

Vocabulary in English Textbooks for Swedish Secondary School

A Study about the Frequency of Vocabulary in Textbook Wordlists and the Selected Vocabulary Exercises

By: Taner Tanriverdi

Supervisor: Harriet Sharp
Södertörn University | School of Culture and Education
Bachelor's essay 15 credits
English | Autumn semester 2020



SÖDERTÖRN UNIVERSITY | STOCKHOLM
sh.se

Abstract

This study examines the level of frequency of wordlists, and the selected vocabulary exercise type in two English textbooks aimed at first year (English 5) and third year (English 7) Swedish secondary school students. A quantitative analysis was conducted by inserting the wordlists into an online tool, which generated a frequency list based on lemmas in accordance with the BNC corpus. The analysis of the wordlist indicates a majority of low frequency lemmas and a minority of high frequency lemmas, which studies report that students are not familiar with. The results of the frequency analysis support findings of previously conducted studies regarding insufficient number of high frequency words in EFL textbooks. Vocabulary exercise types were also identified, counted and analysed. The analysis of the vocabulary exercise types indicates a preference for production exercises in the textbook aimed at English 5, and a preference of interpretation exercises in the English 7 textbook. The effectiveness of commonly found exercise types in the textbooks are supported by previous research.

KEYWORDS: Vocabulary, Frequency, EFL, Textbook, Exercise, Lemma, BNC

1. INTRODUCTION.....	1
1.1 Aim and research questions	1
2. BACKGROUND AND LITERATURE REVIEW	3
2.1 Vocabulary frequency: different types of words	3
2.2 Vocabulary learning & exercises	4
2.3 Swedish second language context.....	5
2.4 Literature review	6
2.4.1 A corpus-based study of lexis in L2 English textbooks.....	6
2.4.2 Profiling vocabulary for proficiency development: Effects of input and general frequencies on L2 learning.....	6
2.4.3 Vocabulary in high-school EFL textbooks: Texts and learner knowledge.....	7
2.4.4 The effect of exercise types on EFL learners' vocabulary retention.....	8
2.4.5 The effectiveness of using vocabulary exercises to teach vocabulary to ESL/EFL learners.....	9
2.4.6 Summary of studies.....	10
3. MATERIAL AND METHODS	10
3.1 Material	10
3.2 Tools	11
3.3 Method	12
3.3.1 Data collection	12
3.3.2 Analysis.....	13
4. RESULTS	14
4.1 The frequency of the Blueprint A textbook	14
4.2 The frequency of the Blueprint C textbook:	16
4.2.1 Differences and similarities between the results.....	19
4.3 Selected vocabulary exercises in Blueprint A	20
4.4 Selected vocabulary exercises in Blueprint C.....	22
5. DISCUSSION	23
6. CONCLUSION	25
7. REFERENCES.....	27
8. APPENDIX.....	29

Vocabulary in English Textbooks for Swedish Secondary School

1. INTRODUCTION

Vocabulary learning is an important aspect of second language acquisition, since without words we would not be able to create sentences and communicate effectively with others. Native speakers of English usually know just under 20000-word families (Nation 2013, p.13). A word family is “a set of word forms, sharing a common meaning” (Read 2000, p.19). According to Nation, native speakers learn around 1000 new words a year between the ages 3-25; therefore, he argues that learning new words at a similar or faster pace is an ambitious goal for non-native second language learners (2013, p.13). How and what vocabulary we learn, or rather we are taught at school, is of interest because it affects our communication with others, and our ability to gain knowledge by reading and listening. Since one of the main settings where we learn vocabulary is in school, the acquisition of vocabulary is thought out and planned to suit the context and the need of students. This happens through didactical and pedagogical choices made by teachers and by producers of material that is used in the classrooms by other teachers, which is based on the criteria of the National curriculum.

These criteria of the curriculum are determined by the Swedish National Agency for Education (*Skolverket*), which bases its curriculum on the understood goals and needs of students. The textbook materials that are used in English lessons for different levels of students in Swedish schools are therefore produced to be in accordance with the curriculum, but the approach of different textbook publishers could be different. Since Swedish secondary school students are required to read English 5 & 6 (English 7 is optional), the different levels of textbooks should correlate in accordance with the national standard requirement produced by *Skolverket*.

Based on an observation, it seems as if the English words that are taught at the secondary school level consist of many complex and abstract words. It would be interesting to investigate if there is a pattern in the selection of words (vocabulary) and if there is a difference in the how the words are taught to students of different levels.

1.1 Aim and research questions

The main aim of this essay is to survey the selected vocabulary in two Swedish secondary school English textbooks for English 5 and 7 students (aged 16-19), by investigating their

glossary/wordlists and comparing the frequency of words found to the words in a multi-million-word corpus of British English (BNC). The vocabulary will be categorized into high-, mid- or low-frequency words (lemmas). A lemma is the base word form of a word without inflections. This study will also investigate the most commonly found didactic methods (exercise types) used to teach vocabulary at the two textbook levels, such as “fill in the gap” or “production of sentences”. The research questions of this study are:

1. How common are the words (lemmas) found in English textbook glossary/wordlists for Swedish secondary schools?
2. What types of exercises are used to teach vocabulary in the textbooks, and how do these distribute by level?

Few studies have been conducted regarding the type of vocabulary and vocabulary exercises of English textbooks for Swedish Secondary schools. This essay therefore intends to add information to an under-researched area. It is hoped that this study will give insights into frequency of the selected vocabulary in the wordlists/glossaries, and the common didactic methods (exercises) used to teach vocabulary. This will be investigated by examining the frequency of words (lemmas) selected by producers of the textbooks, and by identifying the types of vocabulary exercises found in the textbooks.

It is important to be aware of the use of the lexical item ‘word(s)’ and the use of ‘lemma(s)’ throughout this essay. The use of ‘lemma’ refers to the basic form of the ‘word’ without inflections after they have been lemmatized i.e., generated into lemmas using tools (see section 3.3 & 3.3.2). It is the form that has been worked with in this study. On the other hand, the use of ‘word’ refers to the non-lemmatized form, i.e., the lexical item as it was originally found in the wordlist.

Furthermore, it is also hoped that teachers will benefit from information about the frequency of vocabulary and the approaches to teaching vocabulary to second language learners in Sweden, such as fill in the gap, using selected word to produce oral or written sentences or working with synonyms, and that this information will assist them in their teaching careers. This study will provide additional insights into approaches based on the different levels of English textbooks of the 2.0 Blueprint series (English 5 & 7).

However, it should be noted that this essay does not claim to provide information which could be generalized to all English textbooks of year 1 and 3 (English 5 and 7). Rather, the results will only pertain to the material analysed. It is hypothesized that the lower levels of English textbooks will contain more high-frequency words (lemmas), since they are used

more commonly in communication with others. The higher levels of English textbooks will have more low-frequency and specialized words (lemmas), due to students learning the high-frequency words (lemmas) at the lower levels. The didactic methods will probably change from “fill in the gap” exercises at lower levels, to students being asked to produce their own sentences in higher levels of English textbooks.

2. BACKGROUND AND LITERATURE REVIEW

2.1 Vocabulary frequency: different types of words

There are words in every language that are used more or less frequently. According to Nation, the frequency of words can be categorized into three levels: high, mid and low frequency words (2013, p.18). Besides these levels, there are specialized words which are more frequent in specific contexts, such as academic texts or texts with specialized topics. These words are categorized as academic and technical words (2013, p.19).

