The comprehensibility of reading texts in National Tests in English

A quantitative study of Swedish EFL learners' vocabulary size and lexical coverage

Yaran Kakaee

Ämneslärarprogrammet

Abstract

Previous research has indicated a positive correlation between vocabulary size and reading comprehension (Grabe, 2009; Hu & Nation, 2000; Laufer, 1989; Laufer, 1992; Stæhr, 2008). Over the past decade, statistics have shown that Swedish upper secondary learners of English as a foreign language (henceforth EFL) are less proficient in reading comprehension as opposed to other language skills in the Nationals tests (NTs). Therefore, this study aims to examine the relationship between Swedish EFL learners' receptive vocabulary size and lexical text coverage of the reading texts in the NTs. A total of 89 students from two upper-secondary courses (English 5 and 6) completed the revised Vocabulary Levels Test (Schmitt, Schmitt & Clapham, 2001). Two NTs from both courses were analyzed in a Lexical Frequency Profile program to determine the vocabulary size needed to reach 98% lexical coverage and consequently adequate reading comprehension (Hu & Nation, 2000; Nation, 2006). The results revealed that a majority of the students do not reach the 8,000–9,000 vocabulary size that is needed in order to achieve adequate reading comprehension of the whole texts. Furthermore, about 14–23% of the students display a vocabulary size below the 95% threshold level suggested by other researchers (Laufer & Ravenhorst-Kalovski, 2010), i.e. more than the 12% that statistically fail the reading comprehension part of the NTs. Suggestions for future research indicate the need to incorporate the learners' reading comprehension scores to further validate the correlation between vocabulary size and the comprehensibility of reading texts in the NTs.

Table of Contents

1 Introduction

When learning a language, there is one common foundation of learning that underpins the ability to master that language in a variety of communicative situations, namely vocabulary knowledge. This central aspect of language learning is necessary for both productive and receptive skills within that language. Vocabulary knowledge has long been proven to be a good predictor of overall language proficiency in second and foreign language learning. A large body of research has shown that vocabulary size, in particular, plays a significant role in learners' communicative competence (Laufer, 1992; Laufer & Nation, 1995; Stæhr, 2008). As to the different skills involved in overall language proficiency (reading, writing, speaking, and listening), reading comprehension has been found to be "the skill most dependent on vocabulary size" (Stæhr, 2008, p. 148). However, the acquisition of vocabulary is also considered the most arduous tasks in second and foreign language learning and with which learners struggle the most (Schmitt, 2008). Consequently, second and foreign language learners are not as successful in reading comprehension as opposed to the other language skills (Stæhr, 2008). This might be due to an insufficient amount of vocabulary needed for this kind of task. Swedish foreign language learners of English are no exception to this phenomenon. This opens up for a closer examination of this issue, and one way of doing so is by investigating the Swedish National tests in English and the vocabulary levels of the students. Prior to such an examination, also the main aim of the present study, a brief overview of the National tests is needed, which will be presented in the next section.

1.1 Background

1.1.1 National Tests and their construction process

In Swedish upper secondary school, different varieties of standardized tests have been used since the 1940s to ensure equal and fair assessment for grading on a national level (Börjesson & Schönberg, 2012; Lundahl, 2012). These tests, today known as the National tests (hereafter NTs), are developed on behalf of the Swedish National Agency for Education (Sw. Skolverket) and are carried out in a specific set of school subjects, amongst them the English subject. The NTs have several functions. Apart from guiding teachers in their grading of students, they primarily function as a concretization of the syllabus of the particular subject. The tests are mainly designed based on the curriculum and the purpose and core content of the syllabus (Börjesson & Schönberg, 2012). In English, they provide information concerning the

learners' proficiency levels of English as a foreign language (henceforth EFL) all over the country and are currently given to students taking two courses in upper-secondary school, English 5 and English 6.

The test materials of the NTs are created in cooperation with different reference groups (e.g. subject teachers, teachers of special education needs, native speakers, and researchers from different disciplines) and are based on a number of research-based principles (Axelsson & Lindqvist, 2017). The test materials are calibrated in test runs with groups of randomly selected students throughout the country. In connection with these test runs, both students and teachers are to give their opinions on the functionality of the test material, and this is then weighed into the final composition of the actual sample test. The development of one national test takes about two years and is based on a number of quality parameters to ensure validity and reliability (see Axelsson & Lindqvist, 2017).

For the NTs in English, four general aspects are tested: speaking, listening, reading comprehension, and writing comprehensibility. The reading comprehension part usually consists of four tasks of different nature: a matching-task, a cloze task, a longer reading text, and a section with short texts (Axelsson & Lindqvist, 2017). In the matching-task, the student is asked to match a word with its correct definition from a word list. The cloze task is an exercise consisting of a text where words have been removed and the student is asked to replace the missing word by choosing between four different alternatives. The longer and shorter texts are mainly accompanied by comprehension questions on their content and overall message; the former has both open and multiple-choice questions whereas the latter has only multiple-choice questions (for more details, see section 3.2 and Appendix 1). According to Nilsson and Schönberg (2018), test developers of the NTs, the word items in the matching-task are generally of high-frequency nature, whereas the alternatives in the cloze task are usually low-frequency words (S. Nilsson and H. Schönberg, personal communication, 17 April, 2018). What is common for all tasks is the importance of involving as much authentic material as possible in the test material. If needed, the authentic material can be shortened down and not presented as a whole; however, the test constructors are utterly restrictive in terms of altering the vocabulary of test materials, i.e. in reading texts (S. Nilsson and H. Schönberg, personal communication, 17 April, 2018).

1.1.2 Results from previous NTs

For the past eight years, results of the NTs have been collected and compiled by Skolverket (2018). These statistics show the distribution of Swedish upper secondary school students' results over the different language skills tested at two different proficiency levels, the English 5 and English 6 course. While Swedish EFL learners are proficient in the speaking part, their reading and writing results are not as comforting which the following figures will display. Note that Figures 1-4 were made based on statistical information from Skolverket (2018).

As shown in Figures 1 and 2, only 1–2% fail the speaking part of the NTs. However, the percentage of failed performances is much higher for the writing and reading tasks. In reading, the failed performances stretch from 5 to 12% in English 5 and 6.
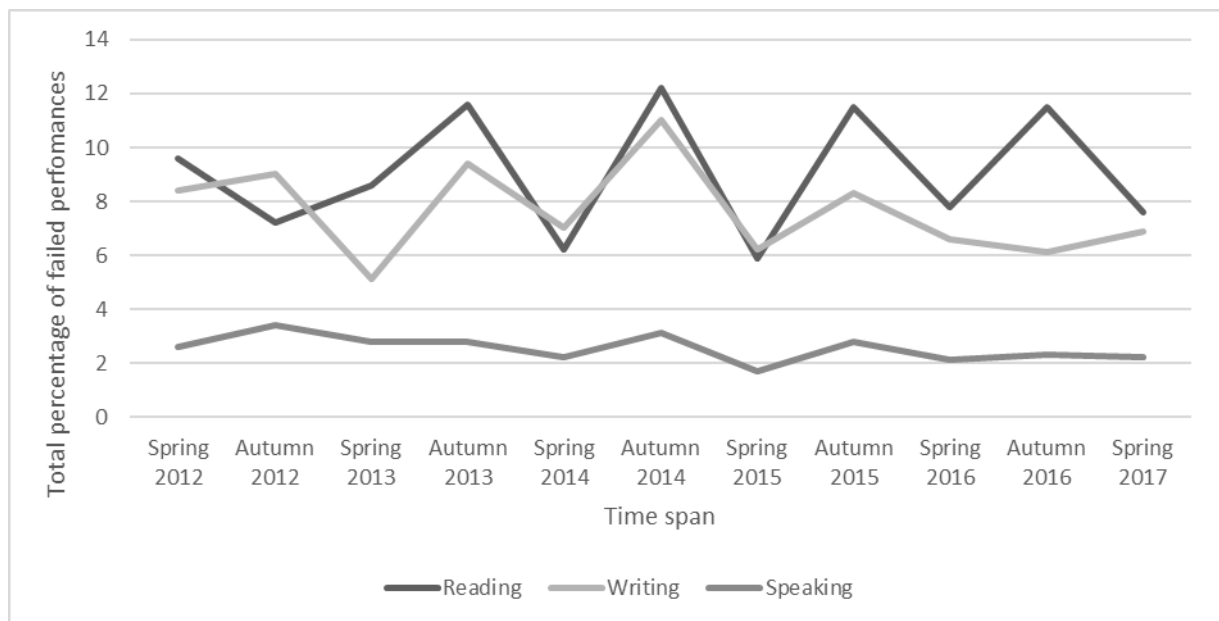
Figure 1: English 5 students' failed performances on NTs, spring term 2012 - spring term 2017
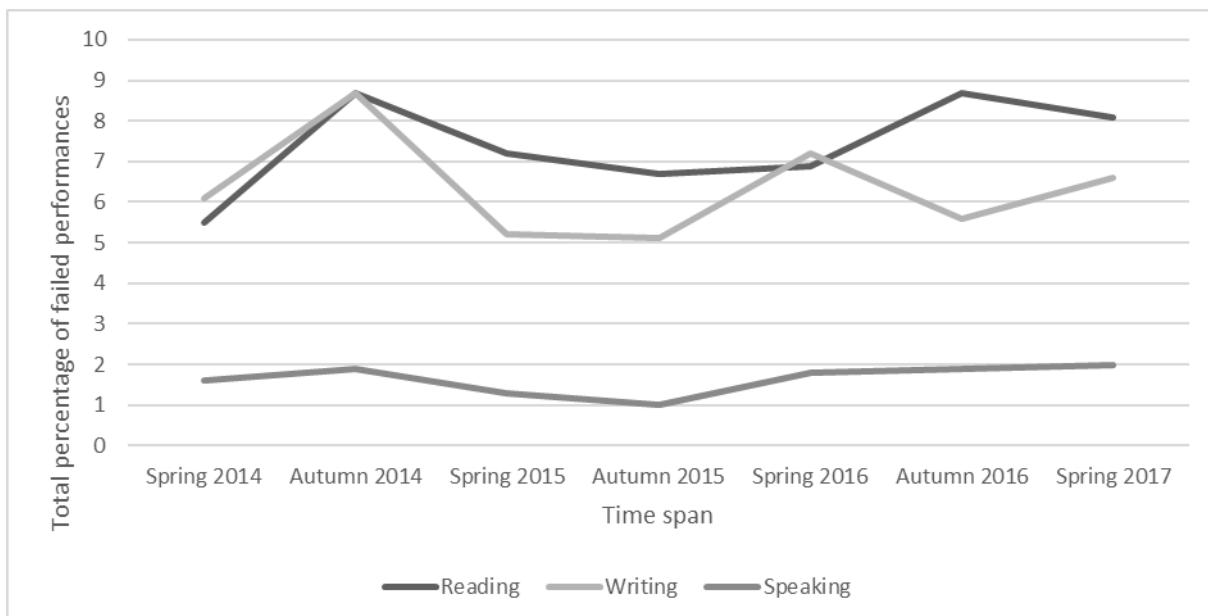
Figure 2: English 6 students' failed performances on NTs, spring term 2014 - spring term 2017

