lexical items in textbooks to be based on word frequency.

Since frequency matters in vocabulary learning, Nation (2006) proposed three consecutive ranges of vocabulary that language learning materials should represent for each level of proficiency: first 1,000 for beginners, second 1,000 for intermediate, and third 1,000 most frequent words for advanced learners. That is, beginner textbooks should cover words in the first range, intermediate textbooks should include words in the second range, and advanced textbooks should present words in the third range. According to Nation (1990), a knowledge of the 3,000 most frequent words would enable learners of English as a foreign language to comprehend reading tasks at the university level.

With respect to vocabulary distribution, it is impossible and unrealistic to present all words in textbooks at once. For each learning hour, Garins and Redman (1986) recommended introducing a reasonable input of 8 to 12 words. That is, about 1,000 words are ideally thought to be presented in 125 hours of schooling. Scholifield (1991) made a similar estimate for presenting an average of nine words per lesson for an imaginary course. Scholifield further illustrated that not every learning hour would include vocabulary input since a number of units in each textbook are dedicated to revision, recycling, testing, among others. Figure 1 presents his hypothesis of vocabulary rate.
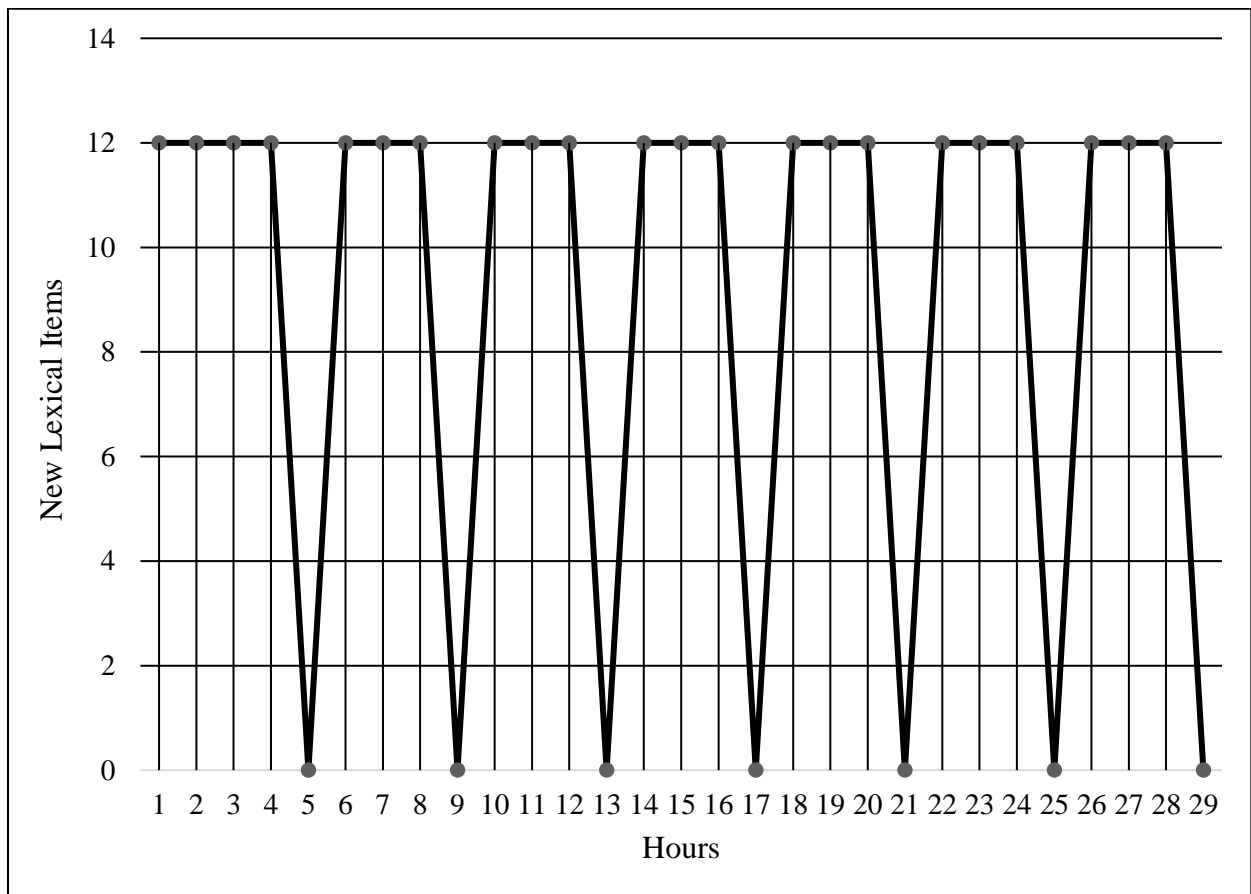
Figure 1. *Scholfield's Vocabulary Rate for an Imaginary Course*

As seen in Figure 1, the course is supposed to be taught over 29 hours of schooling. To help learners of English expand their lexicon, around 12 new words are introduced in each hour of schooling, except for seven hours that are dedicated for reviewing new vocabulary, namely hour numbers 5, 9, 13, 17, 21, 25, and 29. This means that new words are going to be introduced only during 22 hours of schooling throughout the course. To increase vocabulary learning opportunities, the remaining seven hours will be dedicated to review recently learned lexical items.

Similar to what Scholifield proposed, researchers of second language acquisition tend to agree upon the importance of repeating vocabulary in textbooks (Alfotais, 2012; Ghadirian, 2002; Nation, 2001; Pigada & Schmitt, 2006; Rott, 1999; Web, 2007). To enhance the learning of vocabulary items, language learners generally need to be exposed to newly learned vocabulary items a number of times. This would enable the learners to gain a deeper understanding of these new words. Accordingly, "the ability to better memorize a word comes hand in hand with the number of times this word has been recycled" (Alfotais, 2012, p. 20). Nation further (2001) pinpointed that the recycling of vocabulary does not only increase the chances of learning words, but also enables learners to reinforce the meaning of these words.

Researchers have suggested different numbers for recycling vocabulary in textbooks that would enhance the learning of vocabulary (Ghadirian, 2002; Nation, 1990; Pigada & Schmitt, 2006; Rott, 1999; Web, 2007). They strongly indicated that a repeated exposure of more than 10 times will increase the chance of learning a word. Other studies suggested a repetition range of 5 to 20 times would help to facilitate the vocabulary learning. Schmitt (2008) indicated that

> Teachers and materials writers need to think about vocabulary learning in longitudinal terms, where target lexical items are recycled over time in a principled way. From memory research, we know that most forgetting occurs soon after the learning session

and then eventually slows down (Baddeley, 1990), so the first recycling are particularly important and need to occur quickly (p. 343).

To conclude, reducing the number of new vocabulary items would help to allow recently introduced lexical materials to be recycled, which will enable learners of English to acquire the new vocabulary. Yet, despite the importance of vocabulary selection, coverage, and recycling that researchers tend to agree upon, what textbooks include could be far different from the theories. Therefore, it is necessary to examine the types and amounts of vocabulary items presented in language textbooks in order to determine their appropriateness for the intended audience. A number of studies that investigated the lexical input are discussed in the following section.

2.7 RELATED STUDIES ON VOCABULARY INPUT PROVIDED IN BOOKS

During the past few decades, language learning researchers have attempted to measure and investigate the lexical coverage and loading of English language textbooks. A considerable amount of those researchers have utilized corpus linguistic tools as the main methods to analyze the collected data (Alsaif & Milton, 2012; Matsuoka & Hirsh, 2010). McEnery, Xiao, and Tono (2006) defined corpus as "a collection of machine readable authentic texts (including transcripts of spoken data) which is sampled to be representative of a particular language or language variety" (p. 5). Corpus linguistic tools, such as *Range* (Coxhead, Nation, & Heatley, 2002), provide accurate data and descriptions of language use which resulted in greatly increasing its use in vocabulary acquisition research during the recent decades.

Matsuoka and Hirsh (2010) conducted a corpus-based study regarding the lexical coverage of a course book, entitled *New Headway Student's Book Upper-Intermediate*. The computer program Range was used to analyze all 44,877 tokens from the 12 chapters. The focus of the linguistic analysis was based on word frequency and repetition. To analyze the lexical

items, the course book was compared against three different existing word lists, namely K1, K2, and Academic Word List (AWL). The findings showed that about 2,000 words were taken from the most frequent English words. This came as no surprise since learners at the upper-intermediate level of proficiency are expected to know them. These 2,000 words reached a lexical coverage of 93.4% in the book. When counting additional words from the Academic Word List, the lexical converge rose up to 95.5%.

In addition to measuring the repetition of word families in the book, the researchers randomly selected six words out of 187 word families to include in a case study. Then, they looked at the number of their occurrences in the book along with their part of speech and inflectional and derivational forms. *Breakfast*, *invite*, *hotel*, *island*, *crash*, and *conversation* were the randomly chosen words. The presentation of these six words can be seen in Table 4.

Table 4
Results of Matsuoka & Hirsh's (2010) Case Study

Selected Words	Frequency	Part of Speech (Frequency in Parentheses)	Number of Inflections and Derivations	Number of Collocations
Breakfast	5	Noun	1 inflection: breakfast	4
Invite	7	Verb (6) Noun (1)	3 inflections: invite, invites, invited 1 derivation: invitation	5
Hotel	8	Noun	2 inflections: hotel, hotels	9
Island	10	Noun	3 inflections: island, islands, island's	11
Crash	15	Noun (12) Verb (3)	2 inflections: crash, crashed	14
Conversation	70	Noun	2 inflections: conversation, conversations	19

As shown in Table 4, the highest number of occurrences was 70 for the word *conversation*. Yet, the only part of speech used for this word was a noun. In a similar manner, no single occurrence of derivational form was found for this word; only two inflectional forms were presented in the course book, which were *conversation* and *conversations*. As for the rest of the six words, they occurred about 5-15 times in the course book. Despite the fact that the book is intended for upper-intermediate learners, only one of these words had a derivational form, which was *invitation*. The researchers concluded that the book offered great learning opportunities for language learners to deepen their lexical knowledge by knowing words from the GSL and the AWL besides the most frequent words. The shortcomings that the researchers found were related to the repetition of words. A large number of words were only presented in the course book once, which made it extremely difficult for learners to acquire these words. In order to increase the vocabulary learning opportunities, it was suggested to supplement the book with other reading materials in order to allow the learners to encounter these less repeated lexical items.

Scholfield (1991) conducted a vocabulary rate study to investigate the vocabulary loadings of a beginning English language course book, *The Cambridge English* Course by Swan and Walter (1987). The results showed that a total of 1,082 new words were included in the book. The phrasal verbs, idiomatic expressions, and compounds were counted as single lexemes. The book included a total of 33 chapters, and three of those chapters contained the cyclic elements. Figure 2 presents the rate of vocabulary loadings throughout the chapters.
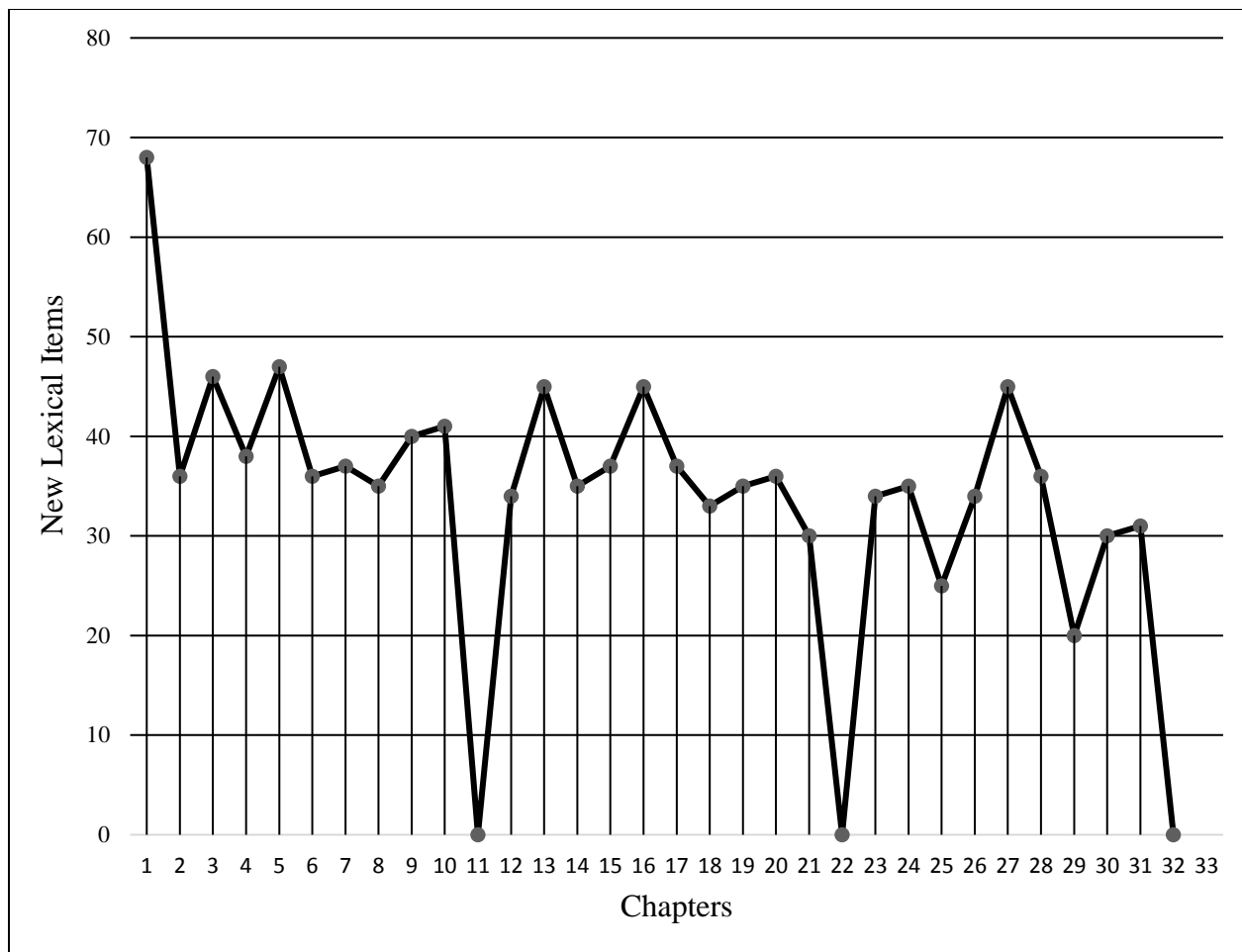
Figure 2. *Scholfield's Vocabulary Rate for an English Course*

As seen in Figure 2, chapters 11, 22, and 32 did not contain any new words since they were dedicated to recycling the lexical items mentioned in previous chapters. The researcher also pointed out that the rate of vocabulary distribution started to decrease beginning in the second cycle, after chapter 11. That is, the plots in the first cycle were above the average amount, whereas most plots in the third cycle were below the average. This way of introducing new vocabulary items gives the chance for learners to deepen their knowledge of recently introduced lexical items.

With regard to the Saudi Arabian context, only one study was conducted to measure the lexical items presented in Saudi K-12 schools. The study was conducted by Alsaif and Milton (2012) who employed a corpus-based methodology to examine the distribution of vocabulary in English language textbooks utilized from grade 6 to 12 in Saudi Arabian public schools. The researchers focused their linguistic analysis on finding the lexical coverage along with the distribution of words and their frequency bands. Two corpus programs were used to carry out the linguistic analysis which were *Range* (Coxhead, Nation, & Heatley, 2002) and *Compleat Lexical Tutor* (Cobb, 2008). Table 5 presents the overall lexical coverage of all materials used from grade 6-12.

Table 5
Overall Frequency Distribution of Lexical Items in the Texts

Word Lists	Word Types	Tokens	Lexical Coverage
K1	2,699	180,551	82.48%
K2	1,524	16,898	90.20%
K3	818	5,543	92.73%
K4	476	3,034	94.11%
K5	304	2,092	95.06%
K6-K15	1147	6,693	98.14%
Additional Words	1229	4,072	100%

As illustrated in Table 5, words accounted from the first 1,000 most frequent English words (K1) and the second 1,000 most frequent English words (K2) reached a lexical coverage of nearly 90%, indicating that more opportunities were provided to learn the most frequent words in English. This coverage was thought to be helpful in facilitating the students' learning since K1 and K2 words occur very frequently in authentic texts. Once students encountered these words in classes, they eventually would be able to use them successfully. The second analysis that Alsaif and Milton (2012) carried out involved investigating the amount of new vocabulary input provided for students each year, as seen in Figure 3.
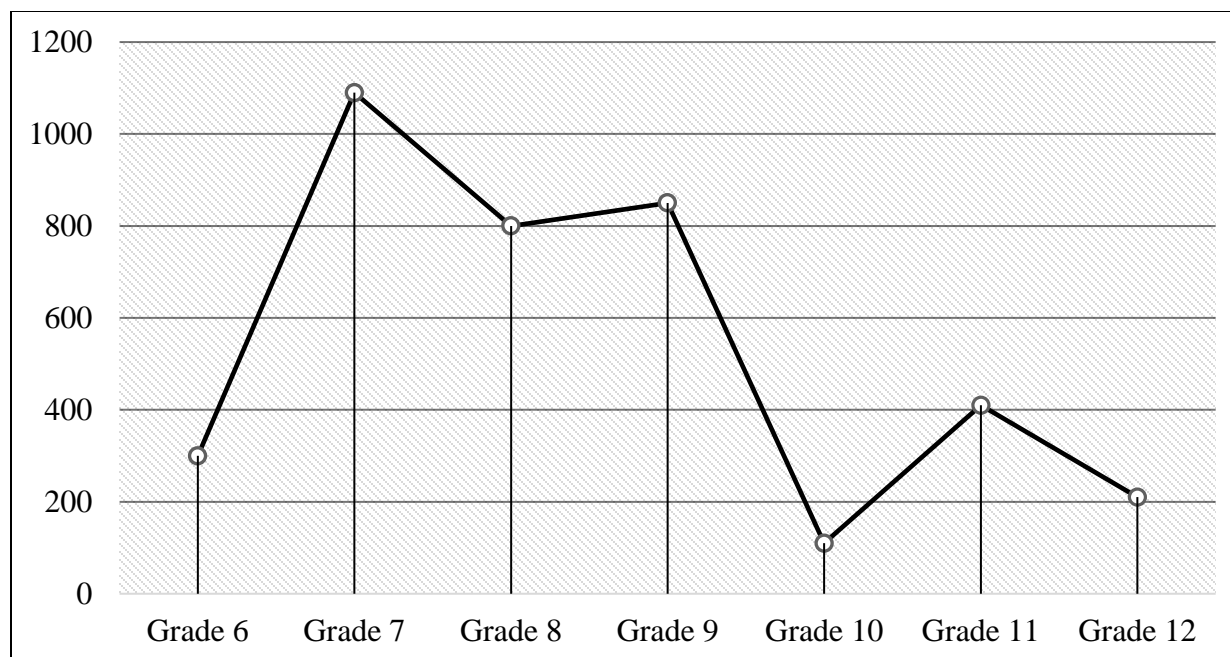
Figure 3. *Number of New Lexical Items Presented Each Year*

Figure 3 illustrates that most new lexical items were presented during the first four years of English instruction, namely from grade 6 to grade 9. In other words, the vocabulary input decreased dramatically in high school, where fewer lexical items were introduced in Saudi EFL textbooks used for grades 10, 11, and 12. In return, this gave high school students the opportunity to encounter more exposures of previously-learned words, which was helpful for long-time learning. However, as a result of introducing less than 400 words in high school, it was estimated that the vocabulary size of Saudi EFL learners would be limited to the number of words introduced mainly in elementary and intermediate schools. Thus, it would be crucial for the Saudi EFL learners to expand their vocabulary size, especially at the high school level.

