The Correlation Between Lexical Richness and Norwegian Lower Secondary School EFL Teachers' Assessment of Written Compositions

MARIKKEN AUENSEN

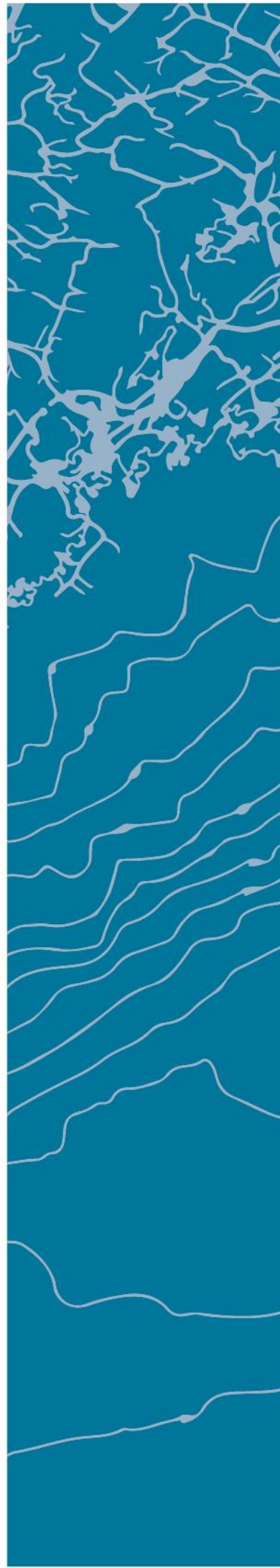SUPERVISOR
Ingrid Kristine Hasund

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Ingrid Kristine Hasund for her excellent help and guidance through the process of writing this master thesis. This thesis would not have been the same without your suggestions and advice. Secondly, I would like to thank my partner Magnus for the invaluable emotional support he has provided throughout my education and in this last year especially. Thank you for always believing in me and for boosting my confidence whenever it is low. In particular, thank you for introducing me to SPSS, without you, I would still be battling other statistical programs. A special thanks goes to my study buddy Jovana for all the fruitful discussions and conversations we have had. Thank you for all advice, revisions and distractions. This thesis and its author are much better than they would have been without your help. The last two parts of the quartet that made up our master group, Siri and Elin deserve a mention as well, as they too have been quick to help and offer great advice whenever it was needed. Lastly, I am sincerely grateful for all the encouragement, help and fun my family, friends and classmates have offered me during these last couple of years. Because of you, my years as a student will be looked back upon with joy and appreciation.

Kristiansand, May 2019

Abstract

This master thesis investigates Norwegian lower secondary EFL teachers' assessment of their students' written compositions. It focusses on the vocabulary skills these students display and aims to reveal whether the level of vocabulary each student reveals is consistent with the grade their texts receive from their teachers. Words are the building blocks of our language (Read, 2000, p. 1), one would not get far trying to communicate in a language without knowing its words. Analysing a texts lexical richness is a way of measuring the vocabulary level of the text. A text's lexical richness describes how developed the vocabulary of the text is (Cobb & Horst, 2015, p. 189). Three components of lexical richness, namely lexical sophistication, lexical diversity and lexical density, were measured in order to analyse the student texts. Previous research on this field has found correlations between students' lexical richness and the holistic quality of the composition (e.g Laufer & Nation, 1995; Roessingh, Elgie & Kover, 2015; Crossley, Salsbury, McNamara, and Jarvis, 2012). This thesis wishes to explore whether such correlations are found in the Norwegian lower secondary school as well. 27 excerpts of Norwegian 9th graders mock exams were analysed by running them through Cobb's (2019) Vocabprofile (VP Compleat, v.2), a lexical profiling program based on Laufer and Nation's (1995) RANGE. The texts had been graded by their teachers and divided into two groups; intermediate (grade 3-4) and high-level (grade 5-6). The texts' type-token ratio and lemma-token ratio, the number of advanced words, and the lexical density were found for each 300-word excerpt. These measurements represented, respectively, the lexical diversity, sophistication and density of the texts. A statistical analysis was then conducted to compare these components to the grade level of each text. The analyses revealed that there was a significant difference between the two groups' lexical sophistication and diversity, though not in lexical density. Positive correlations were also found between the lexical sophistication and diversity of the two groups, while there was no correlation between grade level and lexical density score. Three student texts were chosen for a brief qualitative analysis based on their lexical richness scores. Two of these students received the same grade, despite a large difference in lexical richness score. It was, therefore, concluded that students with a good vocabulary seem to receive higher grades on their written production, though not without exceptions. The brief qualitative analysis revealed that other factors, not surprisingly, have a part in deciding the grade as well as vocabulary. It is suggested that teachers are conscious of how they assess student texts, as these results alongside previous research have shown that a good vocabulary may affect their judgement of other linguistic features.

Contents

List of Tables and Figures

List of Abbreviations

EFL – English as a Foreign Language

ESL – English as a Second Language

L2 – Second Language

LFP – Lexical Frequency Profile

GSL – General Service List of English Words

NGSL – New General Service List

AWL – Academic Word List

NAWL – New Academic Word List

TTR – Type-token Ratio

LTR – Lemma-token Ratio

1.0 Introduction

This master thesis will investigate the correlation between the vocabulary of texts written by English as a foreign language (EFL) students in Norway and the grade their texts are given by their teachers. Before investigating this issue further, some background of the English language in Norway may be fruitful to explore.

1.1 Background

The English language has become a world language, contributing to the current globalization of our world by being the lingua franca, the language most used for international communication (Seidlhofer, 2005, p. 339). The language is taught all over the world as either a native, second or a foreign language, three expressions which vary slightly in their meaning. The definition of English as a second language (ESL) is that it is taught to learners with a different first language in a country where English is either the native language or one of the official languages. The language is taught to immigrants or students who have other native languages (McArthur, 2003). English as a foreign language (EFL), on the other hand, is taught in countries where English is not a part of everyday communication but is of interest in, for example, international communication (McArthur, 2003). By these definitions, English is considered a foreign language in Norway. Even though the Norwegian students do not live in an English-speaking environment, English is very much a part of their everyday lives, just as it is in other countries. Children and teenagers all over the world encounter English daily in music, movies, television-series, games and online interactions (Sayer & Ban, 2014, p. 325; Richards, 2015, p. 6-7). Considering how exposed a normal Norwegian teenager is to this language, one may argue that English can be seen as a second language more than a foreign one in the Norwegian setting.

English is taught in the Norwegian school system from an early age. The English training starts as early as in first grade, when the students are six years old (Utdanningsdirektoratet, 2013, p. 3). After the second year of school, the average student is expected to be able to understand and use simple sentences orally (Utdanningsdirektoratet, 2013, p. 6). By the end of their mandatory education in 10th grade, the students have had almost 600 hours of instruction in English and are supposed to be able to write coherent texts in a number of genres, express themselves orally using appropriate language, and understand and use a general vocabulary connected to different subjects (Utdanningsdirektoratet, 2013, p. 9). This entails that students have learned much about the different parts of the English language by the time they turn 16.

EFL teachers in Norway have to evaluate whether the students have reached the aims set by the curriculum. As language is complex, it cannot be measured by looking at one component alone, as is seen in the Norwegian curriculum of English which uses four main areas (Utdanningsdirektoratet, 2013, p. 2) and five basic skills to define what the students should learn (Utdanningsdirektoratet, 2013, p. 4). EFL teachers must, therefore, weigh the different components against each other in order to find a grade that represents the entire picture of the students' English proficiency. Each teacher must decide on their own which components they believe to be most important at what stage of the education, as there are no set guidelines for this, except an assessment guide for rating the written exam in tenth grade (Utdanningsdirektoratet, 2018, p. 9). This guide will be discussed further later. Because there is no specific framework with criteria on how English proficiency should be assessed, it may be interesting to take a closer look at whether some areas of the language can be used to predict the grades the teachers give their students. Though it is important to know grammar, structure and other aspects of a language in order to use it well, one would not get far without knowing the words of the language. Students' vocabulary is, therefore, an interesting area of language proficiency to take a closer look at.

Vocabulary has been a part EFL teaching in Norway from the beginning, though the teaching of vocabulary has gone through many changes in the last century and has had different roles at different periods. During the grammar-translation method at the start of the 20th century, vocabulary was taught through bilingual word-lists (Simensen, 1998, p. 28). By the mid 20's the Direct Method had made an impact on the Norwegian school system, turning the focus towards using objects, miming and other methods that did not include translation to explain new words (Simensen, 1998, p. 29). Later, science became apparent in the methods of foreign language teaching, creating a different view on vocabulary. The selection, grading, and repetition of vocabulary were approached systematically, introducing frequency lists and including it in syllabus guidelines (Simensen, 1998, p. 59). Today, vocabulary is still counted as an important part of EFL learning and is found in different parts of the curriculum. Students are, as mentioned, expected to possess a general vocabulary of different subjects when graduating from lower secondary school (Utdanningsdirektoratet, 2013, p. 9). They are also expected to develop their vocabulary and to use it in writing (Utdanningsdirektoratet, 2013, p. 2). As previously stated, knowing the words of the language is a key component of learning to understand and use it, it is impossible to master a language without knowing the words. Vocabulary has clearly been an important part of EFL teaching for a long time and is

undoubtedly still a part of the Norwegian curriculum (Utdanningsdirektoratet, 2013, p. 2, 4 & 9). Vocabulary is, therefore, taught in different ways in Norwegian EFL classrooms and is one of the areas of the language which is evaluated by the teachers when they try to pinpoint the students' level of proficiency. The students are assessed on both oral and written performance, including the productive (writing and speaking) and the receptive (reading and listening) sides of these (Utdanningsdirektoratet, 2013, p. 9). Vocabulary is found in all of these parts of the language. Based on this, it may be interesting to take a closer look at the role vocabulary has in Norwegian EFL teacher's assessment of their students' proficiency level. This topic lies at the heart of this study and is an underlying question in its aims.

1.2 Aims and Scope

This study explores the correlation between language proficiency and the vocabulary level of lower secondary school students in Norway. The written production of 27 lower secondary students in year 9 was analysed using Cobb's (2017) online Lexical Profiler, Vocabprofile, to discover their lexical proficiency. The holistic grade given by their teacher on their mock exam determined their overall language proficiency. Holistic assessment describes the overall quality the text or learner has been rated to display in his or her language production, depending on their level of skill in different areas of language such as grammar, vocabulary and structure, employing a single rating scale that provides descriptions of the different levels of performance (Read, 2000, p. 214). There may be many interesting issues to look at when comparing lexical proficiency to the teacher's holistic assessment. This study wishes to investigate the main research question:

> ➢ Is there a correlation between Norwegian lower secondary school teachers' holistic rating of ESL written production and the students' written lexical proficiency?

Lexical richness is a term which needs to be defined further in order for it to be tested. Though there are several characteristics of vocabulary and lexicality which could be used to measure lexical richness, three components have been found to define the term. These components are lexical sophistication, lexical density, and lexical diversity and will be defined and discussed in chapter 2. To explore the research question three sub questions (SQ) have been made:

> ➢ SQ1: Is there a correlation between grades and lexical sophistication?
> ➢ SQ2: Is there a correlation between grades and lexical diversity?
> ➢ SQ3: Is there a correlation between grades and lexical density?

The next paragraph presents how this thesis is structured in order to investigate these questions further.

1.3 Outline

This thesis is divided into 6 chapters, each aiming to explore the issues described above. This first chapter, introduces the topic of the study, places it in context and presents aims and scope. The second chapter presents theoretical issues and a literary review of topics like vocabulary, lexical richness, and previous research. The third and fourth chapter presents the different parts of the study, the method used with discussions about choices made and the results of the research. These chapters map the study itself, showing how it was conducted and what the results are. The fifth chapter discusses the findings and what they may imply, looking back at the theoretical issues in light of the results. The last chapter aims to tie everything together by summarizing and concluding the discussions and research questions.

2.0 Theoretical Perspectives

2.1 What is a word?

There are many ways of defining a word. Some claim that every form of a word should stand as its own word. Others may count a word with all of its inflections as a single unit, yet others would state that a word should be counted in all its derived and inflected forms (Cobb & Horst, 2015, p. 182). Lexical items such as compound nouns, idioms, and phrasal verbs complicate the definition further (Read, 2000, p. 20-21). All these different approaches make it difficult to count words in a text. Putting larger lexical items like compound nouns aside, for now, there are many ways to classify single words in order to count them. One may look at the total number of tokens in the text, which measures the number of words used, without considering repeated words. If one wishes to find the number of unique words in a text one would look at the number of types, though this measure counts each inflected form as different. Another way of counting the number of words in a text is by lemma. A lemma is the base and inflected forms of a word. This means that the word "run", and the inflected form "ran" would be counted as one (Read, 2000, p. 18). A last method of word counting is to divide the words into word families. This type of word division differs from lemma as it counts the derived form of the word as well as the inflected forms. This entails that "unknown", "known" "knowing", and "know" are counted as one word (Read, 2000, p. 18).