Figures 3 and 4 show that the poor results concerning reading and writing skills are a more significant issue at the vocational programmes than at the theoretical ones. In the vocational programmes, the percentage of failed performances in reading is as high as 12–20%
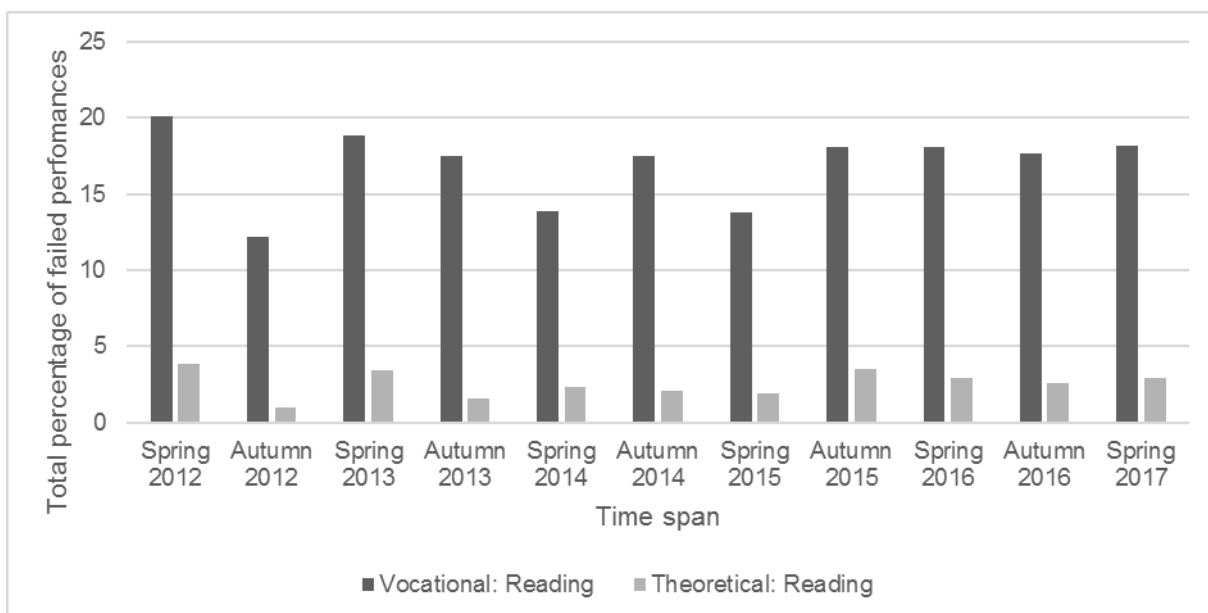
Figure 3: Comparison of English 5 students' failed performances on the reading part of NTs in vocational and theoretical programmes, spring term 2012 – spring term 2017
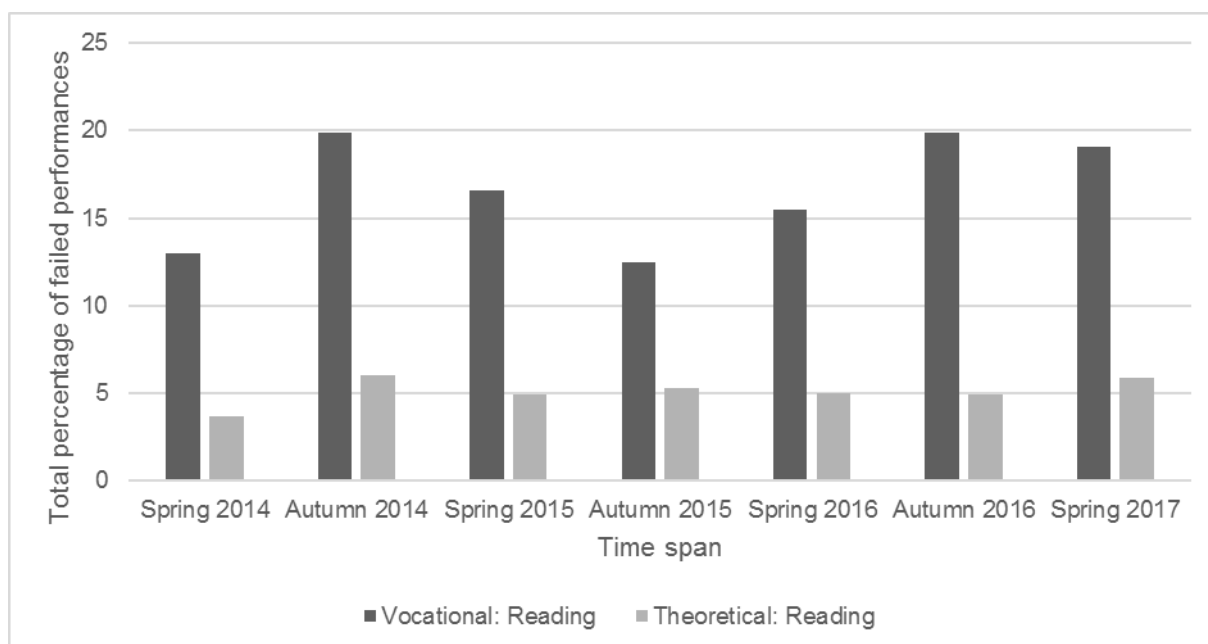
Figure 4: Comparison of English 6 students' failed performances on the reading part of NTs in vocational and theoretical programmes, spring term 2014 – spring term 2017

Figures 1–4 show that a larger number of Swedish EFL students at English 5 and English 6, regardless of their programme, failed the reading part of the test as opposed to the other parts over the past few years. On a national level, both politically and pedagogically, these results have become a worrying and central concern and it has been widely debated how to best approach the problem (Börjesson & Schönberg, 2012; Öhman, 2013; Youcefi, 2012).

Second and foreign language research shows that language comprehensibility in terms of reading is highly correlated to learners' vocabulary knowledge and thus their lexical coverage of the written language (Grabe, 2009; Hu & Nation, 2000; Laufer, 1989; Laufer, 1992; Stæhr, 2008). Concerning reading proficiency, it is suggested that the more developed your vocabulary is and the larger the text coverage is, the less the cognitive load of comprehending the words will be and the better the learner can focus on the meaning and content of the text. Only when the learner has a vocabulary knowledge enough to cover most of the text, approximately 95–98%, can their reading comprehension be considered as adequate (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Nation 2006). Moreover, what kind of vocabulary is presented to the learner in terms of its frequency level distribution also plays a central role in the learners' overall reading comprehension (Agernäs, 2015; Nation, 2006). In order to determine the distribution of a text's vocabulary over different frequency levels or bands, a *Lexical Frequency Profile* (hereafter LFP) can be created (Laufer & Nation, 1995).

In relation to the research findings presented above, the following questions rise to the surface: Could it be that the poor reading results are a consequence of reading texts with too difficult vocabulary in the NTs? Is it possible that the EFL learners' vocabulary size is not sufficient in comparison to what is needed to comprehend the reading texts fully? Since there are no guidelines or numbers in the syllabus of English regarding what vocabulary the learners are expected to have developed, the NTs can be such a guide as they are considered a concretization of the syllabus. Questions like these suggest that a closer examination of the test-takers' vocabulary size in relation to the lexical coverage of the NTs' reading texts might partly explain their low reading proficiency.

1.2 Aim and research questions

The aim of this paper is to explore the relationship between the LFPs of the NTs' reading texts at different levels of proficiency and the vocabulary size of the test-takers. The research questions are, therefore, the following:

1. What is the average receptive vocabulary size of Swedish EFL students in English 5 and English 6?
2. What do the LFPs show concerning the vocabulary of the NTs' reading texts?
3. To what extent are the NTs' reading texts comprehensive in relation to the learners' receptive vocabulary size?

In light of these research questions, this paper will analyze a number of texts used in the NTs at two levels of proficiency, English 5 and English 6, in order to see how these texts correlate to the vocabulary size of the intended test-takers, i.e. the learners in the two English courses. Prior to this paper, there is no research investigating the relationship between reading comprehensibility of the NTs and learners' vocabulary size with the use of LFP as a tool for analysis. Although the NTs are designed on the basis of relevant research and validated test runs of certain tasks and test items, no calibrated measures in relation to the test-takers' vocabulary size have been done earlier (S. Nilsson and H. Schönberg, personal communication, April 16, 2018). This paper might hopefully inspire to make such validated measures in the future when designing the reading texts of the NTs.

1.3   Outline

This paper will first provide an overview of relevant literature and theoretical concepts concerning this field of research with special reference to vocabulary size measurements, lexical coverage, and lexical frequency profiling. Next, the method and material developed for this study is presented before moving on to presenting the results of the EFL learners' vocabulary size as well as the LFPs of the NTs. Key findings are then discussed and some methodological limitations are highlighted. Accompanying pedagogical implications are presented followed by a summative conclusion outlining directions and suggestions for future research.

2 Theory and concepts

In this chapter, relevant literature and some central aspects related to the aim will be explained and discussed: (1) construction of meaning during reading comprehension, (2) lexical threshold levels in relation to reading comprehensibility and lexical coverage, (3) vocabulary frequency levels, and (4) the *Lexical frequency profile.* Since the interest of this paper lies within vocabulary size in relation to reading comprehension, the previous work treated below will mainly focus on learners' receptive (passive) vocabulary, which is generally larger than the productive (active) vocabulary (Laufer, Elder, Hill & Congdon, 2004; Schmitt, 2008). The receptive knowledge requires the ability to distinguish a word, either spoken or written, and to understand it in a variety of contexts (Nation, 2013). Knowing a word in this sense entails the comprehension of its word family, i.e. the head word and all of its inflections and reduced forms as well as closely related forms, for example, with affixes (Laufer & Nation, 1993; Nation, 2013). This is the definition that this paper will adhere to.

2.1 Construction of meaning during reading comprehension

Although this paper focuses on the aspect of lexical content in reading texts, other aspects affect the degree of difficulty when comprehending written language. In their article, Kendeou, McMaster, and Christ (2016) review the theoretical and empirical literature on the construction of meaning during reading comprehension. Apart from vocabulary knowledge, the studies reviewed have shown that factors such as word decoding, reading fluency, prior knowledge of the subject and task at hand, comprehension monitoring, and working memory also affect the level of learners' reading comprehension (Kendeou et al., 2016). While all of these aspects facilitate the construction of meaning during reading comprehension, they, nevertheless, all depend on the fact that the learner understands the vocabulary presented to them. If a learner interacts with too many difficult and unknown words, this will pose a burden on the working memory, thus making it more difficult to comprehend the overall meaning of the text.