High frequency words are the words that are most commonly used by speakers across all different contexts, both informal and formal, such as in conversations with friends or in academic texts. The 50 most frequently used words in the English language are mostly pronouns (e.g., *I, we, you*), prepositions (e.g., *to, of, in*) or conjunctions (e.g., *and, but, or*) (BNC Corpus). These words are essential for English speakers to use in order to communicate properly, since they are function words which either specify words (pronouns) or indicate relations of time, space, and logic between entities (prepositions) or link words and sentences together (conjunctions). Besides the high frequency function words, there are high frequency content words, i.e., nouns, verbs, adjective and adverbs.

Nation describes **medium frequency words** as “[words] that almost got into the high-frequency word lists” (2013, p.18). These words are slightly less common than high-frequency words, but it is possible to find them in high-frequency wordlists. Some examples provided by Nation are “*zoned, pioneering, aired and pastoral*” (2013, p.18). He differentiates mid-frequency words from low-frequency words by pointing out that mid-frequency words similar to high-frequency words are needed to “deal with English without the need for outside support” (Nation 2013, p.18), since when using low frequency words, it could sometimes require an extra explanation of the word.

Low frequency words are rarely used by speakers but could be more common in specific contexts. These words are not essential for English learners to acquire in order to communicate, rather they could sometimes be advanced even for native speakers. The low-frequency words are often technical words (a part of specialized words), more commonly used in specific areas which most people rarely find themselves in (Nation 2013, p.19). Examples are, for instance, “*policy, phase and sustained*” or “*indigenous, regeneration and timber*”, according to Nation (2013, p.19). Nation (2008) argues for the importance of knowing academic vocabulary for people in academia, and the technical vocabulary for the specific disciplines (pp.128;133).

2.2 Vocabulary learning & exercises

Learning new vocabulary is different for native speakers and second language learners, because second language learners usually have a more limited input of the language than native speakers. Therefore, learners become more dependent on lessons taught in school since it creates an environment where second language students can further develop their vocabulary size through more input. Teachers often use materials from textbooks produced in accordance with the curriculum, and there are different types of vocabulary exercises which can be found in secondary school English textbooks. Paribakht and Wesche’s study examined the common types of twelve vocabulary teaching textbooks (Paribakht & Wesche 1997, p.183-184). Vocabulary learning exercises can be divided into the following five types:

- (1) Selective attention
- (2) Recognition
- (3) Manipulation
- (4) Interpretation
- (5) Production

The first type of vocabulary learning exercise is **selective attention** which draws the learner’s attention to target words by, for instance, bolding and italicizing the words, or giving the learners a list of words before starting to read a text containing the words. The second type of exercise is **recognition** which entails working with recognizing the meaning of words. Recognition exercises would consist of, for instance, matching words or synonyms, or matching words with their meaning etc. The third type of exercise is **manipulation**, which deals with grammatically modifying the target words. Learners are expected to, for instance, add affixes to words. The fourth type of exercise is **interpretation**. This type of exercise requires students to analyse words and the context. Examples of exercises are finding the odd words (among related words), classifying and substituting words, multiple-choice cloze exercises (fill in the

gap). The fifth type is **production**, and the exercises deal with open cloze exercises, using the target word in other contexts, etc.

2.3 Swedish second language context

The English language proficiency level in Sweden is among the highest in the world. According to Education First (2020), Sweden ranks fourth among countries described as having “very high proficiency” (p.6). The growing usage of English in Sweden has for a long time been discussed as a potential threat to Swedish. In 1997 to 2003, policies for strengthening the Swedish language were proposed in order to battle the increased usage of English (Hult 2003, p.45-46). One of the published reports for the policy proposal was “Hål i Mun”, which had eight recommendations for protecting the Swedish language. The debate about English as a threat to Swedish has been discussed from various perspectives. It seems as if the common concern of studies debating this issue concerns the perceived prestigiousness of using English in business and academia (Bolton & Meierkord 2013, p.99). However, even though English is seen as a threat by some people, Swedish students are still expected to speak the second language at a high-level.

The Swedish curriculum requires that English should be read by students from the first grade (age 6-7) of elementary school to the first (English 5) and second year (English 6) of secondary school (age 16-17), while reading English 7 in the third year is optional. It is common for Swedish university programs to require a certain level of secondary school English in order to be qualified to apply. Therefore, the ability to speak English at certain level of proficiency could be seen as a standard requirement in the Swedish context, and speaking at a higher level is both necessary and prestigious in contexts such as academia or business.

The motivation behind teaching English in the Swedish school is that it “surrounds us in our daily lives and is used in such diverse areas as politics, education and economics” (Skolverket, n.d., para. 1). Therefore, students are expected to be able to discuss these matters in English at all levels. However, the course requirements and content differ between English 5 and English 7. It seems that English 5 students are expected to have general knowledge of societal issues, while English 7 students are expected to have more complex and theoretical knowledge of societal issues or cultural practices in countries where English is spoken (Skolverket). The difference between the courses seems to be the level of in-depth knowledge students are required to learn and be able to use in the English 7 course, as opposed to English 5 course.

2.4 Literature review

Previous studies such as Norberg & Nordlund (2018), Sakata (2019), Sun & Dang (2020), Hashemzadeh (2012) and Tahir & Mohtar (2016) have been conducted regarding vocabulary selection and the effect of exercise type found in second language English textbooks in foreign non-native contexts. The findings of the studies give general insights into the pedagogical implications of the selected vocabulary, and the most effective vocabulary exercises in second language English textbooks. This section will discuss five studies related to the research questions of this essay.

2.4.1 A corpus-based study of lexis in L2 English textbooks

This study by Norberg & Nordlund (2018) examines the selection of English second language vocabulary in seven Swedish primary school textbooks. The aim of this study was to conduct a corpus-based textbook analysis of multiple English textbooks used in Swedish primary school, in order to see a pattern of similarities regarding construction, number of words for same school level, correspondence with everyday language and common pedagogical idea for vocabulary selection (p.463). The corpus analysis was made using the SWYLC (the Swedish Young Learner Corpus) with its 34,380 words. Other corpora, such as NGSL (New General Service List) and the VP-kids corpus (VocabProfilers) were used for low-frequency words (p.465). The analysis was conducted using computer software programs. The authors focused on content words (nouns, adjectives and verbs), with the motivation that they are more important words since they carry the meaning in communication with others (p.465). The findings show that the Swedish primary school English textbooks vary both in their vocabulary size and their selection of words (p.466). A comparison between the books and the corpora show that textbooks have a large number of low-frequency words. According to the authors, the findings indicate that the textbooks do not have a “common pedagogical idea”, since they vary in vocabulary content and size (p.469). The authors discuss the implications of varying approaches to vocabulary in the textbooks, suggesting that the differences could lead to varying vocabulary size and word exposure for the learners (p.469). The authors conclude by suggesting that the study shows that corpus linguistics is valuable when analysing textbooks (p.470).

2.4.2 Profiling vocabulary for proficiency development: Effects of input and general frequencies on L2 learning

This study by Sakata (2019) explores what type of vocabulary is found in textbooks and what general frequencies are connected to second language proficiency. The aim of this study is to

examine Japanese English learners' vocabulary sizes, and their correlation with textbook word frequencies and general frequencies (found in the COCA, Corpus of American English) (p.3). The method used was testing Japanese university students' vocabulary size and proficiency by conducting three tests: a self-made vocabulary test, a vocabulary size test, and a proficiency test (p.3). The material used for creating a self-made test, and for the purpose of analysing textbook correlation was created by compiling a self-made corpus based on 24 textbooks, all of which were authorized by the government (certified, before they were used in school) (p.3). The data of 52 students were analysed after the test was conducted. The findings show that the students with knowledge of frequent words found in the COCA corpus and infrequent words in textbooks had better proficiency (p.8). According to Sakata, the knowledge of frequent COCA words was more significant for proficiency than vocabulary size (p.8). His finding suggests that the selection of vocabulary found in the English textbooks is insufficient for proficiency, in comparison to the general frequencies of words (p.9). Therefore, Sakata suggests that teachers should be aware of this insufficiency when working with textbooks, since other studies show that teachers are mostly using textbooks when teaching (p.9).