2.8 CHAPTER CONCLUSION

There is a lack of research in regard to investigating the lexical coverage and the characteristics of words presented in EFL books utilized in Saudi Arabian schools. To the best of the researcher's knowledge, there is only one study that quantitatively examined the words

included in the previous EFL curriculum used in Saudi K-12 schools. With a new EFL

curriculum being implemented in these type of schools, it is of a vital importance to do further

research to examine various aspects related to lexical items included in the new curriculum. The

present student aims to gather both quantitative and qualitative data in order to determine the

extent to which input from textbooks contributes to building the students' lexicon. Information

regarding the design of this study is presented in Chapter 3.

CHAPTER 3: METHOD

Considering the fact that there is only one study that examined the coverage of lexical

items of EFL books that were utilized in Saudi schools (Alsaif & Milton, 2012), there is a lack of

research for investigating the vocabulary learning opportunities that Saudi EFL learners receive.

More importantly, there is a need to conduct research that focuses on carrying out an in-depth

investigation regarding opportunities that would assist Saudi EFL learners to deepen their

knowledge of words. Indeed, more qualitative data needs to be gathered in order to better

understand the nature in which words in textbooks are being presented in terms of their parts of

speech, collections, lexical bundles, inflectional and derivational forms, and others. If not, it

would be crucial to determine how words are well presented and covered in textbooks based on

only measuring the number of words and their frequency. To the best of the researcher's

knowledge, there is no study that was conducted to examine the qualitative characteristics of

words presented in books used in the context of the Middle East. Having this in mind, this study

aims to carry out the first in-depth analysis to investigate various aspects of characteristics and

distributions of words covered in EFL textbooks utilized in Saudi K-12 schools.

3.1 RESEARCH QUESTIONS

RQ 1: To what extent are words from K1, K2, AWL, and additional word lists

represented in Saudi EFL textbooks?

RQ 2: How many newly-introduced word types are presented per grade?

RQ 3: How often are newly-introduced words repeated in the textbooks?

RQ 4: What types of learning opportunities do Saudi EFL textbooks present to acquire

new vocabulary?

3.2 MATERIALS

Since the purpose of this study is to investigate the distribution and characteristics of lexical items of EFL textbooks utilized in Saudi K-12 public schools, the selected corpora consisted of 36 EFL textbooks that are currently used in this type of school. With the assistance of the English Language Curriculum Department in the Saudi Arabian Ministry of Education, three different EFL curricula emerged in collaboration with three global organizations: *Macmillan Press*, *Cengage Learning*, *MM Publications* (Ministry of Education, 2016b). These curricula started to be officially implemented during the academic year of 2011-2012 in K-12 public schools. Figure 4 shows all EFL curricula that are used in different regions of Saudi Arabia.
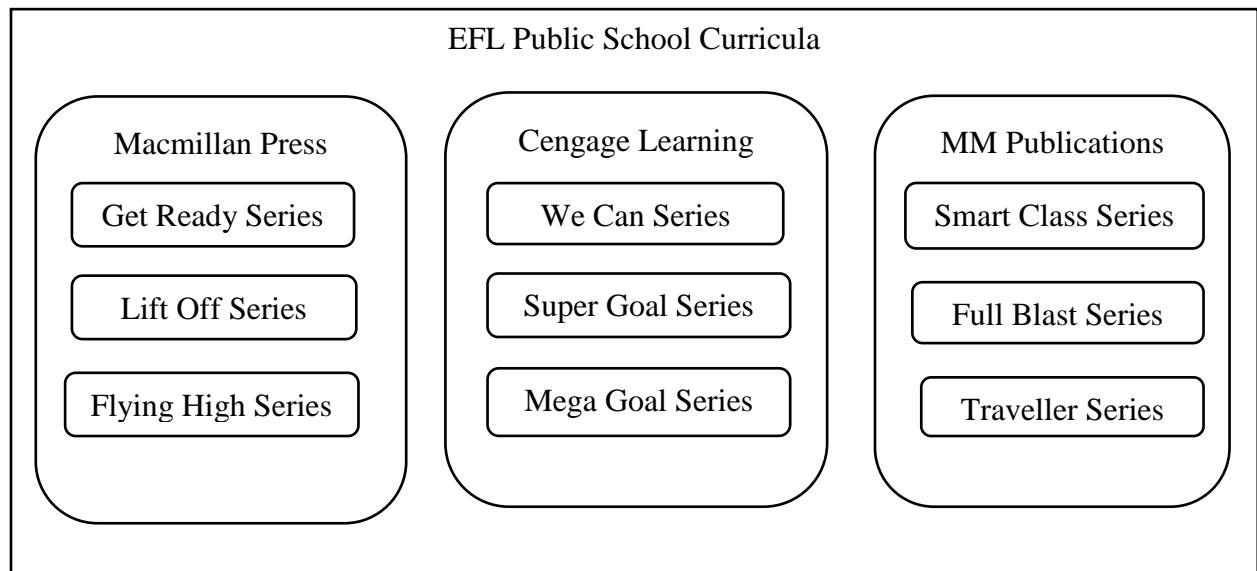
Figure 4. *All EFL Textbooks Utilized in Saudi K-12 Public Schools*

Each organization was asked to develop an EFL curriculum which was further divided into three different textbook series. The first series is for elementary school students from grades 4 to 6, the second series is intended for intermediate school students from grades 7 to 9, and the

last series is developed for secondary school students from grades 10 to 12. Each developed

curriculum is being utilized in different parts of the country in order to investigate which

publisher's textbooks correspond better to the needs of young Saudi EFL students. Since

Macmillan curriculum is used in public schools located in the capital city of Saudi Arabia,

Riyadh, which serves the largest population of students, they are chosen for the analysis in the

present study.

Macmillan developed a custom-made English as a foreign language curricula for Saudi

K-12 schools (Ministry of Education, 2016c). The curricula consist of three series, each

comprising six levels. The first series, *Get Ready*, is specifically intended for elementary

students. The second series, *Lift Off*, is designed for intermediate students. The last series, *Flying

High*, is developed for high school students. Three different books are developed for each level:

a student's book, *a workbook*, and *a teacher's book (teaching manual)*. For the purpose of this

study, only the students' books and workbooks were analyzed. Table 6 shows the number of

textbooks for each school level.

Table 6
Number of EFL Textbooks Used in Saudi Public Schools

School Level	Grade	Number of Student Book	Number of Workbook
Elementary	4	2	2
Get Ready Series	5	2	2
	6	2	2
Intermediate	7	2	2
Lift Off Series	8	2	2
	9	2	2
Secondary	10	2	2
Flying High Series	11	2	2
	12	2	2
Total (=36)		18	18

The textbooks are thematically organized according to various general topics, including family, technology, and law, and integrate all four language skills: reading, writing, listening, and speaking. Since English language is taught as a required subject in grades 4, 5, and 6 for two 45-minute classes per week and in grades 7 to 12 for four 45-mintue classes weekly, elementary school textbooks have fewer units and lessons.

3.3 MODIFICATION OF THE TEXTS

All 36 textbooks were downloaded electronically from the Saudi Arabian Ministry of Education website as PDF files. Since students in each school grade utilize four textbooks per year (two student's books and two workbooks), these four textbooks were compiled as one PDF file on the website. After saving the nine PDF files in which each file represented one grade level, the researcher converted these files to text files.

In order to ensure the accuracy of word counting, the researcher examined each text file to remove function words, including articles, prepositions, and conjunctions, as well as misspellings, such as *Engllish* and *Engli0h*. A further revision was also done to exclude all Arabic scripts, phonetic transcriptions, pictures, numbers, tables, figures, indexes, page numbers, and glossaries. These revisions allowed the researcher to analyze all English content words presented in the texts, which are nouns, verbs, adjectives, and adverbs. Table 7 presents the number of words that are included in the Saudi EFL curriculum.

Table 7
Number of Lexical Items Presented in K-12 Texts

Texts		Word Families	Word Types	Tokens
Elementary School	Grade 4	231	257	1,387
	Grade 5	378	435	2,099
	Grade 6	516	631	3,417
Intermediate School	Grade 7	1,169	1,625	23,851
	Grade 8	1,528	2,277	35,559
	Grade 9	1,828	2,895	37,879
High School	Grade 10	2,670	4,136	52,085
	Grade 11	3,226	5,037	54,191
	Grade 12	2,282	3,643	42,049

As it is illustrated in Table 7 above, English language textbooks utilized in elementary schools had the smallest number of words. When looking at intermediate school textbooks, the number of words rose to an average of 1,508 word families, 2,266 word types, and 32,430 tokens. In a similar manner, the number of words increased dramatically in textbooks utilized in high school levels, with the exception of grade 12 texts. A total of 2,670 word families, 3,114 word types, and 52,085 tokens were presented in grade 10 texts. In grade 11, the number of words rose to 3,226 word families, 5,037 word types, and 54,191 tokens. Surprisingly, the number of words included in grade 12 texts dropped down to 2,282 word families, 3,643 word types, and 42,049 tokens.

3.4 RESEARCH DESIGN

Since the study aims to carry out an in-depth investigation regarding the distribution and characteristics of lexical items presented in the English language curriculum utilized in Saudi schools, Mixed Method Research (MMR) is chosen for the study. Briggs, Coleman, and Morrison (2012) defined MMR as "the broad types of research in which elements or approaches from quantitative and qualitative research are combined or mixed in a research study" (p. 122-123). The generated quantitative and qualitative data from this study helps to provide a better

understanding of the relationship between the lexical coverage of Saudi EFL textbooks and the extent to which they provide vocabulary learning opportunities to the learners.

Therefore, the type of Mixed Method design that the present study followed was Explanatory Sequential, meaning that the study begins with gathering and analyzing quantitative data first, and then moving forward to the collection and analysis of qualitative data. This enabled the researcher to obtain more detailed information regarding the coverage of lexical items. Figure 5 illustrates the process that the researcher followed to collect the data.

| Quantitative Data Collection and Analysis | → | Follow up with | → | Qualitative Data Collection and Analysis | → | Interpret |
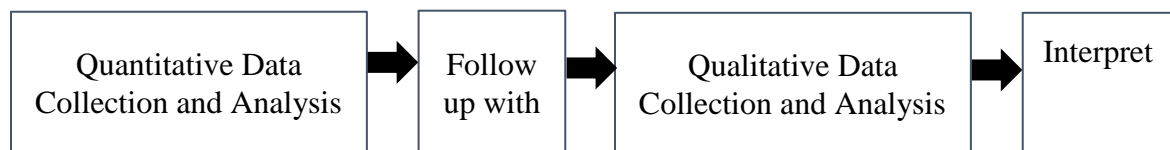
Figure 5. *Sequence of Gathering the Mixed Method Data*

With regard to quantitative information, the type of data that was collected in the present study revolved around the coverage of K1, K2, AWL, and additional words, the distribution of newly- introduced words per grades, and the recycling range of these newly-introduced words. Following up the analysis of quantitative data, the researcher began to collect qualitative data to better determine whether the lexical coverage would assist Saudi EFL learners to deepen their knowledge of words. To do that, case study design was chosen for collecting and analyzing the qualitative data since it helps to "achieve as full an understanding of the phenomenon as possible" (Merriam, 1998, p. 28). A case study, as defined by Creswell (2007) and Merriam (1998), is investigating a phenomenon in a bounded time and place which will help to gain "an in-depth understanding of the situation and meaning for those involved" (p. 19).

The selected case study for the present study revolved around analyzing words from the K1, K2, and AWL word lists that occurred in the selected materials. According to Nation (2001)

and Milton (2009), the chances of learning a word in a second or foreign language increases when it has been recycled five times or more. With that being the case, these words must have been recycled at least five times per text file in order to be considered candidates for the case study. This analysis contributes to better understanding of the depth of word knowledge that these books provide for learners.

The focus of the investigation for each selected word was placed on the frequency, part of speech, derivational forms, inflectional forms, and collections. These words were randomly selected from each text file utilized per grade to better examine the vocabulary learning opportunities that are provided. With nine text files being included in the present study, the total number of the random selected words for the case study analysis was nine. From each text file compiled from elementary school textbooks, three words were selected from the K1 word list since these words are learned first. As for intermediate school texts, three words are selected from the K2 word list. Lastly, three words from the AWL are selected from high school textbooks.

According to the Saudi Ministry of Education (2016a), the overall purpose of teaching English language at Saudi K-12 public schools is to help students improve their skills of English and communicate more effectively in real-life (everyday) situations. Since words from the K1 and K2 word lists typically make up the highest percentage of words in any written or spoken discourse, including everyday conversations and TV programs, the researcher emphasized in investigating these most frequent words presented in the textbooks. In addition, the researcher placed her emphasis in counting words from the AWL, especially in the books used in high school, because most students need to build their academic vocabulary prior to enrolling in universities and colleges in Saudi Arabia.

3.5 DATA ANALYSIS

To carry out the investigation of lexical items, corpus linguistic tools, such as *Range*, were used to analyze the selected materials. Hunston and Francis (2000) defined corpus linguistic-based analysis as "a way of investigating language by observing large amounts of naturally-occurring electronically-stored discourse, using software which selects, sorts, matches, counts and calculates" (p. 15). As Hoffmann, Evert, Smith, Lee, and Berglund-Prytz (2008) indicated, corpus linguistics "is an essentially quantitative method, meaning that corpus linguistics tend to count features of language (or have the computer count them) as part of their analysis of linguistic features" (p. 18).

RANGE program (Coxhead, Nation, & Heatley, 2002) was used to answer the first research question. The purpose of utilizing this program was to identify lexical coverage of elementary, intermediate, and high school textbooks. To carry out the analysis, the researcher compared each text file against eight baseword lists in order to classify all of the words presented in the textbook according to their types. These lists included the spellings of both American and British English, such as *colour* and *color*. Words were grouped in the lists under headwords, also known as word families. For instance, the words *became*, *becomes*, and *becoming* were listed under the word family *become*.

Baseword list 1, 2, and 3 were already existing lists that included the most frequent words in English as well as words commonly found in various academic contexts. Baseword list 4, 5, 6, 7, and 8 were specifically developed for this study in order to classify words presented in the textbooks beyond academic and most frequent words. Table 8 presents information about all eight baseword lists used in this study.

Table 8

Baseword Lists Generated for the Lexical Analysis

Baseword List	Type of Words
1	First 1,000 most frequent words (K1)
2	Second 1,000 most frequent words (K2)
3	Academic words (AWL)
4	Proper nouns
5	Technology words
6	Arabization words
7	Less frequent words
8	Off list words

3.5.1 EXISTING BASEWORD LISTS. Baseword list 1 consists of first 1,000 most

frequent words in English. There are a total of 1,000 word families and 4,090 word types

included in this list. Baseword list 2 is made up of second 1,000 most frequent words in English.

The list includes 1,000 word families and 3,706 word types. Both of these lists are generated

from West's (1953) General Service List (GSL).

Baseword list 3 includes all words from the Academic Word List (AWL). The list is

developed by Averil Coxhead (2002). It includes a total of 570 word families as well as 3,082

word types found in academic contexts.

3.5.2 ADDITIONAL BASEWORD LISTS. Baseword list 4 includes proper nouns

which are specific names of people, organizations, things, and places. The number of word types

found in all texts for proper nouns is 816. Examples of proper nouns included in this baseword

list are *Adam*, *Portuguese*, *Venezuela*, and *Olympics*.

Baseword list 5 includes technology words. A total of 40 word types are included in the

list. Examples of technology terms that are listed are *blog*, *computer*, *website*, and *online*.

Baseword list 6 contains Arabic words that do not have any equivalent in English. *Bedouin* is an example of Arabization words included in the list. This term refers to Arab tribes who live outside the city, mostly in the desert. This list includes a total of 16 word types.

Baseword list 7 is made up of less frequent words that are presented in the text files. A total of 1,865 word types are included in the list from the third 1,000 most frequent word list (K3) to the seventeenth 1,000 (K17). To ensure the accuracy of word counting, the researcher removed all AWL words from the K3 word list. Examples of words from the K17 are *decongestant*, *crick*, and *Aldrin*.

Baseword list 8 consists of all remaining words that do not belong to any of the previous classifications. A total of 259 word types are included in this list. Most of these words are nouns such as *sunglasses* and *backpack*. The rest of words are adverbs, adjectives, and phrasal verbs. Examples of these words are *outstanding*, *forever*, and *check out*.