In alignment with the *Cognitive Load Theory*, which refers to the effort used in the working memory in different learning and comprehension experiences, the cognitive load is likely to increase when there is an interaction with a large number of unfamiliar elements, e.g. words (Sweller, Ayres & Kalyuga, 2011). This kind of cognitive load that deals with the intrinsic nature of the information is called *intrinsic cognitive load*. If this intrinsic cognitive

load exceeds the available working memory resources, e.g. the learner's vocabulary size, the cognitive system will not be able to process the necessary information required for full comprehension (Sweller et al., 2011). In other words, the learner will not manage to construct meaning during reading comprehension if all the working memory resources are needed to deal with the imposed intrinsic cognitive load. Furthermore, the level of intrinsic cognitive load depends on element interactivity, i.e. "elements that must be processed simultaneously in working memory because they are logically related" (Sweller et al., 2011, p. 58). Reading comprehension is one example of such a process where several elements have to be processed and understood simultaneously to derive meaning (Kendeou et al., 2016). The process is complex in itself, and the level of lexical difficulty of a text adds to that complexity, thus imposing a higher burden on the learner's working memory. Researchers have tried to set a threshold level for how much vocabulary is needed to lower the level of difficulty when reading texts. The next part of this section will discuss this aspect further as well as highlighting different estimations and suggestions concerning the so-called *Lexical threshold level.*

2.2 Lexical threshold, lexical coverage, and comprehensibility

The *Lexical threshold level* is a term coined by Laufer and Ravenhorst-Kalovski (2010) and is used to determine the expected level of comprehension of a specific text, i.e. the "minimal vocabulary that is necessary for 'adequate' reading comprehension" (2010, p. 15). *Adequate comprehension* is a vague term since it highly depends on context, purpose, and the proficiency level of the learner. Researchers within this field have interpreted this term in several different ways. For the spoken register of a language, the threshold level for comprehensibility is much lower compared to written language. In a corpus study carried out by Adolphs and Schmitt (2003), it was found that approximately 3,000 word families are needed to adequately understand conversational English, based on the CANCODE (Cambridge and Nottingham Corpus of Discourse in English). Nation (2006) supports these results and confirms in his study of frequency-based lemma lists from the BNC (British National Corpus) that a vocabulary size of 3,000 word families provides a lexical coverage of over 95% in unscripted spoken English. However, reading comprehension requires a significantly larger vocabulary size. One explanation for this difference is that the input of spoken language is usually assisted by other comprehension clues such as body language, intonation, and gestures which is not the case for written input, therefore the different

threshold levels (Laufer & Ravenhorst-Kalovski, 2010). The main reason, however, is that there are fewer difficult words in spoken language as opposed to written (Nation, 2006)

For written language, researchers have different suggestions regarding how many words are needed for adequate reading comprehension; these differences have to do with, as mentioned above, the term 'adequate comprehension' itself. According to Nation (2006), for adequate comprehension, also called "full" or "unassisted comprehension," a lexical coverage of 98% is desirable. The threshold level is then set at around 8,000–9,000 word families. The importance of developing a vocabulary size of this kind is also supported by Schmitt (2008) and Stæhr (2008) if the learner is supposed to read a large variety of texts without any interference due to unknown vocabulary. Laufer and Ravenhorst-Kalovski (2010), however, provide two other lexical threshold levels: one optimal and one minimal. Their optimal threshold level, which provides a lexical coverage of 98%, requires 8,000 word families, whereas the minimal level, with a 95% coverage, is set at around 4,000–5,000 word families. The difference between the two studies above lies within the fact that Laufer and Ravenhorst-Kalovski (2010) argue that the minimal threshold level with a 95% coverage of a text will lead to adequate comprehension whereas Nation (2006) states that this level is not sufficient in terms of adequate comprehension. In other words, there is no agreement on the percentage that is needed for what is concerned as adequate comprehension. Since the NTs' reading tasks require full/unassisted comprehension, i.e. that the lexical information of the text does not constitute such a cognitive load upon the learner that it takes away focus from the message of the text, the present paper will adhere to the 98% limit for what is considered as adequate comprehension in terms of reading proficiency.

2.3 Vocabulary frequency bands and the Vocabulary Levels Test

When determining the lexical threshold levels, researchers often refer to "frequency levels, which are based on corpus studies of how frequently words occur in the English language" (Agernäs, 2015, p. 6). The words are distributed on 1,000-band levels of word families where the first 1,000 word families are the most frequent ones. High-frequency vocabulary, which is found in the first two or three 1,000 bands, provides a lexical coverage of around 80 to 85% of written and spoken texts. This kind of vocabulary is usually easier for EFL learners to understand since they encounter it to a greater extent than any other type of vocabulary (Nation, 2006). The lexical coverage decreases for every progressing frequency band, and the ninth to tenth 1,000 frequency band is usually considered low-frequency vocabulary (Nation,

2006; Schmitt, 2008; Schmitt & Schmitt, 2012). Vocabulary between the 3,000 and 9,000 bands is labeled mid-frequency vocabulary. According to Nation's calculations (2006) (see section 2.2) one would need a receptive vocabulary size of 8,000–9,000 word families to attain 98% lexical coverage for any written or spoken text. This indicates a great deal of focus on mid-frequency vocabulary in vocabulary teaching and learning (Schmitt & Schmitt, 2012). In fact, research has revealed that the best improvement in reading task scores occurs when the vocabulary size increases with words from the 5,000–7,000 levels (Laufer & Ravenhorst-Kalovski, 2010). Similarly, an increase in these levels of word frequency has an equally positive impact on learners' reading fluency in terms of speed, which section 2.1 highlighted as one crucial factor determining successful reading comprehension (Laufer & Nation, 2001). Although reading fluency is not determinant for comprehending the lexical information in a text, it can pose a problem when learners are to demonstrate their knowledge in examination situations, like with the NTs, where time is limited.

Apart from the vocabulary bands mentioned above, there is an additional vocabulary group that also has an essential effect on reading comprehension. This group is known as the academic vocabulary level presented in Coxhead's Academic Word List (AWL) (2000). The corpora upon which the list is based contain texts from four categories of academics: Arts, Commerce, Law, and Science (Coxhead, 2000). The final AWL comprises 570 word families and covers around 10% of the academic corpus used. Interestingly, Coxhead (2000) shows that 64.3% of the word families in the AWL is from the high-frequency vocabulary bands, which implies the usefulness of learning high-frequency words since it provides a large lexical coverage considering the small number of word families they contain. Nevertheless, the rest of the AWL words are distributed on the other levels beyond the 3,000 most frequent words (mid-frequency level). Thus, it is just as important to learn any other types of vocabulary as the high-frequency ones since knowledge across all of the frequency bands yields a broader lexical coverage and thus better comprehension.

A learner's vocabulary size is evidently what determines the lexical coverage and, therefore, to a large extent, his or her comprehension of a text. When testing a learner's vocabulary size, there are a few different tests available. The most common and widely used is the *Vocabulary Levels Test* (VLT) originally designed by Nation in 1983 and later revised by Schmitt, Schmitt, and Clapham (2001). The VLT provides a vocabulary size profile distributed across five frequency bands: 2,000, 3,000, 5,000, 10,000 word families, and a list of academic words. The receptive vocabulary size is tested through a representative selection of 30 words from each of the five levels. The test-takers are asked to match this total of 150

words to their correct descriptions. When having calculated a learner's receptive vocabulary size in this manner, the score can then be compared to the lexical information of any written text to determine whether or not it is comprehensible for the intended learner. One way of analyzing lexical information of texts is by creating a lexical frequency profile for that particular text.

2.4 Lexical frequency profile

When determining the distribution of words in a certain text over the frequency bands, it is helpful to create a so-called *Lexical frequency profile* (LFP) of that text. The lexical frequency profile is a quantitative index that is calculated by a computer program on Tom Cobb's (2018) website *Lextutor* (Laufer & Nation, 1995). Originally, the LFP was created for the purpose of revealing the percentage of high-, mid-, and low-frequency words used in a learner's writing, or, as Laufer and Nation put it: "the relative proportion of words from different frequency levels" (1995, p. 311). However, any piece of written text can be analyzed and thus receive its unique LFP, e.g. the NTs' reading material. One of many strengths of this tool is that it calculates word tokens, word types, and word families in any written composition, which provides the user of the program a clear and rich overview of the text's lexical information. *Tokens* are the total number of words that occur in the text whereas *types* signify the number of different words in the composition (Laufer & Nation, 1995). The authors point out that, although the program calculates the LFP on the proportion of tokens, types, and word families, it is preferred to primarily consider the calculation done on word families since this way of treating words is what best correlates with how learners view words. This proposition stands in alignment with how the VLT tests words, which is why this study will adhere to calculating LFPs on the basis of word families.

By using the LFP, one may obtain an instant overview of the lexical variation of any text. The tool renders an easy and transparent way of making comparisons between learners' vocabulary size and the lexical information of different texts in order to draw conclusions regarding reading comprehensibility.

3 Method and material

With the aim of investigating the relationship between the comprehensibility of NTs' reading texts and EFL learners' receptive vocabulary size, the present study was carried out in a Swedish upper secondary school in western Sweden. In alignment with this aim, three main variables were examined to answer the research questions accordingly: reading comprehensibility of the NTs, the learner's vocabulary size, and, thus, the lexical coverage. All of these aspects were tested and examined quantitatively by the use of a lexical frequency profiler program (Cobb, 2018), and the Vocabulary Levels Test (Schmitt et al., 2001). It is self-explanatory that the quantitative method is the most suitable for this study when dealing with a corpus and vocabulary size study; however, it might be argued that the relationship between comprehensibility and a learner's vocabulary knowledge is not something that can be measured in numbers solely. Although there is some truth to such suggestions, the quantitative approach provides a broader understanding of and insight into the phenomenon considering the limited time frame of the present study.

3.1 Participants

A total number of 89 Swedish EFL learners ranging between the ages of 16 and 18 took part in this study. Around half of them (47 students) took the English 5 course and the other half (42 students) the English 6 course. One of the English 5 classes (with a total of 22 students) took a vocational programme as opposed to the other classes that took a theoretical programme oriented towards social sciences. Prior to their upper secondary education, all but one newly arrived student had completed nine years of English studies at elementary school. All of the participants went to a school where I have had one of my trainee periods; therefore, the choice of participants is a sample of convenience.

3.2 Materials and measures

First and foremost, data was collected through creating a specialized corpus consisting of the NTs' reading material that is accessible for English 5 and 6. The corpus was intended to be large enough to ensure reliable and representative measurements of the vocabulary frequency levels. However, the recent NTs are classified material and the available example tasks are not representative for the NTs as a whole, nor calibrated in terms of their range, level of difficulty or authenticity in the way that the actual sample versions are (Göteborgs universitet,

2018). Thus, the corpus ended up consisting of two NTs for English 5 and 6 respectively: one version from the autumn term 2015 and a replacement test that was released for the public in 2016. Each test consists of four different reading tasks: a longer reading text, a matching-item text, a cloze-item exercise, and a section including mini-texts of approximately one paragraph each (see Appendix 1 for examples).