2.4.3 Vocabulary in high-school EFL textbooks: Texts and learner knowledge

This study by Sun & Dang (2020) examines the vocabulary in EFL textbooks used in Chinese education, also focusing on the high-frequency words and measuring students' vocabulary knowledge. 265 EFL high school students participated in the study conducted in East China (p.4). All students were users of the same textbook series. The material used in the study was a corpus of circa 273.000 words collected from eleven high school textbooks published in 2010 (4p.). The participants were tested in order to measure their vocabulary. Five tests with different frequency levels of words were conducted with the intention of differentiating between students' knowledge of high and low frequency words (p.4). An analysis of the textbooks in relation to COCA/BNC (The British National Corpus of English) frequency lists was conducted in order to learn how many high and low frequency words could be found, but also how many times students would encounter these words (p.4). The findings show that the textbooks have incorporated the 1000 most frequent words with most of them being encountered seven or more times (p.8). The problem with textbooks was that words of the second or third 1000 frequency lists were not represented (p.8). Students had sufficient levels of knowledge regarding the first 1000 frequent words, but failed in knowing the second or third 1000 frequent word lists, i.e., the 2000-3000 most frequent words (p.5). The findings also show that the textbooks were too demanding for students, since they needed to know words which they rarely encountered in the

textbooks (p.9). The authors conclude by suggesting that textbook publishers should take their findings into consideration, and that teachers should be aware of the limitations of textbooks used in their education (p.10).

2.4.4 The effect of exercise types on EFL learners' vocabulary retention

This study by Hashemzadeh (2012) examines the effect of different exercise types on EFL learner's vocabulary retention. The study investigates if EFL students' retention of new words would vary based on the type of vocabulary exercise the students worked with (p.1716). A pre-test, i.e., a pilot study of 92 words was conducted on multiple EFL learners in order to make sure the words were unknown to the learners (p.1720). The words were picked from texts from "Elementary Total English Book". Words that were known to students were excluded from the list, which led to four texts with 40 words being selected (p.1720). Two types of exercises were designed: recognition exercises and production exercises. The recognition exercises were "fill-in-the-blank" and "matching" exercises. The production exercises were "paraphrasing exercises" and "glossing", i.e., working with synonyms (p.1720). Two tests were conducted; the first one, called the *immediate test*, was a test carried out by instructing students during four lessons, during which the students were instructed to read a text and then practice by working with one of the exercises. After each time working with the exercises, students were tested on the vocabulary learnt during the lesson (p.1721). The second test was a *delayed post-test*, a test carried out two weeks after the immediate tests, in order to test the retention of students based on the vocabulary worked; this was done in order to find out which vocabulary was remembered related to the exercises used (p.1721). A total of 65 female Iranian elementary students participated in the study but since some students were absent during lessons where testing and instruction took place, only the data from 46 students were analysed (p.1720). The data collected from the tests were analysed repeatedly using ANOVA, i.e., an analysis of variance which finds differences of variation between two or more groups (p.1721). The analysis showed that the highest retention of vocabulary was with the fill-in-the-blank exercise, followed by the paraphrasing exercise, the matching exercise and then the glossing exercise (p.1722). The analysis indicated that there was a significant difference in results of the test based on the exercise used by participant to practice the vocabulary (p.1722). The results of the analysis showed a significant difference in the following: the recognition exercises were more effective than production exercises regarding vocabulary retention in both tests (p.1723). The author concluded by discussing the importance of teachers being aware of the most effective teaching method for vocabulary (p.1725).

2.4.5 The effectiveness of using vocabulary exercises to teach vocabulary to ESL/EFL learners

This study by Tahir & Mohtar (2016) examines the effectiveness of vocabulary exercises to teach vocabulary in course textbooks for ESL/EFL learners. The study was conducted in Malaysia on two groups (experimental and control) of 30 Form Two textbook students, with a total of 60 participants (p.1655). The methodology used in the study was a quasi-experimental research design, motivated by being a highly controlled method. A word test was conducted to pick out 30 target words out of 90 based on the curriculum document published by the Malaysian Ministry of Education, which specified the words that were to-be-learned by Form Two students (p.1655). Two tests were conducted with the participants, a pre-test and a post-test. In the pre-test, students were instructed to read eleven short texts containing synonyms of, or phrases with the same meaning as the 30 target words (p.1655). The target words were highlighted in the text for students to make them aware of the words that were meant to be learned. After reading the short texts, the participants were instructed to answer a multiple-choice test with 30 questions within a timeframe of 45 minutes (p.1655). The post-test was the same test as the pre-test, but with another arrangement of texts. Between the pre-test and post-test, the students from the experiment group had seven sessions with vocabulary exercises, which were adapted from the Form Two KBSM textbooks and Form Two reference books (p.1655). There were five types of vocabulary exercises: matching, fill-in-the-blanks, spelling, unscramble letters, and crossword puzzle (p.1656). The control group had seven sessions of normal English lessons, without using the exercises (p.1656). After the students were done with the post-test, they were instructed to answer a 3-point Likert-scale questionnaire and also an evaluation form with eight open questions regarding their experiences with regards to the test (p.1656). A data analysis was conducted on the collected data from the pre-test, post-test, questionnaire and evaluation form (pp.1656-1657).

The results show that half of the participating ESL/EFL students failed to know the target words which also were part of the syllabus from an earlier textbook (Form One) (p.1663). The authors argue that this could be an indicator of the ineffective method of implicit vocabulary teaching, but it could also be due to the low proficiency of the participants (pp.1663-1664). Furthermore, the results show that the experimental group had significant improvement in comparison to the control group, which according to the author shows the importance of repetition with exercises for retention (p.1664). Other findings show that the preferred exercise was matching, followed by crossword puzzle, unscramble letters and spelling, with fill-in-the-

blank exercises being the least preferred by students (pp.1664). The authors argue that in order for students to learn they need to be motivated by multiple interesting exercises, and that teachers and curriculum publishers should take this into account (pp.1666-1667).

2.4.6 Summary of studies

The previous studies show that EFL textbooks do not meet the demand for high frequency words in their selected vocabulary, which is widely agreed on to be important for EFL students to know in order to communicate proficiently in English. The conclusion of the different studies was that textbooks lacked a common pedagogical idea, or that the selected words in the textbooks were irrelevant with regards to them being low frequency words which are insufficient for the proficiency of the students. The studies indicate that textbook developers do not follow the recommended pedagogical ideas and teaching methods which are supported by literature, such as working with high-frequency words. Other previous studies regarding vocabulary exercises in EFL textbooks show that the exercises which seem to be effective for students may differ based on the context of the students. One study supports fill-in-the-blank exercises as the most effective exercise type, while the other study supports matching exercises as the more effective due to it being more interesting and therefore more motivating. The previous studies indicate that there is a shared problem with textbook publishers not using enough high-frequency words in the textbooks. The studies regarding vocabulary exercises indicate that fill-in-the-blank and matching exercises could be the most effective way of learning vocabulary.