After comparing the textbooks against the eight baseword lists to classify words and identify their lexical coverage, the researcher measured the type-token ratio (TTR) in order to determine the extent to which words from each baseword list were repeated in the textbooks. In order to obtain the TTR, the total number of word types accounted from each baseword list was divided by its total number of tokens. According to Thomas (2005), "the range falls between a theoretical 0 (infinite repetition of a single type) and 1 (the complete non-repetition found in a concordance)." The results obtained from this TTR analysis helped the researcher make a general estimate regarding the repetition of words.

Although Range software presented quantitative data about new vocabulary words with their range of occurrences among texts, it was impractical and almost impossible to check all lists of words in order to calculate how many new words are introduced per grade and how many

times each new word has been recycled manually. With that being the case, the researcher used *Compleat Lexical Tutor*, a web-based recourse created by Tom Cobb (2008) to answer the second and third research questions. The name of the corpus tool that was used to carry out this lexical analysis is *Text-Lex Compare*. This tool allowed the researcher to subtract one text file or more from another in order to gather information regarding newly-introduced words, shared words, and repetition of words. In order to determine how many new words were introduced in Grade 6 with their degree of repetition for example, the researcher compared Grade 6 text file against Grade 4 and 5 text files. The results obtained from the analysis presented quantitative data regarding the number of not shared or new words in grade 6 textbooks as well as their number of occurrences.

To answer the fourth question, each text file was run separately in *AntConc* (Anthony, 2014) in order investigate words that are newly-introduced per grade. The software enabled the researcher to carry out a qualitative analysis by analyzing each occurrence of target words. This was done by looking at the concordance lines. The software allowed the researcher to investigate different aspects of the depth of word knowledge, including collocates, parts of speech, derivations, inflections, and frequency. All of these tools were used in this study in order to gather qualitative data regarding the nature of words that are being presented in each text file.

CHAPTER 4: RESULTS

All English language textbooks utilized in Saudi public schools were first analyzed quantitatively in order to know how many word types from K1, K2, AWL, and additional baseword lists, as well as the number of new vocabulary input, is presented per grade. The second stage involved a qualitative analysis. This stage helped to investigate the learning opportunities that are provided for Saudi EFL learners to deepen their knowledge of words.

4.1 RESEARCH QUESTION 1

To what extent are words from K1, K2, AWL, and additional word lists represented in the Saudi EFL textbooks?

Prior to carrying out a year-by-year vocabulary profile analysis, all texts used from grade 4-12 (n=9) were compared together against the eight baseword lists: the K1, K2, AWL, proper nouns, technology, Arabization, less frequent, and off-list words. This allowed the researcher to know holistically the total number as well as type of vocabulary words presented to the Saudi EFL learners. The overall coverage of words is seen in Table 9.

Table 9
Cumulative Lexical Coverage in the Whole Corpus

Baseword Lists	Word Families	Word Types	Lexical Coverage	Cumulative Coverage
1 First 1,000 most frequent words (K1)	971	2,731	81.22%	81.22%
2 Second 1,000 most frequent words (K2)	767	1,592	8.86%	90.08%
3 Academic words (AWL)	429	947	3.07%	93.15%
4 Proper nouns	819	819	2.40%	95.55%
5 Technology words	40	40	0.14%	95.69%
6 Arabization words	28	28	0.05%	95.74%
7 Less frequent words	1,432	1,865	3.80%	99.54%
8 Off list words	259	259	0.44%	100%
Total	5,178	8,281	100%	100%

Table 9 revealed that all English language textbooks used from grade 4 to 12 in Saudi public schools presented a total of 5,178 word families and 8,281 word types. The highest lexical coverage of the nine texts were accounted for words in the K1, K2, less frequent words, and AWL, respectively. In terms of word frequency in the texts, the highest percentage of lexical items covered by the lists were accounted from the K1 word list at 81.22%. The second emphasis was placed on covering words from the K2 word list as they accounted for approximately 9% coverage. Words from the baseword list 7 had the top three percentage of lexical coverage at 3.80%. Because of that, the researcher conducted a further analysis to identify frequency word lists that these less frequent words were derived from. Table 10 shows the frequency of words from the baseword list 7.

Table 10
Overall Coverage of Less Frequent Lexical Items

Less frequent words	Word Families	Word Types	Lexical Coverage %
K3	277	433	1.98
K4	324	429	0.65
K5	250	315	0.41
K6	169	210	0.22
K7	98	116	0.16
K8	87	99	0.08
K9	64	77	0.04
K10	31	43	0.06
K11	37	46	0.10
K12	24	28	0.03
K13	17	18	0.01
K14	13	18	0.02
K15	14	15	0.02
K16	6	6	0.01
K17	12	12	0.01
Total	1,423	1,865	3.80

As illustrated in Table 10, numerous levels of low frequency words were included in the textbooks utilized from grade 4 to 12, which ranged from the third 1,000 (K3) to the seventeenth 1,000 frequent words (K17) in English. The analysis revealed that the higher the frequency of words, the higher the coverage of words. That is, words that appear more frequently in English, such as the K3, K4, and K5 word lists, made up the highest percentage of lexical coverage. Words from the K3 word list made up the highest coverage of words at 1.98%. The second top coverage of lexical items was accounted for words from the K4 word list at 0.65%. In comparison, words from the K12 to K17 word lists had the lowest lexical coverage, which varied from 0.03 to 0.01%.

4.1.1 YEAR-BY-YEAR VOCABULARY PROFILE. All English language texts utilized in Saudi public schools from grade 4 to 12 were analyzed separately. Comparing each text file utilized per grade separately against these eight baseword lists assisted in measuring the lexical density of the input provided for students each year. The results of this year-by-year vocabulary profile can be seen in Table 11.

Table 11

Coverage of Baseword Lists

Textbook	BW 1		BW 2		BW 3		BW 4		BW 5		BW 6		BW 7		BW 8	
	TTR	*Tokens %*	*TTR*	*Tokens %*	*TTR*	*Tokens %*	*TTR*	*Tokens %*	*TTR*	*Tokens %*	*TTR*	*Tokens %*	*TTR*	*Tokens %*	*TTR*	*Tokens %*
Grade 4	0.14	78.36	0.34	12.04	0.26	4.03	0.73	1.15	0	0	1	0.06	0.32	3.79	0.37	0.48
Grade 5	0.15	80.94	0.40	8.80	0.30	4.05	0.64	1.63	0	0	0.05	0.09	0.48	3.91	0.58	0.52
Grade 6	0.14	77.97	0.33	10.15	0.15	3.41	0.44	2.41	1	0.05	0	0	0.36	5.39	0.43	0.61
Grade 7	0.04	80.96	0.13	9.18	0.20	1.19	0.10	5.40	0.12	0.20	0.41	0.05	0.23	2.35	0.22	0.66
Grade 8	0.04	81.58	0.14	10.19	0.18	1.54	0.12	3.08	0.18	0.13	0.72	0.03	0.21	2.68	0.19	0.72
Grade 9	0.04	81.54	0.17	9.78	0.20	2.51	0.19	2.73	0.13	0.26	0.15	0.10	0.28	7.44	0.23	0.50
Grade 10	0.04	83.34	0.20	8.44	0.29	2.85	0.28	2.11	0.21	0.15	0.33	0.01	0.40	2.73	0.45	0.34
Grade 11	0.04	83.08	0.23	7.51	0.29	3.60	0.33	1.94	0.36	0.08	0.42	0.01	0.47	3.43	0.49	0.33
Grade 12	0.04	85.93	0.25	6.76	0.29	3.34	0.44	1.02	0.45	0.06	1	0.01	0.42	2.61	0.48	0.27

As indicated in Table 11, words accounted from the baseword list 1 made up the largest parentage of lexical coverage in all texts, ranging from 78.36 to 85.93%. The average type-token ratio (TTR) for the baseword list 1 was 0.07, which indicates a great deal of repetition. The analysis also showed that the second emphasis was placed on introducing words from the baseword list 2 since they accounted for the second largest coverage of lexical items, ranging from 6.76 to 12.04%. Additionally, these words had a relatively low type-token ratio, with an average of 0.24. This means that a sufficient amount of repetition is provided for these types of words.

The table also revealed that words from the baseword lists 3 and 7 had nearly the third top percentage of lexical coverage in the texts. Words accounted from the baseword list 3 had a coverage of 1.19 to 4.05% and an average type-token ratio of 0.24. When looking at words from the baseword list 7, they had an average type-token ratio of .35 and a lexical coverage of 2.61 to 7.44%. Since this baseword list consists of less frequent words, it was imperative to investigate their distribution based on frequency level, ranging from the K3 to K17. Table 12 illustrates the coverage of less frequent words presented each year in the Saudi K-12 schools.

Table 12

Coverage of Less Frequent Words

BW	Elementary School Texts						Intermediate School Texts						High School Texts					
	Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 9		Grade 10		Grade 11		Grade 12	
	Word Type	*Token %*	*Word Type*	*Token %*	*Word Type*	*Token %*	*Word Type*	*Token %*	*Word Type*	*Token %*	*Word Type*	*Token %*	*Word Type*	*Token %*	*Word Type*	*Token %*	*Word Type*	*Token %*
K3	12	1.30	28	1.48	26	1.31	78	1.13	145	1.28	261	4.47	306	1.23	344	1.54	236	1.30
K4	6	0.81	12	0.73	14	0.93	42	0.24	88	0.43	110	1.19	219	0.37	263	0.50	144	0.27
K5	5	1.01	10	0.51	9	0.19	27	0.35	46	0.27	55	0.81	123	0.47	186	0.62	94	0.39
K6	3	0.51	5	0.39	6	0.74	16	0.14	31	0.19	27	0.26	73	0.17	111	0.26	45	0.16
K7	2	0.10	4	0.26	4	0.59	11	0.16	10	0.13	21	0.23	27	0.06	66	0.13	36	0.14
K8	1	0.03	1	0.03	6	0.34	10	0.12	11	0.13	14	0.1	34	0.04	50	0.07	18	0.08
K9	0	0	1	0.03	6	0.34	8	0.06	5	0.02	7	0.05	26	0.07	37	0.1	11	0.05
K10	1	0.03	0	0	6	0.34	2	0.04	6	0.09	12	0.13	9	0.06	17	0.06	18	0.11
K11	0	0	0	0	2	0.25	7	0.03	4	0.03	11	0.14	22	0.1	19	0.05	8	0.04
K12	0	0	2	0.20	1	0.12	4	0.04	5	0.03	2	0.01	11	0.03	14	0.01	4	0.01
K13	0	0	1	0.09	1	0.12	1	0	0	0	1	0	7	0.03	9	0.02	2	0
K14	0	0	2	0.19	1	0.12	2	0.01	2	0.01	4	0.05	10	0.06	9	0.03	8	0.03
K15	0	0	0	0	0	0	6	0.03	3	0.03	0	0	6	0.01	6	0.01	1	0.01
K16	0	0	0	0	0	0	1	0	2	0.01	1	0	1	0	3	0.01	3	0.01
K17	0	0	0	0	0	0	1	0	3	0.03	1	0	4	0.03	5	0.02	3	0.01
Total	30	3.79	66	3.91	82	5.39	216	2.35	361	2.68	527	7.44	878	2.73	1,139	3.43	631	2.61

As seen in Table 12, words from the K3 word list had the largest number of word types as well as the highest coverage of lexical items in all texts, ranging from 1.13 to 4.47%. Words from the K16 word list, on the other hand, had the smallest number of word types and coverage. As the table illustrated, very low levels of frequency words, such as K16 and K17, were included only in texts used in grades 7 through 12. When looking at the distribution of less frequent words, it appeared that the lower the frequency of words in English, the lower the amount of coverage and word types in the texts. That is, words from the K5 word list had higher coverage and number of types compared to the K6 words in all texts. In a similar manner, the higher the level of school texts, the higher the number of low frequency words, with the exception of grade 12 texts. For instance, the text utilized in grade 5 included 66 less frequent words where grade 4 text only included 30 less frequent words.

Given the fact that hundreds of hours of English instruction are provided for students from grades 4 to 12, students at higher grade levels are thought to be more familiar with the English language. With that being stated, it would be extremely useful to take a deep look regarding the coverage of academic words accounted from grade 10 texts and higher. The academic word list (AWL) includes a total of 570 word families and 3,082 word types. All of these academic words are divided between 10 sublists. The first sublist includes the first most common academic words, such as *available* and *research*. The second sublist consists of the second most frequent academic words like *acquire* and *conclude*, and so on. Knowing which level of frequency is well covered in grades 10, 11, and 12 texts helped to determine the learning opportunities that are provided for the Saudi EFL learners. The coverage of AWL sublists can be seen in Table 13.

Table 13
Frequency Distribution of AWL in High School Texts

AWL Sublists	Grade 10		Grade 11		Grade 12	
	Word Types	Tokens %	Word Types	Tokens%	Word Types	Tokens%
1	79	0.49	97	0.79	71	0.78
2	76	0.46	80	0.65	63	0.63
3	52	0.31	62	0.38	41	0.25
4	54	0.58	59	0.44	52	0.36
5	35	0.24	49	0.32	37	0.2
6	28	0.09	33	0.17	34	0.19
7	33	0.15	43	0.17	27	0.16
8	29	0.21	28	0.26	22	0.34
9	17	0.13	20	0.17	14	0.16
10	9	0.19	12	0.25	7	0.27

As illustrated in Table 13, words from AWL Sublist 1 were accounted for the highest percentage of coverage. More specifically, these words reached a coverage of 0.49% in grade 10, 0.79% in grade 11, and 0.78% in grade 12 texts. The AWL Sublist 2 were accounted for the second highest coverage of academic lexical items, ranging from 0.46-0.65%. When moving down to less common academic words presented in sublists 3-10, the number of word types decreased as well as their coverage in the texts, indicating that prior attention is given to the introduction of the most common academic words.

4.2 RESEARCH QUESTION 2

How many newly-introduced word types are presented per grade?

To answer this question, each text was compared against texts used in previous grades in order to determine how many new words are introduced each year. To illustrate this, words that were shared between the text and previous text, or in two or more text files, were eliminated. Words that were unique in the text (being presented in only one text file for the first time) were

included in this analysis. This comparison allowed the researcher to gather quantitative data

regarding the number of newly-introduced words per grade, which is seen in Figure 6.
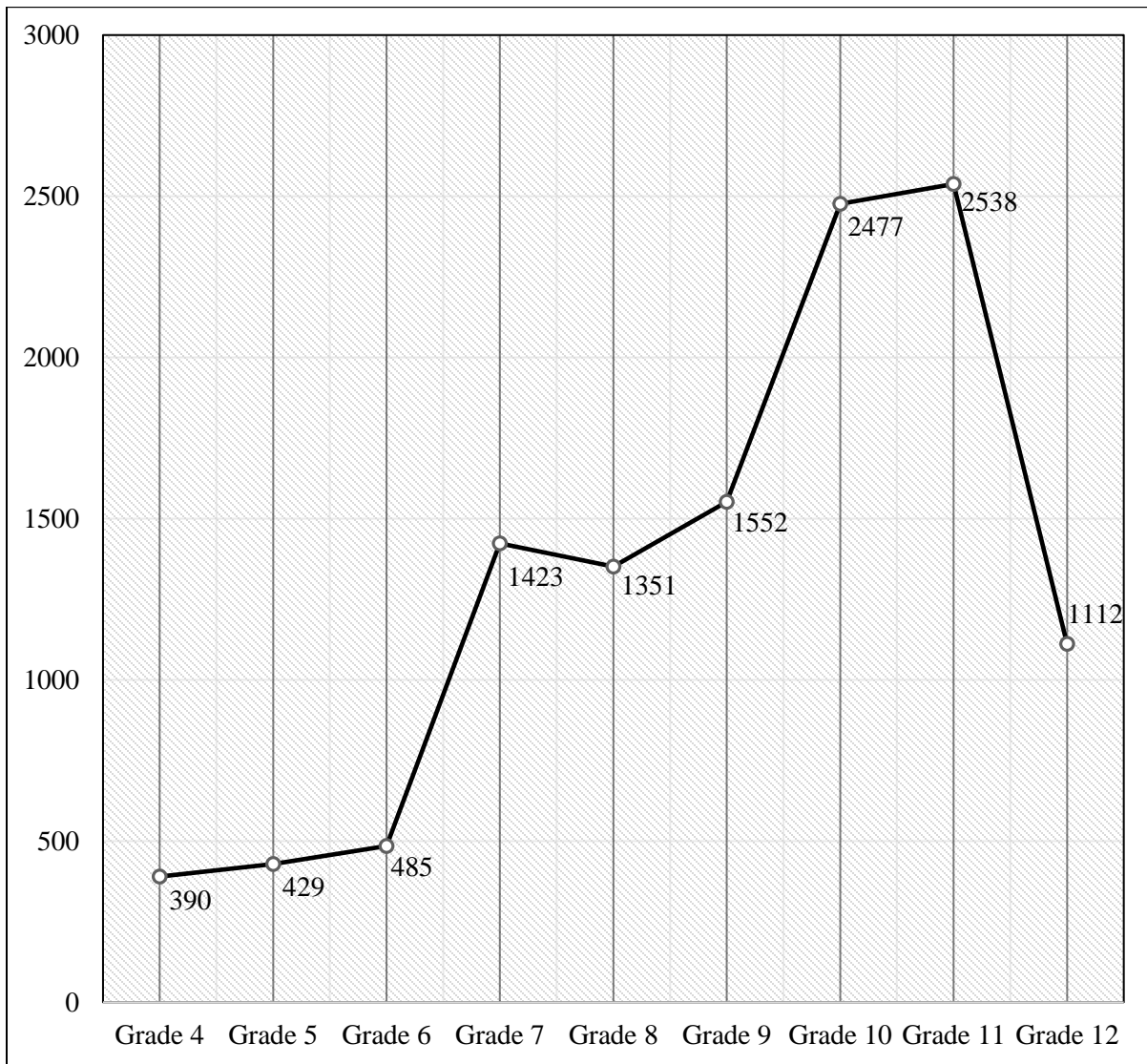
Figure 6. *Number of Newly-introduced Word Types per Grade*

As Figure 6 above shows, there were around 500 new word types presented each year in

elementary schools. Specifically, 390 word types were introduced in grade 4. The number of

newly-introduced words rose to 429 word types in grade 5. The number of new word types

introduced in grade 6 also increased to 485. The figure also revealed that there was a dramatic

rise in the introduction of new word types in intermediate school levels where nearly 1,500

words were introduced in each grade. Moving up to high school, the number of newly-introduced word types spiked in grades 10 and 11 to around 2,500. However, the number of newly-introduced words included in the last grade of high school, grade 12, dropped down to 1,112, which was less than the number of new words presented in intermediate school.