With the help of Cobb's computer program (2018), each test received two unique LFPs, one on the language in the text and one on the language in the questions/tasks related to that text. Then, a general LFP was assigned to the test as a whole for English 5 and 6, respectively, thus providing two mean LFPs. These LFPs answered the question regarding what is needed in terms of vocabulary for the texts to be comprehensible. A total of 11,297 tokens were included in the corpus.

Furthermore, the participants of the study took a receptive vocabulary size test. For this part, the revised version of Nation's (1983) VLT was used (Schmitt et al., 2001). As briefly mentioned in section 2.3, the VLT contains a total amount of 150 test items evenly distributed at four frequency levels (2,000, 3,000, 5,000, and 10,000 word families) and in a list of academic words which includes words beyond the 3,000-word level (Schmitt & Schmitt, 2012). The test format is of multiple-choice character where the test items are presented in clusters of six accompanied with three different definitions, thus, leaving out three words in each cluster as shown in Figure 5:

1 copy	
2 event	_____ end or highest point
3 motor	_____ this moves a car
4 pity	_____ thing made to be like another
5 profit	
6 tip	

Figure 5: Vocabulary Levels Test example from the 2,000-word frequency band (Schmitt et al., 2001)

Since the VLT tests a learner's vocabulary knowledge over different frequency levels, it provides a profile of the learner's vocabulary rather than a specific estimate of their receptive vocabulary size. Nevertheless, a rough vocabulary size can be measured when scoring the learner's correct answers on each level. Undeniably, no vocabulary test can provide a precise

estimate of a learner's vocabulary size. There is no guarantee that a learner will know all words within a frequency level based on the fact that they know the ones being tested. This is why the VLT result should be considered as an approximate estimate of a learner's vocabulary size. This does not, however, mean that the VLT as a method for measuring vocabulary size would be invalid for the aim of this study, a discussion raised in the next section.

3.3 Validity and reliability of the VLT

Criticism has been raised towards the reliability and validity of the VLT as a testing method. Milton (2010) states that the format of the VLT is questionable since "success relies not just on a learner's knowledge of the test words (on the left hand side) but also on the knowledge of the words in the explanations (on the right side)" (2010, p. 222). Although there might be some truth to this assumption, Milton fails to mention the fact that the words used in the explanations are always of a more frequent nature than the target words. For example, Schmitt et al. (2001) explain that the 1000 level words are used to define the 2000-level words to ensure that the learner's ability to show their word knowledge does not depend on a misunderstanding of the words used in the definitions.

Milton (2010) also raises the issue of guessing in multiple-choice tests. As he puts it, a learner's knowledge of some word items can have an effect on the ability to deduce answers to other unknown word items. Such educated guesses might, therefore, compromise the test results and lower their reliability. There is always the possibility of guessing with this kind of testing format. However, the extensive work of validating the VLT has shown that guessing is not a central concern regarding low proficient learners since they are generally unsuccessful at their guessing (Schmitt et al., 2001). On the other hand, the higher proficiency learners were more successful in their guesses. However, the authors claim that this could not have been the case if these students did not have a sufficient knowledge of at least some of the other words (distractors) which are not being tested. Therefore, they argue that this type of guessing is also an indication of the students' vocabulary knowledge at the intended frequency level.

3.4 Procedure

In advance of the VLT, all participants were informed that they would take a diagnostic vocabulary test as part of a university study. Participants were immediately informed that

their results would only be used for research purposes and not as assessment material in the course that they were currently attending. Before the test was handed out, I explained to the students what the purpose of the test was and thus the purpose of this study. This was done in Swedish to ensure that all students fully comprehended the purpose of participating in and taking this vocabulary test. Thereafter, a consent form was distributed to all students under the age of 18 where their guardian was asked to give consent to their results being used (anonymously) in my study (see Appendix 2). In this consent form, it was also clearly stated what the purpose of the test was and what their results will be used for (in Swedish).

After having read through the consent form and signed it, the test was then handed out to the participants. I requested of the students to silently read the cover page of the test where the instructions were given (in English), and not to open the actual test before receiving my permission. When all had received their tests and read through the instructions, I recalled their attention to carefully go through the instructions on the cover sheet together (see Appendix 3). During this time, I also let the students ask questions if something seemed unclear. As for the time frame of the test, I explained to the students that they would have a maximum of 60 minutes to complete the test and that everybody must remain in the classroom during that hour. The reason for this was to avoid the risk of students neglecting the test situation as well as stressing their peers by leaving early and, therefore, affect the test results. If someone, however, finished the test before the end of the hour, they were told to stay seated an do other school-related tasks (exercises on paper).

For correcting and calculating the learner's performance on the VLT, one could settle with collecting and analyzing the scores at each frequency level. On the other hand, if one prefers calculating an estimate of the learner's actual receptive vocabulary size, which this study aims to do, one needs to go about calculating the scores differently. The following procedure that I have used when calculating a learner's vocabulary size is a rough way of doing it. It will probably also reveal an overestimation on knowledge, i.e. learners marking words that they do not know (because of the guessing frequency) (Schmitt et al., 2001). It is thus vital to clarify the fact that the estimations of the learners' vocabulary sizes are approximate.

For each level, 30 items are tested. When calculating a participant's vocabulary size score on a given level, their correct score on a specific level is divided with the number of words included within that frequency level; e.g. if a learner has 25 out of 30 correct answers on the 2,000 level, the calculation looks like the following: $(25/30) \times 2,000 = 1,666.66$ words, roughly, are expected to be known within that level. The 3,000 level tests another 30 items

from the next 1,000 words above the previous level, thus, if a learner answers 20 items correctly, the calculation is: (20/30) x 1,000 = 666.66 words. Since the VLT does not test items from each and every frequency level up to the 10,000 frequency band, some rough adjustments of the calculation were made. The test items from the 5,000 level can be taken as a rough guide to the whole range from 4,000 – 5,000, therefore, the calculation is: (number of correct items/30) x 2,000. The group of academic words tests items from a list of 570 words, hence, the following calculation: (number of correct items/30) x 570. Lastly, the 30 words from the 10,000 level are taken as a rough estimate of all the words between those covered above and the 10,000, i.e. the remaining 4430 words: (number of correct items/30) x 4430. The total vocabulary size of a specific learner is the entirety of these five different scores on each level. As mentioned above, there are limitations with this rationale of estimating the learner's vocabulary size, which will be elaborated on below.

3.5 Limitations

Some methodological limitations can be identified in the present study. First of all, the corpus is of a rather small size for any significant generalizations to be made upon the lexical nature of the reading texts in the NTs. It would have been desirable to include a minimum of ten test versions; however, this is not possible due to confidentiality. Nonetheless, there may be a value of such a study, especially since the NTs have not been examined in this manner before.

Second, it would have been ideal to have the participants take the reading tests included in the NT corpus to strengthen the claims upon their comprehensibility further. In this way, the reading test scores could have been compared to the learners' different vocabulary sizes. Unfortunately, the limited amount of time prevents the implementation of such a procedure.

Third, the students took the VLT two months before taking the NTs. It is possible that the students have increased their receptive vocabulary size during this amount of time, which also needs to be considered when analyzing the results of this study. However, one can probably assume that any drastic developments have not occurred concerning vocabulary growth.

Lastly, the fourth limitation considers the way in which a learner's unique vocabulary size was calculated. As previously mentioned, the VLT does not test a learner's knowledge on the 4,000 and 6,000–9,000 frequency bands. Although this was solved in an adjusted manner, it is of utter importance to emphasize the fact that scores at for example the 10,000 level is

based upon probable average scores on the preceding levels. Thus, the estimated vocabulary sizes need to be treated with caution. There is an additional new Vocabulary Size Test (Nation & Beglar, 2007) that measures learners' knowledge of the 6,000–8,000 vocabulary frequency levels. Although this one could have been included in the measurements for this study, this test is of a different character in terms of the number of items tested on each level (10 items per frequency level) and the way in which the items are presented (one word accompanied by four definitions). In the present paper, this test was not included partially because of the time frame but mostly because it is not comparable with the number of words per level in the VLT.

3.6 Ethical considerations

As mentioned in section 3.4, a consent form was distributed before the participation in the VLT. In this form, it was clearly stated that the participation was completely voluntary and that their results would not affect the students' grading in the English course they were taking. If the students wished to receive their test scores, this could be arranged after the trainee period. For the students who wanted to receive their test scores, these were sent to them via their English teacher. Although the English teachers of the classes that I have been in contact with were fully aware of the ethical considerations of not letting students' VLT results affect their grades, it is probably difficult to disregard what has once been seen. This is only the case for those students asking for their test scores. From what I have experienced in terms of these teachers' professionalism, it is unlikely that it would have an effect on their grading. Nonetheless, the potential outcome of such a possibility needs to be raised.

4 Results

In this part of the study, the NT corpus size and the participants' VLT scores will be presented. The VLT scores will then be compared to the LFPs of the NTs for each English course in alignment with the research questions. This will provide tables on students' lexical coverage and comprehension of the tests.

4.1 The NT corpus

The four different NTs used in this paper will be named: ENG5-1, ENG5-2, ENG6-1, and ENG6-2. The number 1 represents the replacement tests and the number 2 the autumn 2015 tests. When the texts were processed into LFPs, some automatic adjustments were made in the output: all punctuation marks were eliminated, figures (1, 2, 3, et cetera) were replaced by the word *number*, and contractions were replaced by the words that they consist of (e.g. *I'm* → *I am*). Also, single letters were eliminated except for *a* and *I*. The final corpus size on which this study is based is presented in Table 1:

Table 1: Total and separate corpus sizes of the four different NTs.

NTs	Tokens	Types	Word families
ENG5-1	2,173	996	826
ENG5-2	2,819	1,262	1,035
ENG6-1	3,166	1,246	1,002
ENG6-2	3,139	1,290	1,050
Total	**11,297**	**4,794**	**3,913**

Table 1 shows that the ENG6-tests are larger in terms of tokens, types, and families as opposed to the ENG5-tests. Also, the two ENG5-tests differ in size. This has to do with the fact that the longer reading texts and short texts in ENG5-2 are longer than the ones in ENG5-1 which section 4.3 will illustrate. Had there been more NTs included in the corpus, it would have been possible to determine which of the two is representative for an ENG5 NT. Moreover, the table above represents all language that is present in the tests, i.e. both reading texts and the questions and instructions accompanying them. The token, type and word family ratio of actual reading text versus questions and instructions will be presented in section 4.3 and 4.4.

4.2 Participants' VLT scores

For sections 4.3 and 4.4, the participants' VLT scores will be used when examining their potential lexical coverage of the NTs within each English course. 89 students from four different classes took the VLT. The number of participants is presented in Table 2; the results of each of the four classes have been summarized in Appendix 4.