To illustrate, the sentence "Even if the man's message was unclear, it could be heard clearly by all the men in the crowd" would have 19 tokens, 17 types, 15 lemma, and 14 word families. As there is no correct way of counting words, each researcher must choose their preferred approach. When investigating learner's vocabulary, it is important to consider what each of these approaches entails. If one uses word families, one assumes that a learner using the word "believe" also knows the words "disbelief", "believer", and "unbelievable" as well, while a lemma count only would assume that the learner knows inflected forms like "believes" and "believed" (Cobb & Horst, 2015, p. 187). This issue raises a new question; What does it mean to know a word?

2.1.1. The aspects of knowing a word

Does knowing a word entail knowing the base with all its inflections? Is it enough to be able to recognise a word, or does one have to be able to use it correctly as well? Nation (2001, p. 24-25) discusses the difference between receptive and productive knowledge of a word. He explains that this distinction often is linked to receptive and productive use of language, where reading and listening are viewed as receptive, while writing and speaking are viewed as productive. Productive vocabulary use involves the intent to express something through speaking or writing, retrieving and producing the appropriate word (Nation 2001, p. 25). Studies have shown that learners' receptive vocabulary may be much larger than their productive (Laufer, 1998, p. 263-263). This means that it is important to know which type of vocabulary one studies, as the same person may score better on receptive vocabulary tests than productive vocabulary tests.

Productive language is found in both written and spoken form. There are differences concerning these forms which are important to note. Correct spelling is an important part of the written productive vocabulary, just as pronunciation is important in spoken productive vocabulary. The English language is filled with exceptions and inconsistencies when it comes to spelling and pronunciation (Schmitt, 2000, p. 48). The ability to spell is often connected with how language users represent the phonological structure of the language (Nation, 2001, p. 45). It has been shown that the inconsistencies of the English language create difficulties for native speakers compared to native speakers of other languages (Moseley, 1994, quoted in Nation, 2001, p. 45). Many spelling errors are often correct phonologically (Schmitt, 2000, p. 48). This may indicate that the writer knows the word but uses the wrong letters to describe the sounds the word is made of. With English being such an irregular language, spelling by sound can often create mistakes. With this in mind, is it enough to know the spoken form of

the word, or should one also be able to spell it in order to claim knowledge of it? This paper uses written corpora to investigate vocabulary, which creates the need to take a stance as to what should count as knowing a word. Spelling errors are not uncommon in second or foreign language writing. If one chooses to define knowledge of a word as knowing all of its forms, a misspelt word would be counted as unknown, even if it were spelt in a way that makes sense phonetically.

There is also another issue that needs to be considered when trying to define what knowing a word means. Some words have the same form, but different meanings depending on the context they are put in. Such words are called homonyms (or homographs if it is the written form of the word which is identical) (Nation, 2001, p. 49). One example of a homonym is the word bat, which can mean the animal or the sports equipment used in baseball. As the different uses of a homonym differ from each other, they have to be sorted into different word families (Read, 2000, p. 20). It is therefore important to note that if someone knows one meaning of a homonym, he or she does not necessarily know the other meanings.

2.2 Vocabulary

The discussion has so far revolved around the single word and what one must know about the word to be able to claim knowledge of it. One's vocabulary, or lexicon, consists of all the words one knows. As discussed above, there are many aspects of knowing a word, which can create uncertainty if one wishes to measure the size of someone's vocabulary. The issue of defining a word itself can be another problem one must consider when trying to measure vocabulary. This next section will take a closer look at what defines a vocabulary, how many words a vocabulary can consist of, as well as how to measure it.

2.2.1 What is a vocabulary?

There have been many discussions surrounding what makes up a good vocabulary, and how one can measure the quality of the vocabulary. Meara (1996) suggested two measurable dimensions. The first one was vocabulary size. He believes that vocabulary size can estimate language proficiency (Meara, 1996, p. 37). When one's vocabulary reaches a certain size, the second dimension becomes more important. This dimension refers to the lexical organisation of the vocabulary and describes the connections found between the lexical items in a vocabulary (Meara, 1996, p. 48). Other attempts of describing the different aspects of vocabulary have been made. Nation mentions vocabulary depth (how well one knows the different words) and vocabulary breadth (how many words one knows) as two aspects of vocabulary (Nation, 2001, p. 354). Cobb and Horst (2015, p. 192) use the term "lexical

diversity" to explain what Nation refers to as breadth, and Meara vocabulary size. They also mention lexical sophistication which consists of how advanced the words found in a vocabulary is (Cobb & Horst, 2015, p. 189-190). A text with a high density of advanced words would score high on tests measuring lexical sophistication.

Another issue to consider is whether a vocabulary consists of only single words, or if it should include other words as well. As previously mentioned larger lexical items like compound nouns (firefighter), phrasal verbs (move out) and idioms (a piece of cake) complicates the definition (Read, 2000, p. 20-21). The meaning of the two words "fire" and "fighter" changes somewhat when put next to each other. This could mean that such terms and phrases should be counted as different from the words they are made up of and thereby become a part of someone's vocabulary.

2.2.2 How many words does the English language consist of?

In order to measure a learner's vocabulary, it may be interesting to take a look at how many words the English language is estimated to consist of, as well as how many words it is common for native speakers to know. The English language, which is the one this study looks at, has been claimed to consist of everything from 400 000 words (Clairborne, 1983, p. 5: cited in Schmitt, 2000, p. 2-3), to over 2 million (Crystal, 1998, p. 32: cited in Schmitt, 2000, p. 2-3). These large variations are due to how the scholars choose to define a word. Goulden, Nation and Read (1990, p. 348) used word families to count the number of words in *Webster's Third New International Dictionary* (1963) and its later additions and found an estimate of 58,000 word families. Though this is an estimate of the total amount of words in the English language, it is not common to know every word in one's native language. A learner should therefore never be expected to reach a vocabulary consisting of 58,000 word families. In fact, in the same study, Goulden et al. (1990, p. 356), found that on average, an adult native speaker of English knows approximately 17,000 base words. When acquiring language in early life, native speakers learn roughly 1000 new words per year. This is not an impossible feat for learners of English, though it can be a bit unrealistic if one learns the language as a foreign rather than a second language (Nation, 2001, p. 9).

2.2.3 How many words must one know to use another language?

One important part of learning another language is to learn the words of the language. But how many words should one know? Nation (2006) investigated how many word-families one needs to know to be able to read and listen. He created multiple frequency lists based on the British National Corpus (BNC) and compared novels, newspaper articles, graded readers and

the script from a children's movie to the lists. Nation argues that a reader is able to understand a text if less than 1 in every 50 words is unfamiliar, equalling approximately 98% text coverage (Nation, 2006, p. 61). With this in mind, he found that one needs a vocabulary of between 8000-9000 words in order to read a novel (p. 71), approximately the same to read a newspaper article (p. 72), while it only takes a vocabulary of 3000 words to read a graded novel (p. 72). In order to understand the content of a children's movie, in this case *Shrek*, one's vocabulary should consist of around 7000 words, though it is stated that a movie gives substantial visual support that creates comprehension even when the watcher has a smaller vocabulary size (Nation, 2006, p. 75-76). Nation's study focused on receptive vocabulary, which is not the focus of this study. This does not make these numbers irrelevant, as comprehension is necessary for production. One cannot produce a text one does not understand. It is therefore evident that an English language learner should have a receptive vocabulary of around at least 3000 words to be able to understand simple English texts, though ideally, they should aim to have a vocabulary of at least 8000 words. As is shown here, the needed vocabulary is not close to the language's total vocabulary size. A much smaller vocabulary is needed to read authentic texts and understand the language. The question then is, how can one know which of the possible 58 000 word families one should learn in order to become proficient in a language?

2.2.4 Different Kinds of Vocabulary

Although the English language is made up of thousands of words, a learner does not need to know all of them in order to be proficient in it, as is shown in the last paragraph. Some words are more commonly used than others and may, therefore, be more essential to know. The most frequently used words in a language cover most of the words used in both spoken and written language (Nation, 2001, p. 13). This high-frequency group usually includes the 2000 most frequent words in the language and consists of both function words and content words. These words are more essential for a foreign language learner to learn, as they are what make up a large part of the used language. The words found on the other end of the scale, the low-frequency words, are not as important to learn in the beginning (Nation, 2001, p. 16). They consist of a large number of words but cover a small proportion of a given text. These words range from the words which are close to making it to the high-frequency word lists to those which are generally used rarely (Nation, 2001, p. 19-20). Some of the low-frequency words are often sorted into other types of word-lists, namely specialized vocabulary. Technical vocabulary, that is words connected to special topics and subject area, is one example of

specialized vocabulary, but there are others as well (Nation, 2001, p. 12). Some words are more common in one text type than another, for example; academic texts, fantasy, diary entries or newspaper articles. These words can be placed into lists of their own, which is just what Averil Coxhead (2000) did when she investigated the vocabulary of academic texts and gathered words which were especially frequent in such texts in her *New Academic Word List* (AWL). Technical vocabulary lists can be made in every genre, as there might be words that are particular to some genres. Genre can affect the vocabulary in more ways than this, though. This issue will be looked at next.

2.2.5 Genre and Vocabulary

There is evidence that shows that the assessment of the vocabulary in written production can be affected by the genre the text is written in. There are differences in which components of lexical richness are most apparent as a predictor of holistic quality in persuasive texts, informative texts, and stories (Olinghouse & Wilson, 2013, p. 60). In stories, lexical diversity has been found to be the greatest lexical predictor of the holistic quality of the text, alongside maturity, or sophistication. When it comes to persuasive texts, the best predictor seems to be the number of content words specific to the theme and the register (i.e. the number of words of Latin origin). Content words have also been found to be the best predictor for informative texts, as well as maturity (Olinghouse & Wilson, 2013, p. 60-61). When comparing stories written by 5th-grade native speakers of English with the other two genres, stories tend to have more diversity than informative texts and more sophistication than persuasive texts, though they lack in elaboration (i.e the number of modifiers per noun phrase) (Olinghouse & Wilson, 2013, p. 60). This means that if one wishes to compare the lexical richness of student texts, the genre should be the same, or else it is difficult to say anything about the differences in lexical richness. For example, a student writing a story may be rated higher than a student writing a persuasive text if his or her story has a better diversity, even though the persuasive text may have more elaboration and content words of Latin origin.

These last paragraphs have discussed issues which should be considered before deciding how someone's vocabulary is measured. How one can measure someone's will be addressed next.

2.2.6 Lexical Richness

The level of development of a learner's lexicon is often referred to as lexical richness (Cobb & Horst, 2015, p. 189). To measure the lexical richness of a learner, one may choose to look at different aspects of the vocabulary. Lexical originality, lexical density, lexical sophistication, and lexical diversity are all parts of a learner's lexical richness. This paper

focuses on the last three of these measurements. A text's lexical density describes the percentage of lexical words found (Laufer & Nation, 1995, p. 309). The lexical words are often the carriers of the information in a text, contrary to the function words which are mostly needed to maintain the grammatical structure of a text. Texts are therefore often considered to be more literate if they contain a high density of content words (Read, 2000, p. 196). Waller (1993), found that texts with a lexical density of over 50% (0.50) are either written by a native speaker or perceived that way by native readers (quoted in: Read, 2000, p. 207). This entails that EFL learners who are able to write texts with a high lexical density are seemingly closer to writing like a native than those with a low lexical density score. While lexical density describes the ratio of content words in a text, lexical diversity (or variation) describes how varied the vocabulary is, that is, how much the writer varies his/her vocabulary. A high proficiency writer is often assumed to have a bigger vocabulary and is therefore able to vary the language more when using it (Read, 2000, p. 200)

The last component of lexical richness, lexical sophistication, is described above. This component can be measured by sorting the most frequently used words in a language into frequency lists and comparing learners' productions with these lists. By using this measurement, one may find an estimate of how advanced or sophisticated their vocabulary is (Cobb & Horst, 2015, p. 190). This is one of the most commonly used methods of investigating lexical sophistication, and will, along with the measurements of the other components of lexical richness, be looked at in further detail next.