Since it is impossible to introduce all new vocabulary input in class at once, the number of these newly-presented word types per grade was divided throughout the English instructional hours. According to the Saudi Ministry of Education (2016), the total number of English instructional hours offered per year for elementary school students (from grades 4 to 6) is 48. Intermediate school students receive 96 hours, and high school students receive 96. This resulted in having a total of 720 hours of schooling. After eliminating all of the hours that are assigned by the MOE for revisions and testing, the total number of hours offered for instruction in public schools from grades 4 to 12 is 630. More specifically, there are 42 hours of learning that are offered each year for students from grades 4 to 6, and 84 hours for students from grades 7 to 12. Figure 7 presents information about the number of new words presented in each hour of schooling.
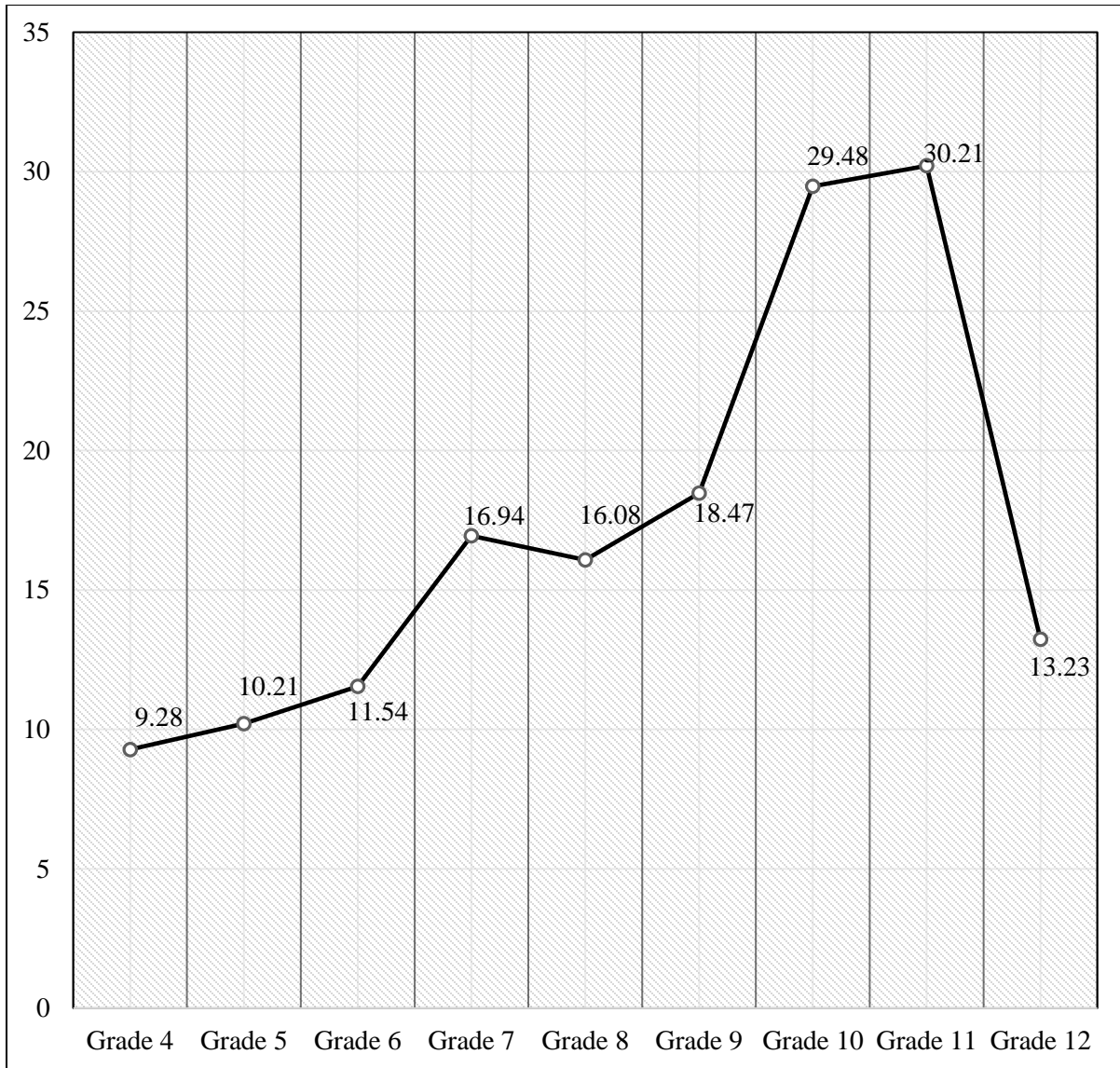
Figure 7. *Number of New Word Types Introduced per Instructional Hour*

As Figure 7 illustrates, the result of dividing the total number of new words, introduced

to each grade by the total number of instructional hours offered per grade, allowed the researcher

to measure how many new word types were thought to be introduced in each instructional hour.

Beginning with elementary school, the results of dividing the 390 words by 42 schooling hours

revealed that around 9 words are introduced each hour in grade 4. When measuring the average

number of new words introduced in elementary school from grade 4 to 6, the analysis revealed

that about 10 words are introduced for the first time during each hour of instruction. Moving up to intermediate school, around 17 words are presented for the first time each instructional hour in grade 7, 8, and 9. Lastly, nearly 30 new words are introduced per schooling hour in each high school grade, except grade 12, in which around 13 words are introduced hourly.

4.3 RESEARCH QUESTION 3

How often newly-introduced words are repeated in the textbooks?

After measuring how many new word types were introduced in each grade, it was imperative to investigate the number of occurrences of these newly-introduced words in each text. This allowed the researcher to assess whether the number of word occurrences would enable Saudi EFL learners to encounter these words. Figure 8 shows the number of newly-introduced word types per text, used from grades 4 to 12, with their recycling range. Each word type was categorized according to its number of occurrences in the text: *once*, *four to twice*, *five to fifteen*, or *above fifteen*.
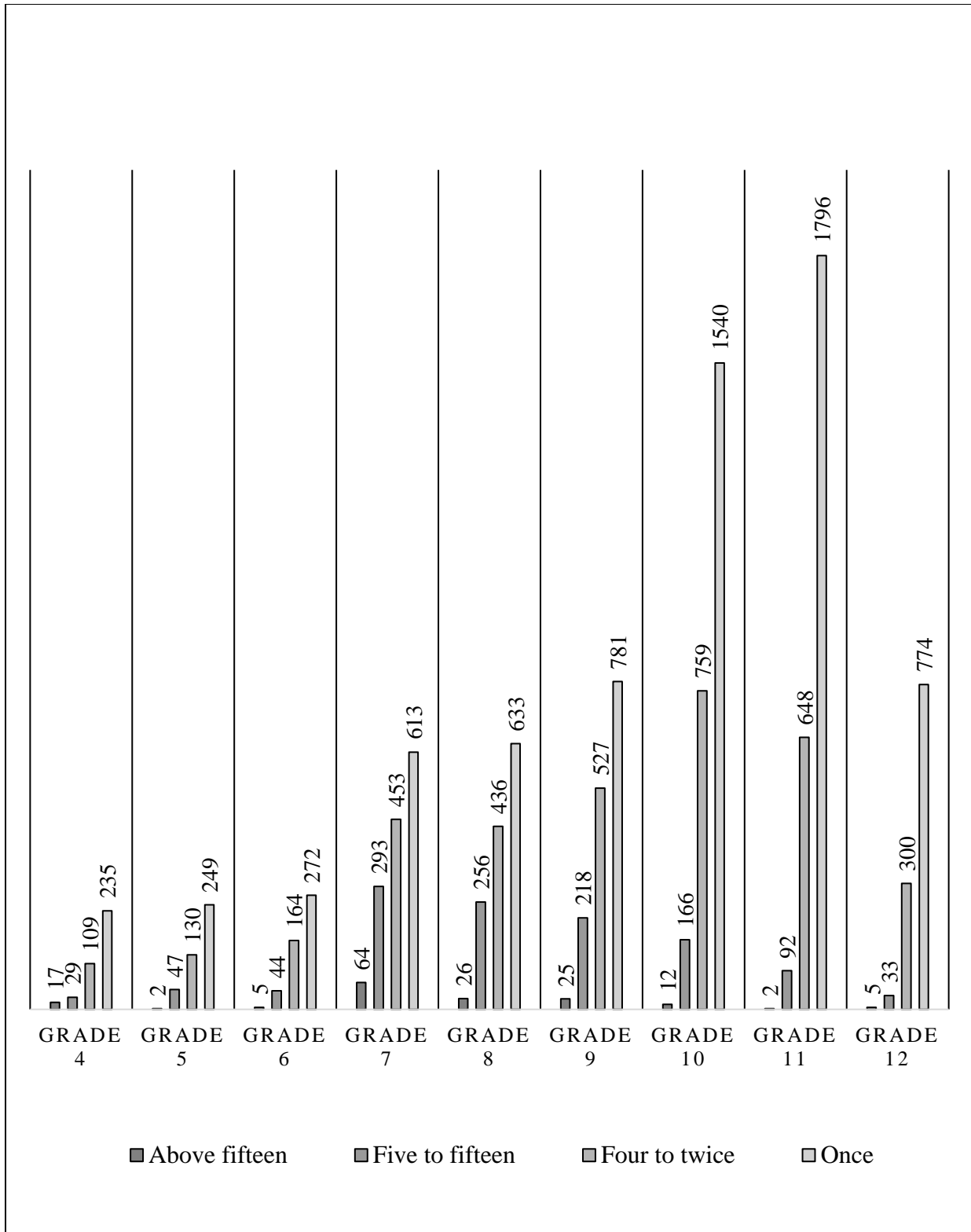
Figure 8. *Recycling Range of Newly-introduced Word Types*

As Figure 8 indicates, the vast majority of word types introduced in each grade fall within the category of the "*once*" occurrence. The second recycling category in which newly-introduced word types tends to occur was in "*four to twice*". Continuing to move down, a smaller number of newly-introduced word types occurred between "*five to fifteen*" times in the texts. Lastly, the smallest number of word types presented in each text was accounted for in the category of "*above fifteen*" occurrences. Table 14 presents quantitative data in regards to the percentage of recycled words included in texts used from grade 4 to 12.

Table 14
Percentage of Recycling Recently-introduced Words

Grade	Words repeated four times or less	Words repeated five times or more
4	88%	12%
5	88%	12%
6	89%	11%
7	74%	26%
8	79%	21%
9	84%	16%
10	92%	8%
11	96%	4%
12	96%	4%

As Table 14 illustrates, words that have repeated exposures of five times or more made up the smallest percentage, while words repeated less than five times had the largest. As stated previously, around 2,500 word types were introduced in high school texts each year, which was the highest amount of input. However, high school texts accounted for the smallest percentage of word repetition, indicating that a negative correlation exists between the two. In other words, when more new words are introduced, there are fewer opportunities for repetition.

4.4 RESEARCH QUESTION 4

What types of learning opportunities do Saudi EFL textbooks present to acquire vocabulary?

Nine newly-introduced words that occurred five times or more in the corpora were selected randomly to be used in the case study. Numerous aspects related to the depth of word knowledge were considered in this analysis: *word lists*, *part of speech*, *collocations*, *number of occurrences*, and *inflectional and derivational forms*. The results of this qualitative corpus-based analysis is illustrated in Table 15.

Table 15
Results of Case Study

Grade	Selected Words	Number of Occurrences	Word List	Number of Inflections and Derivations	Part of Speech	Number of Collocations
4	Shapes	8	1K	9 Inflections: Shapes, shape	Noun	2
5	Using	15	1K	21 Inflections: Using, used, use	Verb	9
6	Carry	6	1K	15 Inflections: Carry, carries, carried	Verb	5
7	Excited	7	2K	12 Inflections: Excited, exciting	Adjective	4
8	Invitation	16	2K	18 Inflections: Invite, inviting, invitation(s) 16 Derivation: Invitation	Noun Verb	23
9	Prefer	30	2K	31 Inflections: Prefer, prefers 1 Derivation: Preference	Noun Verb	4
10	Stress	21	AWL	32 Inflections: Stress, stressed 7 Derivation: Stressful	Adjective Noun	12
11	Consumers	8	AWL	11 Inflections: Consumed, consumer(s) 10 Derivations: Consumption, consumer	Adjective Noun Verb	14
12	Achieve	9	AWL	17 Inflections: Achieve, achieves, achieving 5 Derivation: Achievement	Noun Verb	14

Table 15 revealed that the word *prefer* was the most frequent word since it occurred 30 times in all texts utilized in grade 9. This word belongs to the second 1,000 most frequent words in English. Two inflectional forms were found for this word which were *prefer* and *prefers*. Only one derivational form of *prefer* was presented which was *preference*. The word, *prefer*, occurred 29 times as a verb in the present tense (*prefer* and *prefers*) and only once as a noun (*preference*). Four collocations were presented in grade 9 texts for the word *prefer* with its all inflectional and derivational forms. The patterns of these collocations were (*modal verb + prefer + direct object*), (*noun phrase + prefer + gerund*), (*noun phrase + prefer + infinitive verb*) and (*noun phrase's + preference*). Examples of the collocations found in the texts were *I prefer swimming*, *I prefer to have juice*, *would prefer studying*, and *Adam's preference*.

On the other hand, the word *carry* was the least frequent out of the nine case study words. It occurred six times in the form of a verb in the present and past tense (*carry*, *carries*, and *carried*). There were zero occurrences of a derivational form. This was also the case for the case study words presented at the beginning school levels, from grades 4 to 7. No single occurrence of a derivational form was found for the words *shape*, *use*, and *excite*. Possible derivations of these words would be *shapeless*, *useless*, *misused*, *useful*, *usefulness*, *usefully*, *user*, *excitement*, and *excitedly*. Although these derivations are very frequent in English, in which they were included in the K1 and K2 word lists, none of these words were found in the texts. Another important finding of this in-depth lexical analysis was the number of parts of speech. A very large number of the selected words occurred in the form of nouns or verbs. Only three out of nine words occurred as adjectives. Although adverbs, adjectives, nouns, and verbs are considered open classes of morphemes in which new words can be created, there was no occurrence of adverbs in the results.

57

CHAPTER 5: DISCUSSION

The findings of the first question revealed that words accounted from the K1 and K2 word lists had the highest percentage of lexical coverage in all texts used from grade 4 to 12, which was an average of 90.72%. Table 9 shows the total number of words from the K1 and K2 word lists included in all texts. A total of 971word families and 2,731 word types were presented from the K1 word list and 767 word families and 1,592 word types from the K2 word list. Since not all word types accounted from these two lists are covered in the books, this indicates that the Saudi EFL learners are most likely familiar with headwords from the K1 and K2 lists. Nation and Bauer (1993) stated that once learners learn headwords, they can easily understand the meaning and form of the rest of words listed under them. In other words, knowing the word *engage* will help learners to easily identify and understand the rest of its derivational and inflectional forms, such as engag*ing*, engage*d*, engage*ment*, and engage*s*.

This way of identifying words might be easy for native speakers and maybe even for ESL learners, but definitely not for EFL learners. This is evident when looking at the findings of the Schmitt and Zimmerman (2002) study. These researchers asked 106 non-native English speakers who study at the college level in the United States to identify different derivational forms of stem words that belong to major word classes: *Verb*, *noun*, *adjective*, and *adverb*. The results of the study showed that the majority of participants were not able to understand the derivational forms of the word family. It is thus possible to argue that if students in an ESL setting only have a partial knowledge of word forms, despite all of the language exposure they receive, it would be especially difficult for EFL students if they have never encountered these word forms before. In order to help build the Saudi EFL learners' lexicon, there is a necessity to provide explicit instruction regarding the roles that roots, inflections, and derivations play in forming any word in

English. After students formally learn that adding '-s' at the end of a noun changes the number to plural, for example, it will hopefully strengthen their noticing abilities and aid their acquisition of various members of word families.

When contrasting the lexical coverage of the current EFL curriculum with the former one, the coverage of words accounted from the K1 and K2 word lists, about 91%, are in line with the findings of the study conducted by Milton and Alsaif (2012). However, it is of great importance to note that more less frequent words were presented in the former Saudi EFL curriculum. To further this point, the overall coverage of words from the K4 to K15 reached nearly 6%, whereas about 3% were allocated for less frequent words in the current EFL curriculum. This limited coverage of less frequent words in the current curriculum is thought to better facilitate the students' learning of words since the acquisition of very frequent words should occur first. To elaborate, the most frequent English words make up the largest coverage of tokens of any authentic written or spoken text, as noted previously in the literature. Having a deep knowledge of the most frequent words would enable non-native speakers to interpret the meaning of fewer less frequent words from the context. With that being given, it is possible to claim that spending a sufficient amount of time to train EFL students to use various vocabulary learning strategies, such as guessing the meaning from the context, would be more useful than introducing numerous less frequent words that students might never encounter in their real lives.