Table 2: Number of students from each class

Class	Number of students
English 5: Vocational programme	22
English 5: Theoretical programme	25
English 6: Theoretical programme 1	20
English 6: Theoretical programme 2	22
Total	**89**

4.3 Vocabulary of English 5 students compared to the NTs

In this section, the VLT results of 47 EFL learners in English 5 will be presented in connection to the NTs within that proficiency level. Whether or not a test is comprehensible could be answered merely by analyzing the actual text that is to be read by the students. However, comprehensibility is also dependent upon the fact that learners understand what is asked of them when presenting their comprehension in accompanying instructions, questions, and tasks (from now on abbreviated IQT). The token, type, and word family ratios for these two groups of language (texts and IQT) are displayed in Table 3:

Table 3: Analysis of the text language and the IQT language in ENG5-1 and ENG5-2 tests

	Tokens	Types	Word families
ENG5-1 text	1,559	656	539
ENG5-2 text	2,201	943	763
Total text	**3,760**	**1,599**	**1,302**
ENG5-1 IQT	614	340	287
ENG5-2 IQT	618	319	272
Total IQT	**1,232**	**659**	**559**

What can be drawn from the table above is that around 25% of the tests contain IQT language

(1,232 tokens divided with the total tokens in the tests which is 4,992), and thus the necessity to understand that language is important. However, this language is usually of high-frequent character, hence not posing a significant problem for students mastering these frequency levels of vocabulary knowledge.

Table 4 presents the data from Table 3 distributed over ten frequency bands and shows the coverage on these levels. Proper names, i.e. personal and geographical names and other low-frequency words are included in the "off-list" words.

Table 4: Coverage of the ENG5-1 and ENG5-2 tests

Frequency level	Coverage % ENG5-1	Coverage % ENG5-2	Average cumulative coverage
K1	79.48	83.12	81.66
K2	10.40	7.06	90.11
K3	4.05	3.19	93.65
K4	1.70	1.81	95.37
K5	1.38	0.96	96.51
K6	0.14	1.03	97.15
K7	0.74	0.53	97.77
K8	0.05	0.28	97.95
K9	0.09	0.14	98.07
K10-20	0.56	0.72	98.71
Off-list	1.33	0.96	~100

First and foremost, Table 4 shows that around 93% of the words are within the high-frequency levels (K1–K3). Approximately 4% of the words can be found in the mid-frequency vocabulary levels (K3–K9) and the rest in the low-frequency category (K9–K20) and off-list.

Because a majority of the off-list words are proper names, one might assume that these words are familiar to the learner. In that case, these words would be included in the calculation of the lexical coverage. Consequently, a 98% coverage could be achieved by the knowledge of 6,000 words, which cover 97.15% and the off-list words which cover an additional 1.29%. However, it is not a vast majority of the off-list words that is of this character (proper names); some are also low-frequent. Therefore, this group will not be

included in the calculations below.

Since the aim of the paper is to find out the relationship between vocabulary size and coverage based on LFPs, the data on the coverage from Table 4 will be presented with the data on learners' vocabulary size from the VLT in Table 5. The learners' vocabulary scores are divided into intervals of 1,000 words and the frequency level column has been replaced by the learners' vocabulary size. If, for example, 8,000–9,000 words cover 98% of a text, as shown in Table 4, then a learner with a vocabulary size of 8,000–9,000 words can understand a comparable percentage of that text. In the fifth column of Table 5, the total number of students within each vocabulary knowledge interval is presented.

Table 5: Vocabulary size and lexical coverage of EFL learners in English 5

Approximate vocabulary size	Lexical coverage	Vocational programme students	Theoretical programme students	Total number of students
1,000	81.66	1	0	**1**
2,000	90.11	0	0	**0**
3,000	93.65	2	2	**4**
4,000	95.37	4	2	**6**
5,000	96.51	1	7	**8**
6,000	97.15	3	3	**6**
7,000	97.77	6	9	**15**
8,000	97.95	4	1	**5**
9,000	98.07	1	1	**2**

As mentioned earlier, a minimum of 8,000–9,000 words is needed for a 98% coverage of the ENG5-1 and ENG5-2. This means that 7 students out of the 47 will reach that threshold level. For these students, vocabulary size is no issue and they would most likely have passed the NTs. The other 40 students, however, are below the 98% level, i.e. 85% of the group. In comparison to the 12% that statistically fail the NTs, an even larger number of learners with an insufficient vocabulary size can be found in Figure 6:
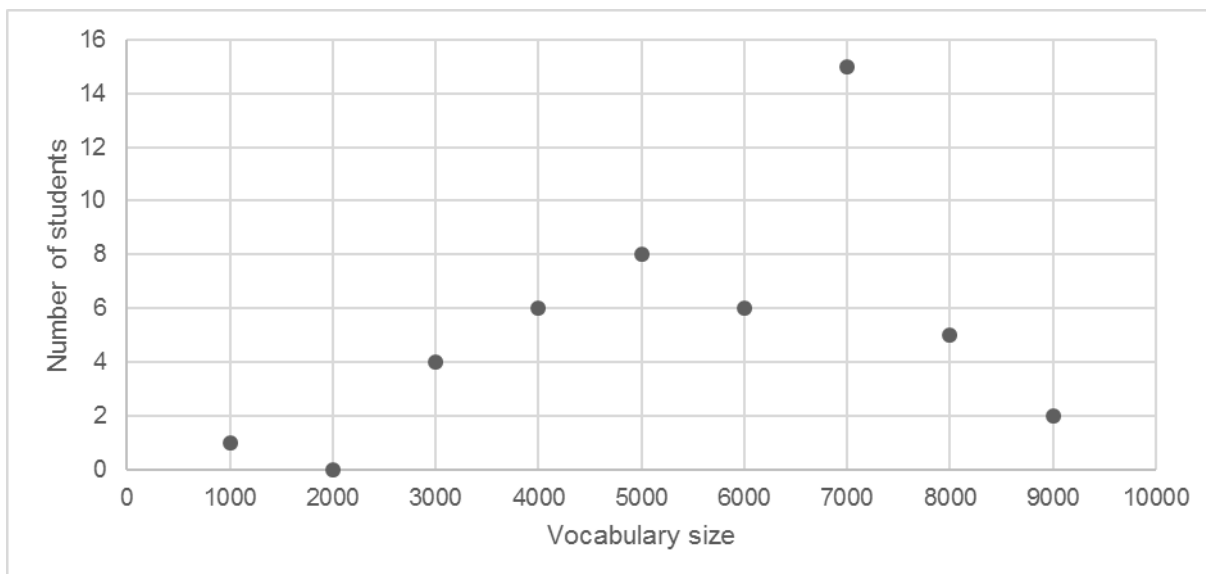
Figure 6: The number of English 5 students distributed on different vocabulary size levels

As shown in Figure 6, the VLT scores of the English 5 students resemble a normal distribution. Out the 40 students below the 98% level, 29 students can be found in the middle of the bell curve (vocabulary sizes of 5,000–7,000 words) and have about 95–98% coverage. 11 students (approximately 23%), have a vocabulary size of. no more than 4,000 words. This means that they have about 80-95% lexical coverage, which is too low to understand the meaning of the texts. These students would most probably be the ones receiving the lowest reading scores and thus also risk failing the reading part of the NTs, if they were to take the tests. When calculating the average of all of the students' vocabulary size, the following numbers are revealed:

Table 6: Average of English 5 learners' vocabulary size and standard deviation (SD) of data values

	Number of students	Average vocabulary	SD
Vocational programme	22	6,306	2,054
Theoretical programme	25	6,303	1,540
Total	**47**	**6,305**	**1,778**

The averages of the two groups are nearly identical: the vocational programme with an average of 6,306 and the theoretical with an average of 6,303. The standard deviation of the

whole group indicates a spread of about 1,800 words around the mean. A vocabulary size of 6,300 words yields a lexical coverage of just above the 97% threshold level. However, when taking a closer look at the two groups vocabulary sizes distributed on levels of 1,000, the following chart can be distinguished.
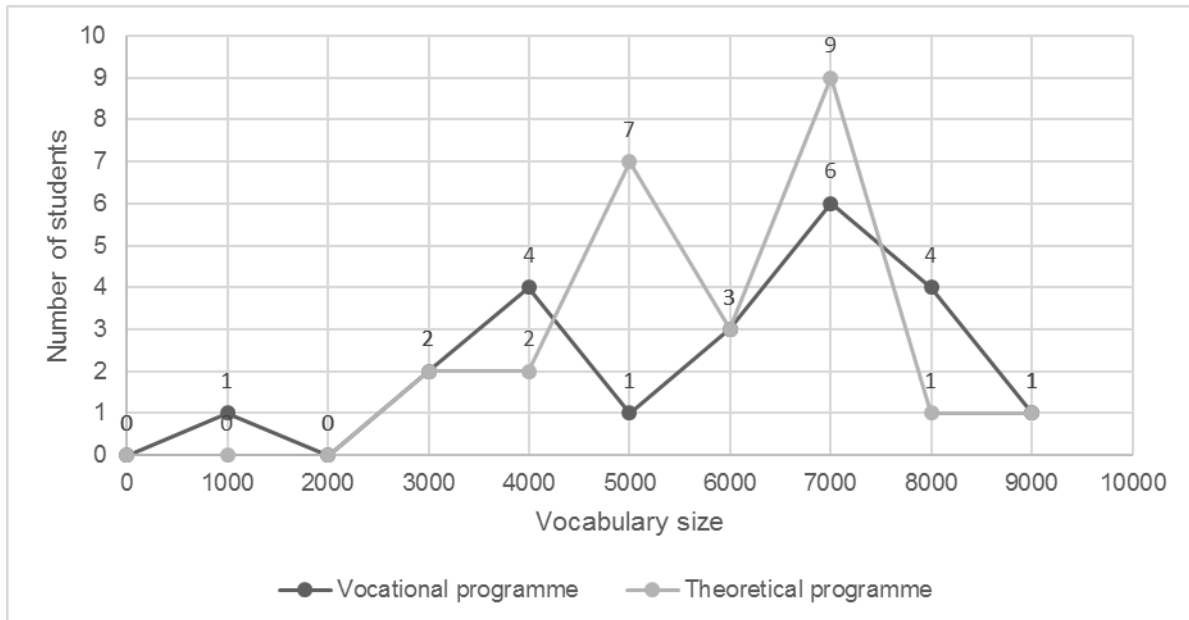

Figure 7: The number of English 5 students within the vocational and theoretical programme distributed on different vocabulary size levels

As shown in Figure 7, the distribution curves are somewhat different on certain vocabulary size levels. For instance, four students from the vocational programme had a vocabulary size of 8,000 words whereas only one student from the theoretical programme managed to reach this level. Additionally, 5 out of 22 vocational programme students had a vocabulary size reaching up to the low-frequency levels (8,000–9,000), whereas only two of out 25 theoretical students reached this level. However, the fact that the vocational programme consists of 22 students and the theoretical programme of 25 needs to be readdressed. Therefore, the variance needs to be treated with some caution.