2.3 Measures of Lexical Richness

2.3.1 Lexical Sophistication

Of all the words existing in a language, there are some that are more frequently used than others. The words which are less common are often looked upon as more advanced, or sophisticated. Lexical sophistication refers to how advanced someone's vocabulary is. A common way of measuring lexical sophistication is, as discussed above, by making frequency lists over the most common words and sorting the words of a text into where they belong in these lists. The lexical sophistication is then found by looking at how many low-frequency words there are (Read, 2000, p. 203-204). *The General Service List of English Words* (West, 1953, cited in Nation, 2001, p. 11) is perhaps the most classic frequency list, where the 2000 most commonly used word families have been identified (Nation, 2001, p. 15). This list has later been updated by both Breznia and Gablasova (2013) and Browne, Culligan, and Phillips (2013). It is the latter of these two updates which will be used in this study, which is why this

is the one discussed further. Browne et al. created the *New General Service List* (NGSL) using data from the Cambridge English Corpus, which consists of contemporary samples of written and spoken American, British and other varieties of English (Browne et al., 2013). The list was made with the aim to update and increase the size of the GSL, as well as to try to achieve a greater text coverage with as few words as possible (Browne, 2014, p. 2). The purpose of the list was to find a core vocabulary for second language learners (Browne, 2014, p. 1). Lists such as these can be used to find the lexical sophistication of a given text, as any word not found, or found amongst the least frequent words on the list could be counted as advanced. These lists are not the only types of frequency lists that can be used to find lexical sophistication. Vocabulary can, as previously discussed, be divided into other types of frequency lists as well, such as technical and academic vocabulary. One example of this is the *Academic Word List* (AWL) which replaced Nation's original *University Word List*. These lists contain words that are typically found in academic texts (Coxhead, 2000, p. 214). Browne et al. have made an updated version of this list as well, in their *New Academic Word List* (NAWL) (Browne et al. 2013b). As Coxhead's list was made by excluding any words found in the GSL, Browne et al. wished to do the same with their NGSL, in order to create a supplement of academic words to their list as well (Browne et al., 2013b). These lists include advanced words and can be used to discover whether the students use appropriate vocabulary when writing academic texts. These two new versions will be discussed further in chapter 3.2.4.

2.3.2 Lexical Density

Lexical density measures the percentage of content words in a text. As previously mentioned, a high proportion of lexical items in a text is believed to indicate literacy (Read, 2000, p. 196). By looking at the lexical density of the different proficiency levels, one may, therefore, get an indication of whether the density of content words is a factor which can indicate the grade level of a text. To calculate lexical density, one uses this formula:

$$\text{Lexical Density} = \frac{\text{Number of lexical words}}{\text{Total number of words}}$$

(Read, 2000, p. 203).

There are some issues concerning lexical density. First of all, some lexical words convey more meaning than others. For example, the verb "to have" is often used similarly to a function word and conveys little meaning on its own, contrary to other lexical words like "laugh", "dog" or "beautiful" (Read, 2000, p. 211). High use of lexical items such as "to have" may therefore not be considered as literate by the reader, though they may score the same at a lexical density measure as a text using more meaningful words. Secondly, Schmitt argued that such a basic measure of lexicality was unable to capture the complexity of lexis in texts, implying that more sophisticated measures should be used (Schmitt, 2000, p. 75). Lastly, counting the number of lexical words in a text may not measure lexis as it does not take into account other factors such as the text's syntactic and cohesive properties (Laufer & Nation, 1995, p. 309). A text with a high ratio of lexical words could therefore also get that score because the writer uses many subordinate clauses, which would lower the number of function words used. Though it may not be a sufficient measurement on its own, lexical density can be an interesting supplement to the analysis of the texts, as it does show if there is any difference between the proficiency levels in how many lexical items they use in their texts. As lexical words often convey the meaning of a text, it could indicate whether the higher proficiency students have denser texts than lower or intermediate students.

2.3.3 Lexical Diversity

Lexical variation or diversity is one characteristic of lexical richness which will be looked at in this paper. Lexical variation measures how varied the vocabulary of a text is, aiming to discover how varied someone's language is. A proficient user with a more substantial vocabulary is more likely to avoid repetition of words by using synonyms, superordinates and other forms of variation (Read, 2000, p. 200). Type-token ratio (TTR) has been the most common way to measure lexical variation and is perhaps the most transparent one (Jarvis, 2013, p. 18). However, it is not a perfect measurement as it tends to decrease as a text grows longer. This issue arises because we have a limited number of function words in our vocabulary. The longer a text is, the more function words would be repeated (Cobb & Horst, 2015, p. 192). Another problem with the TTR is brought up by Jarvis (2013, p. 20) who claims that measures that focus on repetition of words only describe the lexical variability of a text, not the diversity, claiming that lexical diversity consists of more than what ratios like the TTR show. He defines lexical diversity to consist of six parts, where variability is only one. The last four are volume, evenness, rarity, dispersion, and disparity. The volume refers to the size of the sample, i.e. how many words the sample is made up of. Evenness describes how many times each word is repeated. A text with a TTR of 0.45 could have a few types which

are repeated many times or many types which are repeated a few times (Jarvis, 2013, p. 23). Rarity is connected to how advanced the words used are, which is also called lexical sophistication and has been seen as a separate measurement by others (Cobb & Horst, 2015, p. 189-190; Laufer & Nation, 1225, p. 309-310). Dispersion describes how the repeated words are spread throughout the text. If there are many similar types clustered together in one part of a text it may seem more redundant to the reader than if the same amount of similar types is spread more evenly through the text (Jarvis, 2013, p. 24-25). Lastly, disparity refers to the fact that similar lexical types in a text, like synonyms or paraphrasing, may also lead to the feeling of redundancy (Jarvis, 2013, p. 25). All of these elements may give a deeper view of what lexical diversity means, but all are not easy to test. Though TTR may not measure all of these elements of lexical diversity, it can tell us something about a learner's ability to vary his or her language, which is why it is included in this study. One issue which will be discussed next is whether it is more relevant to use a "Lemma-token ratio" (LTR) instead, as this study uses lemma as the definition of words.

2.3.4 TTR vs. LTR

TTR is the most common and widespread method of measuring lexical diversity. LTR has not been used as much but is arguably a better measure for lexical diversity as it uses lemma instead of types as its definition of a word (Granger & Wynne, 2000, p. 3). This study has already defined a word to be the word base with all its inflected forms, i.e. a lemma. A student who uses five different versions of the base word "to swim" (swim, swam, swum, swims, swimming) is hardly varying his or her language a lot, but a TTR would identify these five inflected forms as five different words. An LTR, on the other hand, would define these as one word. This creates a smaller number, but a more correct picture of the student's language variation and vocabulary knowledge (Granger & Wynne, 2000, p. 3-4). One may imagine comparing two texts, where one has more types, but less lemma than the other. The TTR would be highest for the first text, while the LTR would be higher for the second. This indicates that both texts have the highest lexical diversity, depending on how one defines a word. Because of this ambiguity, this study will include both measurements.

The calculations for these to measurements are:

$$TTR \; = \; \frac{\text{Number of word types}}{\text{Number of word tokens}}$$

$$\text{LTR} = \frac{\text{Number of word lemma}}{\text{Number of word tokens}}$$

(Granger & Wynne, 2000, p. 1, adapted by me).

A text without repetition would have a TTR/LTR = 1.0, the closer the score is to 1.0, the more varied the text is.

2.4 Assessing Foreign Language Written Compositions

Vocabulary is, as previously stated, a natural part of language teaching. This means that it is also a part of language assessment. In the introduction of this thesis, it was clearly shown that having a diverse vocabulary is a part of the Norwegian curriculum of English (Utdanningsdirektoratet, 2013, p. 9). Though the curriculum sets aims and guidelines for what knowledge students should have gained at certain points in their education, it does not give any suggestions or guidelines to how one should assess students, nor what to focus on when assessing for example written production. There are, however, other guidelines made which aims to provide teachers with assistance and a common set of criteria when assessing foreign language learners.

2.4.1 CEFR

The Common European Framework of Reference for Languages (CEFR) was developed by the Council of Europe as a reference guide meant to be a basis for development of curriculums, textbooks and language learning material, as well as to help teachers assess students' language proficiency (Council of Europe, 2001a, p. 1). The common reference scale is divided into six levels of proficiency, ranging from A1 – C2 (see figure 1) where the A-levels define a basic user, the B-levels an independent user and the C-levels a proficient user (Council of Europe, 2018). The scale defines descriptors for understanding (reading and listening), speaking (spoken interaction and spoken

Figure 1 CEFR levels, 2001, by The Council of Europe

production), and writing for each of the six levels (Council of Europe, 2001b). The descriptors for writing may be the most important ones to look closer at in this thesis. A basic language user is described to be able to write short, simple texts concerning personal information and
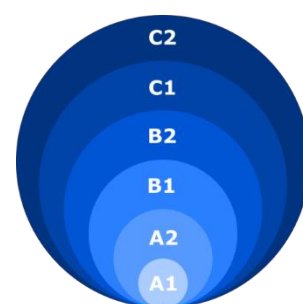
immediate needs. The independent user should write increasingly more coherent and argumentative texts, while the proficient user should be able to write articles, papers and reports using logical structures and appropriate style (Council of Europe, 2001b).

There have been attempts made to make the CEFR scale relevant to young language learners in primary and lower secondary school in Norway, particularly in two projects; the Bergen "Can-do" project and computerized national tests (Hasselgreen, 2005, p. 346). The national tests are meant to map students' proficiency in English, Norwegian and Mathematics. The English national test is based on the CEFR scale, though a modified version of it where there is a ceiling of B2 for primary school and C1 for lower secondary (Hasselgreen, 2005, p. 348-349). The work with these two projects has shown that the CEFR can be adapted to assess young learners' English proficiency, though it is argued that the descriptors used in the CEFR scale are not enough by themselves to describe learners' abilities in English (Hasselgreen, 2005, p. 351-352).

2.4.2 Norwegian English School Exam

There is no common standard for how each English teacher in Norway should evaluate his or her students' written production throughout the year. The only guide one may refer to when correcting texts is the guide meant for the written English exam in 10th grade. Though this guide is not meant for the evaluation of 8th or 9th graders, it can give an idea of how secondary school teachers in Norway assess students' written production. The English exam for 10th graders in Norway is evaluated by external raters. The guide they use is issued by the Norwegian Directory of Education and gives directions to how the raters should grade the exam papers, listing characteristics that follow each grade (Utdanningsdirektoratet, 2018, p. 9, see appendix 1). The grades represent a holistic assessment of texts and range from 1-6, where 1 is a fail grade. The rest of the grades are divided into three groups, high level of proficiency (5-6), medium level (3-4) and low level of proficiency (2) (Utdanningsdirektoratet, 2018, p. 9). These grades are the same students receive on other compositions they produce. There are four competence areas that are described in the guide; Content, subject knowledge and the use of sources, textual structure, language, and formal skills. Within these areas, different characteristics are described according to the different levels. One example is:

The exam answer

- uses central patterns of spelling
- uses central patterns of conjugation

- uses mostly an English sentence structure
- cites used sources in a mostly testable way

(Utdanningsdirektoratet, 2018, p. 9, my translation, see appendix 1)

This extract is from the formal skills area and shows the characteristics of an intermediate text. As one can see, there are many aspects of written competence the teacher is supposed to consider when evaluating the exam. The expected proficiency for the intermediate learners is that the exam is satisfying for their level and age, though not excellent or extraordinary. These expectations are found at the highest grades, where the exam should be well written and relevant to the situation and audience (Utdanningsdirektoratet, 2018, p. 9). Vocabulary is included as one of these criteria. It is found under "language":

Low level: Has a simple vocabulary of some subjects.
Intermediate level: Has a vocabulary that covers known subjects.
High level: Has a general vocabulary of different subjects.

(Utdanningsdirektoratet, 2018, p. 9, my translation, see appendix 1).

The difference between the highest and the lowest grade when it comes to vocabulary seems to be in both the size and variation of the vocabulary. Variation in this sense refers to the extent of vocabulary knowledge the student has in a varied landscape of subjects, not necessarily the type-token ratio of his or her text. These criteria are made in order to give the external raters of the exam a common set of criteria to use when grading each exam so that every student is assessed according to the same scale. There are some issues concerning whether two different raters would give the same exam the same grade. These issues will be discussed next.

2.4.3 Rater Reliability

In her doctorate Garshol (2019) argued that "… as grading is at best a partially subjective matter, it is unsuitable for a quantitative analysis of the students' linguistic development" (Garshol, 2019, p. 126). Even when provided with guidelines as the ones described above, there are differences in how teachers assess student texts. Although they follow the same criteria, there is still room for the raters' subjective judgement. One rater may emphasize structure as a descriptor of text quality, while another may see grammar as more important when rating a text. This is one of the major issues of holistic assessment of texts. The holistic score given a text represents a compromise between competing considerations such as

grammar, structure or lexical errors (Read, 2000, p. 214). These kinds of compromises cause differences between teachers' and raters' evaluation and grading of students' proficiency (Skar & Jølle, 2017, p. 15). One teacher may decide to emphasize the vocabulary, while another finds correct grammar to be more important. A text with good vocabulary and many grammatical mistakes may therefore be rated higher by the first teacher than the second. This is called inter-rater reliability. The inter-rater reliability is often found to be low, meaning that there are large variations between how strictly raters assess written production (Skar & Jølle, 2017, p. 14; Midtbø, Rossow & Sagbakken, 2018, p. 22). The intra-rater reliability, on the other hand, is often found to be consistent (Skar & Jølle, 2017, p. 14; Modtbø, Rossow & Sagbakke, 218, p. 22). This means that even though teachers and other raters often assess written production differently between each other, the same teacher does not give one text a 3 and then other texts of the same level a 5. Using grades as a tool for dividing groups by level may, therefore, be unreliable, as the strictness of the rating varies from one teacher to another. With this in mind, it may be interesting to take a closer look at what features of written production have been found to have a large impact on the assessed holistic quality of a text.