The findings also indicate that the smallest number of word types and tokens are presented in elementary school English textbooks. Since the teaching of English as a foreign language starts in elementary schools at grade 4, numerous visual illustrations are included in the books in place of written articles or conversations. This is typically found in language books designed specifically for beginning learners in order to help facilitate their acquisition of new

words. This allows learners to easily connect the form of a newly-presented word with its meaning. An example of visual aids that are included in elementary school textbooks is found in Figure 9.
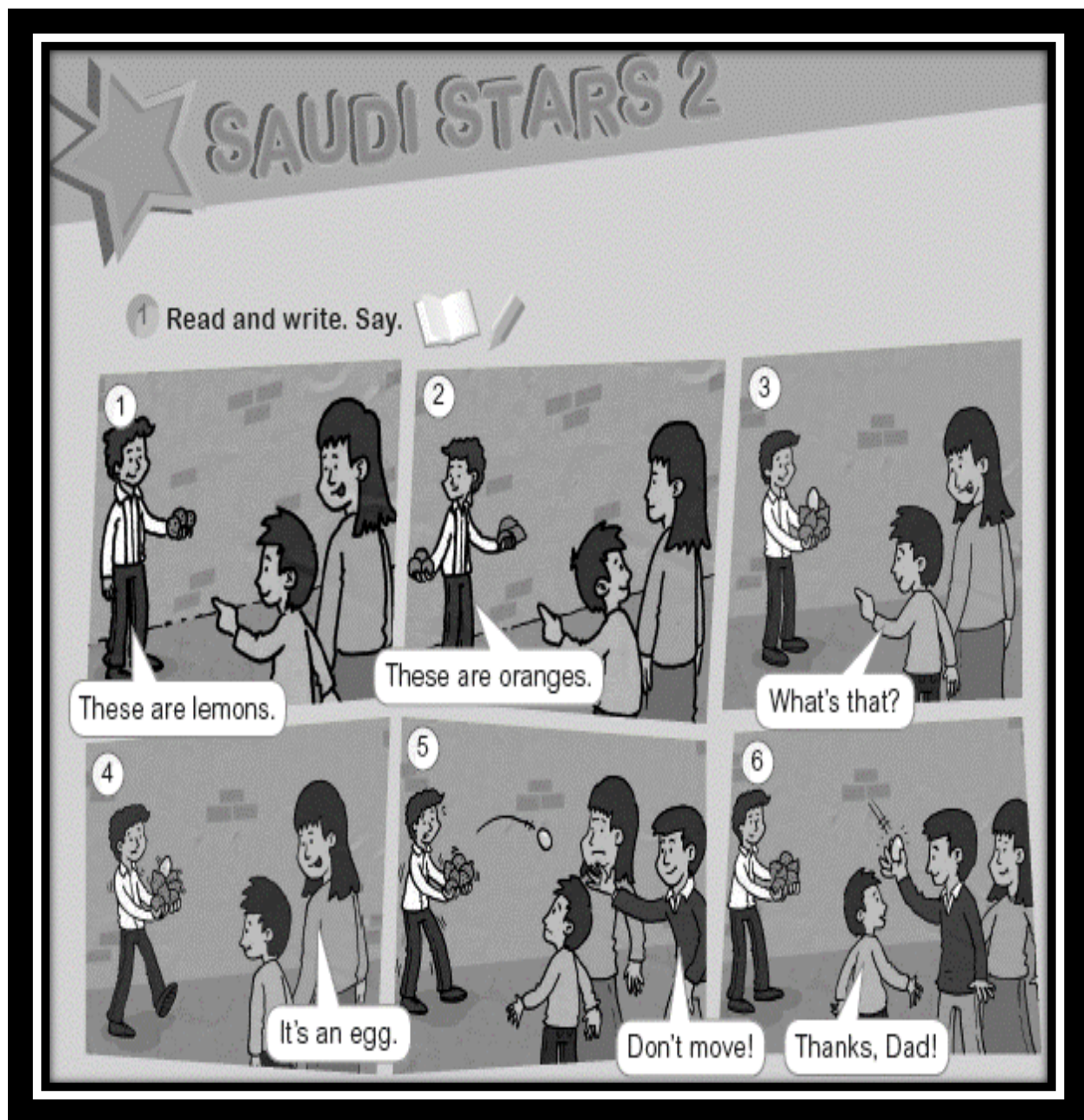
Figure 9. *An Example of Using Pictures to Introduce Words*

Figure 9 illustrates the use of visuals in elementary school books to introduce new words. In this particular example, three food items are introduced for the first time: lemons, oranges, and eggs. This helps learners at low proficiency levels to quickly make a connection between the form of a word and its meaning. It is rare to find a page in any of the books utilized in elementary schools that do not contain large visual illustrations. However, this is not the case with books used at intermediate and high school levels, which could explain why they include such a large percentage of word types and tokens.

Another possible factor that led to introducing fewer word types in elementary school books is related to instructional hours. Fewer instructional hours are offered for English at this school level, where students receive a total of 48 instructional hours per grade. Both intermediate and high school students, on the other hand, are offered a total of 96 hours of English instruction each year. Because of that, it would not be surprising why more English vocabulary words are presented for intermediate and high school students.

However, there is still a major difference with respect to the number of words included in intermediate and high school books, despite the fact that they have the same number of schooling hours. To elaborate, the last grade in intermediate school, grade 9, includes 2,339 words from the K1, K2, and AWL, and 273 words from the additional baseword lists. When moving forward to the first grade in high school (grade 10), the number of words increased dramatically, in which a total of 3,114 word types are included from the K1, K2, and AWL, and roughly 423 word types from the additional baseword lists. There is no possible explanation for this rapid growth of word types introduced between these school levels other than the students' proficiency levels, in which they reach a stage of being familiar with the language, giving them the opportunity to process

words faster. If that is the case, it is crucial to understand why the number of word types

presented in books utilized in the last grade in high school, grade 12, dropped down dramatically.

The second question of this study is concerned with the number of new vocabulary input

provided for the Saudi EFL learners per grade. In second language acquisition, it is strongly

suggested to introduce 8 to 12 words per schooling hour in order to ensure the learning of these

newly-presented words (Garins & Redman, 1986; Scholifield, 1991). This number is thought to

be realistic in terms of providing language learners the needed time to process the new

vocabulary input. When measuring how many new words are introduced per hour in the present

study, around 9 words are counted in grade 4, 10 words in grade 5, and 12 words in grade 6. The

findings of words introduced for the first time during each hour of schooling in elementary

school are in line with the principles and requirements of vocabulary acquisition.

However, the number of words introduced during each instructional hour in both

intermediate and high school level is far from what researchers recommended. Around 17 words

are introduced per hour in intermediate school and about 30 new words presented in high school,

with the exception of grade 12. The newly-introduced words presented each hour from grades 7

to 11 is contrary to the literature noted. Because of that, it is thought to hinder the learners from

acquiring new words. The number of newly-introduced words per hour in the last grade of high

school decreased to 13 new words. This is similar to the findings of Scholfield's (1991) study

discussed in chapter 2; the last chapters of the beginning English language course book, *The

Cambridge English*, included the lowest number of words in order to provide more opportunities

for revision and recycling of words.

The third research question deals with another important principle of introducing new vocabulary input, which is repetition. Studies in second language acquisition have stressed the importance of recycling newly-learned words (Nation, 1990; Pigada & Schmitt, 2006; Web, 2007). Providing learners with the needed repeated exposures will help to improve their learning of words. The suggested number for recycling newly-introduced words ranges from 5 to 20 times. Yet, the findings of the third research question revealed that new words are repeated very poorly. At least 74% of the total new words introduced per grade are repeated less than five times. This shortcoming is also found in the study conducted by Matsuoka and Hirsh (2010), in which they carried out a lexical analysis of a course book, titled *New Headway Student's Book Upper-Intermediate*. The findings illustrated that a large number of words were only presented in the course book once. This limited exposure of words decreases the chances of learning, raising the need for pedagogical interventions. Developing supplemental materials that target the use of these infrequently repeated words could potentially assist learners in recognizing the words and to become familiar with their features, including their form, part of speech, and meaning.

When looking at the case study findings, the textbooks offered good learning opportunities for words repeated at least five times. To begin, various collocations are included in the texts which helps learners gain a deeper knowledge regarding not only the form and meaning of words, but also how these words are typically used. When learners are explicitly informed how to form infinitive phrases, for instance, it can be very challenging for them to be familiar with the structure. It is much easier and simpler to introduce collocations (Lewise, 1997), as "*to talk about*", "*to read about*", and "*to hear from*". This leads learners to learn implicitly how to form the target patterns by using "*to*" followed by "*a verb in the base form*"

then "*a preposition*". Using patterns is a great learning strategy that enables learners of languages to improve their "*lexicogrammar*" knowledge more easily.

Similarly, the analysis of the case study shows that the textbooks include different word families for each word entry. This gives learners the opportunity to know how different affixes can be added to words to either change the form or the lexical category, such as changing a verb to a noun. This helps in facilitating the learning of vocabulary, especially in EFL contexts. Oftentimes, there is not enough exposure to the language, where the only source of input for students occurs inside the classroom. Thus, presenting various inflectional and derivational forms, such as *–s, -ing, -er, -ence, -tion, -ed, -ful, and –ion*, helps the learners deepen their knowledge of words.

When comparing the findings of the case study with those obtained from Matsuoka and Hirsh's (2010) study, there is a difference in terms of the number of parts of speech as well as the derivations and inflections. Out of the six most frequent words, only one derivation was found in the text. Additionally, only nouns and verbs were found in their lexical analysis. The present study, on the other hand, includes more opportunities to deepen the knowledge of word parts and major word classes, including adjectives. Furthermore, the analysis of the present case study shows that various derivational forms are included in texts, starting in grade 8.

It is imperative to note that the case study analysis revealed that there are more inflections than derivations. This comes as no surprise since the formation of inflectional forms can be easily understood. There are specific grammatical rules that need to be applied in order to have an inflectional form. To illustrate this point, if a verb ends in –y and is proceeded by a vowel, the suffix –ing can be easily added to the verb, such as with *enjoy* to *enjoying*. These forms are only added at the end of the words as suffixes, changing the type, but never the lexical

category. Because of this, they appear very frequently in English, and they are mostly added to verbs. Conversely, there is no specific regulation regarding the formation of derivational forms. According to Laufer (1997), there is irregularity in English when it comes to forming derivations. This explains why the corpora analyzed in this study includes more opportunities to learn different inflectional forms since they are generally acquired earlier.

CHAPTER 6: CONCLUSION

6.1 PEDAGOGICAL IMPLICATIONS

The distributional patterns of lexical items and the characteristics of case study words have several implications for teaching and learning vocabulary in an EFL environment. To begin, it is of great importance to present new vocabulary words in sufficient quantities in order to lead learners to encounter these words and improve their knowledge of words. However, the Saudi EFL textbooks show a shortcoming in the recycling of words in which 74 to 96% of words occurred four times or less. This decreases the chances of learning new words. One of the main implications for Saudi EFL teachers is to take responsibility in facilitating the learning of these infrequently repeated words by providing opportunities for repetitions in class.

Considering the type and nature of less frequently repeated words would help teachers to identify words that require more repetition. Words like *apple*, *orange*, and *camera* are listed in the K2 word list, but they can be easily learned (Reed & Dick, 1968) by just looking at flash cards, for example. Recycling these concrete words through visuals would lead students to a successful self-word learning. This allows teachers to have more time to introduce activities that target less repeated words that refer to abstract concepts, such as *victory* and *kindness*, which are harder for students to learn.

However, it is important to note that a number of less frequent words are introduced explicitly at the beginning of lessons, although they do not hold central meanings to the context they are included in. A word like *youth* is introduced in a grade 11 textbook before the reading passage. After a careful reading of the passage, the word is not used in any context included in the lesson. Oftentimes, publishers highlight words even if they are rarely used in textbooks (Feldman & Kinsella, 2005). This raises the needs for teachers to proofread any reading passage

prior to the class in order to identify new words that carry central lesson concepts. Emphasizing on the teaching of these types of words would give learners the opportunity to have the needed vocabulary words. This in turn will allow them not only to comprehend their readings, but also to be able to infer the meaning of new words from the contexts.

The qualitative-based analysis also shows that the textbooks provides good opportunities for deepening the knowledge of words that are repeated five times or more. To further this point, the selected nine words from the K1, K2, and AWL were found to have various inflectional and derivational forms, a range of collocations, and different parts of speech. Although these selected words are considered to be the best examples of deepening word knowledge, due to their high number of occurrences, learners might still struggle to move from receptive knowledge (i.e., passively recognizing words in listening or reading tasks) to productive knowledge (i.e., actively applying these words in speaking and writing tasks).

In Nation's (2001) framework, 'word parts' is one of the nine aspects of word knowledge. Explicit instruction of word parts is needed to reinforce the knowledge of word form, meaning, and use. According to Lee (2003), "explicit vocabulary instruction helps to convert recognition vocabulary into productive vocabulary." To enhance the students' learning, teachers can explain explicitly how affixes can be added to word entries to change either the form or the lexical category of the word. Once learners understand that adding *–ment*, *-ence*, or *–tion* to the end of a word changes the part of speech to a noun, it would enable them to apply the rules of word parts to other word entries, which then leads them to expand their vocabulary size.

Finally, another approach that teachers of English can follow to facilitate the learning of words is data-driven learning (DDL) which is "the use in the classroom of computer-generated concordances to get students to explore regularities of patterning in the target language, and the

development of activities and exercises based on concordance output" (Johns, 1991). Using

computer/online-generated concordances, such as *AntConc* and *Compleat Lexical Tutor*, is very

useful to motivate learners to investigate the nature of words introduced either in their texts or

even in authentic materials. Through going over the concordance lines, which are the lines in the

texts that include the target words, learners are able to explore the usage of vocabulary. Since not

all Saudi public schools have internet access or are equipped with computers, teachers can bring

copies of concordance lines to their classrooms and ask students to investigate different features

of new words, including word parts. Figure 10 is an example of a DDL activity that teachers can

incorporate in their lessons.
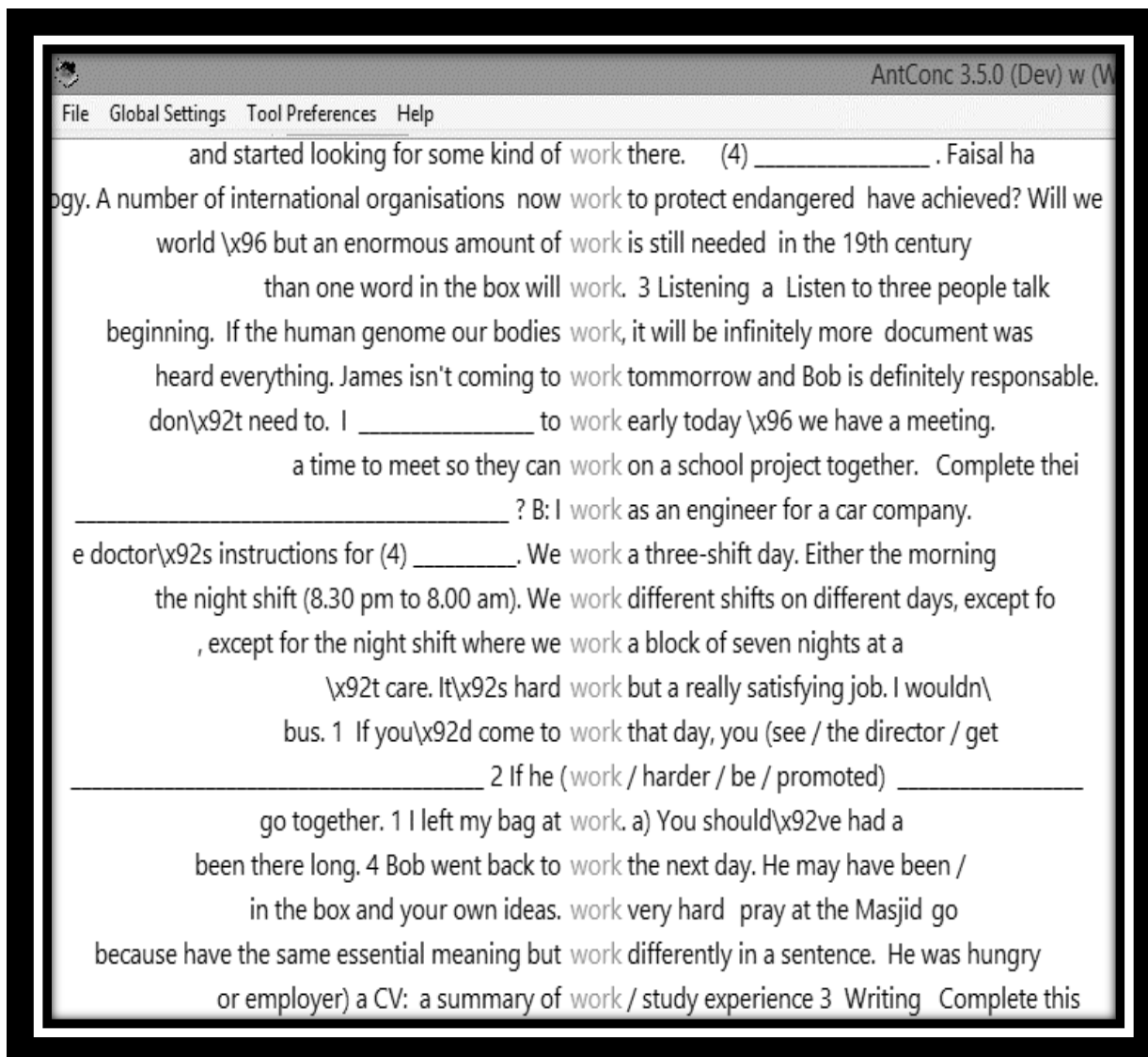
Figure 10. *An Example of a Data Driven Learning Task*

This example is generated from a qualitative lexical analysis of the text used in grade 11 in high school. *AntConc* (Anthony, 2014) was used to generate all of the concordance lines that include the target word, which was *work*. Once the copies of the analysis are distributed, teachers then can ask their students to identify either the grammatical patterns, parts of speech, or collocations. The results of this analysis can help the Saudi EFL learners deepen their vocabulary knowledge.

6.2 LIMITATIONS OF THE PRESENT STUDY

Collecting both quantitative and qualitative data in the present study helped the researcher obtain more detailed information, providing a better understanding of the nature of words included in the texts. Despite the fact that employing a mixed method design in the research helped to overcome weaknesses typically found in one method, a number of important limitations were found in the study. To begin, there are different aspects in which the researcher could have looked into to determine how well the repetition of words included in the texts help facilitate students' learning. Instead, the researcher only made the attempt to analyze the frequency of word occurrences, whether words are presented five time or more in the texts. No attention was given towards measuring the spreading of words throughout the texts. Numerous researchers in memory and vocabulary acquisition have agreed that introducing new words over a longer period of time will help learners build their vocabulary as well as learn words more easily (Gathercole & Baddeley, 1990; Bloom & Shuell, 1981). Since all of the four books utilized in each grade were compiled in one PDF file, it was impossible for the researcher to go over each unit and lesson to determine if new words are introduced in one chapter or more.