3. MATERIAL AND METHODS

3.1 Material

The material used in this study is two English textbooks at different course levels of the Blueprint series. Blueprint A is aimed at meeting the requirements to teach first year secondary school students at the English 5 level. The other textbook, Blueprint C, is aimed at the third-year secondary school students reading the voluntary English 7 level. Both of the textbooks are from the second edition (2.0 version), but they differ in the year they were published. Blueprint A was published in 2007, and Blueprint C was published in 2011. Both of the textbooks were published by Liber AB, but authorship of the textbooks differs. The authors of Blueprint A are

Lundfall, Nyström and Clayton, while the authors for Blueprint C are McKay, Brodin, Clayton and Webster.

The material that was used specifically from the textbooks were those pages or parts that dealt with wordlists and vocabulary exercises, in order to collect data for the analysis and answer the research questions. The textbook wordlists consisted of marked and unmarked words. The words that were considered to be among the core words of English were marked with blue or grey dots by the authors. The unmarked words were therefore not considered to be as essential for students and teachers to work with, in comparison to the marked words.

The two textbooks differ regarding the level of complexity (English 5 and 7). Based on observation, one could claim that the content of the Blueprint A seems to be more general because the texts of the textbook deal with societal issues, such as norm criticism of gender roles. Contrary to the Blueprint A textbook, the Blueprint C textbook explores the general issues of society more in-depth by discussing them through the use of English fictional texts, e.g., Bram Stoker's *Dracula*, or Virginia Woolf's *To the Lighthouse*. The textbook teaches students about literary criticism by containing chapters about popular contemporary perspectives in literary criticism, such as Modernism, Postmodernism, Postcolonialism and Feminist criticism.

3.2 Tools

Two main tools were used for the analysis of the data collected from the material: **Excel** and **SketchEngine**. Excel is a computer program that enables users to insert data and analyse it using the tools available. The tool mostly used in Excel was the “create rules” tool which would either compare two columns of separate data by marking the duplicate or the odd and unique words. Excel was also used to categorize and sort the words into categories, such as the single words, hyphenated words and multiwords.

SketchEngine was the main tool for analysis of the data. It is an online program that is used by researchers, professionals and students etc. around the world and gives users multiple tools to analyse language, and, in particular, words. The tool that was used in the analysis was the Wordlist tool, which provides a list of the most frequent words or lemmas based on the corpus that the user chooses. Based on their preferences, users could also change the settings and the way the data were presented. Examples of some of the settings are minimum or maximum frequency, adding a subcorpus, and focusing on a word class. The settings chosen for this analysis were, for instance; minimum of 1 frequency, whole corpus, and generating and analysing as so-called **lemmas**. A lemma could, according to Nation (2013), be described as “consist[ing] of a headword and its inflected forms and reduced forms (n't)” (p.10). He goes on

to explain that “The English inflections consist of plural, third person singular present tense, past tense, past participle, -ing, comparative, superlative, possessive.” (p.10). For example, eat and *eating* are identified as the same lemma.

It was also possible to paste a list of words into the tool (SketchEngine), which the tool then analysed and sorted from the most frequently found in the corpus to the least. The frequency was presented using the term (formula) called ‘Absolute Frequency’ and ‘Absolute Frequency per million’. The Absolute Frequency is the total number of occurrences that the word was found in the whole corpus, while the Absolute Frequency per million was how many times the word was found per million in the corpus. The latter is useful to determine the frequency of words from corpuses of different sizes. It should be noted that there could be some exceptions to the lemmatization of words in the corpus found in SketchEngine, if the corpus was obtained from an external source.

3.3 Method

3.3.1 Data collection

The wordlists of Blueprint A (English 5) and Blueprint C (English 7) were manually entered into Excel documents, categorized into three groups: Single words, hyphenated words and multiword. The words were carefully spellchecked by using the Word-program in order to identify mistakes or differences of spelling (British or American). The words that had a spelling mistake were corrected after examining how they were spelt originally in the textbook wordlist. By using the tools found in the Excel program, duplicates or triplicates with the same meaning, slang and archaic words, but also multiwords expressions (consisting of two to seven words) were removed for the following reasons:

- a) it was not necessary to analyse the same words twice or thrice,
- b) the word was not spelled in a proper way (i.e., slangs words) or in a way not spelt by users of the English language today (archaic words),
- c) the multiwords required additional tools in order to be analysed and using other tools could cause problems when fit in with data of the single words.

Out of the identified duplicates (in pairs of two), one was deleted, and out of a group of three, either two or one was deleted based on the meaning of words. If words had the same meaning, one or two was deleted in order to have one word with one meaning. If two or three words had different meanings but were spelt the same way (homonyms), they were sorted by their word class and then analysed separately.

3.3.2 Analysis

An analysis of the single and hyphenated words found in both textbooks was conducted using SketchEngine. The remaining words after the categorization of the textbook wordlist were inserted into the program and a frequency wordlist was generated based on lemmas. The option of using lemmas instead of words was chosen in order to generate data close to other frequency wordlists of the literature. According to Paquot (2018), there has been an increased production of frequency lists out of different corpora, which have used lemmas instead of word families (p. 360). One example that is mentioned by Paquot is Brezina and Gablasova's *New General Service List* (2015). Therefore, for the purpose of simplifying data and defining frequency, the data generated were in the form of lemmas.

The choice of corpus was the British National Corpus (BNC), which consisted of circa 464 thousand items (lemmas) and with the total of circa 96.8 million frequencies (rounded up to 97 million). According to SketchEngine, the BNC corpus consists of circa 90 % written text (academic papers, newspapers etc.) and circa 10 % speech (radio shows, etc.). The lowest frequency of the BNC corpus allowed in the generated list was the frequency of 1.

Once the SketchEngine analysis was conducted, the frequency list was downloaded as an Excel file and manually categorized into three frequency groups: high-, mid- and low, based on definitions I created. The definitions were based on the Absolute Frequency per million from the SketchEngine program and are outlined in the bulletpoints below.

- **High frequency lemmas** were defined as having an Absolute Frequency per million value of 97 or more, which would mean a minimum of circa 10.000 frequency per million.
- **Mid frequency lemmas** were defined as having an Absolute Frequency per million value between 96.9 and 9.7, which would mean circa 9999 – 100 frequency per million.
- **Low frequency lemmas** were defined as having the maximum Absolute Frequency per million value of 9.7, which means having a frequency of circa 100 per million.

The following section will give a description of the results obtained through the analysis of the texts.

4. RESULTS

The results will answer both research questions regarding how common the words (lemmas) are in the Blueprint English textbook series (both A and C), and the selected (more commonly found) vocabulary exercise type in both of the textbooks. The frequency of the textbook wordlists will be answered based on the analysis of the SketchEngine tool using British National Corpus (BNC) (see Appendix) and according to a self-defined categorization of high-, mid and low frequencies of words (lemmas). The question concerning selected vocabulary exercise type will be answered using Paribakht and Wesche's categorization of vocabulary exercise types, i.e., selective attention, recognition, manipulation, interpretation and production (see p. 4).

4.1 The frequency of the Blueprint A textbook

The wordlist of Blueprint A English textbook (English 5) for first year secondary school students between the ages 16-17, was analysed. The number of words (and expressions i.e., multiword combinations of words) found in the wordlist was originally 1498. Regarding duplicates, triplicates and multiwords, a total of 148 duplicates and 241 expressions were deleted. A total of three slang or archaic words were identified and therefore deleted.

The exclusion of a total of 389 duplicates and expressions, and three slang words led to the total of 1106 words being analysed using SketchEngine wordlist frequency tool. The tool provided a frequency list sorted by highest to lowest frequency, i.e., in order of lemmas with the highest number of occurrences in the corpus to lemmas with the lowest number of occurrences. Out of 1106, a total of sixteen lemmas were not found in the list. This could, for instance, be due to not being found in the BNC corpus. One example is '*thorn-studded*'. Therefore, the list provided by SketchEngine had a total of 1090 lemmas.