2.4.4 The Importance of Lexical Errors and Text Length in the Assessment of Written Production

A holistic rating of written production assesses more than just the lexical qualities of a text. As one may read from the exam assessment guide and the CEFR scale; content, grammar, and structure are all criteria that are taken into account when rating (Utdanningsdirektoratet, 2018, p. 9; Council of Europe (3), 2001). One element has been shown to be a reliable predictor of the quality of L2 texts, namely lexical errors (Bestgen & Granger, 2011, p. 249; Engberg, 1995, p. 149). Lexical errors include using the wrong word, borrowings from one's native language, coinage of native words into L2 writing, and spelling errors (Llach, 2007, p. 8). Spelling errors are often the largest group of lexical errors found in L2 texts (Llach, 2007, p. 9). Though not all research has found that lexical errors have a significant correlation with the holistic rating of a text (Llach, 2007, p. 15), many have (Bestgen & Granger, 2011, p. 249; Engberg, 1995, p.149). In their study of how spelling errors affect both the overall rating of a text and the teachers' comments on it, Vögelin, Jansen, Keller & Möller (2018, p. 9) discovered that not only did the texts with the most spelling errors receive a lower holistic score, the teachers' comments on these texts were more negative towards both the vocabulary and other grammatical mistakes as well, indicating what they call a halo effect. This means that lexical errors such as spelling errors could affect the holistic score of written production.

As discussed above, being able to spell words correctly is part of knowing the written form of a word. This creates the question of how spelling errors and other lexical errors should be treated in learner texts. If the percentage of lexical errors is an important part of how a text is rated, should potential errors be removed, or corrected?

Text length is another factor that has been shown to affect rating scores. Longer texts are often rated higher than shorter ones, with only a few exceptions (Jarvis, Grant, Bikowski & Ferris, 2003, p. 400). This means that longer texts are not necessarily just less concise than shorter ones, they are often viewed as better. This may also have implications for the data gathered in this study, as there were no criteria of maximum length for the texts used. Both of these issues will be discussed later. These two components of assessment have an impact on the quality of the text. This study looks at whether vocabulary has any impact as well. As the last topic in this presentation of theoretical perspectives, previous research on this field will be presented.

2.4 Previous research Concerning Lexical Richness and Language Proficiency

Lexical richness and its effect on the holistic quality of both written production and overall proficiency has been studied by others previously. Some of these studies will be presented next, in order to explore what others have found before.

2.4.1 Lexical Richness and the Quality of Written Production

Roessingh, Elgie & Kover (2015, p. 71) investigated the lexical richness in the written production of 3rd-grade native speakers of English and compared it the holistic rating made by two of the authors (who were both experienced teachers). By running the texts through Vocabprofile for kids, a version of the program aimed at younger speakers and learners (Cobb, 2019), they gained a lexical profile describing how the words in each text distributed across 10 frequency bands (Roessingh et al., 2015, p. 71-72). The results revealed a correlation between the holistic grade of the text and the amount of low-frequency words used (Roessingh et al., 2015, p. 79). This implies that the students who were rated as high proficiency writers had a more sophisticated language than the other students.

These results are similar to the results Laufer and Nation (1995) gained in their study of foreign learners of English who attended a university in New Zealand. They developed the original version of the aforementioned program, Vocabprofile, and wished to investigate whether it was reliable as a source for testing vocabulary (Laufer & Nation, 1995, p. 314). They analysed the written production of the students, discovering that the students ranked as

advanced learners used more academic and off-list words (i.e words not found on the three frequency lists used by the program) and less from the high-frequency bands than the intermediate and low-level learners (Laufer & Nation, 1995, p. 316). This indicates that high proficiency learners have a more advanced vocabulary than low and intermediate proficiency learners.

Another study defined three dimensions of lexical proficiency, quantity or breadth, quality or connectivity amongst lexical items in the mental lexicon and metacognitive awareness (Zareva, Schwanenflugel & Nikolova, 2005, p. 572). They tested the correlation between the lexical proficiency and overall proficiency of English of L2 students at an American university and discovered that there was a positive correlation between the level of proficiency and the lexical quantity and quality of the vocabulary and the overall lexical proficiency (Zareva et al., 2005, p. 584-588). The metacognitive awareness, on the other hand, did not differ between the groups, meaning that the relationship between students self-reported degree of vocabulary knowledge and their actual knowledge was equal in all groups (Zareva et al., 2005, p. 590).

Crossley, Salsbury, McNamara, and Jarvis (2012, p. 572) sought to create a model of lexical proficiency, by finding lexical indices that could predict the variance of human ratings of lexical proficiency in written production. By using the computational program Coh-Metrix that analyses texts and gives scores for different indices, the scholars were able to identify three lexical indices which predicted 44% of the variance in human ratings. These indices were lexical diversity, word hypernymy, and word frequency, which are all elements of vocabulary breadth and depth (Crossley et al., 2010, p. 572). This reveals that these three components of lexical richness explain almost half of the differences between the holistic rating of texts.

Aiming to discover whether highly proficient L2 writers have a greater linguistic sophistication than less proficient L2 writers, Crossley & McNamara (2012, p. 119) used their aforementioned program, Coh-Metrix to analyse the lexical elements such as diversity, word familiarity and amount of low-frequent words. They refer to linguistic sophistication as "the production of infrequent and more complex linguistic features" (Crossley & McNamara, 2012, p. 119). Their results indicated that highly proficient L2 writers used more low-frequency words, had a greater lexical diversity, and scored better on other lexical devices as well (Crossley & McNamara, 2012, p. 130-131). They conclude by stating that high-

proficiency learners display a more sophisticated language than lower-proficiency learners (Crossley & McNamara, 2012, p. 131).

Engberg is another scholar who aimed to explore the correlation between lexical richness and the holistic quality of written compositions (Engberg, 1995, p. 143). She investigated the compositions of University students who had English as their second language. Their essays were assessed by English teachers and analysed in order to establish the lexical richness of each text (Engberg, 1995, p. 145). She calculated the TTR, with and without lexical errors, the lexical density and the error percentage, which has been discussed before (Engberg, 1995, p. 145-147). The results gave a significant positive correlation between the holistic quality and the TTR, both with and without errors (Engberg, 1995, p. 149-150). Lexical density did not correlate with the holistic quality, causing the author to conclude that "simply piling up lexical words did not have any effect the quality scores assigned" (Engberg, 1995, p. 148).

Interestingly, research has found that lexical components such as lexical sophistication and diversity have an impact on how teachers perceive other aspects of the text as well, such as grammar (Vögelin, Jansen, Keller & Möller, 2018, p. 9). This was discovered by altering the lexical sophistication and diversity of different essays, before letting teacher-graduates grade and comment on them (Vögelin et al. 2018, p. 4-5). The comments show that the essays where the vocabulary was better got more positive comments concerning not only vocabulary but grammar as well (Vögelin et al., 2018, p. 9). In a later study aiming to explore the influence of lexical features, such as sophistication and diversity, on teachers' holistic assessment, the scholars found some of the same results (Vögelin, Jansen, Keller, Macht & Möller, 2019, p. 51). Texts which were altered to have a greater lexical sophistication and diversity were generally viewed as better and received more positive comments concerning vocabulary, grammar, and frame of essay (the introduction and conclusion) (Vögelin et al., 2019, p. 59). These studies indicate that a halo-effect has occurred, as the lexical sophistication and diversity of the text affected the teachers' assessments of other textual features outside of vocabulary (Vögelin et al, 2018, p. 9; Vögelin et al., 2019, p. 59)

2.4.2 Vocabulary Development

Studies of vocabulary development can be interesting to look briefly at, as they do tell us something about the differences in lexical richness from one age group to another, as well as how the lexical richness changes when students become more proficient. They are also interesting as their methods are comparable to the methods used to investigate the difference between high and intermediate learners' vocabulary.

A study from Norway investigated, amongst other issues, the development of Norwegian primary school children's productive vocabulary skills from 5th to 7th grade. The study analysed texts from each year using Cobb's Vocabprofile for kids (Cobb, 2019) and found signs which pointed to the development of productive vocabulary, though not an even development (Langeland, 2012, p. 142).

Other studies have explored whether there is an increase in lexical richness in written production as English as a second language learners develop their English skills (Horst & Collins, 2006, p. 88). By analysing lexical components of French-speaking 6th-grade students' texts from one year to another, Horst & Collins found that there was a change in the lexical richness, though not in the way one may assume. From writing short, simple texts filled with French words where their English vocabulary faltered, French-speaking 6th-grade students in Quebec showed a decrease in the use of French words, and an increase in the diversity of words used within the K1 list (the 1000 most commonly used word families). They did not, however, increase the use of advanced words, as the scholars expected (Horst & Collins, 2006, p. 100).

Another study located in Norway analysed texts from a group of 7th graders and compared them to texts written by 10th -graders assessed to be at the same proficiency level as measured by the CEFR scale (Helness, 2012, p. 146). The TTR, lexical density and text length were measured on all the texts in each grade put together (Helness, 2012, p. 149-150). The results revealed that the lexical diversity as measured by TTR was higher in the 10th-grade texts than in the 7th-grade texts, the lexical density was not significantly higher in either grade. The 7th-graders wrote longer texts, which the author argues may have been a contributing factor to the lower TTR, as this measure is highly affected by text length (Helness, 2012, p. 153-154). These results suggest that the lexical richness is relatively the same in groups who are rated as being on the same level of proficiency in the CEFR scale, regardless of differences in age.

Ishikawa (2015, p. 204) investigated how certain lexical components developed as learners' L2 proficiency developed. By measuring lexical diversity, density, complexity (number of letters per word), fundamentality (high-frequency words), and noun orientation (noun/verb ratio) he discovered that the lexical diversity, as well as the learners' lexical fundamentality decreased before increasing when the proficiency developed, while lexical density remained nearly unchanged. Lexical complexity increased steadily while noun orientation decreased (Ishikawa, 2015, p. 209). This suggests that the components of lexical richness develop differently.

3.0 Method

The aim of this study is to explore the role of lexical richness in EFL teachers' holistic assessment of written production. In order to investigate this, a quantitative approach was made in order to analyse the different components of lexical richness and compare them to the holistic grades given by teachers. By choosing to conduct a quantitative analysis of the data, one wishes to discover how much or how many there is/are of something. The data is quantifiable, meaning that it can be put into numbers (Rasinger, 2013, p. 10). This way of conducting research is deductive, one creates a hypothesis on the basis on already known theory and tries to prove or disprove it (Rasinger, 2013, p. 11). Qualitative research methods, on the other hand, are inductive, meaning that theory is derived from the results of the research. Qualitative analyses aim to the discover how something is (Rasinger, 2013, p. 10-11). Three measures of lexical richness will be measured; lexical sophistication, lexical density and lexical diversity. In addition, the number of spelling errors will be counted and analysed. Though the main analyses are quantitative, a brief qualitative approach will be taken on a few texts, in order to get a better insight into how vocabulary is assessed in these texts.

The reliability and validity of the research methods are discussed throughout this section. A study which has a strong reliability can be reproduced by others, not depending on time, place, mood or experience, and still get the same, or approximately the same results (Rasinger, 2013, p. 28). In other words, "reliability refers to a method repeatedly and consistently measuring whatever it is supposed to measure" (Rasinger, 2013, p. 28). The validity of a method, on the other hand, refers to whether the method measures what it is supposed to measure (Rasinger, 2013, p. 29).

3.1 Defining a word

This study chooses to define a word as the base and its inflections, that is a lemma. This means, as previously mentioned, that one assumes that a student who writes for example "believe" knows "believes" and "believed", but not "belief" or "unbelievable" as he or she would if a word was defined as its word family. There are several reasons for this choice, one of them being that the version of the VocabProfile (Cobb, 2019) program (which will be discussed later) chosen used lemma (Cobb, 2019). Lemma is also used because it is the definition of a word which was found to be the closest to what a student knows. It was decided that it was to assume too much knowledge of a student to say that they know every inflected and derived form of a word. For example, to say that a student who uses "part"

knows "particle" as well may be to give them too much credit. On the other hand, one may also have used types, but this was deemed as giving the students too little credit for their word knowledge.