Limitations also exists in the present study for the generalizability of the qualitative based analysis. Only nine words that occurred at least five times in the texts were selected randomly for the case study. The findings obtained from this analysis revealed that great learning opportunities are provided for Saudi EFL learners to deepen their knowledge of these nine newly-introduced words. However, even with these findings, it is not practical to generalize these findings to other words occurring very frequently in the texts, as some may have one form of an inflection or one collocation, for instance.

6.3 CONCLUSION

Second language acquisition research has recognized vocabulary learning as an integral part of learning any second/foreign language. Having a deep knowledge of thousands of words would enable language learners to achieve a high level of proficiency and use the target language adequately. It has been estimated that learners of English as a foreign language need to have a vocabulary size of at least 3,000 words to be able to comprehend texts, written or spoken, effectively. Textbooks are considered to be the most accessible resource for vocabulary learning in an EFL context; thus, it is of a vital importance to investigate words included in EFL textbooks in order to ensure their appropriateness. The present study aims to examine the nature of words included in the Saudi EFL texts, applying a mixed method corpus-based analysis.

In the first stage of the analysis, texts used in grades 4 to 6 had a sufficient amount of newly-introduced words, presenting 9 to 12 new words during each hour of schooling. Conversely, texts used from grades 7 to 11 had an inappropriate amount of new vocabulary input, where nearly 17 words were introduced during each hour of instruction in intermediate school levels, and approximately 30 words in grades 10 to 12. When it comes to the frequency of word occurrences, the findings indicate that there is a shortcoming in the recycling of new words, in which at least 74% of newly-introduced words appeared less than five times in the texts. During the second stage of the analysis, nine words were selected randomly based on their high frequency to be included in the case study. Results indicate that good vocabulary learning opportunities are provided in the texts. This in turn helps learners acquire new words more easily as well as adequately comprehend the texts.

The findings of the present study open many questions regarding what another analysis of lexical items would reveal about the vocabulary input provided in the Saudi EFL texts. To begin,

future researchers could use Nation's (2001) framework to identify how many words Saudi EFL learners know from the texts as well as how well these learners understand various aspects of word knowledge. Obtaining this type of information helps to shed light on Saudi EFL learners' productive knowledge of words learned from the texts. Another direction for research could look at the spaced repetition of new words included in corpora similar to the ones used in the present study, where researchers extract chapters in textbooks to identify all occurrences of the target new words. Findings obtained from this type of analysis will help to answer an important question: *does input provided in EFL texts aid the acquisition of new words and promote long-term learning*?

REFERENCES

Abdan, A. (1991). An exploratory study of teaching English in the Saudi elementary public schools. *System, 19*, 153–266.

Adolphs, S. & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics, 24*(4), 425–438.

Ahola, J. (2011). A corpus-driven approach to the infinitival complementation patterns of the verb HELP in recent centuries.

Alakloby, S. (2001). *Teaching and learning English vocabulary in Saudi Arabian public schools: An exploratory study of some possible reasons behind students' failure to learn English vocabulary* (Unpublished doctoral dissertation). University of Essex, UK.

Alamri, A. (2008). *An evaluation of the sixth grade English language textbook for Saudi boys' schools* (Unpublished master thesis). King Saud University, Saudi Arabia.

Albogami, A. (1995). *Teaching English vocabulary to EFL male students at intermediate and secondary public schools in Riyadh* (Unpublished master thesis). King Saud University, Saudi Arabia.

Alfahadi, A. (2014). *Saudi teachers' views on appropriate cultural models for EFL textbooks: Insights into TESOL teachers' management of global cultural flows and local realities in their teaching worlds* (Unpublished doctoral dissertation). University of Exeter, Exeter, UK.

Alfotais, A. (2012). *Investigating textbooks' input as a possible factor contributing to vocabulary knowledge failure among Saudi EFL learners at Taif University*. Department of Language and Linguistics, University of Essex, UK.

Alharbi, H. A. (2015). Improving students' English speaking proficiency in Saudi public schools. *International Journal of Instruction*, *8*(1), 105-116.

Alhazemi, G. (1993). *Low-level EFL vocabulary tests for Arabic speakers* (Unpublished doctoral dissertation)*. University College of Swansea, Wales, UK.

Aljohani, H. M. (2009). *Finding a way forward the impact of teachers strategies, beliefs and knowledge on teaching English as a foreign language in Saudi Arabia* (Unpublished doctoral dissertation). University of Strathclyde, Glasgow, Scotland, UK.

Almutairi, N. H. (2008). *The influence of educational and sociocultural factors on the learning styles and strategies of female students in Saudi Arabia* (Unpublished doctoral dissertation). University of Leicester, Leicester, UK.

Alnasser, A. S. (2015). Problems of English language acquisition in Saudi Arabia: An exploratory-cum-remedial study. *Theory and Practice in Language Studies, 5*(8), 1612.

Alotaibi, G. (2004). *Language learning strategy use among Saudi EFL students and its relationship to language proficiency level, gender and motivation* (Unpublished doctoral dissertation). Indiana University of Pennsylvania, Indiana, USA.

Alqahtani, M. (2011). *An investigation into the language needs of Saudi students studying in British postgraduate programmes and the cultural differences impacting on them* (Unpublished doctoral dissertation). University of Southampton, Southampton, UK.

Alrashidi, O., & Phan, H. (2015). Education context and English teaching and learning in the kingdom of Saudi Arabia: An overview. *English Language Teaching, 8*(5), 33-44.

Alsaif, A. (2011). *Investigating vocabulary input and explaining vocabulary uptake among EFL learners in Saudi Arabia* (Unpublished doctoral dissertation)*. Swansea University, Wales, UK.

Alsaif, A., & Milton, J. (2012). Vocabulary input from school textbooks as a potential

contributor to the small vocabulary uptake gained by English as a foreign language

learners in Saudi Arabia. *Language Learning Journal, 40*(1).

Alseghayer, K. (2005). Teaching English in the Kingdom of Saudi Arabia: Slowly but steadily

changing, In G. Braine (Ed), *Teaching English to the world: History, curriculum, and

practice* (pp. 125-134). Mahwah, N.J: L. Erlbaum Associates.

Alshammari, M. (2011). The use of the mother tongue in Saudi EFL classrooms. *Journal of

International Educational Research, 7*, 95–102.

Alshumaimeri, N. (2003). *A study of classroom exposure to oral pedagogic tasks in relation to

the motivation and performance of Saudi secondary learners of English in a context of

potential curriculum reform* (Unpublished doctoral dissertation). University of Leeds,

Leeds, UK.

Anthony, L. (2014). *AntConc*. Tokyo: Waseda University.

Badawood, O. (2003). *The development of leadership skills of private high school principals in

Jeddah, Saudi Arabia* (Doctoral dissertation). Retrieved from ProQuest dissertations and

theses database. (UMI No. 3117743).

Bauer, L., & Nation, P. (1993). Word families. *International journal of Lexicography*, *6*(4), 253-

279.

BinObaid, R. (2016). An evaluation of the second intermediate Saudi English language textbook

from the teachers' point of view. *Advances in Language and Literary Studies*, *7*(2), 231-

248.

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research, 74*(4), 245-248.

Briggs, A. R., Morrison, M., & Coleman, M. (2012). *Research methods in educational leadership and management*. Sage Publications.

Cobb, T. N. D. (2008). *The compleat lexical tutor for data-driven learning on the web*. Retrieved from: http://lextutor.ca.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238.

Creswell, J. W. (2007). *Qualitative enquiry and research design: Choosing among five approaches*. Sage publications.

Creswell, J. W. (2015). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Boston: Pearson.

Ebad, R. (2014). The role and impact of English as a language and a medium of instruction in Saudi higher education institution: Students-instructors perspective. *Study in English Language Teaching, 2*(2), 142-143.

Fareh, S. (2010). Challenges of teaching English in the Arab world: Why can't EFL programs deliver as expected? *Procedia-Social and Behavioral Sciences*, *2*(2), 3600-3604.

Feldman, K., & Kinsella, K. (2005). *Narrowing the language gap: The case for explicit vocabulary instruction*. Scholastic.

Gairns, R., & Redman, S. (1986). *Working with words*. Cambridge.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection?. *Journal of Memory and Language, 29*(3), 336-360.

General Authority for Statistics. (2015). *Population statistics*. Retrieved from:

    http://www.cdsi.gov.sa/en/node

Ghadirian, S. (2002). Providing controlled exposure to target vocabulary through the screening

    and arranging of texts. *Language, Learning & Technology 6*(1), 147–164.

Global Language Monitor. (2014). *Number of words in the English language: 1,025,109.8*.

    Retrieved from: http://www.languagemonitor.com/number-of-words/number-of-words-

    in-the-english-language-1008879/

Heatley, A., Nation, P., & Coxhead, A. (2002). *Range computer program*. Retrieved from:

    http://www.victoria.ac.nz/lals/about/staff/paul-nation

Hoffmann, S., Evert, S., Smith, N., Lee, D., & Berglund-Prytz, Y. (2008). *Corpus linguistics*

    with BNCweb-a practical guide, 6. Peter Lang.

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical*

    grammar of English. Amsterdam: John Benjamins.

Johns, T. (1991). Should you be persuaded: Two examples of data driven learning. *ELR Journal*,

    4, 1-16.

Khan, I. (2011). Learning difficulties in English: Diagnosis and pedagogy in Saudi Arabia.

    Educational Research, *2*(7), 1248-1257.

Konstantakis, N., & Alexiou, T. (2012). Vocabulary in Greek young learners' English as a

    foreign language course books. *The Language Learning Journal*, *40*(1), 35-45.

Laufer, B. (1989). *What percentage of text-lexis is essential for comprehension?* Clevedon:

    Multilingual Matters.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. *Second language vocabulary acquisition: A rationale for pedagogy, 1*, 20-34.

Laufer, B. (2000). *Task effect on instructed vocabulary learning: The hypothesis of 'involvement'*. Japan: Waseda University Press.

Lee, S. H. (2003). ESL learners' vocabulary use in writing and the effects of explicit vocabulary instruction. *System*, *31*(4), 537-561.

Lewis, M. (1997). Pedagogical implications of the lexical approach. *Second Language Vocabulary Acquisition,* 255-270. New York: Cambridge University Press.

Maiguashca, R. (1993). Teaching and learning vocabulary in a second language: Past, present, and future directions. *The Canadian Modern Language Review, 50* (1), 83-100.

Matsuoka, W., & Hirsh, D. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a foreign language*, *22*(1), 56.

McCarten, J. (2007). Teaching vocabulary. *Lessons from the Corpus, Lessons forthe Classroom. Cambridge: CUP*.

McCarthy, M. J. (2004). *Touchstone: from corpus to course book*. Cambridge: Cambridge University Press.

McCrostie, J. (2007). Examining learner vocabulary notebooks. *English Language Teachers Journal, 61*(3), 246-255.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Taylor & Francis.

Melka, F. (1997). *Receptive vs productive aspects of vocabulary.* Cambridge: Cambridge University Press.

Merriam, S. B. (1998). *Qualitative research and case study applications in education*. Jossey

Bass Publishers. San Francisco, CA.

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

Matters.

Milton, H., & Benn, T. (1933). Study of the vocabulary of thirty first-year French courses.

Modern Languages, 14(3), 11-148.

Ministry of Education. (2016a). احصائيات عامة – التعليم العام (*General statistics for K-12 schools*).

Retrieved from: http://www.moe.gov.sa/Pages/StatisticalInformation.aspx

Ministry of Education. (2016b). تدريس اللغة الانجليزية (*English language teaching*). Retrieved from:

http://ed.edu.sa/archives/category/%d9%88%d9%83%d8%a7%d9%84%d8%a9-

%d8%a7%d9%84%d8%aa%d8%ae%d8%b7%d9%8a%d8%b7-

%d9%88%d8%a7%d9%84%d8%aa%d8%b7%d9%88%d9%8a%d8%b1/%d8%a7%d9%

84%d8%a5%d8%af%d8%a7%d8%b1%d8%a9-

%d8%a7%d9%84%d8%b9%d8%a7%d9%85%d8%a9-

%d9%84%d9%84%d9%85%d9%86%d8%a7%d9%87%d8%ac

Ministry of Education. (2016c). مناهج اللغة الانجليزية (*English language curriculum*). Retrieved

from: https://www.moe.gov.sa/ar/Alfasly/books/LiftOff_L1_STWB.pdf

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

Nation, P., & Waring, R. (1997). *Vocabulary size, text coverage and word lists*.

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

Press.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian

Modern Language Review, 63*(1), 59–81.

O'Dell, F. (1997). Incorporating vocabulary into the syllabus. In N. Schmitt & M. McCarthy

    (Eds.), *Vocabulary and language teaching*. Cambridge: Cambridge University Press.

Pigada, M. & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study.

    Reading in a Foreign Language, *18*(1), 1-28.

Rahman, M., & Alhaisoni, E. (2013). *Teaching English in Saudi Arabia: Prospects and*

    challenges. Retrieved from http://www.ahmedv.com/wp/wpcontent/

    uploads/2014/05/20134.1-11.pdf

Rajab, H. (2013). Developing speaking and writing skills of L1 Arabic EFL learners through

    teaching of IPA phonetic codes. *Theory and Practice in Language Studies*, *3*(4), 653-659.

Reed, H. B., & Dick, R. D. (1968). The learning and generalization of abstract and concrete

    concepts. *Journal of Verbal Learning and Verbal Behavior*, *7*(2), 486-490.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental

    vocabulary acquisition through reading. *Studies in Second Language Acquisition, 21*(4),

    589-619.

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

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know?.

    Tesol Quarterly, 145-171.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching*

    Research, *12*, 329–63.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave

    Macmillan.

Scholfield, P. (1991). *Vocabulary rate in course books: living with an unstable lexical economy*.

    Thessaloniki: Aristotle University.

Sutarsyah, C. P., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based case study. *RELC Journal 25*(2), 34-50.

Thomas, D. (2005). *Type-token ratios in one teacher's classroom talk: An investigation of lexical complexity*. United Kingdom: University of Birmingham.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46-65.

Webb, S. & Rodgers, M. H. (2009). Vocabulary demands of television programs. *Language Learning*, *59*(2), 335-366.

Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of text and picture annotation types. *CALICO journal, 20*(1), 33-58.

APPENDIX A

HEADWORDS OF THE FIRST 1,000 MOST FREQUENT ENGLISH WORD LIST

ABLE	FRIDAY	RATE
ABOUT	FRIEND	RATHER
ABOVE	FROM	REACH
ACCEPT	FRONT	READ
ACCORD	FULL	READY
ACCOUNT	FURNISH	REAL
ACCOUNTABLE	FUTURE	REALISE
ACROSS	GAIN	REALLY
ACT	GAME	REASON
ACTIVE	GARDEN	RECEIPT
ACTOR	GAS	RECEIVE
ACTRESS	GATE	RECENT
ACTUAL	GATHER	RECOGNIZE
ADD	GENERAL	RECORD
ADDRESS	GENTLE	RED
ADMIT	GET	REDUCE
ADOPT	GIFT	REFUSE
ADVANCE	GIRL	REGARD
ADVANTAGE	GIVE	RELATION
ADVENTURE	GLAD	RELATIVE
AFFAIR	GLASS	RELIGION
AFTER	GO	REMAIN
AGAIN	GOD	REMARK
AGAINST	GOLD	REMEMBER
AGE	GOOD	REPLY
AGENT	GREAT	REPORT
AGO	GREEN	REPRESENT
AGREE	GROUND	REPUBLIC
AIR	GROUP	RESERVE
ALL	GROW	RESPECT
ALLOW	HALF	REST
ALMOST	HAND	RESULT
ALONE	HANG	RETURN
ALONG	HAPPEN	RICH
ALREADY	HAPPY	RIDE
ALSO	HARD	RIGHT
ALTHOUGH	HARDLY	RING
ALWAYS	HAVE	RISE
AMONG	HE	RIVER
AMOUNT	HEAD	ROAD

ANCIENT	HEAR	ROCK
AND	HEART	ROLL
ANIMAL	HEAT	ROOM
ANOTHER	HEAVEN	ROUGH
ANSWER	HEAVY	ROUND
ANY	HELP	ROYAL
APPEAR	HERE	RULE
APPLY	HIGH	RUN
APPOINT	HILL	SAFE
ARISE	HISTORY	SAIL
ARM	HOLD	SALE
ARMY	HOME	SALT
AROUND	HONOUR	SAME
ARRIVE	HOPE	SATURDAY
ART	HORSE	SAVE
ARTICLE	HOT	SAY
AS	HOUR	SCARCE
ASK	HOUSE	SCENE
ASSOCIATE	HOW	SCHOOL
AT	HOWEVER	SCIENCE
ATTACK	HUMAN	SEA
ATTEMPT	HUNDRED	SEASON
AVERAGE	HUSBAND	SEAT
AWAY	IDEA	SECOND
BACK	IF	SECRET
BAD	ILL	SECRETARY
BALL	IMPORTANT	SEE
BANK	IN	SEEM
BAR	INCH	SELL
BASE	INCLUDE	SEND
BATTLE	INCREASE	SENSE
BE	INDEED	SENSITIVE
BEAR	INDEPENDENT	SEPARATE
BEAUTY	INDUSTRY	SERIOUS
BECAUSE	INFLUENCE	SERVE
BECOME	INSTEAD	SERVICE
BED	INTEREST	SET
BEFORE	INTO	SETTLE
BEGIN	INTRODUCE	SEVEN
BEHIND	IRON	SEVERAL
BELIEVE	IT	SHADOW
BELONG	JOIN	SHAKE
BELOW	JOINT	SHALL