4.4 Vocabulary of English 6 students compared to the NTs

For this part of the study, the VLT results of 42 English 6 students, all within theoretical programmes, will be presented. The analysis of the token, type and word family ratio for the ENG6-1 and ENG6-2 tests is shown in Table 7:

Table 7: Analysis of the text language and the IQT language in ENG6-1 and ENG6-2 tests

	Tokens	Types	Word families
ENG6-1 text	2,623	983	783
ENG6-2 text	2,373	903	719
Total text	**4,996**	**1,886**	**1,502**
ENG6-1 IQT	543	263	219
ENG6-2 IQT	766	387	331
Total IQT	**1,309**	**650**	**550**

As with the ENG5 tests, Table 7 shows that an approximate of 20% percent of the ENG6 tests contains IQT language, also here of high-frequent nature. Compared to the ENG5-1 and ENG5-2, these tests are larger in terms of the number of tokens (6,305 compared to 4,992 in English 5). This has to do with the progressing proficiency level, thus requiring the ability to read longer texts (S. Nilsson and H. Schönberg, personal communication, 17 April, 2018).

In Table 8, a similar compilation of the total data from the previous table is presented. The data is distributed on 10 frequency levels of coverage and the "off list" words which contain the same kind of vocabulary as in the ENG5 tests, i.e. personal and geographical names as well as low-frequency vocabulary. If using the same rationale of calculating the 98% threshold level as was used with the ENG5 tests, a minimum of 8,000 words is needed. Here, an approximate of 94% of the words are of high-frequency character whereas only 2% belong to the mid-frequency group. The rest are found in the off-list and the low-frequency level.

Table 8: Coverage of the ENG6-1 and ENG6-2 tests

Frequency level	Coverage % ENG6-1	Coverage % ENG6-2	Average cumulative coverage
K1	83.23	78.66	80.99
K2	7.96	8.47	89.17
K3	4.70	5.51	94.24
K4	0.88	2.04	95.70
K5	0.69	1.40	96.75
K6	0.41	0.96	97.43
K7	0.44	0.25	97.78
K8	0.19	0.32	98.03
K9	0.09	0.10	98.13
K10-20	0.21	0.34	98.48
Off-list	0.82	0.86	~100

In Table 9, data on learners' vocabulary size from the VLT is presented in relationship to the lexical coverage percentages. As earlier, the learners' vocabulary scores are divided into intervals of 1,000 words.

Table 9: Vocabulary size and lexical coverage of EFL learners in English 6

Approximate vocabulary size	Lexical coverage	Theoretical programme 1 students	Theoretical programme 2 students	Total number of students
1,000	80.99	0	0	0
2,000	89.17	0	0	0
3,000	94.24	1	1	2
4,000	95.70	1	3	4
5,000	96.75	7	2	9
6,000	97.43	3	8	11
7,000	97.78	5	4	9
8,000	98.03	1	3	4
9,000	98.13	2	1	3

If a minimum of 8,000 words is needed for a 98% coverage of the ENG6-1 and ENG6-2, this means that 7 students out of the 42 will reach that threshold level. For these students, vocabulary size is no issue and they would most likely have passed the NTs. The other 35 students, who constitute 83% of the group, are below the 98% level, however. In comparison to the 12% that statistically fail the NTs, a similar percentage of learners with an insufficient vocabulary size can be found in Figure 8:
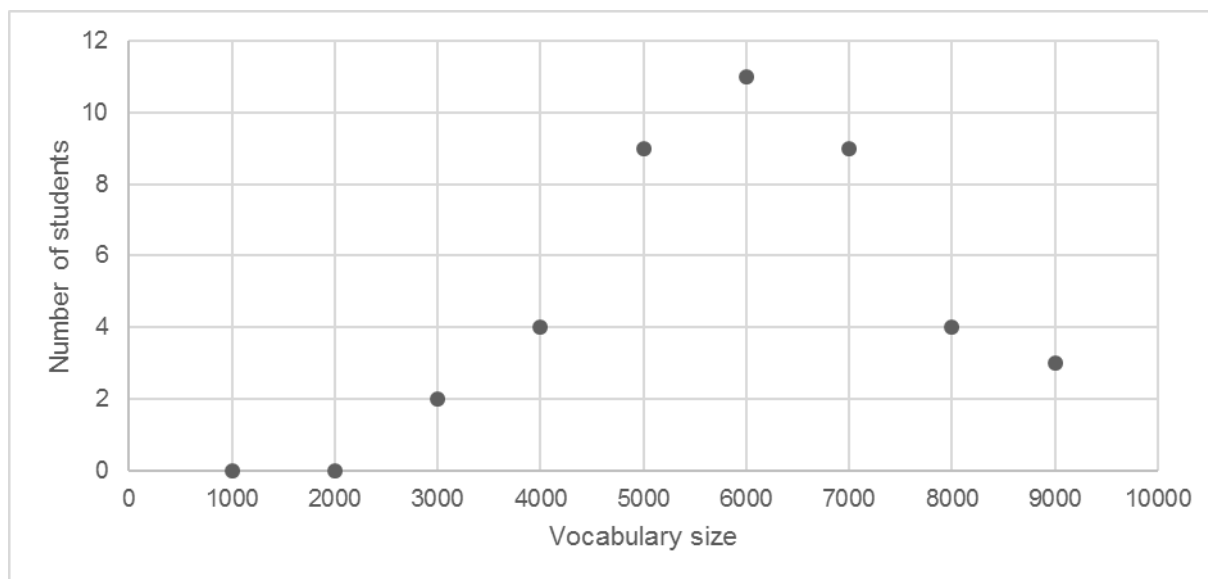

Figure 8: The number of English 6 students distributed on different vocabulary size levels

As shown in Figure 8, the VLT scores of the English 6 students in Figure 8 also resemble a normal distribution. Out the 35 students below the 98% level, 29 students can be found in the middle of the curve (vocabulary sizes of 5,000–7,000 words) and have about 95–98% coverage. 6 students (approximately 14%), have a vocabulary size of no more than 4,000 words. This means that they have about 81-95% lexical coverage, which is too low to understand the meaning of the texts. As was the case with the low-risk group in English 5, these students would most probably also be the ones receiving the lowest reading scores and thus also risk failing the reading part of the NTs, if they were to take the tests. When calculating the average of all of the English 6 students' vocabulary size, the following numbers are revealed:

Table 10: Average of English 6 learners' vocabulary size and standard deviation (SD) of data values

	Number of students	Average vocabulary	SD
Theoretical programme	20	6,527	1,559
Theoretical programme	22	6,606	1,434
Total	**42**	**6,566**	**1,477**

Table 10 shows that the averages of the two groups are similar: the first programme with an average of 6,527 words and the other with an average of 6,606 words. The standard deviation of the whole group indicates a spread of about 1,500 words around the mean. A vocabulary size of 6,500 words yields a lexical coverage of just below the 98% threshold level. In comparison to the average vocabulary size in English 5, this suggests an approximate receptive increase of approximately 250 words after one additional year of instructed EFL teaching.

5 Discussion and pedagogical implications

In this section, the vocabulary size results from the two proficiency level groups (English 5 and 6), as well as the LFPs will be discussed. Some pedagogical implications regarding vocabulary knowledge, learning, and teaching will also be raised throughout each section.

5.1 EFL learners' vocabulary size

In alignment with previous research within the area (Hu & Nation, 2000; Nation, 2006), the present paper has shown that a vocabulary size of 8,000–9,000 word families is needed for a lexical coverage of 98%. Furthermore, this study has shown that a majority of the students are below the 98% lexical threshold level of their respective proficiency levels. In the English 5 group, 85% of the students were below the threshold level; in English 6 the percentage was 83%. This does not mean that the same percentage would fail the reading tasks in the NTs. It means that the insufficient vocabulary size probably plays a significant role for the results in the reading part for most students. The fact that 5-12% statistically fail in the reading part of the NTs may be due to the vocabulary of about 14–23% of the students only giving a lexical coverage of 80–95%, which is insufficient for reading comprehension.

There are other factors determining reading comprehension success than what numbers can mediate solely. Some learners might benefit from other reading comprehension-related skills in order to succeed in the NTs. As with all reading material, the texts in the NTs include contextual clues which may facilitate for learners to derive meaning in relation to unknown words. Even if a specific word is unknown, the overall passage in which the word is presented may help in the meaning construction process. Additionally, the comprehension questions may not always require students to know the word meanings in the text. Nevertheless, statistics still reveal a percentage of up to 12% failing the reading tasks in NTs. The probability that this is due to an insufficient vocabulary size cannot be neglected after the results of this study. Although displaying results in percentages when the sample size is rather small is problematic, it is still interesting to see how a noteworthy number within the two learner groups does not manage to reach the 98% level. Thus, one can ask if the NTs are too difficult in terms of its lexical information and should be made easier or if the learners' vocabulary sizes are too small and more focus should be put on vocabulary learning. This paper suggests the latter possibility, which will be elaborated upon further.

The fact that the average receptive vocabulary size of an English 6 student is just

around 6,500 words towards the end of the semester can cause problems. Many of these students might probably not take the last English 7 course before proceeding to the university since this course is not obligatory in order to qualify for higher education. Presumably, their receptive vocabulary size will not increase significantly until they enter academic studies. Obviously, new vocabulary can be and probably will be learned outside of school and also in the years between graduation and entering higher education. For the sake of simplicity, one can, however, assume that such a vocabulary learning process is of random and uncontrolled character. Therefore, the probability of facing difficulties when dealing with academic texts on advanced levels is high. This poses an additional problem when dealing with productive tasks at university levels since these students' productive vocabulary is much smaller than their receptive vocabulary of 6,500 words. They will, therefore, probably not have the tools necessary for coping with a variety of advanced tasks in their higher education.

Another interesting aspect that was revealed was the fact the learners from the vocational programme in English 5 performed equally as the ones within the theoretical programme. In light of the low results on reading scores that statistics have shown, one would expect the learners from the vocational programme to perform worse in comparison to the theoretical programme students. However, their average vocabulary size was, to a small degree, better than the theoretical group. Had there been an equal amount of students within the vocational group as the theoretical one (i.e. 25 students), the difference might even have been larger. In some individual cases, the vocational programme students performed significantly better than the theoretical programme students, especially in regards to having vocabulary sizes reaching up to the low-frequency levels (8,000–9,000). From the results, it has been shown that a total of six English 5 students had this kind of vocabulary size, where all but two students belonged to the vocational programme. One might thus wonder why statistics still state the fact that the vocational programme students are the ones with higher percentage of failed performances on the NTs, despite the fact that their vocabulary size is equally as large, and in some cases larger, as the vocabulary size of the theoretical programme students? One suggestion might be that they face other non-vocabulary related difficulties when taking the reading tests in the NTs, e.g. reading fluency problems. Evidently, no clear and definite conclusions can be drawn from just one vocational class. It might even be the case that this particular vocational class is a group that performed above the average of a typical vocational class in terms of vocabulary size. Further examinations of such kinds are thus required in order to draw general conclusions upon the matter.