3.2 The Lexical Frequency Profile

The testing of the three aspects of lexical richness, sophistication, density and diversity, were done in VocabProfile (v. 2; Cobb, 2019) a program deriving from Laufer and Nation's RANGE computer program (1995), which has been adapted to an internet software by Tom Cobb on his Lextutor website (Cobb, 2019). This tool processes texts resulting in a lexical frequency profile. The profile consists of measurements like TTR and lexical density. More importantly, it sorts the words of the texts into frequency lists, showing how many lemmas the text has within each list. There are several options as to which frequency lists the program can use. The original Vocabprofile (VP-Classic) uses the GSL (West, 1953) and the AWL (Coxhead, 2000) in its analysis. These two word lists are out-dated and have been renewed in later years, and as there were other, more updated options, this version of the programme was deemed unsuitable. A second version is VP-Kids, a version that uses lists based on a corpus made of texts from American children. The list is aimed at grades 0-3 and was therefore found to be some levels below the lower secondary students tested in this study (Cobb, 2019). The last version of the program is the VP-Complete, where one can choose between several different word lists, such as COCA (Corpus of Contemporary American English), BNC (British National Corpus) and the *New General Service List* (NGSL). The NGSL is, as previously mentioned, an updated version of the GSL. This version uses lemma instead of word families and includes modern words like "internet" and "TV" (Browne, 2014, p. 2-3). The list is also made with the intention of creating a core vocabulary for English as a second language learners (Browne, 2014, p. 1). This means that it is very suitable for analysing the lower secondary texts used in this study. The choice fell on using the VP-Complete version of the program with the NGSL and NAWL. These are both updated versions of two classic word lists (Browne et al., 2013a). This version of Vocabprofile uses, as previously stated, lemma and has four word lists; the NGSL1 level (1-1000 most frequent lemmas), NGSL2 (1001-2000), NGSL3 (2001-2802) and NAWL (963 lemmas). Words that are not found in any of these lists are gathered in as "off-list" words. The texts used in this study are all stories. One may not expect there to be many academic words used in stories, therefore the number of words from the NAWL list is not expected to be high in any of the students. It is still

included, as it can be interesting to see whether the more proficient learners use more academic words than their peers.

There are many benefits with using the Vocabprofile software. It is accessible, easy to use and analyses many parts of the text. There are, however, some weaknesses to the program. First of all, the program does not recognize homonyms, meaning that an important dimension of word knowledge disappears. This is not considered a major issue, as the sample sizes are small (300 words) and the chance of a student using both meanings of a homonym is unlikely. Second, the software is unable to distinguish spelling errors from words not found on the frequency lists, creating a possible miscount in the number of low-frequency words. Lastly, although it recognizes some proper nouns and sorts them into the high-frequency list, it does not recognize all, and usually only English ones. A student who uses Norwegian names in his or her story would, therefore, get the wrong number of off-list words.

The data found in the Lexical Frequency Profiler reveals how many words the texts have within the five frequency lists. As previous research has shown that the difference between high and intermediate students is found in the number of advanced or low-frequency words they use, not the number of high-frequency words (Laufer & Nation, 1995, p. 318), the only words counted in this study was NGSL3, NAWL, and Off-list. This means that the results will show whether the amount of advanced words used in the text has any correlation with the holistic grade given by the teacher. The program will also find the TTR and lexical density scores of each student (Cobb, 2019). The LTR, on the other hand, has to be calculated in Excel, using the lemma counts found in Vocabprofile. Some issues arise from using these measurements. First of all, the validity of the text may be weakened by the use of TTR and LTR as the measure of lexical diversity. As previously discussed, Jarvis (2013, p. 20) argues that TTR only measures a part of lexical diversity, what he calls lexical variability. Secondly, the frequency lists define words such as "two", "girlfriend", and "birthday" as off-list and NGSL3 words. This may not be words which are typically considered advanced words by teachers or students, though they are kept, as they do represent words not found amongst the 2000 most frequent lemma.

3.2.2 Qualitative analysis of Student Texts

In addition to the quantitative analyses made of the data collected from Vocabprofile (Cobb, 2019), a few texts will be selected for a more qualitative analysis to exemplify and enlighten the results. These analyses will not be particularly deep, as the quantitative data is the main focus of this study. The focus of the qualitative analyses will be on what kind of advanced

words the student uses, what other aspects may have affected the holistic grade and what the teacher comments reveal. The students with the highest and lowest lexical richness score will be examined, alongside any students whose text stand out in one way or another.

3.3 Data Collection

The data is a part of a corpus that is under development by a research group at the University of Agder called "Elevspråk i Transitt", translated "Pupil's Language in Transit" (ESIT). ESIT is a part of a bigger project called "Tracking Written Learner Language" (TRAWL). The corpus is currently being developed, but several texts have already been gathered and made ready for the finished corpus. The aim of TRAWL is to create an authentic, longitudinal corpus of students' written production. The corpus is to be used in research but will also be open for teachers and teacher educators to use (Dirdal, Drange, Graedler, Guldal, Hasund, Nace, & Rørvik, 2017).

3.4 The Tasks

At the end of each semester, the students have a mock exam in English. It was from one of these tests the texts used in this study were gathered. The mock exams are a part of the students normal testing in lower secondary school, meaning that the tests were not crafted or executed by researchers but by their teachers. This creates an authentic situation, where the students try their best to achieve a good result, contrary to what might have happened if the test was something they did for a research project, where their results would be unimportant for their performance in class. However, it may weaken the reliability of the method some, as another aiming to replicate the study may end up with different tasks then the ones these students had. The mock exam consists of 3 tasks, two short-answer tasks that all students have to answer, and one long answer where the students can choose between a number of different tasks (the mock exam with its tasks can be found in appendix 2). This means that the students may have chosen different genres when writing the last task. As the texts are gathered from two different classes and schools, the mock exams were different in each class, but are built similarly. There are some issues which should be considered because of this. First of all, the students were allowed to use a dictionary, which creates the issue of whether they know the words they use or if they have found them in the dictionary. They also had a booklet with information and texts they could use to write. There is, therefore, a possibility that some words have been found there. This is a drawback and could affect the end results negatively, as the students may use words they find in the dictionary or booklet to make their vocabulary become more advanced and more varied, without actually knowing these words. This affects

the validity and reliability of the results. Unless the test had been made and administered by researchers, these kinds of weaknesses are hard to avoid, as dictionaries are common to use in mock exams. The fact that the texts are gathered from an authentic testing situation may in fact increase the validity of the study, as it gives a real picture of the students' performance and the teachers' grading. In addition, if the results reveal a significant difference between the high-level and intermediate students, the fact that they were allowed to use a dictionary would only suggest that, at the most, the high-level students were better at utilizing the available aids than the intermediate students. Lastly, the texts used in the analyses are only an excerpt taken from the last task. This means that some students may have gotten a lower or higher grade than the long answer suggests, depending on how they answered the first two tasks, which weakens both the validity and the reliability. Though this is an issue, the last task is also the longest one, and the one where the students are allowed to write freely, which is why this part was chosen to represent the rest of the test.

3.5 Participants

The participants in this study were 30 9th grade students from two different lower secondary schools in Kristiansand. They had all agreed on participating in the ESIT project. It is voluntary to participate in the project, which means that there only are texts from a portion of the classes. This could mean that the participants are not representative of the entire class. For example, one may imagine that it is more difficult to recruit students who have a low proficiency and subsequently get low grades, than students who are at an intermediate or high level of proficiency. This turned out to be the reality, as none of the students from these two schools had been given the lowest passing grade.

Before the texts could be used, they had to be checked for genre. As there are differences in which components of lexical richness raters put weight on depending on genre (Olinghouse & Wilson, 2013, p. 60-61) the texts had to be in the same category of genre. The mock exam allows them to choose between different tasks. In order to keep as many texts as possible, the definition of genre used could not be too narrow. By choosing texts that were fictional, though set in the real world, 27 of the original 30 texts could be kept, the three last having written argumentative or factual texts.

The students were divided by grade, following the system set by the Norwegian Directory of Education in their Exam Guide (Utdanningsdirektoratet, 2018, p. 9). There is a total of six grades one may receive, ranging from 1-6, where 6 is the best and 1 is a fail. The Norwegian

Directory of Education divides these grades into three levels of competence; high (5-6), intermediate (3-4) and low (2) (Utdanningsdirektoratet, 2018, p. 9). The texts are divided by the grade the class' English teacher gave them. This means that half the texts were graded by one teacher, and the rest by another. This increases the validity of the method, as they represent a larger community than if only one class from one school had participated. One issue which occurs when using grades as a measure of the students' proficiency is that, as previously discussed, holistic assessment is a compromise of the strengths and weaknesses a student displays in his or her production and that, at best, the grade is a partially subjective score of the student's proficiency. This entails that a student's vocabulary can be better or worse than what the holistic grade suggests, depending on the level of the other components assessed by each teacher (Garshol, 20129, p. 126; Read, 2000, p. 214). This may affect both the validity and reliability of this study. However, this study's focus is on whether there is a correlation between the teachers holistic grading and the students' vocabulary as measured by the three components of lexical richness, which means that this issue is a part of what the study aims to investigate.

The texts are not graded on a pure scale of 1-6. The teachers use plus, minus, and half-grades as a method of showing whether the grade is a strong or a weak one. As there were no texts rated as a 2 or 1, the students were divided into two groups, one intermediate group for those who got a grade between 3/2 and 4/5, and one high proficiency group for those who received grades ranging from 5/4 to 6.

Table 1: Participants divided by grade

Grade	Number of Students
3-4	20
5-6	7

As can be seen from table 1, there is not an even distribution of grades. Most students fall into the intermediate level, which is not surprising, considering how this is the "average" grade. The high-level students are supposed to be exceptional, which may be why so few have received such a grade (Utdanningsdirektoratet, 2018, p. 8). Naturally, most students will be average, though this does create some challenges in this study.

3.6 Data Processing

Before running the texts through Vocabprofile, they had to be read through in order to avoid any flaws in the data. Vocabprofile has some weaknesses, one of which is that it does not differ between non-existing words and words not found on the frequency lists (Cobb, 2019). Spelling errors were, therefore, an important issue to address. The choice was made to correct any word which was spelt wrong but was used correctly and made sense phonetically. For example, student P60201 wrote; "it is the fear of school that's Pusch me down". The word "push" is spelt wrong but used correctly. As previously mentioned, English is a language with great inconsistencies when it comes to spelling and pronunciation, which creates difficulties especially for non-native speakers (Schmitt, 2000, p. 48; Nation, 2001, p. 45). Although knowing a word's written form is an important part of knowing the word, the students do show that the word is not unfamiliar to them, that they know it to some degree. In comparison, student P60207 wrote "summing" instead of "buzzing" when writing "the bees are summing". In the first text, "pusch" was corrected, while "summing" was removed completely as the student clearly did not know the correct word or the meaning of "summing". All the lexical errors each student made was noted in a separate table (see appendix 3). In addition, the total number of mistakes each group made was counted, and the means were found. This was done so that the eventual results could be discussed in the light of the number of lexical errors the groups made.

The controlling of text length is another issue that the measurements done by Vocabprofile does not take into account in some of its measures. Difference in text length is something which affects especially the type-token ratio of a text. The longer the text is, the more repetition of function words there is (Cobb & Horst, 2015, p. 192). To try to control for this, the texts were all cut down to 300 words (the length of the shortest text) starting from the beginning of each. This removes the issue of the type-token ratio becoming invalid but removes possibly valuable data from some of the samples. Cutting the texts from the beginning is the most natural procedure, though it is debatable whether the first paragraphs of long texts are representative of the rest of the text. Research similar to this one has done much of the same when it comes to spelling errors (Laufer & Nation, 1995, p. 315; Horst & Collins, 2006, p. 90; Helness, 2012, p. 149; Langeland, 2012, p. 134) and text length (Laufer & Nation, 1995, p. 315).

A last revision made was to remove all proper names. Vocabprofile has a function which sorts all proper names into the high-frequency list. The problem is that foreign names,

months, weekdays and holidays are either not recognized, or sorted differently. For example, Saturday is considered a low-frequency word (found in the NGSL3 list), while Monday and Wednesday are considered to be high-frequency words (found in the NGSL1 list) (Cobb, 2019). Though they may be used more or less frequently by native speakers and writers, it would be strange for an EFL learner to learn only half the weekdays. As they are all proper names, the decision was made to remove any proper noun, be it a weekday, a holiday, a language or a city. There have been different practices concerning proper names in student texts. Some choose to remove them as I have (Laufer & Nation, 1995, p. 315; Horst & Collins, 2006; Helness, 2012, p. 149) while others choose to re-categorize them to frequency list 1 (Langeland, 2012, p. 134), as Vocabprofile allows.

The data from one of the schools were also used in another master thesis, written by Jovana Dašić (2019). Her study looks at the connection between gaming and vocabulary development as well as between gaming and grades. As we were going to do the same analyses on these texts, we were able to work together to clear the texts of any lexical errors or proper nouns. This means that there were two sets of eyes that read through each of the texts in order to spot spelling errors and proper nouns. Though only half of the texts were checked by both of us, I had the opportunity to discuss possible issues in the rest of the texts with her as well. Working together on this meant that the choices made were thoroughly discussed. Spelling errors were corrected, and proper nouns removed before the texts were shortened to 300 words. The texts were then run individually through Vocabprofile in order to find the type-token ratio, lexical density, and frequency lists. The choices made may have affected the reliability and validity of the study. Some errors may have been made when processing the texts, though the fact that two people searched through at least half of the texts will, hopefully, have excluded most of these mistakes. Nevertheless, this may have weakened the reliability some. Someone aiming to replicate this study would have to make the exact same choices in order to receive the results. Even if they do, some differences may occur, as most of the choices made when processing the texts were subjective. This affects the validity of the text.