BENEATH	JOINTED	SHAPE
BESIDE	JOY	SHARE
BEST	JUDGE	SHE
BETWEEN	JUST	SHINE
BEYOND	JUSTICE	SHIP
BIG	KEEP	SHOOT
BILL	KILL	SHORE
BIRD	KIND	SHORT
BLACK	KING	SHOULD
BLOOD	KNOW	SHOULDER
BLOW	LACK	SHOW
BLUE	LADY	SIDE
BOARD	LAKE	SIGHT
BOAT	LAND	SIGN
BODY	LANGUAGE	SILENCE
BOOK	LARGE	SILVER
BOTH	LAST	SIMPLE
BOX	LATE	SINCE
BOY	LATTER	SING
BRANCH	LAUGH	SINGLE
BREAD	LAUGHTER	SIR
BREAK	LAW	SISTER
BRIDGE	LAY	SIT
BRIGHT	LEAD	SITUATION
BRING	LEARN	SIX
BROAD	LEAVE	SIZE
BROTHER	LEFT	SKY
BUILD	LENGTH	SLEEP
BURN	LESS	SMALL
BUSINESS	LET	SMILE
BUT	LETTER	SNOW
BUY	LEVEL	SOCIAL
BY	LIBRARY	SOCIETY
CALL	LIE	SOFT
CAN	LIFE	SOLDIER
CAPITAL	LIFT	SOME
CAPTAIN	LIGHT	SON
CAR	LIKE	SOON
CARE	LIKELY	SORT
CARRY	LIMIT	SOUL
CASE	LINE	SOUND
CASTLE	LIP	SOUTH
CATCH	LISTEN	SPACE

CAUSE	LITERATURE	SPEAK
CENTRE	LITTLE	SPECIAL
CERTAIN	LIVE	SPEED
CHANCE	LOCAL	SPEND
CHANGE	LONG	SPIRIT
CHARACTER	LOOK	SPITE
CHARGE	LORD	SPOT
CHIEF	LOSE	SPREAD
CHILD	LOSS	SPRING
CHOOSE	LOVE	SQUARE
CHURCH	LOW	STAGE
CIRCLE	MACHINE	STAND
CITY	MAIN	STANDARD
CLAIM	MAKE	STAR
CLASS	MAN	START
CLEAR	MANNER	STATE
CLOSE	MANUFACTURE	STATION
CLOUD	MANY	STAY
COAL	MARK	STEEL
COAST	MARKET	STEP
COIN	MARRY	STILL
COLD	MASS	STOCK
COLLEGE	MASTER	STONE
COLONY	MATERIAL	STOP
COLOUR	MATTER	STORE
COME	MAYBE	STORY
COMMAND	MEAN	STRANGE
COMMITTEE	MEASURE	STREAM
COMMON	MEET	STREET
COMPANY	MEMBER	STRENGTH
COMPLETE	MEMORY	STRIKE
CONCERN	MENTION	STRONG
CONDITION	MERE	STRUGGLE
CONSIDER	METAL	STUDENT
CONTAIN	MIDDLE	STUDY
CONTENT	MIGHT	SUBJECT
CONTINUE	MILE	SUBSTANCE
CONTROL	MILK	SUCCEED
CORN	MILLION	SUCH
COST	MIND	SUFFER
COTTON	MINER	SUGGEST
COULD	MINISTER	SUMMER
COUNCIL	MINUTE	SUN

COUNT	MISS	SUNDAY
COUNTRY	MISTER	SUPPLY
COURSE	MODERN	SUPPORT
COURT	MOMENT	SUPPOSE
COVER	MONDAY	SURE
CROSS	MONEY	SURFACE
CROWD	MONTH	SURPRISE
CROWN	MOON	SURROUND
CRY	MORAL	SWEET
CURRENT	MORE	SWORD
CUT	MOREOVER	SYSTEM
DANGER	MORNING	TABLE
DARK	MOST	TAKE
DATE	MOTHER	TALK
DAUGHTER	MOTOR	TAX
DAY	MOUNTAIN	TEACH
DEAD	MOUTH	TEAR
DEAL	MOVE	TELL
DEAR	MUCH	TEMPLE
DECIDE	MUSIC	TERM
DECLARE	MUST	TEST
DEEP	NAME	THAN
DEFEAT	NATION	THE
DEGREE	NATIVE	THEN
DEMAND	NATURE	THERE
DEPARTMENT	NEAR	THEREFORE
DEPEND	NECESSARY	THEY
DESCRIBE	NECESSITY	THING
DESERT	NEED	THINK
DESIRE	NEIGHBOUR	THIRTEEN
DESTROY	NEITHER	THIRTY
DETAIL	NEVER	THIS
DETERMINE	NEW	THOUGH
DEVELOP	NEWS	THOUSAND
DIE	NEWSPAPER	THREE
DIFFERENCE	NEXT	THROUGH
DIFFICULT	NIGHT	THROW
DIRECT	NOBLE	THURSDAY
DISCOVER	NONE	THUS
DISTANCE	NOR	TILL
DISTINGUISH	NORTH	TIME
DISTRICT	NOT	TO
DIVIDE	NOTE	TODAY

DO	NOTICE	TOGETHER
DOCTOR	NOW	TON
DOG	NUMBER	TOO
DOLLAR	NUMERICAL	TOP
DOOR	NUMEROUS	TOTAL
DOUBT	OBJECT	TOUCH
DOWN	OBSERVE	TOWARD
DRAW	OCCASION	TOWN
DREAM	OF	TRADE
DRESS	OFF	TRAIN
DRINK	OFFER	TRAVEL
DRIVE	OFFICE	TREE
DROP	OFFICIAL	TROUBLE
DRY	OFTEN	TRUE
DUE	OH	TRUST
DUTY	OIL	TRY
EACH	OLD	TUESDAY
EAR	ON	TURN
EARLY	ONCE	TWELVE
EARTH	ONE	TWENTY
EAST	ONLY	TWO
EASY	OPEN	TYPE
EAT	OPERATE	UNDER
EFFECT	OPINION	UNDERSTAND
EFFICIENT	OPPORTUNITY	UNION
EFFORT	OR	UNITE
EGG	ORDER	UNIVERSITY
EIGHT	ORDINARY	UNLESS
EITHER	ORGANIZE	UNTIL
ELECT	OTHER	UP
ELEVEN	OTHERWISE	UPON
ELSE	OUGHT	USE
EMPIRE	OUT	USUAL
EMPLOY	OVER	VALLEY
END	OWE	VALUE
ENEMY	OWN	VARIETY
ENGLISH	PAGE	VARIOUS
ENJOY	PAINT	VERY
ENOUGH	PAPER	VESSEL
ENTER	PART	VICTORY
EQUAL	PARTICULAR	VIEW
ESCAPE	PARTY	VILLAGE
EVEN	PASS	VIRTUE

EVENING	PAST	VISIT
EVENT	PAY	VOICE
EVER	PEACE	VOTE
EVERY	PEOPLE	WAGE
EXAMPLE	PER	WAIT
EXCEPT	PERHAPS	WALK
EXCHANGE	PERMIT	WALL
EXERCISE	PERSON	WANT
EXIST	PICTURE	WAR
EXPECT	PIECE	WATCH
EXPENSE	PLACE	WATER
EXPERIENCE	PLAIN	WAVE
EXPERIMENT	PLAN	WAY
EXPLAIN	PLANT	WE
EXPRESS	PLAY	WEALTH
EXTEND	PLEASE	WEAR
EYE	POINT	WEDNESDAY
FACE	POLITICAL	WEEK
FACT	POOR	WELCOME
FACTORY	POPULAR	WELL
FAIL	POPULATION	WEST
FAIR	POSITION	WESTERN
FAITH	POSSESS	WHAT
FALL	POSSIBLE	WHEN
FAMILIAR	POST	WHERE
FAMILY	POUND	WHETHER
FAMOUS	POVERTY	WHICH
FAR	POWER	WHILE
FARM	PREPARE	WHITE
FAST	PRESENT	WHO
FATHER	PRESIDENT	WHOLE
FAVOUR	PRESS	WHY
FEAR	PRESSURE	WIDE
FEEL	PRETTY	WIFE
FELLOW	PREVENT	WILD
FEW	PRICE	WILL
FIELD	PRIVATE	WIN
FIGHT	PROBLEM	WIND
FIGURE	PRODUCE	WINDOW
FILL	PRODUCT	WINTER
FIND	PROFIT	WISE
FINE	PROGRESS	WISH
FINISH	PROMISE	WITH

FIRE	PROOF	WITHIN
FIRST	PROPER	WITHOUT
FISH	PROPERTY	WOMAN
FIT	PROPOSE	WONDER
FIX	PROTECT	WOOD
FLOOR	PROVE	WORD
FLOW	PROVIDE	WORK
FLOWER	PROVISION	WORLD
FLY	PUBLIC	WORTH
FOLLOW	PULL	WOULD
FOOD	PURPOSE	WOUND
FOR	PUT	WRITE
FORCE	QUALITY	WRONG
FOREIGN	QUANTITY	YEAR
FOREST	QUARTER	YESTERDAY
FORGET	QUEEN	YET
FORM	QUESTION	YIELD
FORMER	QUITE	YOU
FORTUNE	RACE	YOUNG
FREE	RAISE	YOUTH
FRESH	RANK	

HEADWORDS OF THE SECOND 1,000 MOST FREQUENT ENGLISH WORD LIST

ABROAD	FAT	PROGRAMME
ABSENCE	FATE	PROMPT
ABSENT	FAULT	PRONOUNCE
ABSOLUTE	FEAST	PROUD
ABSOLUTELY	FEATHER	PUMP
ACCIDENT	FEMALE	PUNCTUAL
ACCUSE	FENCE	PUNISH
ACCUSTOM	FEVER	PUPIL
ACHE	FIERCE	PURE
ADMIRE	FILM	PURPLE
ADVERTISE	FINGER	PUSH
ADVICE	FIRM	PUZZLE
AEROPLANE	FLAG	QUALIFY
AFFORD	FLAME	QUARREL
AFRAID	FLASH	QUART
AFTERNOON	FLAT	QUICK
AGRICULTURE	FLAVOUR	QUIET
AHEAD	FLESH	RABBIT
AIM	FLOAT	RADIO
AIRPLANE	FLOOD	RAIL
ALIKE	FLOUR	RAIN
ALIVE	FOLD	RAKE
ALOUD	FOND	RAPID
ALTOGETHER	FOOL	RARE
AMBITION	FOOT	RAT
AMUSE	FORBID	RAW
ANGER	FORGIVE	RAY
ANGLE	FORK	RAZOR
ANNOY	FORMAL	RECOMMEND
ANXIETY	FORWARD	REFER
APART	FRAME	REFLECT
APOLOGIZE	FREEZE	REFRESH
APOLOGY	FREQUENT	REGRET
APPLAUD	FRIGHT	REGULAR
APPLAUSE	FRUIT	REJOICE
APPLE	FRY	RELIEVE
APPROVE	FUN	REMEDY
ARCH	FUNERAL	REMIND
ARGUE	FUR	RENT
ARRANGE	GALLON	REPAIR

ARREST	GAP	REPEAT
ARROW	GARAGE	REPLACE
ARTIFICIAL	GAY	REPRODUCE
ASH	GENEROUS	REPUTATION
ASHAMED	GLORY	REQUEST
ASIDE	GOAT	RESCUE
ASLEEP	GOVERN	RESIGN
ASTONISH	GRACE	RESIST
ATTEND	GRADUAL	RESPONSIBLE
ATTENTION	GRAIN	RESTAURANT
ATTRACT	GRAM	RETIRE
AUDIENCE	GRAMMAR	REVENGE
AUNT	GRAND	REVIEW
AUTUMN	GRASS	REWARD
AVENUE	GRATEFUL	RIBBON
AVOID	GRAVE	RICE
AWAKE	GREASE	RID
AWKWARD	GREED	RIPE
AXE	GREET	RISK
BABY	GREY	RIVAL
BAG	GRIND	ROAR
BAGGAGE	GUARD	ROAST
BAKE	GUESS	ROB
BALANCE	GUEST	ROD
BAND	GUIDE	ROOF
BARBER	GUILTY	ROOT
BARE	GUN	ROPE
BARELY	HABIT	ROT
BARGAIN	HAIR	ROW
BARREL	HALL	RUB
BASIN	HAMMER	RUBBER
BASKET	HANDKERCHIEF	RUBBISH
BATH	HANDLE	RUDE
BATHE	HARBOR	RUG
BAY	HARM	RUIN
BEAK	HARVEST	RUSH
BEAM	HASTE	RUST
BEAN	HAT	SACRED
BEARD	HATE	SACRIFICE
BEAST	HAY	SAD
BEAT	HEAL	SADDLE
BEG	HEALTH	SAKE
BEHAVE	HEAP	SALARY

BEHAVIOUR	HESITATE	SAMPLE
BELL	HIDE	SAND
BELT	HINDER	SATISFY
BEND	HIRE	SAUCE
BERRY	HIT	SAUCER
BICYCLE	HOLE	SAWS
BILLION	HOLIDAY	SCALE
BIND	HOLLOW	SCATTER
BIRTH	HOLY	SCENT
BIT	HONEST	SCISSORS
BITE	HOOK	SCOLD
BITTER	HORIZON	SCORN
BLADE	HOSPITAL	SCRAPE
BLAME	HOST	SCRATCH
BLESS	HOTEL	SCREEN
BLIND	HULLO	SCREW
BLOCK	HUMBLE	SEARCH
BOAST	HUNGER	SEED
BOIL	HUNT	SEIZE
BOLD	HURRAH	SELDOM
BONE	HURRY	SELF
BORDER	HURT	SENTENCE
BORROW	HUT	SEVERE
BOTTLE	ICE	SEW
BOTTOM	IDEAL	SHADE
BOUND	IDLE	SHALLOW
BOUNDARY	IMAGINE	SHAME
BOW	IMITATE	SHARP
BOWL	IMMEDIATE	SHAVE
BRAIN	IMMENSE	SHEEP
BRASS	IMPROVE	SHEET
BRAVE	INFORM	SHELF
BREAKFAST	INFORMAL	SHELL
BREATH	INFORMALLY	SHELTER
BREATHE	INN	SHIELD
BRIBE	INQUIRE	SHILLING
BRICK	INSECT	SHIRT
BROADCAST	INSIDE	SHOCK
BROWN	INSTANT	SHOE
BRUSH	INSTRUMENT	SHOP
BUCKET	INSULT	SHOUT
BUNCH	INSURE	SHOWER
BUNDLE	INTEND	SHUT

BURIAL	INTERFERE	SICK
BURST	INTERNATIONAL	SIGNAL
BURY	INTERRUPT	SILK
BUS	INVENT	SINCERE
BUSH	INVITE	SINK
BUSY	INWARD	SKILL
BUTTER	ISLAND	SKIN
BUTTON	JAW	SKIRT
CAGE	JEALOUS	SLAVE
CAKE	JEWEL	SLIDE
CALCULATE	JOKE	SLIGHT
CALM	JOURNEY	SLIP
CAMERA	JUICE	SLOPE
CAMP	JUMP	SLOW
CANAL	KEY	SMELL
CAP	KICK	SMOKE
CAPE	KILOGRAM	SMOOTH
CARD	KILOMETRE	SNAKE
CARRIAGE	KISS	SOAP
CART	KITCHEN	SOCK
CAT	KNEE	SOIL
CATTLE	KNEEL	SOLEMN
CAUTION	KNIFE	SOLID
CAVE	KNOCK	SOLVE
CENT	KNOT	SORE
CENTIMETRE	LADDER	SORRY
CENTURY	LAMP	SOUP
CEREMONY	LAZY	SOUR
CHAIN	LEAF	SOW
CHAIR	LEAN	SPADE
CHALK	LEATHER	SPARE
CHARM	LEG	SPELL
CHEAP	LEND	SPILL
CHEAT	LESSON	SPIN
CHECK	LIBERTY	SPIT
CHEER	LID	SPLENDID
CHEESE	LIMB	SPLIT
CHEQUE	LIQUID	SPOIL
CHEST	LIST	SPOON
CHICKEN	LITRE	SPORT
CHIMNEY	LOAD	STAFF
CHRISTMAS	LOAF	STAIN
CIVILISE	LOAN	STAIRS

CLAY	LOCK	STAMP
CLEAN	LODGING	STEADY
CLERK	LOG	STEAL
CLEVER	LONE	STEAM
CLIFF	LOOSE	STEEP
CLIMB	LOT	STEER
CLOCK	LOUD	STEM
CLOTH	LOYAL	STICK
CLUB	LUCK	STIFF
COARSE	LUMP	STING
COAT	LUNCH	STIR
COFFEE	LUNG	STOCKING
COLLAR	MAD	STOMACH
COLLECT	MAIL	STORM
COMB	MALE	STOVE
COMBINE	MANAGE	STRAIGHT
COMFORT	MAP	STRAP
COMMERCE	MAT	STRAW
COMPANION	MATCH	STRETCH
COMPARE	MEAL	STRICT
COMPETE	MEANTIME	STRING
COMPETITION	MEANWHILE	STRIP
COMPLAIN	MEAT	STRIPE
COMPLICATE	MECHANIC	STUFF
COMPOSE	MEDICINE	STUPID
CONFESS	MELT	SUCK
CONFIDENCE	MEND	SUDDEN
CONFUSE	MERCHANT	SUGAR
CONGRATULATE	MERCY	SUIT
CONNECT	MERRY	SUPPER
CONQUER	MESSAGE	SUSPECT
CONSCIENCE	MESSENGER	SUSPICION
CONSCIOUS	METRE	SWALLOW
CONVENIENCE	MILD	SWEAR
CONVERSATION	MILL	SWEAT
COOK	MILLIGRAM	SWEEP
COOL	MILLILITRE	SWELL
COPPER	MILLIMETRE	SWIM
COPY	MINERAL	SWING
CORK	MISERABLE	SYMPATHY
CORNER	MISTAKE	TAIL
CORRECT	MIX	TAILOR
COTTAGE	MODEL	TALL