The results of the analysis of the lemmas showed that the top five lemmas from the Blueprint A textbook wordlist in terms of the highest frequency according to the BNC Corpus had the Absolute Frequency ranging between 56948 – 46496 (see Table 1 & Appendix). This means that they could be found between circa 57.000 to 46.500 times in the whole BNC corpus. The lemma 'course' was at the top of the list among the lemmas found in the Blueprint A wordlist but was at the 174th spot in the whole corpus. The Absolute Frequency per million (out of circa 97 million frequencies) of the top five list was between 506.8 – 413.8 (see Table 1 below)

Table 1: Top five most frequent lemmas in Blueprint A

Item	Absolute Frequency	Absolute Frequency (per million)
course	56948	506.8
form	54266	483.0
report	51249	456.1
run	47157	419.7
move	46496	413.8

Table 1 above indicates, for instance, that the fifth lemma in the list with the highest frequency ‘move’ could be found circa 413- 414 times for every million frequency of a corpus. This amounts to having circa 46.500 occurrences in the whole corpus.

The results of the frequency list also showed that there were eleven lemmas with the lowest possible frequency of being included in the list which is by having the frequency of 1. The frequency of 1 means that the lemma occurs once. Five examples of lemmas with the lowest frequencies are shown in Table 2.

Table 2: Top five least frequent lemmas in Blueprint A

Item	Absolute Frequency	Absolute Frequency (per million)
segregational	1	0.00890
nurturing	1	0.00890
hunched	1	0.00890
disillusioned	1	0.00890
cupless	1	0.00890

Table 2 above shows that the top five least frequent lemmas in Blueprint A all had the Absolute Frequency per million of 0.00890.

The analysis of the Blueprint A textbooks wordlist shows that out of the 1090 lemmas analysed by SketchEngine, a total of 85 high frequency lemmas were found, since they had the Absolute Frequency per million higher than the value of 97. The analysis also shows that a total of 437 lemmas were between the Absolute Frequency of 96.9 – 9.7, which is the criteria for

mid frequency lemmas. The majority of the lemmas were low frequency lemmas, with a total number of 568. Diagram 1 below shows the percentage of frequency lemmas of Blueprint A.

Diagram 1: Percentage of frequency in Blueprint A

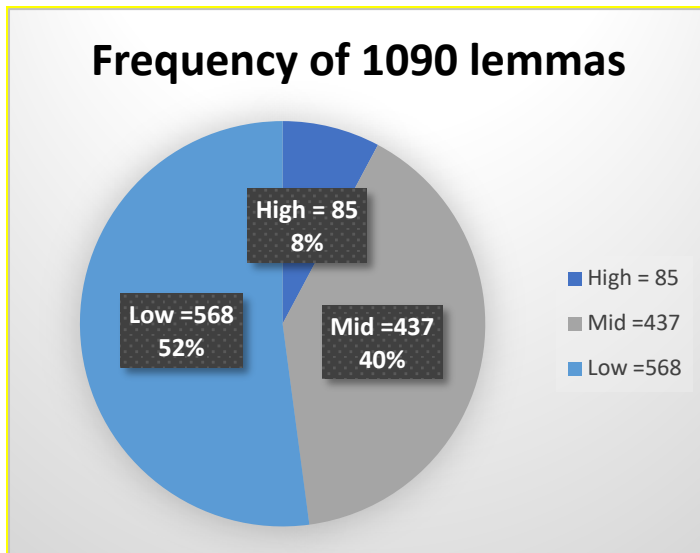


Diagram 1 above show the total number of analysed lemmas which was found in the Blueprint A wordlist, i.e. A total of 1090 lemmas. These are categorized into 85 (8%) high frequency lemmas, 437 (40%) mid frequency lemmas, and 568 (52%) low frequency lemmas which thus make up the majority of lemmas.

The overall results of the analysis indicate that the Blueprint A textbook, which is aimed at English 5 students (first year secondary school), have a majority of low frequency and mid frequency lemmas, while having but a minority of high frequency lemmas. The Absolute Frequency of the highest and lowest lemma ranged between 49067 and 1, with the Absolute Frequency per million ranging between circa 506.8 and 0.00890.

4.2 The frequency of the Blueprint C textbook:

The Blueprint C English textbook, which is two levels more advanced than the Blueprint A textbook (English 5), is aimed at English 7 students. The course is taken in the last year of secondary school as a voluntary course and aims at preparing students for further academic studies at university.

The Blueprint C textbook wordlist contained 1059 lemmas and expressions (multiwords), which were manually entered into Excel. The duplicates and triplicates were

identified and there was a total of 53 duplicates and four triplicates. The lemmas with the same meaning were deleted. Those with the same spelling and sound (homonyms) but with different meanings, were identified and analysed separately. A total of nine lemmas with different meanings remained out of 118, which means that 109 were deleted. A total of ten slang and archaic words were identified and deleted. Examples of slang and archaic words are ‘*crackalackan*’ and ‘*maketh*’. A total of 130 multiword expressions (consisting of two to seven words), for instance ‘*cut to the chase*’ or names such as ‘*Big Brother*’ were identified and deleted due to them requiring a separate and additional tool to analyse the frequency, which would be problematic when matching it with the frequency list of single and hyphenated lemmas. Finally, a total of 855 lemmas remained out of 1059, which were analysed by SketchEngine. The tool produced a list based on frequency on 827 lemmas out of 855. A total of 28 could not be analysed due to different reasons. One of these reasons, for instance, could be that the lemma was not found in the corpus. One example is ‘*gloom-haunted*’.

The analysis showed that the top five lemmas with the highest frequency found in the textbook wordlist had the Absolute frequency ranging between 336445 and 19405 (see Table 4 & Appendix). This means that they could be found between circa 336.500 – 19.400 times in the BNC corpus. The lemma ‘*will*’ has the highest frequency among the lemmas in the Blueprint C wordlist with circa 336.500, and is the 32nd most frequently used lemma in the whole BNC corpus (see Appendix).

The top five lemmas also had the Absolute Frequency per million (out of a total of circa 97 million frequencies) ranging between circa 2994.7 and 172.7 (see Table 3 below). This means that they could be found circa 3000 – 173 times for every million frequency of a corpus.

Table 3: Top five most frequent lemmas in Blueprint C

Item	Absolute Frequency	Absolute Frequency per million
will	336445	2994.7
line	35219	313.4
minister	28237	251.3
range	23589	209.9
benefit	19405	172.7

Table 3 above shows that the lemma ‘will’ had the highest Absolute Frequency per million of circa 2994.7. The fifth lemma in the Blueprint C wordlist with the highest frequency was ‘benefit’ with a circa 19.405 Absolute Frequency, which means that it could be found circa 19.400 times in the whole BNC corpus. The lemma ‘benefit’ also had the Absolute Frequency per million of 172.3, which means that it could be found circa 173 times for every million frequency of a corpus. Even though it was fifth with regards to frequency among the Blueprint C wordlist, the lemma ‘benefit’ is in 594th place for most frequently used lemma in the whole BNC corpus.

The analysis of the Blueprint C textbook wordlist also showed that there were nine lemmas with the lowest possible frequency required to be included into the wordlist, which is the Absolute frequency of 1. The Absolute Frequency of the lemmas per million was 0.00890 for all of the nine lowest lemmas. Out of the all the lemmas with the frequency of 1, there are five examples listed in the Table 4 (see below).