3.7 Ethical Considerations

The texts used in this study were all gathered from the ESIT corpora, as previously mentioned. As the students are under the age of 15, they have all gotten parental consent to the participation of this project (NSD, 2019). No student or teacher names were left on any of the texts, other personal information which could be used to identify students had been removed according to the Norwegian national guidelines as well (NSD, 2019). Any

information connecting the student names to codes are available to authorized personnel only. The participation in the project is optional for all students and teachers and the texts gathered are all from tests the students would have to take as a part of their education. This means that no additional tests were given participants, avoiding any potential extra stress this may have caused, though some students may feel pressured to perform due to the knowledge that their texts may be examined by researchers.

3.8 Statistical Analysis

The data found after running the texts through Vocabprofile were sorted according to the grade. The data were separated into the three components of lexical richness; lexical diversity (LTR and TTR), lexical density, and lexical sophistication (frequency lists). The statistical software program SPSS for Windows (v. 25; SPSS Inc, Chicago II, USA) was used to test for normality, significance, and correlation. The following measures were then applied in order to compare the results.

3.8.1 Central Tendency

The mean score of each group was calculated for each of the lexical components and their measurements. Comparing the numbers of the groups without finding the mean would not make sense due to the large difference in group size. The results would, in that case, be faulty and incomparable (Rasinger, 2013, p. 120-121). The mean summarizes the data and gives one the opportunity to compare them, though some weaknesses should be noted. The mean does not reveal how large the difference between the highest and lowest score of the groups is, or how dispersed they are (Rasinger, 2013, p. 132). It shows the central tendency of the data but can be affected by extreme values and give an inaccurate representation of the data (Rasinger, 2013, p. 124-125). Because of these flaws, other measures of central tendencies were applied, the median being one of them. The median shows whether the mean score is also the one in the middle of the values when listed from lowest to highest. If the median and mean are approximately the same, the same amount of values are found on each side of the mean. This reveals that the data is evenly dispersed (Rasinger, 2013, p. 127). The median and mean are not always the same, which can be an indicator of large differences in the sample values. The mean and median are both measures of central tendency, while the range and standard deviation show how dispersed the data is.

3.8.2 Dispersion

As described above, the mean is largely affected by extreme values. To check whether there are large differences between the data in the two groups, measures of dispersion were applied.

By finding the highest and lowest value of each dataset one is able to calculate the range of the data. If the range is large, the data may not be evenly dispersed which may affect the mean (Rasinger, 2013, p. 133-134). The standard deviation was also found. The standard deviation reveals how dispersed the data is. The smaller it is in relation to the mean, the less dispersed the data is (Rasinger, 2013, p. 136). The standard deviation can also be used to read whether the data is normally distributed.

3.8.3 Normal Distribution

A dataset can have different kinds of distributions. Data that show a normal distribution will have many observations around the mean, and fewer below and above. The curve of a normally distributed dataset will be bell-shaped, like this:
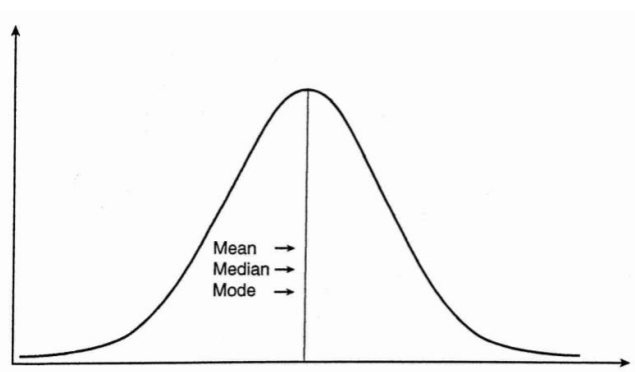
Figure 2 Normal Distribution. From Quantitative Research in Linguistics by S. M. Rasinger, 2013, p. 139.

In order to determine whether one's data is distributed normally one may look at the median and standard deviation of the data, as well as its histogram. If the histogram looks normally distributed and the median is the same as the mean, it is a good indicator of normally distributed data. If in addition 68% of our data is within one standard deviation (i.e. the mean +/- 1 standard deviation) of the mean, and 95% are within two standard deviations, it is likely that the data is normally distributed. All of these measures of normality were applied by using SPSS. The results showed that the TTR and LTR were normally distributed as the histogram resembled the bell-shaped curve, and the mean and median were approximately equal. The lexical density and LFP were a bit more uncertain, as the histogram revealed high values in both ends of the graph. On the other hand, the mean and median were approximately the same, which indicated that they were normally distributed. This may be because of the small sample size, as the problems were mostly related to the high-level group which only has 7 texts. Because everything except the histogram was a good match for normal distribution, the lexical richness data were all deemed to be normally distributed. The number of lexical errors

were not found to be normally distributed as the mean and median of the high-level group were not the same and the histogram was skewed. As one of the groups was not normally distributed, this measure will be counted as non-parametric. This is important to know as it decides which tests of significance and correlation should be used.

3.8.4 Statistical Significance and Correlation

To find out whether any results are significant, that is, if the differences between the two groups are big enough for it to be of significance, hypothesis tests will be executed on the data. An independent t-test will be applied to all the variables except the lexical errors, as they are normally distributed (Rasinger, 2013, p. 196-197 & 230). The independent t-test allows one to compare the arithmetic means of two different groups (Rasinger, 2013, p. 196). The confidence level is set to $p = 0.05$, meaning that there is a 5% chance of the result being wrong. Any value below the p-value reveals significance (Rasinger, 2013, p. 198). A Mann Whitney U test was conducted on the lexical errors. Mann Whitney U is the non-parametric equivalent of the independent t-test (Rasinger, 2013, p. 230).

The correlation between the grades (in this case high or intermediate grades) will be checked in order to see if there are any associations between the students' grade level and the lexical components. Pearson correlation reveals a coefficient r which lies between 1 and -1, where 1 is a perfect positive correlation and -1 is a perfect negative one. This means that if the result of the correlation test is $r > 0$ the two values compared increases in a straight line. A result were $r < 0$ reveals data were one variable decreases as the other increases (Rasinger, 2013, p. 163). In addition, a p-value is found which indicates how significant the correlation is. The smaller the p-value the more significant the correlation (Rasinger, 2013, p. 166). It is important to note that the size of the correlation coefficient is strongly affected by the sample size (Robson, 2002, p. 424). For the lexical errors, a non-parametric correlation test is needed. The Spearman Rank correlation test was, therefore, applied to these data (Rasinger, 2013, p. 221).

4.0 Results

This chapter will take a look at the results gained from the statistical analyses described in the previous chapter. The results will be presented individually according to which aspect of lexical richness they represent. Table 2 presents the students' results and grades;

4.1 Lexical Sophistication

The analysis of the student texts shows that there is a difference between the high-level students and the intermediate level students when it comes to the number of advanced words they use in their texts. On average the high-level students use approximately 5,5 more advanced words, i.e words found in the NGSL3, NAWL or Off-list frequency lists, in their writing. The results with descriptive data are presented below in table 2;

Table 2: Lexical Frequency

LFP	Mean	Median	Std. Dev	Min	Max	Range
3-4	14	15.0	5.31	4	22	17
5-6	19.57	19.0	5.86	13	27	14

The mean and median are nearly the same, which suggests that the data is evenly distributed and that the mean represents the data well. The range between the highest and the lowest score in both groups is high, which means that the data may have some extreme high or low values, or that it is very dispersed. This may affect the mean, but as one can see, the difference between the lowest score in the intermediate level group, and the one in the high-level group is higher than the difference between the highest scorer in each group. This may suggest that there is one student who has good vocabulary skills but had other issues that caused him or her to receive a lower grade.

As the data were deemed to be normally distributed, an independent t-test was applied in order to find whether this difference was significant. The test revealed a p-value of 0.034, which means that the difference is significant, as it is larger than the confidence level of p = 0.05. The results of the Pearson correlation test gave an r = 0,410, with a p = 0.034. This means that there is a positive correlation between the grades and the number of advanced words. As the grades increase, the number of advanced words increases as well. Note that this is related to the two groups of intermediate (grade 3-4) and high (grade 5-6), and not to the individual grades. The correlation is only significant on the 0.05 level, which means that it is not particularly strong, though it is significant.

4.2 Lexical Density

When it comes to lexical density, the mean scores of each group are almost the same, with the intermediate group actually scoring one point higher. There does not seem to be any difference in lexical density between the groups, as can be seen in table 3;

Table 3: Lexical Density

LD	Mean	Median	Std. Dev	Min	Max	Range
3-4	0.45	0.45	0.025	0.40	0.49	0.09
5-6	0.44	0.44	0.022	0.41	0.46	0.05

The descriptive data shows that the mean is reliable as a summary, as the median is equal to it in both groups, the standard deviation is small, and the range between the highest and lowest scorer in each group is not too high. It is interesting to note that both the highest and lowest score of the entire group of participants are found in the intermediate group. The statistical analyses confirm that the difference between the groups is too small to count, as the independent t-test reveals a p-value of 0.207, which is above the confidence level of $p = 0.05$. Not surprisingly, there is no significant correlation between the grade level and the lexical density of the text either, with $r = - 0.251$ and $p = 0.207$.

4.3 Lexical Diversity

Although they measure the same component of lexical richness, the TTR and LTR will be presented separately so they do not get confused. The TTR is the most traditional way to investigate lexical diversity, while the LTR is a newer and less used method. The advantage of the LTR is that it uses lemma as its definitions of a word, which is closer to the reality of how people conceive the concept of words than type (Granger & Lynne, 2000, p. 3-4).

4.3.1 Type-Token Ratio (TTR)

The mean TTR score of the high-level group is higher than the intermediate group by 0.05 points. The results are centred around the mean, as the median is equal to it in both groups. The range of the intermediate group is a bit large, which suggests that the group may have some extreme values which may affect the mean. This does not seem to be the case though, as the standard deviation is relatively small, and the mean and median is, as mentioned, the same. The reason for the large range may be the same as mentioned above, one student who has good vocabulary skills but has other issues with his or her essay which causes the teacher to give the text a lower grade. The results are presented in table 4 below;

Table 4: Type-token Ratio

TTR	Mean	Median	Std. Dev	Min	Max	Range
3-4	0.47	0.47	0,033	0,41	0,54	0.13
5-6	0.52	0.52	0,025	0,48	0,56	0.08

The difference between the intermediate level group and the high-level group is significant, as the p-value was $p = 0.001$. This means that there is only a 1% chance that the significance is a fluke. The results of the Pearson Correlation test showed a positive correlation between the grade level and the TTR score, with an $r = 0.587$ and $p = 0.001$. With such a low p-value, the correlation seems to be very significant, with a small chance of it being wrong. This means that the TTR scores increase as the grade level increase.

4.3.2 Lemma-Token Ration (LTR)

The lemma token ratio was not calculated automatically by Vocabprofile. These values were therefore found by using the frequency numbers found by the program and calculated in Excel. The mean scores of the two groups show a similar difference to the difference found between the TTR scores, which is not surprising, considering how they measure approximately the same aspect of lexical richness. The difference between the means is at 0.05. There is some difference between the means and medians of each group, and the standard deviation is passably small, though the range is a bit large in the intermediate level group here as well. Despite this issue, the data seems to be centred around the mean, entailing that the mean gives a good picture of the data and that it is evenly dispersed. The results are presented in table 5;

Table 5: Lemma-token Ratio

LTR	Mean	Median	Std. Dev	Min	Max	Range
3-4	0.41	0.42	0.022	0.36	0.48	0.13
5-6	0.46	0.45	0.027	0.42	0.5	0.08

The data were normally distributed, which means that the independent t-test was applied. The test showed that the LTR scores are similar to the TTR, as $p = 0.001$. The difference in the

LTR score between the two groups was thereby a very significant one, as the confidence level was set to p = 0.05. There was also a positive correlation between the grade level and the LTR score, as the r = 0.583. The correlation is significant even at the 0.01 level as p = 0.001.

As the TTR and LTR scores has a highly significant correlation, it was deemed interesting to check the correlation between individual grades and the TTR/LTR scores. A non-parametric correlation test (Spearman's rho) was conducted on both measures. The test revealed a significant positive correlation between TTR and grade (r = 0.605, p = 0.001) and between LTR and grades (r = 0.543, p = 0.003).

The TTR scores and the LTR scores both gave similar results and revealed that the lexical diversity of a text does have a correlation with the teacher's rated holistic quality. In addition, the correlation between the LTR and the TTR was near perfect, with a score of r = 0.959 with a significance of p = 0.00. This is an interesting observation, which will be discussed later.