After having contrasted the two proficiency groups' average receptive vocabulary

sizes, which should be treated as rough estimates, a difference of 250 words was manifested; 6,300 words in English 5 and 6,550 in English 6. This means that the difference between their productive vocabularies is even smaller. It is important to note the fact that the English 5 and 6 groups are two different groups of learners, thus it is not possible to determine a definite vocabulary increase from the former proficiency level to the other. Also, it might be the case that the vocabulary sizes of the English 5 and English 6 students of this study are not representative for a typical student within that course. The sample size of this study is too small to draw such conclusions. However, if treating the suggestion of a possible vocabulary increase of 250 words with caution, some valuable estimations can be made. If one presumes that Swedish EFL learners in general have about 2 hours of instructed teaching of English each week, this would mean that they receive about 80 hours of teaching each year. When dividing the increase of 250 words on these total hours, an average of 6 words is expected to be learned, receptively, each week. One interesting question arising from this estimation is if a maximum of 6 words per week is what the students are cognitively capable of learning during this amount of time or if it is a result of inefficient vocabulary teaching. Inefficient vocabulary teaching could signify several things. One example could be that explicit vocabulary teaching is not dealt with to a great extent. Another possibility is that the vocabulary is not taught the right way, i.e. in combined forms by using a variety of methods and contexts where new vocabulary is presented (Schmitt, 2008). It could also be possible that the wrong type of vocabulary is being taught, both in terms of frequency levels and individualized vocabulary teaching. Learners will evidently differ in vocabulary sizes, as key findings of this study has shown (see SD numbers in Tables 6 and 10 in sections 4.3 and 4.4). Thus, it is valuable checking the receptive vocabulary size levels of the learners and provide them with the relevant training according to their needs.

5.2 The LFPs of the NTs

Since it was not possible to include more than two NTs for each English course in the present study, the following numbers have to be treated with caution. They should not be treated as absolute truths for all of the NTs that have been created the last decade. Moreover, it would be most optimal to have the students take the reading tests in the NTs that were analyzed to draw more solid correlations between vocabulary size and reading comprehension. Also, it is important to note that the analyzed texts are of different nature; some are longer reading texts whereas others are short matching-items exercises. The text type difference might therefore

affect the frequency level percentages in the LFPs of the tests. Despite these affecting factors, there is still some value to the discussion below.

Lexical analyses of the NTs show that 4% of the words in the ENG5-tests belongs to the mid-frequency group whereas the ENG6-tests only include 2% of mid-frequency words. In other words, the opportunity of meeting mid-frequency words is extremely low. One significant problem could most likely arise from this fact. Since the NTs, among other things, serve as a concretization of the syllabus of English, teachers can tend to treat the different texts as a guideline to what texts are relevant and suitable to work with in the respective proficiency levels. Therefore, it could lead to teachers choosing reading texts of the similar lexical nature, which reproduces the low chances for students to meet a sufficient amount of mid-frequency vocabulary.

As recently mentioned, a small number of words in the NTs belonged to the mid-frequency vocabulary level. As with all kinds of teaching, learners' should receive instruction in areas that are within their proximal zone of development. In terms of these learners' proximal zone of vocabulary development, the vocabulary that teachers should focus on is mid-frequency vocabulary, specifically words from the 6,000 level and above. One way of treating mid-frequency words is to work with the AWL words compiled by Coxhead (2000). Not only do they treat mid-frequency words on several levels, but they also prepare students for higher education with a larger and more varied vocabulary. In fact, working with frequency lists in vocabulary teaching is useful when deciding on which type of vocabulary is relevant to teach in a certain group or with certain students. Another strategy is to teach morphological knowledge in relation to vocabulary instruction. A larger amount of vocabulary learning and vocabulary size developments can occur with the background knowledge of prefixes, suffixes, and stems (Laufer & Nation, 1993; Schmitt, 2008). It does not lay in the nature of this paper to recommend in what ways vocabulary teaching best affects vocabulary learning and growth. However, it is undoubtedly so that the complex process of vocabulary learning needs to be treated effectively and in various ways considering the limited amount of time that learners have for encountering new vocabulary (Laufer & Nation, 2001).

6 Conclusion

Key findings of the present study have confirmed that Swedish upper-secondary EFL learners' receptive vocabulary size is too small in comparison to the requirements of the NTs in terms of lexical coverage. In alignment with national statistics, evidence has shown that a risk-group of 14–23% students would fail the NTs due to insufficient vocabulary knowledge below the 95% threshold level. Although the sample and corpus size are too limited for any general conclusions to be drawn, it is probably not the NTs that are too difficult. It is more likely the case that the students do not practice vocabulary enough in order to reach a 98% lexical coverage. It has been shown that the average vocabulary size of the Swedish EFL learner stretches from 6,300 word families in English 5 to 6,550 word families in English 6 which is not enough to achieve adequate reading comprehension of the NTs.

The results of the study have direct implications for vocabulary teaching. As the average EFL learner's vocabulary size is somewhere within the mid-frequency levels (6,000 word families), the present paper emphasizes the need to focus more on teaching vocabulary within these frequency bands. Learners need relatively large vocabularies (8,000–9,000 word families) to use a second language or foreign language at a successful level, especially concerning reading comprehension (Laufer, 1992; Laufer & Nation, 1995; Stæhr, 2008). This fact clearly suggests the need to set and pursue high vocabulary targets in language teaching. If the English 6 course is supposed to prepare Swedish EFL learners for higher education and they are expected to meet the standards within that advanced context of learning, it is vital that one year of language instruction yields more than a receptive increase of 250 words.

While this study has only treated one dimension of vocabulary development, namely the vocabulary size, the aspect of vocabulary quality and depth needs to be given attention as well. Not only do learners need to have a large vocabulary, but they also need to know a great deal about the words in order to be able to use them for different purposes. Therefore, when teachers focus on teaching the words closest to EFL learners' zone of proximal development, it is vital that they treat other aspects than just the receptive knowledge of the form-meaning link. Instruction regarding different word parts, collocations, semantic relationships between words, and constraints on use is also of importance when planning direct and indirect vocabulary teaching (Nation, 2013).

Hopefully, this paper can be seen as a pilot study introducing a new way of analyzing the NTs and consequently giving directions and suggestions for further validations of the tests as well as examinations of Swedish EFL learners' vocabulary size in English 5 and 6. For

future research it is suggested that more NTs are investigated, if possible, and that larger groups of students (both vocational and theoretical) are tested on their vocabulary size levels. This will yield a better understanding of the lexical nature of the NTs but most importantly Swedish EFL learners' vocabulary size levels; a field of research that is yet not examined enough. In addition, it would be desirable to let the participants take the NTs that are being analyzed to include an additional variable to the analysis, namely their reading scores in relation to vocabulary size and lexical coverage. Another proposition for future research is to analyze the learners' written compositions in the NTs to examine their productive vocabulary within different frequency levels. Consequently, a comparison can be drawn in order to determine what kind of receptive vocabulary has been activated in the learners' written production, a skill that is most useful for higher education.

From the present study, it is once again confirmed that vocabulary is, in fact, a great determiner for successful reading comprehension. As much as vocabulary development is a complex phenomenon, it is also the necessary core of overall language success and should thus receive the time and focus that is needed for further growth.

Reference list

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

Agernäs, E. (2015). *Vocabulary size and type goals in advanced EFL and ESL classrooms. A review of research on lexical threshold, lexical coverage, reading and listening* (Kandidatuppsats). Göteborg: Institutionen för språk och litteraturer. Available at: https://gupea.ub.gu.se/handle/2077/38439

Axelsson, P., & Lindqvist, A.K. (2017). *Kursprovet i engelska 5 vårterminen 2017.* Göteborg: Institutionen för pedagogik och specialpedagogik. Göteborgs universitet.

Börjesson, L., & Schönberg, H. (2012). *Kursproven i gymnasieskolan – Engelska A och Engelska B 2001-2012.* Skolverket.

Cobb, T. (2018). *VocabProfile Home* [computer program]. Retrieved 23 April 2018 from https://www.lextutor.ca/vp/comp/

Coxhead, A. (2000). A new Academic Word List. *TESOL Quarterly, 34*(2), 213–238. Retrieved 2 April 2018 from http://search.proquest.com/docview/62327492?accountid=11162

Grabe, W. (2009). Teaching and testing reading. In M. H. Long, & C. J. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 441–462). Chichester: Wiley & Blackwell.

Göteborgs universitet. (2018). Projektet nationella prov i främmande språk. Engelska 5 – Exempel på uppgiftstyper. Retrieved 18 May 2018 from http://nafs.gu.se/prov_engelska/exempel_provuppgifter/engelska_5_exempeluppg

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403−430.

Kendeou, P., McMaster, K.L., & Christ T.J. (2016). Reading Comprehension. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 62–69.

Laufer, B. (1989). What percentage of text leis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Clevedon: Multilingual Matters.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Bejoint & P. Arnaud (Eds.), *Vocabulary and Applied Linguistics* (pp.126–132). Basingstoke & London: Macmillan.

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

Laufer, B. & Nation I. S. P. (1993). Word families. *International Journal of Lexicography 6*(4), 253–279.

Laufer, B. & Nation, I. P. S. (1995). Vocabulary size & use: Lexical richness in L2 written productions. *Applied Linguistics, 16*(3), 307–322.

Laufer, B. & Nation, I. S. P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? In S. Foster-Cohen & A. Nizegorodcew (Eds.), *EUROSLA Yearbook 1,* (pp. 7–28). Amsterdam: Benjamins.

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

Lundahl, C. (2012). Bedömning – att veta vad andra vet. In U.P. Lundgren, R. Säljö & C. Liberg (Eds.), *Lärande skola bildning: [grundbok för lärare].* (2nd ed.). (pp. 257–293). Stockholm: Natur & kultur.

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. *EUROSLA Monograph Series 1. Communicative proficiency and linguistic development: Intersections between SLA and language testing research*, 211–232.

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

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

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

Öhman, L. (2013, November 17). *Sverige "världsbäst" på engelska – igen*. Skolvärlden. Retrieved May 13, 2018 from http://skolvarlden.se/artiklar/sverige-varldsbast-pa-engelska-igen.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363.

Schmitt, N., & Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching 1. *47*(4), 484–503.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88.

Skolverket. (2018). *SIRIS kvalitet och resultat i skolan. Statistik avseende gymnasieskolan.* Retrieved 2 April 2018 from https://siris.skolverket.se/siris/f?p=SIRIS:114:0::NO:::

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language*

Learning Journal, 36(2), 139–152.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory* (Vol. 1, Explorations in the Learning Sciences, Instructional Systems and Performance Technologies Ser). New York, NY: Springer.