Table 4: Top five least frequent lemmas in Blueprint C

Item	Absolute Frequency	Absolute Frequency per million
re-mould	1	0.00890
plutonian	1	0.00890
leech-like	1	0.00890
hybridity	1	0.00890
footlight	1	0.00890

In accordance with the previously explained criteria and categorization for high-, mid- and low frequency lemmas, the result of the analysis of the wordlist found in the Blueprint C textbook show that there is a majority of low frequency lemmas in the textbook. Only thirteen of the 827 lemmas analysed could be placed into the category of high frequency lemmas, which had the criterion of having an Absolute Frequency per million higher than 97. These lemmas amount to circa 2 % of the 855 lemmas analysed (see Diagram 2 below).

Diagram 2: Frequency of lemmas in Blueprint C

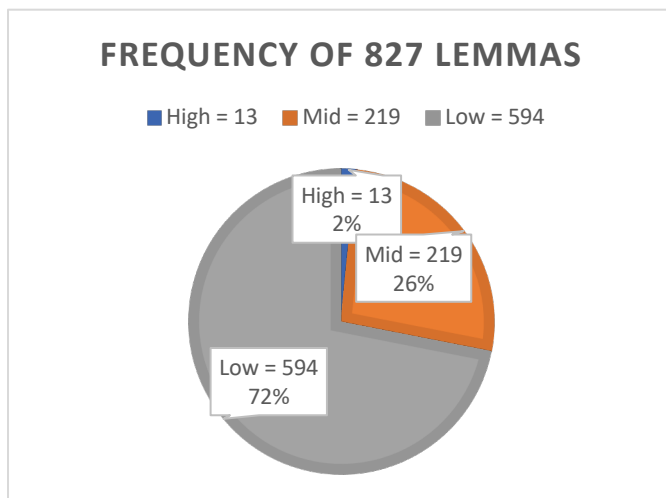


Diagram 2 above shows the total number of analysed lemmas which were found in the Blueprint C wordlist, i.e., a total of 827 lemmas. These lemmas consist of thirteen (2%) high frequency lemmas, 219 (26%) mid frequency lemmas, and 594 (72%) low frequency lemmas. The results indicate that the Blueprint C textbook, which is aimed at third year secondary school students reading the English 7 course, have a large majority of low frequency lemmas which make up 72%. The high and mid frequency lemmas only make up 2% and 26% respectively, which makes a total of 28%. The low frequency lemmas are more than twice the number of both high and mid frequency lemmas.

4.2.1 Differences and similarities between the results

The Blueprint textbooks (A and C) for English 5 and 7 were analysed based on their frequency of lemmas, but other results were also obtained in the process of analysis. When analysing the frequencies in both of the textbooks, one of the first differences between the two textbooks that became clear concerned the number of words in the wordlists. The Blueprint A textbook originally had 1498 words (lemmas) and multiword expressions, while the Blueprint C textbook originally had 1059 words (lemmas) and multiword expressions. The difference between the two wordlists of the textbooks amounts to a total of 439 more words in the Blueprint A textbook. Both textbooks also differed in the number of duplicates and triplicates that were removed. Among the Blueprint A lemmas, a total of 148 (9.8% of 1498) were deleted consisting of duplicates and triplicates, while in the Blueprint C the total number of deletions were 109

lemmas (10.2% of 1059). The multiword expressions identified and removed had a total of 241 (16% of 1498) and 130 (12.2% of 1059) in the two textbooks; i.e., Blueprint A contained 111 more multiword expressions than Blueprint C. Lastly, there was also a difference between the number of slang and archaic words (lemmas) identified and removed in the textbooks. A total of three (0.2% of 1498) slang or archaic words (lemmas) were identified and removed in Blueprint A, while a total of ten (0.9% of 1059) slang words (lemmas) were identified and removed in Blueprint C.

The textbooks had similar results when comparing the percentage of high-, mid- and low frequency lemmas in the wordlist (see Table 5 below). Both wordlists of the two textbooks had a majority of low frequency lemmas, which were 568 (52%) and 594 (72%). The mid frequency words were both the second largest frequency group with 437 (40%) and 219 (26%) (see Table 5). Furthermore, the high frequency lemmas were significantly lower than the other categorizations. In other words, the percentage of the Blueprint A high frequency lemmas (8%) were four times the percentage of Blueprint C high frequency words (2%).

Table 5: Difference in number of lemmas in the two textbooks

	Blueprint A	Blueprint C
High frequency	85 (8%)	13 (2%)
Mid frequency	437 (40%)	219 (26 %)
Low frequency	568 (52%)	594(72%)

The comparison of the textbooks indicates that even though there are differences in the number of lemmas, multiwords and other types, the textbooks contained similarities in the level of frequencies.

4.3 Selected vocabulary exercises in Blueprint A

The analysis of the Blueprint A textbook was conducted by first identifying the vocabulary exercises. Selective exercises were found three times in the exercise section as a direct instruction to look for certain words, but there was an indirect instruction, or support, in the form of bolded glossary words next to the texts in the textbook. Please note that the following detailed data concerning the number of selected vocabulary exercises is summed up in Table 6 on p.21. A total of 185 words (bolded words at the side of the text) were counted, which appeared in nineteen instances both in the context of a text, i.e., short story, but also in the context of explanatory texts next to the instructions. Therefore, the total number of selective

attention exercises that was identified is 22 exercises (23.4%) (three direct and nineteen indirect). Recognition exercises were also found 22 times (23.4%) in the exercise section. These were usually in the form of matching synonyms or meaning with the targeted words. One example of such an exercise found in the textbooks is: “Combine the words in the box to the left with the synonyms from the text in the box to the right” (Blueprint A, page 82). The manipulation exercise type was found seven times (7.4%) in the textbook. Many of these exercises were about creating new words or changing the tense of the word in the context of fill-in-the-blank exercises. Some of the exercises were connected to other types. One example of a manipulation exercise is:

“b) Then look at the other sentences and decide how you need to change the words so that they fit. Think of the meaning and what part of speech (ordklass) it is. Remember to add negative prefixes (*unimportant*, *disadvantage*) where necessary” (Blueprint A p. 179)

The interpretation exercise was found nineteen times (21.2%) in the textbook. Most of the nineteen exercises consisted of filling in the blank exercises, but also finding words in the text and understanding their meaning. One example of this kind of exercise is: “Read the sentences and fill in the missing verbs. Use the past tense.” (Blueprint A p. 81). The most commonly found exercise type was the production exercise type, which was found 24 times (25.5%). The exercises were mostly connected to other exercises in the textbook because after students were instructed to complete a synonym matching exercise, the next exercise was to use the same words when writing a short story.

Table 6: Type and number of vocabulary exercises in Blueprint A

Types of vocabulary exercise	Number of exercise types identified
Selective attention:	22 (23.4%) = 3 direct (3,2%) + 19 indirect (20,2%)
Recognition:	22 (23.4%)
Manipulation:	7 (7.4%)
Interpretation:	19 (20.2%)
Production:	24 (25.5%)
Total:	94 (100%)

As is shown in Table 6, the most commonly found exercise in the Blueprint A textbook was the production exercise type. Based on how one would like to view the direct and indirect selective

attention exercises, it could be the shared second most found exercise in the textbook with the recognition exercises which were found 22 times (23.4%). On the other hand, if one only counts the three direct instructions, it would be the least found exercise type among the five types (3.2%). The third most commonly found exercise would then be the interpretation exercises which appeared nineteen times (20.2%). Fourth on the list would then be the manipulation exercises which were found seven times (7.4%).