4.4 A Closer Look at some Student Texts

In addition to the statistical analyses made on the two groups, three of the texts were chosen for a closer look. It should be noted that this is only a brief analysis of the three texts, as this thesis' main focus is on the quantitative analyses. This qualitative examination is merely meant as a supplement to the quantitative data in order to enrichen the discussion. The two first texts were picked because they had some of the highest and lowest scores of the entire student group, while the last text is a special case, where the teacher gave it an intermediate grade, despite the fact that it has one of the highest scores of lexical richness. All the texts are found in the same class and has been graded by the same teacher. The texts in their entireties with the teacher's comments can be found in appendix 4. Words from the three frequency bands defined as showing advanced vocabulary are marked with different colours in order to distinguish which frequency band they are located in. Words marked in red are off-list words, those on yellow are from the NAWL, while the ones in green are from NGSL3. Note that the texts may seem a bit strange as all the proper nouns are removed, in addition to them being cut abruptly where the word count reached 300.

The first excerpt to be examined belongs to student P60213, from now on named Rebekka. Her excerpt is the highest scorer on lexical richness of all the students participating.

Monster! You are a monstrosity! A wild animal, run!

These were all normal comments for as he would walk along the market. Everywhere he went, there would be a crowd of people waiting to throw their moldy food, rotten tomatoes and other unpleasant substances at him. Guards would even put him in the pillory, just for fun if they felt like it. And then they would just leave him there for the rest of the day, humiliating him. I would rather die he always think when he see the guards at the end of the street. But still, he manages to live through every day. He stands in front of a pastry shop, not that he could afford any of it, or that the owner would even sell him anything, he just appreciates the smell. But he can also see if any guards or angry people are coming for him behind, because of the strange amount of mirrors in the pastry shop. The owner's name is she looks like the offspring of a donkey, and sounds like a cold and sick bird. Somehow, her daughter looks like an angel, and smells like newly bloomed flowers. She is also very kind to him, but sadly, she is never home, because her goody two shoes of a boyfriend, who reminds of someone, whom he cannot remember, always takes her on to great adventures. just stands the pastry shop for as long as he can until shouts him away, because he is scaring all her customers. Never thinking it might be herself who scares them away.

Get off me shop, you blasted animal, she shouted at him as she noticed him from inside. And then comes a sneaky trick What if I do not want to? he provoked. face fired up all

This text was graded a 6, the highest achievable grade. It clearly has a great vocabulary and scored highest on all the components of lexical richness. As one can see by the marked words, there is a high density of low-frequency words in this excerpt. The lexical diversity of this text is also high, with a TTR score of 0.56 and an LTR score of 0.5. In addition to these measures, the length of the entire long answer was measured (by number of words) and the number of lexical errors found in the excerpt were counted. These investigations revealed a word count of 2029 and no lexical errors. This student is a clear example of a proficient language user, which is reflected in the teacher's comments and marks. There are few mistakes corrected overall, this excerpt has no lexical errors (see appendix 3) and the comments are generally positive. The few remarks which are made concerns concord and some structural issues in the first two tasks.

The second excerpt is found in the mock exam answer of student P60209, from now named Susanne. Her text represents some of the lowest lexical richness scores of all the participants.

Hi, you are welcome to my birthday. Hi, you are welcome to my birthday. has birthday today and she has invited the whole class. I am invited, but I do not know if I am want to go. Everyone has for sure buy a big surprise for her, but my family has not that kind of money. I am just glad if I get something to eat before I went to sleep. Nobody knew that my family is poor and I want it to continue like that.

Are you coming to my birthday? asked me. I do not know yet, I said but I was really sure that I does not want to go. I hope you can come, all the other boys in the class are coming, she said and walked away.

The boys in the class has fast cars a lot of friends, some of them has also a girlfriend. Me, whatever have not friends so than I will just stand for myself. I am a completely loser. My parents has told be to be more social and make some friends, but I guess no one will be my friend, but it is worth a try. I should go to the party and make some friend. I should go to the party and make some new friends. It feel that I has convinced myself.

At the end of the day, a boy from my class asked me if I am going to birthday. I have not decided yet. Not me either, but I think I will go and we can just go home earlier. he said and look me in the eye. Yes, I will guess I see you then, I said and walked away. I always be really nervous when someone looks me in the eye. I

This text had the lowest amount of low-frequency lemma, a total of only 4. The TTR and LTR scores (0.42 and 0.37 respectively) were among the lowest as well. As can be seen, the text does not contain many low-frequency words. In addition, the low-frequency words used are often repeated, giving the impression that the student has a higher density of advanced lemma than the results show. The text was graded a 4, which was not amongst the lowest graded texts. The entire long answer is 824 words long, less than half of the length of Rebekka's text. The excerpt had no lexical errors which needed to be corrected. The teacher comments revealed that the student switches between tenses and has some concord mistakes, as well as some poorly written sentences. Any comments concerning content were generally good though, and the teacher did comment on some sections as well written. The excerpt contained only 2 lexical mistakes (see appendix 3), which may have had an impact on the holistic grade as well.

The last excerpt selected is that of student P60205, from now on named Johanne, whose lexical richness score is much higher than what one would expect judging by her grade.

The school bell rings and I see the other kids running inside, but all I wanna do is keep hiding. Keep hiding forever, never be seen ever again. To be invisible has to be better than this, everything has to be better than this. There is no one left outside and it is so quiet now. I know I am gonna be late to class, but honestly, I would much rather go home. But I force my legs to go slowly towards the school door, and suddenly much faster then I wanted, I am standing right outside the classroom. I am both cold and warm and the last thing I want to do is to go inside that door, but that is exactly what I am doing.

The door creaks and slowly opens. Everyone is staring at me, the teacher looks at me with a harsh look, but she does not say anything. That is almost worse then when she yells at me. I hurry down to my seat with my eyes fasten to the floor. Even though, I can feel smiling and when I am walking by him, he puts something in my hand. I sit down and sees that it is a piece of paper. has drawn two dead adults and a kid, and written with red script orphan. My stomach hurts and I feel sick. I should be angry, I should be pissed, but all I feel is sadness.

The bell rings and it is break time. A cold hand grabs my shoulder, and before any normal person would react, I turn around in full alert with my pulse up in. It is just me says and my pulse calms down. Why did you scare me like that? I try to sound angry, but she can hear

By taking a closer look at this student, one finds that she displays a very good vocabulary compared to most of her peers. She uses a total of 22 different advanced lemmas in the 300 words excerpt of her paper, 11 of them not found in any the four frequency lists. The excerpt has a TTR of 0.54 and an LTR of 0.48, some of the highest scores in both classes. Despite this, the paper was graded as a 4, placing it amongst the intermediate texts. A quick overview of the teacher comments and marks on the entire paper (see appendix 4), shows that there are other issues concerning this text which may overshadow the great vocabulary use. One such thing is verb tense and subject-verb concord. There is no written general comment on content or language (though there might have been an oral one), creating the possibility that structural or content related issues may have been contributing factors to the holistic grade as well. A quick count of the number of words her long answer consists of reveals that it is 893 words long, close in length to Susanne. In addition, the number of lexical errors found in the excerpt was 7 errors, the highest found in any text (see appendix 3).

4.5 Lexical Errors

Appendix 3 presents each student's lexical errors, as well as other words which were removed, mainly due to them being expressions of sounds such as "aaah" and "oh". The intermediate group had a total number of 54 errors (not counting "ohs" and aahs"), while the

high-level group had 3 errors in total. As the difference between the number of students within each group is large, the mean scores were calculated. The intermediate group had an average of 2.84 lexical errors for each student, while the high-level group 0.57 errors per student. As can be seen, the intermediate group had over 2 errors more per student than the high-level group. The descriptive statistics of the spelling errors are presented in table 6;

Table 6: Lexical Errors

Lexical Errors	Mean	Median	Std. Dev	Min	Max	Range
3-4 (High-level)	2.70	3	2.08	0	7	7
5-6 (Intermediate)	0.43	0	0.53	0	1	1

A Man Whitney U test of significance and a Spearman rho correlation test was conducted. The difference between the groups was strongly significant, as $p = 0.007$. The correlation test revealed a negative correlation ($r = -0.528$), which was strongly significant ($p = 0.005$). As the grades increase, the lexical errors decrease. This may have implications for the rest of the results and will, therefore, be included discussion.

5.0 Discussion

Before diving into the discussion of the results, a reminder of the aims of this study is necessary:

Is there a correlation between Norwegian lower secondary school teachers' holistic rating of ESL written production and the students' lexical proficiency?

> ➢ SQ1: Is there a correlation between grades and lexical sophistication?
> ➢ SQ2: Is there a correlation between grades and lexical diversity?
> ➢ SQ3: Is there a correlation between grades and lexical density?

The discussion will try to answer each of the sub questions in turn, before looking at what the results may say about the main research question.

5.1 Lexical Sophistication

A writer who uses many advanced words in his or her writing has a sophisticated language and would score well on measures of lexical sophistication (Read, 2000, p. 200). The results of the analyses made on the students' texts showed that there was a significant difference in

the number of advanced words used between the students who received the grade 3 or 4 and those who received a 5 or a 6. This reveals that higher-level students tend to use more low-frequency words when writing. As there was a significant positive correlation between the level of grade the students were assessed to belong to and the number of advanced words they used, it would seem like the teachers take this component of lexical richness into account when assessing students' written production. This corresponds well with previous research on the field. Laufer and Nation (1995, p. 316) found evidence that showed that high proficiency students tended to use more advanced words than intermediate and low-level students. They also discovered that the low-level students mostly used high-frequency words found on the first and, partially, the second level of the GSL (Laufer & Nation, 1995, p. 314). Though a different and more updated version of the frequency lists used by Laufer and Nation was used in this study, the findings are mainly the same. One difference in the approach to discovering these differences is that Laufer and Nation (1995, p. 312) looked at the percentage of the total number of words the low-frequency words made up. This thesis chose to focus on comparing the number of advanced words found in each text. The reason this was possible is that the texts were cut to an equal length, allowing plain numbers to be compared instead of ratios or percentages.

While Laufer and Nation tested university students who learned English as a foreign language, Roessingh, Elgie, and Kover (2015, p. 71) tested the correlation between Canadian 3rd graders' lexical sophistication and proficiency level by analysing their written production in Vocabprofile for Kids. Their results show that the students who were rated as the most proficient learners used fewer of the high-frequency words and more of the off-list words (Roessingh et al., 2015, p. 72-73). Compared to the 9th graders in this study, the results are similar. This illustrates that the difference in lexical sophistication between student texts which are rated high, intermediate and low is found in both primary school children, lower secondary school students, and university students. Though Roessingh et al's (2015) study investigated native speakers, it is comparable to studies of second or foreign language learners of English, as it, amongst other things, looks at the differences between proficiency groups and holistic quality. It complements the other research and reveals that the results found in this study have been found for every form of English learning, whether it is one's foreign, second or native language.

Previously it has been discussed that some elements of lexical richness were more prominent in distinguishing between the level of proficiency depending on the genre (Olinghouse &

Wilson, 2013, p. 60-61). Lexical diversity and sophistication were found to be the most prominent predictors of the holistic quality of a story. This means that these two components of lexical richness were found to predict the overall quality of a story, while in other genres, such as persuasive and informative texts, other elements such as the number of content words specific to the theme and register, were found to correlate more with the overall quality (Olinghouse & Wilson, 2013, p. 60-61). As all of the texts used in this study were different kinds of stories, or at least realistic fiction (some were made up letters), the results are interesting to compare to this study. Though only three elements of lexical richness were measured, the results clearly showed that lexical sophistication and diversity correlated with the grades given by the teacher. It seems like using words beyond the 2000 most frequent when writing stories may improve the holistic quality of the text. Of course, as Laufer and Nation (1995), as well as Roessingh et al. (2015) has shown above, the holistic score of other genres are positively affected by the number of low-frequency words as well, though there may be other elements of lexical richness which affects these genres more, e.g. the aforementioned number of content words, lexical complexity (number of letters per word), or the number of lexical errors.

5.2 Lexical Density

Lexical density has been stated to indicate literacy in texts, the higher density of content words, the more literate the text is (Read, 2000, p. 196). This statement may be true, but it does not seem like there is any correlation between the lexical density of a text and its holistic quality, at least not according to the results of this study. As shown above, the correlation between grade level and lexical density is a slightly negative, non-significant one. This shows that contrary to the statement above, the higher the lexical density of a text, the lower the grade level is. The difference between the groups is not a significant one though, revealing that there is little to no difference between the lexical density of students assessed to present a high level of proficiency and those who present an intermediate level of proficiency in their writing. These findings are supported by Engberg (1995, p. 148), who found small and insignificant results concerning the correlation between lexical density and the holistic rating of texts. She investigated L2 learners from mixed language backgrounds who wished to enrol into American universities. She concluded that the students did not gain any higher holistic scores by "simply piling up lexical words" (Enberg, 1995, p. 148). Ishikawa (2015, p. 209) found similar results when he measured how lexical components developed as the language proficiency developed. The lexical density did not change, even though the learners became

more proficient in the language. Compared to the results of this study, the results reveal that lexical density does not differ between advanced learners and intermediate learners, even as the learner becomes more proficient.