COUGH	MODERATE	TAME
COURAGE	MODEST	TAP
COUSIN	MONKEY	TASTE
COW	MOTION	TAXI
COWARD	MOUSE	TEA
CRACK	MUD	TELEGRAPH
CRASH	MULTIPLY	TELEPHONE
CREAM	MURDER	TEMPER
CREATURE	MYSTERY	TEMPERATURE
CREEP	NAIL	TEMPT
CRIME	NARROW	TEND
CRIMINAL	NEAT	TENDER
CRITIC	NECK	TENT
CROP	NEEDLE	TERRIBLE
CRUEL	NEGLECT	THANK
CRUSH	NEPHEW	THEATRE
CULTIVATE	NEST	THICK
CUP	NET	THIEF
CUPBOARDS	NICE	THIN
CURE	NIECE	THIRST
CURIOUS	NOISE	THORN
CURL	NONSENSE	THOROUGH
CURSE	NOON	THREAD
CURTAIN	NOSE	THREAT
CURVE	NOUN	THROAT
CUSHION	NUISANCE	THUMB
CUSTOM	NURSE	THUNDER
CUSTOMER	NUT	TICKET
DAMAGE	OAR	TIDE
DAMP	OBEY	TIDY
DANCE	OCEAN	TIE
DARE	OFFEND	TIGHT
DEAF	OMIT	TIN
DEBT	ONWARDS	TIP
DECAY	OPPOSE	TIRE
DECEIVE	OPPOSITE	TITLE
DECREASE	ORANGE	TOBACCO
DEED	ORGAN	TOE
DEER	ORIGIN	TOMORROW
DEFEND	ORNAMENT	TONGUE
DELAY	OUTLINE	TONIGHT
DELICATE	OVERCOME	TOOL
DELIGHT	PACK	TOOTH

DELIVER	PAD	TOUGH
DESCEND	PAIN	TOUR
DESERVE	PAIR	TOWEL
DESK	PALE	TOWER
DESPAIR	PAN	TOY
DEVIL	PARCEL	TRACK
DIAMOND	PARDON	TRANSLATE
DICTIONARY	PARENT	TRAP
DIG	PARK	TRAY
DINNER	PASSAGE	TREASURE
DIP	PASSENGER	TREAT
DIRT	PASTE	TREMBLE
DISAPPOINT	PATH	TRIBE
DISCIPLINE	PATIENT	TRICK
DISCUSS	PATRIOTIC	TRIP
DISEASE	PATTERN	TRUNK
DISGUST	PAUSE	TUBE
DISH	PAW	TUNE
DISMISS	PEARL	TWIST
DISTURB	PECULIAR	TYPICAL
DITCH	PEN	UGLY
DIVE	PENCIL	UMBRELLA
DONKEY	PENNY	UNCLE
DOT	PERFECT	UNIVERSE
DOUBLE	PERFORM	UPPER
DOZEN	PERMANENT	UPRIGHT
DRAG	PERSUADE	UPSET
DRAWER	PET	UPWARDS
DROWN	PHOTOGRAPH	URGE
DRUM	PICK	VAIN
DUCK	PIG	VEIL
DULL	PIGEON	VERB
DURING	PILE	VERSE
DUST	PIN	VIOLENT
EAGER	PINCH	VOWEL
EARN	PINK	VOYAGE
EARNEST	PINT	WAIST
EASE	PIPE	WAKE
EDGE	PITY	WANDER
EDUCATE	PLANE	WARM
ELASTIC	PLASTER	WARN
ELDER	PLATE	WASH
ELECTRIC	PLENTY	WASTE

ELEPHANT	PLOUGH	WAX
EMPTY	PLURAL	WEAK
ENCLOSE	POCKET	WEAPON
ENCOURAGE	POEM	WEATHER
ENGINE	POISON	WEAVE
ENTERTAIN	POLICE	WEED
ENTIRE	POLISH	WEIGH
ENVELOPE	POLITE	WET
ENVY	POOL	WHEAT
ESPECIAL	POSTPONE	WHEEL
ESSENCE	POT	WHIP
ESSENTIAL	POUR	WHISPER
EVIL	POWDER	WHISTLE
EXACT	PRACTICAL	WICKED
EXAMINING	PRACTISE	WIDOW
EXCELLENT	PRAISE	WINE
EXCESS	PRAY	WING
EXCITE	PREACH	WIPE
EXCUSE	PRECIOUS	WIRE
EXPLODE	PREFER	WITNESS
EXPLORE	PREJUDICE	WOOL
EXTRA	PRESERVE	WORM
EXTRAORDINARY	PRETEND	WORRY
EXTREME	PRIDE	WORSE
FADE	PRIEST	WORSHIP
FAINT	PRINT	WRAP
FALSE	PRISON	WRECK
FAN	PRIZE	WRIST
FANCY	PROBABLE	YARD
FASHION	PROCESSION	YELLOW
FASTEN	PROFESSION	ZERO

APPENDIX C

HEADWORDS OF THE ACADEMIC WORD LIST

ABANDON	ENABLE	OVERALL
ABSTRACT	ENCOUNTER	OVERLAP
ACADEMY	ENERGY	OVERSEAS
ACCESS	ENFORCE	PANEL
ACCOMMODATE	ENHANCE	PARADIGM
ACCOMPANY	ENORMOUS	PARAGRAPH
ACCUMULATE	ENSURE	PARALLEL
ACCURATE	ENTITY	PARAMETER
ACHIEVE	ENVIRONMENT	PARTICIPATE
ACKNOWLEDGE	EQUATE	PARTNER
ACQUIRE	EQUIP	PASSIVE
ADAPT	EQUIVALENT	PERCEIVE
ADEQUATE	ERODE	PERCENT
ADJACENT	ERROR	PERIOD
ADJUST	ESTABLISH	PERSIST
ADMINISTRATE	ESTATE	PERSPECTIVE
ADULT	ESTIMATE	PHASE
ADVOCATE	ETHIC	PHENOMENON
AFFECT	ETHNIC	PHILOSOPHY
AGGREGATE	EVALUATE	PHYSICAL
AID	EVENTUAL	PLUS
ALBEIT	EVIDENT	POLICY
ALLOCATE	EVOLVE	PORTION
ALTER	EXCEED	POSE
ALTERNATIVE	EXCLUDE	POSITIVE
AMBIGUOUS	EXHIBIT	POTENTIAL
AMEND	EXPAND	PRACTITIONER
ANALOGY	EXPERT	PRECEDE
ANALYSE	EXPLICIT	PRECISE
ANNUAL	EXPLOIT	PREDICT
ANTICIPATE	EXPORT	PREDOMINANT
APPARENT	EXPOSE	PRELIMINARY
APPEND	EXTERNAL	PRESUME
APPRECIATE	EXTRACT	PREVIOUS
APPROACH	FACILITATE	PRIMARY
APPROPRIATE	FACTOR	PRIME
APPROXIMATE	FEATURE	PRINCIPAL
ARBITRARY	FEDERAL	PRINCIPLE
AREA	FEE	PRIOR
ASPECT	FILE	PRIORITY

ASSEMBLE	FINAL	PROCEED
ASSESS	FINANCE	PROCESS
ASSIGN	FINITE	PROFESSIONAL
ASSIST	FLEXIBLE	PROHIBIT
ASSUME	FLUCTUATE	PROJECT
ASSURE	FOCUS	PROMOTE
ATTACH	FORMAT	PROPORTION
ATTAIN	FORMULA	PROSPECT
ATTITUDE	FORTHCOMING	PROTOCOL
ATTRIBUTE	FOUNDATION	PSYCHOLOGY
AUTHOR	FOUNDED	PUBLICATION
AUTHORITY	FRAMEWORK	PUBLISH
AUTOMATE	FUNCTION	PURCHASE
AVAILABLE	FUND	PURSUE
AWARE	FUNDAMENTAL	QUALITATIVE
BEHALF	FURTHERMORE	QUOTE
BENEFIT	GENDER	RADICAL
BIAS	GENERATE	RANDOM
BOND	GENERATION	RANGE
BRIEF	GLOBE	RATIO
BULK	GOAL	RATIONAL
CAPABLE	GRADE	REACT
CAPACITY	GRANT	RECOVER
CATEGORY	GUARANTEE	REFINE
CEASE	GUIDELINE	REGIME
CHALLENGE	HENCE	REGION
CHANNEL	HIERARCHY	REGISTER
CHAPTER	HIGHLIGHT	REGULATE
CHART	HYPOTHESIS	REINFORCE
CHEMICAL	IDENTICAL	REJECT
CIRCUMSTANCE	IDENTIFY	RELAX
CITE	IDEOLOGY	RELEASE
CIVIL	IGNORANT	RELEVANT
CLARIFY	ILLUSTRATE	RELUCTANCE
CLASSIC	IMAGE	RELY
CLAUSE	IMMIGRATE	REMOVE
CODE	IMPACT	REQUIRE
COHERENT	IMPLEMENT	RESEARCH
COINCIDE	IMPLICATE	RESIDE
COLLAPSE	IMPLICIT	RESOLVE
COLLEAGUE	IMPLY	RESOURCE
COMMENCE	IMPOSE	RESPOND
COMMENT	INCENTIVE	RESTORE

COMMISSION	INCIDENCE	RESTRAIN
COMMIT	INCLINE	RESTRICT
COMMODITY	INCOME	RETAIN
COMMUNICATE	INCORPORATE	REVEAL
COMMUNITY	INDEX	REVENUE
COMPATIBLE	INDICATE	REVERSE
COMPENSATE	INDIVIDUAL	REVISE
COMPILE	INDUCE	REVOLUTION
COMPLEMENT	INEVITABLE	RIGID
COMPLEX	INFER	ROLE
COMPONENT	INFRASTRUCTURE	ROUTE
COMPOUND	INHERENT	SCENARIO
COMPREHENSIVE	INHIBIT	SCHEDULE
COMPRISE	INITIAL	SCHEME
COMPUTE	INITIATE	SCOPE
CONCEIVE	INJURE	SECTION
CONCENTRATE	INNOVATE	SECTOR
CONCEPT	INPUT	SECURE
CONCLUDE	INSERT	SEEK
CONCURRENT	INSIGHT	SELECT
CONDUCT	INSPECT	SEQUENCE
CONFER	INSTANCE	SERIES
CONFINE	INSTITUTE	SEX
CONFIRM	INSTRUCT	SHIFT
CONFLICT	INTEGRAL	SIGNIFICANT
CONFORM	INTEGRATE	SIMILAR
CONSENT	INTEGRITY	SIMULATE
CONSEQUENT	INTELLIGENCE	SITE
CONSIDERABLE	INTENSE	SO-CALLED
CONSIST	INTERACT	SOLE
CONSTANT	INTERMEDIATE	SOMEWHAT
CONSTITUTE	INTERNAL	SOURCE
CONSTRAIN	INTERPRET	SPECIFIC
CONSTRUCT	INTERVAL	SPECIFY
CONSULT	INTERVENE	SPHERE
CONSUME	INTRINSIC	STABLE
CONTACT	INVEST	STATISTIC
CONTEMPORARY	INVESTIGATE	STATUS
CONTEXT	INVOKE	STRAIGHTFORWARD
CONTRACT	INVOLVE	STRATEGY
CONTRADICT	ISOLATE	STRESS
CONTRARY	ISSUE	STRUCTURE
CONTRAST	ITEM	STYLE

CONTRIBUTE	JOB	SUBMIT
CONTROVERSY	JOURNAL	SUBORDINATE
CONVENE	JUSTIFY	SUBSEQUENT
CONVERSE	LABEL	SUBSIDY
CONVERT	LABOUR	SUBSTITUTE
CONVINCE	LAYER	SUCCESSOR
COOPERATE	LECTURE	SUFFICIENT
COORDINATE	LEGAL	SUM
CORE	LEGISLATE	SUMMARY
CORPORATE	LEVY	SUPPLEMENT
CORRESPOND	LIBERAL	SURVEY
COUPLE	LICENCE	SURVIVE
CREATE	LIKEWISE	SUSPEND
CREDIT	LINK	SUSTAIN
CRITERIA	LOCATE	SYMBOL
CRUCIAL	LOGIC	TAPE
CULTURE	MAINTAIN	TARGET
CURRENCY	MAJOR	TASK
CYCLE	MANIPULATE	TEAM
DATA	MANUAL	TECHNICAL
DEBATE	MARGIN	TECHNIQUE
DECADE	MATURE	TECHNOLOGY
DECLINE	MAXIMISE	TEMPORARY
DEDUCE	MECHANISM	TENSE
DEFINE	MEDIA	TERMINATE
DEFINITE	MEDIATE	TEXT
DEMONSTRATE	MEDICAL	THEME
DENOTE	MEDIUM	THEORY
DENY	MENTAL	THEREBY
DEPRESS	METHOD	THESIS
DERIVE	MIGRATE	TOPIC
DESIGN	MILITARY	TRACE
DESPITE	MINIMAL	TRADITION
DETECT	MINIMISE	TRANSFER
DEVIATE	MINIMUM	TRANSFORM
DEVICE	MINISTRY	TRANSIT
DEVOTE	MINOR	TRANSMIT
DIFFERENTIATE	MODE	TRANSPORT
DIMENSION	MODIFY	TREND
DIMINISH	MONITOR	TRIGGER
DISCRETE	MOTIVE	ULTIMATE
DISCRIMINATE	MUTUAL	UNDERGO
DISPLACE	NEGATE	UNDERLIE

DISPLAY	NETWORK	UNDERTAKE
DISPOSE	NEUTRAL	UNIFORM
DISTINCT	NEVERTHELESS	UNIFY
DISTORT	NONETHELESS	UNIQUE
DISTRIBUTE	NORM	UTILISE
DIVERSE	NORMAL	VALID
DOCUMENT	NOTION	VARY
DOMAIN	NOTWITHSTANDING	VEHICLE
DOMESTIC	NUCLEAR	VERSION
DOMINATE	OBJECTIVE	VIA
DRAFT	OBTAIN	VIOLATE
DRAMA	OBVIOUS	VIRTUAL
DURATION	OCCUPY	VISIBLE
DYNAMIC	OCCUR	VISION
ECONOMY	ODD	VISUAL
EDIT	OFFSET	VOLUME
ELEMENT	ONGOING	VOLUNTARY
ELIMINATE	OPTION	WELFARE
EMERGE	ORIENT	WHEREAS
EMPHASIS	OUTCOME	WHEREBY
EMPIRICAL	OUTPUT	WIDESPREAD

OFF-LIST WORDS FOUND IN THE CORPORA

aftershave	headaches	seawater
aftershocks	headband	seaweed
airbus	headline	shelfs
airplane	headlines	shopkeeper
airport	headscarf	sightseeing
airports	headscarves	silversmith
autopilot	healthcare	skydivers
backdrop	homestay	spacecraft
backlash	hometown	spaceship
backpack	homework	spaceships
backstroke	housewife	sportsman
bathroom	housework	sportsmen
bathtub	hydro	sportsperson
bedroom	inabilities	stallholder
bedrooms	intensifier	stallholders
bedtime	intensifiers	steamboat
birdlife	interfaith	stomachache
blackboards	lampshades	storyteller
bodyguards	landfill	suitcase
bookmark	layout	suitcases
bookshop	lifeboat	sunglasses
breakups	lifestyle	sunset
businessman	lifestyles	sunshade
businesswoman	lifetime	superhuman
checklist	lightyear	superjumbos
childcare	lunchtime	supermarket
childminders	microenterprises	supermarkets
classmate	microfinance	superstar
classmates	middleman	swimsuits
classroom	milestones	tai
classrooms	motorbike	takeaway
coffeehouse	motorcycle	teabag
coffeehouses	motorway	teamwork
countryside	newfound	teapot
coursework	newsletter	textbook
craftsman	newsroom	timekeeper
craftsmen	notebook	timekeeping
crewman	notebooks	timeline
crewmen	paperwork	timetable
crossroads	phrasebook	timetables

darkroom	playground	toothache
daytime	playgrounds	toothbrush
dishwasher	policeman	toothpaste
doorbell	policemen	townie
doorstep	policewoman	underground
downside	possessives	upstairs
downstairs	postcard	viewpoint
downtown	postcards	wallpaper
earrings	postman	wallpapers
fisherman	proofreader	waterproof
fishermen	quantifier	weatherproof
fishmonger	quantifiers	weekdays
folktales	racquetball	weekend
foodstuffs	railway	weekends
foots	raincoat	weeknights
footwear	rainforest	whiteboard
freestyle	rainforests	whiteboards
futurologists	ringtones	wholegrains
gill	roommate	wildlife
goalkeeper	salesman	windmill
guidebook	sandcastle	windmills
guidebooks	sandstorm	woodland
gunpowder	schoolbook	woodlands
gunshot	schoolchild	woodwork
gunshots	scrapbook	woodworms
haircare	seafood	wordlist
halfway	seagull	workbook
handwriting	seagulls	workforce
handwritten	seamen	workplace
headache	seatbelts	workshops

APPENDIX E

NATION'S (2001) SEQUENCE LIST OF DERIVATIONAL AFFIXES FOR

LEARNERS OF ENGLISH

Stage 1	-able, -er, -ish, -less, -ly, -ness, -th, -y, non-, un- (all with restricted uses)
Stage 2	-al, -ation, -ess, -ful, -ism, -ist, -ity, -ize, -ment, -ous, in- (all with restricted uses)
Stage 3	-age (leakage), -al (arrival), -ally (idiotically), -an (American), -ance (clearance), -ant (consultant), -ary (revolutionary), -atory (confirmatory), -dom (kingdom, officialdom), -en (wooden), -ence (emergence), -ent (absorbent), -ery (bakery, trickery), -ese (Japanese), -esque (picturesque), -ette (usherette, roomette), -hood (childhood), -i (Israeli), -ian (phonetician), -ite (Paisleyite), -let (coverlet),-ling (duckling), -ly (leisurely), -most (topmost), -ory (contradictory), -ship (studentship), -ward (homeward), -ways (crossways), -wise (endwise), anti- (anti-inflation), ante- (anteroom), arch- (archbishop), bi- (biplane), circum- (circumnavigate), counter- (counter-attack), en- (enslave), ex- (ex-president), fore- (forename), hyper- (hyperactive), inter- (inter-Afican), mid- (mid-week), mis- (misfit), neo- (neo-colonialism), post- (post-date), pro- (pro-British), semi- (semi-automatic), sub- (subclassify), un- (untie)
Stage 4	-able, -ee, -ic, -ify, -ion, -ition, -ive, -th, -y, pre-, re-
Stage 5	-ar (circular), -ate (compassionate, captivate, electorate), -et (packet, casket), -some (troublesome), -ure (departure, exposure), ab-, ad-, com-, de-, dis-, ex- ('out'), in- ('in'), ob-, per-, pro- ('in front of'), trans-