4.4 Selected vocabulary exercises in Blueprint C

As was the case for Blueprint A, the following detailed data concerning the number of selected vocabulary exercises is summed up in Table 7 below. The analysis of the Blueprint C textbooks showed that the selective attention exercises were found five times (5.6%) as direct exercises and 34 times (38.6%) as indirect exercises, i.e., the number of times glossary words were next to a text. The total number of selective attention exercises was 39 times (44.3%). Recognition exercises were found eleven times (12.5%) in the textbook, often in the form of instructions to students to find synonyms or matching the meaning of target words, as previously mentioned in regard to Blueprint A. The manipulation exercises were the least found exercise types, and the total number of exercises of this type was found three times (3.4%) among all the different exercises. Interpretation type exercises were found 27 times (30.6%), and therefore the most commonly found type in the exercise pages (selective attention were mostly next to texts, i.e., not in the exercise pages). Most of the interpretation exercises were instructions to find answers to fill-in-the blank by analysing the text, i.e., the context of the target words. The production exercise type was found eight times (9%) in the textbook. As was the case in the Blueprint A textbook, these exercises were often connected to other exercises since the students were instructed to use the target words in their own text after working with them through other exercise types. See Table 7 below for all of the numbers and percentages.

Table 7: Type and number of vocabulary exercises in Blueprint C

Types of vocabulary exercises:	Number of exercise types identified
Selective attention:	39 (circa 44.3%) = 5 direct (5.6) + 34 indirect (38.6)
Recognition:	11 (12.5%)
Manipulation:	3 (3.4%)
Interpretation:	27 (30.6%)
Production:	8 (9%)
Total:	88 (100%)

The result of the analysis of both textbooks shows that there was a difference in the more prevalent exercise types. Besides the selective attention exercises, which could be understood as either the total of both direct and indirect exercises or only direct exercises, the most common type seems to be the interpretation exercise type which was found most times when adding the results of the two textbooks together. The ranking of the frequency of the identified exercise types in Blueprint A and Blueprint C would then be the following:

- (1) Interpretation exercise: 46 times total
- (2) recognition exercise: 33 times total
- (3) production exercises: 32 times total
- (4) manipulation exercises: ten times total
- (5) Or (1) Selective attention exercise: Either eight times total (counting only direct instructions), or 61 times total (counting both direct and indirect instructions).

5. DISCUSSION

This study has examined the word frequency and the selected exercise type of two English Blueprint second version textbooks (A and C) for the Swedish secondary school. Blueprint A was aimed at first year English 5 students, and Blueprint C is aimed at third year English 7 students. The wordlists of both textbooks were analysed using SketchEngine, which generated a frequency list based on lemmas. The lemmas were then analysed and grouped based on the proposed criterion for categorization of high-, mid- and low frequency lemmas. The selected vocabulary exercise types were counted and analysed using Excel.

The results of the analysis of the Blueprint A textbook wordlist showed that out of the 1090 lemmas analysed by SketchEngine from the Blueprint A wordlist, a total of 85 (8%) high frequency lemmas, 437 (40%) mid frequency lemmas, and 568 (52%) low frequency lemmas were found. The results of the analysis of the Blueprint C wordlist showed that out of the 827 lemmas analysed by the SketchEngine, a total of thirteen (2%), high frequency lemmas, 219 (26%) mid frequency lemmas, and 594 (72%) low frequency lemmas were found.

The hypothesis of the study was correct with regards to lower levels (Blueprint A) consisting of more high frequency words than higher levels (Blueprint C) which would consist of more low frequency and specialized words. It should be noted that this was hypothesized

before deciding to use lemmas. However, the low number of high frequency lemmas and high number of low frequency lemmas were not predicted.

The results of the analysis of both textbook wordlists partly supports results of previous studies. In comparison to Norberg & Nordlund (2018) (see p. 6), the results of this study confirm the finding regarding a high number of low frequency words in textbooks. The fact that their study was conducted using Swedish elementary school textbooks suggests that there is a continuation of high number of low frequency words in English textbooks from lower levels to the higher levels of the Swedish school. One of the reasons for this could be due to the expectation of being able to speak English at a high proficiency level in Sweden, which could be related to Bolton & Meierkord's (2013) suggested idea of the prestigiousness of speaking English at a high proficiency level (see p. 5).

Other previously conducted studies such as Sakata (2019) and Sun & Dang (2020) (see p. 6-7.) also seem to be similar to the results of this study. Their studies suggest an insufficiency in textbook vocabulary with regards to high frequency words being found, which they base on the idea of high frequency words being more essential for EFL students. The number of high frequency lemmas found in Blueprint A and C could be seen as being low and therefore insufficient, especially with regards to Blueprint A which is aimed at first year students. It could also be argued that the low number of high frequency lemmas and high number of mid and low frequency lemmas found in Blueprint C was a deliberate choice by publishers, since at English 7 level students are expected to know high frequency lemmas and therefore learning low and mid frequency lemmas should be their goal.

The results of the analysis of the vocabulary exercise types found in the textbooks showed that the following exercises types were the most commonly found when combining the results of both textbooks:

- Interpretation exercise (46 times)
- recognition exercise (33 times),
- production exercises (32 times),
- manipulation exercises (ten times),
- selective attention exercise (eight times as direct instructions) or (61 times as both direct and indirect instructions).

The selective attention exercise could be the most or least commonly found exercise based on choice of calculation. If one counts the most common exercise type in the textbooks separately, production exercises are the most common type in the Blueprint A, while interpretation exercises are the most common type in the Blueprint C.

If one compares the results of the exercise type analysis with previous studies, there are some aspects that are similar between the most effective exercise type and the most commonly found type. Hashemzadeh's (2012) study about the effectiveness of vocabulary exercises (see p. 8), showed that recognition exercises (fill-in-the-blank or matching) were more effective for vocabulary retention than production exercise (paraphrasing and glossing). These results seem to match with the findings of Blueprint A and C, since recognition (matching) and interpretation (fill in the gap) exercises were among the commonly found types (even though production exercise was the most common type in the Blueprint A textbook and the least effective exercise type in the previous study).

The results of this study also partly agreed with Tahir & Mohtar's (2016) results (see p. 9) which indicated that the most selected exercise for students were the matching exercise, which is commonly found in both Blueprint textbooks. It should also be mentioned that Tahir & Mohtar argued that implicit exercises were an ineffective method of teaching, which would suggest that the large number of indirect selective attention exercises found in both textbooks are less effective.

6. CONCLUSION

The aim of the study was to examine the frequency level (high-, mid- and low) of vocabulary found in Blueprint English textbook wordlists for English 5 and 7 in Swedish secondary schools, and to examine which vocabulary exercise types were more commonly found in the textbooks. Previous studies indicated that there was lack of high frequency words in English textbooks, even though they are essential. The previous studies also indicated that fill-in-the-blank and matching exercises are the most effective type to learn vocabulary.

The textbook wordlists were categorized and inserted into SketchEngine, which generated a frequency list based on lemmas. The results show that both textbook wordlists consisted of a high number of low frequency lemmas (52% and 72%) and a low number of high frequency lemmas (8% and 2%). These results correlate with the findings of previous studies, such as having an insufficient number of high frequency words (in this case, lemmas) in the English textbooks. The results also show that the interpretation exercise type (i.e., fill in the gap) was the most commonly found exercise, if one does not count the indirect selective attention exercises. Therefore, the results of the analysed sample of wordlists would suggest that the Blueprint (2.0) series do not follow the recommended use of high frequency lemmas,

but consists of a commonly found large number of vocabulary exercises which are suggested to be effective by previous research.

One example of a weaknesses of this study is the usage of the spoken part (10%) of the BNC corpus. Further studies should be conducted without using the spoken part. Another example of a weakness is not being able to incorporate the relatively large number of multiword expressions into the total data. Further research should therefore examine the results of the total data including multiword expressions. Finally, it would also be interesting to find out whether the large number of indirect selective attention exercises found in both textbooks are indeed less effective from a learning vocabulary perspective.