Youcefi, F. (2012, June 21). *Svenska elever talar utmärkt engelska.* SVT Nyheter. Retrieved May 13, 2018 from https://www.svt.se/nyheter/inrikes/sa-bra-ar-svenkar-pa-engelska

Appendices

Appendix 1: Four text examples from the NTs in English 5

Example A: A longer reading text and accompanying questions

Grand Central Terminal

Read the following article from the magazine National Geographic and then answer all the questions either by writing a few words or by marking the best alternative – A, B, C or D.

Every day more than 700,000 people hustle through Manhattan's Grand Central Terminal. It's one of the biggest, deepest – and indeed grandest – train stations in the world.

In Grand Central Terminal, people are lost and found. At the information booth – a gilded cupcake of a structure in the middle of the main concourse – they're mostly lost. They need to find their way to Poughkeepsie, Mamaroneck, the theater district, the bathroom. They've misplaced their cell phones, their maps, their tickets, their friends. They need someone, please, to show them how to get through Grand Central and how to get out.

On a recent day Audrey Johnson was one of the two information officers working in the booth. Johnson is small and solid, with marcelled curls, carefully polished lips, and nearly perfect recall. The lost souls approach her window and lob a destination at her, and **she returns it before it even bounces.** As many as a thousand people an hour come to the booth, and only occasionally does Johnson peek at a schedule or reference book before setting them straight. The exchanges are fast and steady. They produce a kind of boogie-woogie rhythm. While she's working, Johnson jiggles her leg back and forth, keeping time.

"Garrison?"
"2:09, track 7."

A businessman, neat as a pin: "The next train to Stamford?"
"3:47 p.m., track 104."

A Hasidic teenager: "The way to Poughkeepsie?"
"Hudson line."

An elderly woman in a Donald Duck T-shirt saying I'M THE BOSS; her husband in a Michael Jordan jersey and support hose: "Brewster?"
"Track 39, five minutes."

A blonde, eating breath mints: "I'm meeting a guy tomorrow, and all he said was, 'Meet me on 43rd Street'." She looks pleadingly at Johnson, then walks away.

People are found at Grand Central too. It's rumored that most tourists and nearly all New Yorkers use the information booth as their fail-safe meeting place in case of disaster or dislocation. It is, after all, the center point of the place that feels most central in Manhattan – the place that is both totally public and yet still protected, that is both a main entry into Manhattan and a main way out.

1   In what **city** is Grand Central situated?

_____

2   Where do people go when they are lost?

_____

3   What is said about Audrey Johnson?
   A   She is in charge of the booth ☐
   B   She has social competence ☐
   C   She has a good memory ☐
   D   She is physically strong ☐

4   What is meant by the phrase **"she returns it before it even bounces"**?

_____

5   How often does Audrey have to look for information when helping people?

_____

6   What is exchanged in a way that is compared to a boogie-woogie rhythm?

_____

7   What is meant when a train is said to leave at 3:47 **"p.m."**?

_____

8   What is suggested about the blonde customer?
   A   She seems confused ☐
   B   She reacts angrily ☐
   C   She feels pressed for time ☐
   D   She has the wrong address ☐

9   What other use, apart from asking about destinations, is made of the information booth?

_____

☞

Example B: Short texts/Mini-essays and accompanying multiple-choice questions

Short Texts

Read the following short texts and decide which of the four alternatives agrees best with the text. Show that you have understood by writing one of the letters – A, B, C or D – in the boxes below.

1 Urban Cycling

Bike riders beware. Urban cycling is not for the faint of heart. You respectfully share the road, of course, but you *will* be confronted by reckless drivers and the occasional angry pedestrian. But you're a warrior, so ride like one. Don't forget to bring your battle gear: a helmet, some LED lights and, of course, a dose of healthy aggression.

According to this text…

A   bad temper is common among drivers
B   cycling could be a risky undertaking
C   cycling makes your heart stronger
D   bikers should avoid busy roads

2 Twitter Behaviour

A small but vocal subculture has emerged on Twitter of grammar and taste vigilantes who spend their time policing other people's tweets – celebrities and nobodies alike. These are people who build their own algorithms to sniff out Twitter messages that are distasteful to them and then send scolding notes to the offenders. They see themselves as the guardians of an emerging behaviour code: Twetiquette.

This text tells us that…

A   Internet censorship is increasing
B   Twitter has become a police matter
C   some people have set rules for tweets
D   chatting is a growing phenomenon

Answers

1 ☐   2 ☐

Example C: Matching-item text

What Is Explained?

Below there are 14 definitions of words. Read each definition and decide which word is explained. Choose the word from the list that matches the description and write the letter of that word in the appropriate box. There are many more words than definitions so you cannot use all the words.

1 Perform and say the words of a particular character in a theatre performance ☐

2 Change your position or make something do this; go to a different place ☐

3 Actively use your mind to do something; have a particular belief or mental attitude to a problem ☐

4 To reach the place you are going to at the end of your trip ☐

5 Give out or produce bright light; direct a torch or another light somewhere in order to see something in the dark ☐

6 Attempt to find something or desire to obtain something ☐

7 Say "yes" to an offer or proposal; agree to do something that someone asks you to do ☐

8 Give something to someone or make it available to them because it is useful or necessary for them ☐

9 Move a spoon or other tool round and round in a liquid or other substance in order to mix it thoroughly ☐

10 Make smaller or less in amount, degree or size ☐

11 Ask for something or insist on having something, especially because you think you have a right to do this ☐

12 Risk a sum of money or valued item against someone else's based on the result of an unpredictable event ☐

13 Talk or write about something, usually quickly and without saying very much or going into detail ☐

14 Link or connect something with something else; become a member of an organisation, society, or group ☐

A	accept
B	act
C	arrive
D	bet
E	beg
F	demand
G	depart
H	journey
I	join
K	lie
L	lose
M	mention
N	move
O	prefer
P	provide
R	read
S	reduce
T	remove
V	seek
X	shine
Y	stir
Z	think

Points

14

Example D: Cloze-item exercise with vocabulary alternatives

Cheeky Monkeys

Read the following text where a number of words have been left out. In the margin you will find four alternatives for each gap. Choose the best alternative and mark your answer A, B, C or D in the appropriate box.

One day in 1996, the film-maker Karen Bass was discussing a project in one of the city parks in Durban, South Africa. One moment, she had her lunch in her hand, the next, it had ___1___ .

The sandwich stealer was a vervet monkey, one of the hardened criminals of urban Durban. Now city dwellers, the monkeys ___2___ food from fruit bowls and rubbish bins, from bird tables and delivery vans. If they had a criminal ___3___ , it would include breaking and entering, ___4___ off car alarms and common theft. Nor would they ___5___ to rehabilitation. These hungry monkeys can no more resist an open window leading to ___6___ than a pickpocket can leave a bulging wallet. "They are in the cities because they are so ___7___ ," says Bass. "They are capable of getting through razor wire and burglar bars. South Africa is probably the most security-conscious ___8___ in the world, and in Durban, everybody has dobermans and high wire ___9___ . But they cannot keep out the monkeys."

1
A fallen
B hidden
C vanished
D left

2
A pinch
B reserve
C help
D deserve

3
A look
B record
C mind
D protest

4
A chewing
B putting
C going
D setting

5
A respond
B object
C reply
D turn

6
A capture
B housing
C money
D food

7
A generous
B charitable
C agreeable
D adaptable

8
A landscape
B country
C ground
D view

9
A fences
B cutters
C checks
D locks

Appendix 2: Consent form for participation in the VLT

Samtyckesblankett för studie på Göteborgs universitet

2018-02-20

Kära vårdnadshavare,

Mitt namn är Yaran Kakaee och jag är lärarstudent på Göteborgs universitet. För närvarande läser jag mitt femte och sista år på lärarprogrammet och har nu min praktikperiod på [X] mellan 5e februari och 20e mars.

Under denna praktikperiod ska jag samla material för min kommande studie i engelska som jag ska skriva denna vår. Syftet med denna studie är att undersöka svenska elevers ordförråd i engelska i jämförelse med vad som krävs i Nationella provens läsförståelseuppgifter.

Anledningen till att jag skriver till er är för att fråga huruvida det är möjligt för ert barns resultat att användas i min studie. Eleven kommer att få skriva ett ordkunskapstest (ca 60 minuter) under skoltimmar för att inte störa deras hemarbete och fritidsaktiviteter.

Elevens medverkan är frivillig, hålls konfidentiellt och kommer inte på något sätt att påverka deras betyg. Elevens resultat kommer att anonymiseras i studien. Om eleven vill ta del av sitt provresultat är detta möjligt.

Jag hoppas ni samtycker till detta och att ni vill hjälpa mig med mitt projekt. Om ni samtycker till detta, signera denna blankett och låt eleven ta med sig den till skolan.

Vid eventuella frågor, kontakta mig via mejl, guskakya@student.gu.se. Tack för din tid och hänsyn.

Vänliga hälsningar,

Yaran Kakaee

Jag samtycker härmed till att mitt barns _____ resultat används i den nämnda studien.   (namn)

JA ( ) NEJ ( )

_____

Vårdnadshavares underskrift

Appendix 3: Instructions for the VLT

Instructions for vocabulary size test

For this multiple choice test, you will be required to match a word with its correct description by filling in the number in the blank space. You will be presented with 6 words at a time (on the left side) where 3 of them have to be matched with the word descriptions to the right. Notice that one chosen item cannot be used twice.

The words that you will be tested on increase in difficulty, where the 2,000-word level is the easiest (most frequent words in the English language) and the 10,000-word level and the Academic Vocabulary are the most difficult (less frequent in the English language).

If you wish to receive your test score (vocabulary size), please tick the box below.

( ) Yes, I wish to receive my test score.

Thank you and good luck!

Yaran Kakaee

Student's signature (full name)

Appendix 4: Participants' VLT scores

English 5: Vocational programme

Students	Vocabulary size
1	4162
2	4028
3	7638
4	8604
5	5014
6	7185
7	6690
8	6452
9	8219
10	7914
11	8781
12	3862
13	7576
14	7085
15	4133
16	3685
17	6004
18	9305
19	8714
20	7066
21	1976
22	4648

English 5: Theoretical programme

Students	Vocabulary size
1	7657
2	7038
3	5757
4	7319
5	5066
6	5352
7	8800
8	9119
9	7048
10	5071
11	7085
12	4047
13	6280
14	5376
15	7399
16	5133
17	6995
18	5266
19	7947
20	3519
21	6076
22	3985
23	7828
24	7881
25	4533

English 6: Theoretical programme 1

Students	Vocabulary size
1	9367
2	7204
3	5482
4	7423
5	3295
6	8038
7	5599
8	6438
9	7614
10	7085
11	5523
12	5719
13	6657
14	5528
15	5547
16	6682
17	7628
18	5952
19	4123
20	9638

English 6: Theoretical programme 2

Students	Vocabulary size
1	7328
2	4685
3	6266
4	5380
5	5461
6	6723
7	8328
8	4780
9	7838
10	3833
11	6218
12	6938
13	6490
14	6642
15	9257
16	8776
17	8657
18	7561
19	4966
20	6166
21	7009
22	6023