7. REFERENCES

- Bolton, K., & Meierkord, C. (2013). English in contemporary Sweden: Perceptions, policies, and narrated practices. *Journal of Sociolinguistics*, 17(1), 93-117.
- Education First,. (2020). EF EPI: EF English Proficiency Index [PDF file]. Retrieved from:https://www.ef.se/assetscdn/WIBIwq6RdJvcD9bc8RMd/legacy/_/~/media/centralefcom/epi/downloads/full-reports/v10/ef-epi-2020-english.pdf
- Hashemzadeh, M. (2012). The Effect of Exercise Types on EFL Learners' Vocabulary Retention. *Theory and Practice in Language Studies*, 2(8), 1716-1727.
- Hult, F. (2003). English on the streets of Sweden: An ecolinguistic view of two cities and a language policy. *Working Papers in Educational Linguistics*, 19(1), 43-63.
- Lundfall, C., Nyström, R. & Clayton, J. (2007). *Blueprint A*. (2. uppl... ed). Stockholm: Liber.
- McKay, C., Brodin, M., Clayton, J. & Webster, C. (2011). *Blueprint. [steg 7] C* (2. ed.). Stockholm: Liber.
- Nation, I.S.P., (2008). *Teaching vocabulary : Strategies and techniques*. Boston, MA: Heinle.
- Nation, I.S.P., (2013). *Learning vocabulary in another language* (2.nd ed., The Cambridge applied linguistics series). Cambridge: Cambridge University Press.
- Norberg, Cathrine, & Nordlund, Marie. (2018). A Corpus-based Study of Lexis in L2 English Textbooks. *Journal Of Language Teaching And Research*, 9(3), 463-473.
- Paquot., M. (2018) Corpus research methods for language teaching and learning. In Phakiti, A., DeCosta, Peter editor, Plonsky, Luke editor, & Starfield, Sue, editor.

(2018). *The Palgrave handbook of applied linguistics research methodology*. London: Palgrave Macmillan.

Paribakht, T. S & Wesche, M .,(1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In Coady, J., & Huckin, T. N. (1997). *Second language vocabulary acquisition : A rationale for pedagogy* (The Cambridge applied linguistics series). Cambridge: Cambridge Univ. Press.

Read, J. (2000). *Assessing vocabulary* (The Cambridge language assessment series). Cambridge: Cambridge Univ. Press.

Sakata, N. (2019). Profiling vocabulary for proficiency development: Effects of input and general frequencies on L2 learning. *System (Linköping)*, 87, System (Linköping), December 2019, Vol.87

Skolverket., (n.d), *English* [PDF file]. Retrieved from <https://www.skolverket.se/download/18.4fc05a3f164131a74181056/1535372297288/English-swedish-school.pdf>

Sun, Y., & Dang, T. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System (Linköping)*, 93.

Tahir, M., & Mohtar, T. (2016) The Effectiveness of using Vocabulary Exercises to Teach Vocabulary to ESL/EFL Learners. *Pertanika J. Soc. Sci. & Hum.* 24 (4): 1651 - 1669.

8. APPENDIX

Top 50 lemmas from Blueprint A

Ranking	Item (lemma)	Absolute Frequency	Absolute Frequency (per million)
1	course	56948	506.89959
2	form	54266	483.02685
3	report	51249	456.17224
4	run	47157	419.74896
5	move	46496	413.86534
6	lead	44394	395.15523
7	support	40811	363.26261
8	development	36961	328.99339
9	market	36476	324.67636
10	offer	35622	317.07482
11	return	33108	294.69747
12	stand	32603	290.20242
13	subject	32073	285.48484
14	involve	31726	282.39616
15	remain	31672	281.91550
16	authority	30800	274.15374
17	cause	30123	268.12770
18	require	29289	260.70419
19	experience	28912	257.34847
20	major	28831	256.62748
21	increase	28435	253.10265
22	research	27567	245.37650
23	develop	27250	242.55485
24	value	26863	239.11013
25	community	26564	236.44870
26	walk	25908	230.60958
27	patient	25401	226.09673
28	sense	24341	216.66157
29	agree	24082	214.35618
30	condition	24069	214.24047
31	range	23589	209.96794
32	note	23514	209.30036
33	charge	23478	208.97992
34	project	21742	193.52762
35	evidence	21419	190.65256

36	drive	21391	190.40333
37	event	20535	182.78400
38	raise	20314	180.81685
39	apply	19715	175.48510
40	complete	19622	174.65730
41	award	19622	174.65730
42	game	19469	173.29543
43	concern	18792	167.26939
44	similar	18313	163.00576
45	contain	17966	159.91708
46	pound	17350	154.43401
47	join	17223	153.30357
48	former	17042	151.69247
49	supply	16899	150.41961
50	director	16120	143.48566

Top 50 lemmas in Blueprint C

Ranking	Item (lemma)	Absolute frequency	Absolute frequency (per million)
1	will	336445	2994.72907
2	line	35219	313.48768
3	minister	28237	251.34023
4	range	23589	209.96794
5	benefit	19405	172.72576
6	establish	17840	158.79554
7	press	17791	158.35939
8	scheme	17063	151.87939
9	various	15295	136.14226
10	grant	15069	134.13061
11	determine	13455	119.76424
12	lack	13120	116.78237
13	display	11079	98.61524
14	average	10817	96.28315
15	reveal	10219	90.96029
16	sentence	9855	87.72030
17	favour	9755	86.83019
18	relative	8406	74.82261
19	survive	8211	73.08690
20	emerge	8169	72.71305
21	advance	8141	72.46382
22	decline	8082	71.93865

23	master	7929	70.57679
24	spring	7767	69.13481
25	escape	7528	67.00745
26	settle	7433	66.16184
27	blow	7210	64.17690
28	vehicle	7144	63.58943
29	respond	7129	63.45591
30	acquire	6995	62.26316
31	threat	6910	61.50657
32	head	6791	60.44734
33	victim	6651	59.20119
34	yard	6627	58.98756
35	manner	6577	58.54251
36	promote	6544	58.24877
37	substantial	6175	54.96426
38	decade	6134	54.59932
39	crown	6039	53.75372
40	slip	6028	53.65580
41	defeat	6014	53.53119
42	challenge	5703	50.76295
43	assumption	5497	48.92932
44	imply	5391	47.98581
45	urban	5377	47.86119
46	remark	5357	47.68317
47	contemporary	5337	47.50515
48	entitle	5288	47.06899
49	cast	5094	45.34218
50	expand	4951	44.06932

Top 50 lemmas in the British National Corpus (BNC)

Ranking	Item (lemma)	Absolute Frequency
1	the	6054950
2	be	4130873
3	of	3049448
4	and	2624170
5	to	2599454
6	a	2176259
7	in	1945698
8	have	1303104
9	that	1120808

10	it	1054366
11	for	880815
12	i	872841
13	not	769041
14	on	731272
15	you	668429
16	with	659976
17	as	655198
18	he	641310
19	do	535498
20	at	524083
21	by	513429
22	this	454536
23	but	446798
24	from	425966
25	they	420511
26	his	410291
27	or	370096
28	which	366198
29	she	353270
30	we	351145
31	an	338820
32	will	336445
33	there	319774
34	say	319322
35	one	306543
36	her	304308
37	would	283119
38	all	277710
39	can	262031
40	their	254608
41	if	253827
42	what	240789
43	so	239685
44	go	229111
45	no	226430
46	make	210667
47	when	209964
48	more	209754
49	get	208578
50	up	208232