Laufer and Nation (1995, p. 309) critiqued the use of lexical density as a tool for measuring lexical richness, as they claim it is too affected by the number of function words, as well as structural devices such as participle clauses, ellipsis or subordinate clauses. The measurement was still applied as it was deemed interesting to investigate alongside the other components of lexical richness. The results of the analyses showed that lexical density does not have any correlation with the quality of the text, meaning that it is an inadequate measure of both lexical richness and the holistic quality of a text.

5.3 Lexical Diversity

A writer who presents texts with a high lexical diversity has a vocabulary big enough for them to be able to vary which words they use (Read, 2000, p. 200). A limited vocabulary makes them dependent on few words when wishing to express themselves. Lexical diversity can, therefore, be said to not only measure how varied one's language is, but also indicate how large the vocabulary is. As the high-level students had higher TTR and LTR scores, it would seem like the have larger vocabularies than the students in the intermediate group. As previous research has shown, one needs a vocabulary of only 8000-9000 words in order to understand newspaper articles, authentic novels, and children's movies (Nation, 2006, p. 71-72.). This is receptive vocabulary knowledge though, which is usually more extensive than productive vocabulary knowledge (Laufer, 1998, p. 263-263). This study's findings suggest that the high-level students on general are better equipped to comprehend such authentic sources of English. The results even showed a significant positive correlation between each grade and the TTR/LTR score, not only from one grade level (intermediate) to another (high), strengthening this statement, as well as the fact that lexical diversity seems to be a good predictor of holistic quality. Zareva et al's (2005, p. 591) findings supports these conclusions, as they found that advanced learners had a larger vocabulary than intermediate learners. In addition, Zareva et al's. (2005, p. 591) study imply that a high lexical competence is one of the determining characteristics of language proficiency. Their research focuses on EFL proficiency in general, while this study's focus is angled towards written EFL proficiency. This serves to show that lexical diversity may be an important component for raters and teachers when they assess EFL learners. Several other researchers confirm this notion as they have achieved similar results in their studies. Lexical diversity has even been shown to be the

strongest lexical indicator of holistic quality in story as a genre (Olinghouse & Wilson, 2012, p. 60). The texts in this study are not all exclusively stories, but they are all fictional, closer to the story genre than for example the persuasive one. The strongest predictor of holistic quality would seem to be diversity in these texts as well, as the TTR and LTR ratio had the highest correlation with the grade level of the three components measured. In addition, the TTR and LTR ratio had a positive correlation with each grade when divided into decimal numbers. Crossley et al. (2010, p. 572) found lexical diversity to correlate well with human ratings of L2 texts. A later study found lexical diversity to be greater in L2 written production provided by advanced learners, than the lexical diversity of intermediate students (Crossley et al., 2012, p. 130). All of these previous findings support and strengthen the claim that lexical diversity is an important part of language proficiency. Both the teachers in this study and the raters in previous studies seem to assess written production to be of a higher quality if the lexical diversity is higher. This may not be surprising considering Vögelin et al's (2019, p. 59) discovery of how lexical diversity and sophistication affected the teachers' attitude towards other linguistic aspects, such as grammar and "frame of essay", of written production. Such a halo effect may have affected the teachers in this study and caused them to look more positively at texts with a high lexical richness than on texts which displayed a simple language with little variation. All of these previous studies on the impact of lexical diversity on written or overall L2 language proficiency support the results of this study. It would seem like lexical diversity affects how the teacher or rater assesses L2 and EFL written production. It would also seem like advanced L2 learners have a better vocabulary than intermediate ones, which is also one of the conclusions one may draw from the results of this study. This is not surprising, as students who are rated as displaying a high proficiency in a language are generally better. Because words are the building blocks of a language (Read, 2000, p. 1), it is reasonable to assume that vocabulary is one of the areas of which an advanced student outperforms an intermediate one.

Although the evidence seems to point towards lexical diversity being a reliable predictor of the holistic quality of texts, there are some issues that should be kept in mind. First of all, the students were allowed to use dictionaries, opening up for the possibility of them finding synonyms for words they felt were overused. Secondly, spelling errors were corrected, which may add more unique words to the text. Lastly, TTR is a heavily discussed measurement, as it is affected by text length. Though this issue was resolved, other aspects of TTR has been

claimed to make it an unfit measure of lexical diversity, as previously discussed (Jarvis, 2013, p. 20).

5.3.2 LTR vs TTR

As discussed above, TTR is the most traditional measure of lexical diversity and has been used by scholars for many years, though it may not be the most precise way of measuring lexical diversity as lemma may be a more precise definition of a word (Granger & Lynne, 2000, p. 3-4). This study chose to use both measurements to control for any possible ambiguity which may have occurred, as discussed in chapter 2.3.4. The analysis of the results gave an interesting addition to this debate. The correlation between the TTR and LTR score was nearly a perfect one, as r = 0.959 with a significance level of p = 0.000. This entails that the LTR and TTR scores give the same picture of the lexical diversity of the texts. It can, therefore, be debated whether there is any point in using LTR instead of TTR, though as lemma is generally seen as a more accurate representation of a word (Granger & Lynne, 2000, p. 3-4) it may be considered to replace TTR with LTR. It is a possibility that a larger dataset may have changed these results, though as Granger and Lynne (2000) found similar correlations in their study, it would seem like the two forms of measuring lexical diversity gain comparable results.

5.4 Lexical Errors

As lexical errors such as spelling mistakes and wrong use of word were corrected and counted as a part of the data processing, it may be interesting to take a brief look at how they may have impacted the results. Though the choice of correcting lexical errors was a carefully considered one, it does have some consequences. As previously discussed, written knowledge of a word includes the ability to spell it correctly (Schmitt, 2000, p. 48). Research has shown that lexical errors such as spelling errors affect holistic assessment negatively (Engberg, 2015; Vögelin, 2018; Bestgen & Granger, 2011). The correction of these mistakes may, as discussed above, cause some students to score higher on the lexical sophistication charts despite not having full knowledge of all of the words. After correcting all the spelling errors and removing any wrongly used words, the list showed that the most errors were found in the intermediate group where the average student had 2.65 errors which were corrected or removed. In the high-level group, on the other hand, only 0.6 errors were corrected per student on average (see appendix 3). The difference is significant, and a significant negative correlation was found. The fact that the intermediate group has made more lexical mistakes than the high-level group means that correcting spelling errors would benefit the intermediate

group the most. As the results showed that the high-level group scored significantly better on lexical sophistication and diversity, the fact that such mistakes were corrected would therefore not have any large impacts on the results. However, the significant difference and negative correlation between the grade level and the number of lexical errors suggest that lexical errors may have had an impact on the teachers' assessment of the texts and its lexical richness. Though this is not a part of this study's original focus or aims, it is an interesting tendency to note nonetheless.

5.5 A Closer Look at Lexical Richness as a Predictor of Holistic Quality

Most research has found a correlation between various components of lexical richness and the holistic quality of a text, as has been discussed above. As can be seen from the results of the lexical error analysis, the vocabulary is not the only element of a text which affects the holistic quality. The Norwegian curriculum of English names coherency, text and sentence structure, as well as grammar to be important (Utdanningsdirektoratet, 2013, p. 9). As has been discussed in chapter 2, the teachers' only guidelines for correcting written production in English is the curriculum with its aims and the Exam assessment guide (Utdanningsdirektoratet, 2018, p. 9). The CEFR scale is possible to use in assessment as well, though it is focused on language proficiency in general (Council of Europe, 2001b). The CEFR scale and the national curriculum are not specific enough for the teachers to apply directly to their assessment. The exam guide, on the other hand, is, though it is made to assess the final exam given at the end of 10th grade (Utdanningsdirektoratet, 2018, p. 9). Therefore, one cannot expect 9th graders to display the same competence as this exam guide requires. As there are no proper guidelines made for assessing written production in 9th grade, the teachers are left to use their own subjective opinion of the text's quality when grading their students' production. It may be common to modify the exam guide to the age group, though as no teachers were interviewed about this, no certain statement can be made. Even with a common set of criteria such as the exam guide, teachers are left to decide which of the criteria they deem more important, perhaps forcing them to make compromises (Read, 2000, p. 214). As previously discussed, the subjectivity of a teacher's rating creates low inter-rater reliability (Midtbø et al., 2018, p. 22; Skar & Jølle, 2017, p. 14), as there are several components which teachers consider when grading written compositions. The grade is supposed to show the holistic quality of the text. This does not mean that every element of the text is assessed to be at the same grade level. A text can have content which is good enough for a 5, grammar

which equals a 3, vocabulary which is rated as a 6 and structure which adds up to a 2. The teacher's job is to take all of these components and weigh them against each other, deciding on a grade that they believe represents the text as a whole. Some teachers may find correct grammar and structure to be the most important and give such a text a 3-. Another teacher may think that communication is key, and weigh vocabulary and content as more important. The text may then be given a 5.

As has been discussed before, research has found that lexical sophistication and diversity have an impact on teachers' holistic judgment of written compositions. Texts with high lexical diversity and sophistication tend to be rated higher. A halo effect has been found in these texts, as teachers view other aspects, namely grammar and frame of essay, as better in addition to the vocabulary. (Vögelin et al., 2018, p. 9; Vögelin et al., 2019, p. 59). This is consistent with the results found in this thesis, as the teachers seem to at least grade the lexically rich texts higher than those less lexically rich. In addition, a closer look was taken on some of the texts, revealing interesting results about the teacher's grading. Student P60213, Rebekka displayed the highest score on lexical diversity and sophistication, the two components of lexical richness that were found to correlate with grade level. In addition, she had no lexical errors in the 300-word excerpt. She was graded a 6, which fits well with the results of this study, as well as the results other scholars have found (e.g: Laufer & Nation, 1995; Crossley et al., 2012; Crossley et al., 2005; Vögelin et al., 2018). Although the results show that there is a positive correlation between both lexical diversity and grade-level, as well as between lexical sophistication and grade level, this characteristic does not apply to every student text. The most obvious example is found by looking at the highest scoring text from the intermediate group, student P60205, who was named Johanne. Johanne scores higher on both TTR and LTR than the other students of her group. In addition, she uses the most advanced words in this group as well. Her text has the second highest TTR score (0.54) and the third highest LTR (0.48) and LFP score (22 words) of all texts which were analysed. Despite these high scores in lexical richness, Johanne's text was graded as a 4, placing her in the intermediate group. Another teacher may have given Johanne a higher grade, given that they considered vocabulary to be one of the most important components of a text. Johanne's grade could, therefore, be intermediate because other components of his text proved not to be as good as its lexical richness. P60209, named Susanne, is another text which does not fit the scale completely, as it is amongst the lowest scorers on lexical diversity (TTR = 0.42, LTR = 0.37) and sophistication (4 advanced lemmas). The text was graded the same as Johanne's,

which means that they were considered to be of equal quality, despite the large difference in lexical richness. The comments on these two texts yielded little information concerning the teacher's evaluation of their vocabulary, though this feedback may have been given orally. Any critical comments surrounded the use of tense and sentence structure, the rest of the comments being mainly positive.

There are many factors to consider for a teacher when he or she assesses written production. Lexical richness is, as previously stated, only one of them. Research has shown that factors such as text length and lexical errors have an impact on the assessed quality of a text (Jarvis, Grant, Bikowski & Ferris, 2003; Engberg, 1995; Vögelin et al., 2018; Bestgen & Granger, 2011). Though the total length of these texts was not a part of this study, a quick count was made to reveal that the two texts had almost the same amount of words, Johanne's text reaching 893 words, while Susanne's text counted 824 words. Without measuring the rest of the texts in this study, a claim as to whether these texts are long or short cannot be made, though it does reveal that these two texts are close enough in length that the teacher probably has rated them equally in that regard. Other components could explain how two texts with such differences in lexical richness could receive the same grade, though. The number of lexical errors could be one such component, as Johanne's text had a total of 7 mistakes, while Susanne's had 2 mistakes spread across the 300-word excerpt. This could be one of the reasons why Johanne received an intermediate grade, as research has shown that lexical errors can produce a halo effect, making teachers look more negatively at the grammar and vocabulary as well as the lexical errors (Vögelin et al., 2018, p. 9). The opposite can be said of Susanne who had few lexical mistakes. She may have used a less sophisticated and diverse vocabulary when writing, but the vocabulary used was accurate and almost free of errors. This may have affected the grading positively, the same way as Johanne's spelling errors affected the teacher negatively. As Engberg (1995) discovered, the amount of lexical errors correlates negatively with the holistic quality of the text, entailing that texts with more lexical errors are generally rated lower than those with few errors. It may therefore not be strange that Johanne received the same grade as Susanne, as Johanne was the student with the most lexical errors, while Susanne had amongst the fewest mistakes in her group. Beside these students, the rest seem to score relatively close to the group mean, Rebekka being a good example.

5.6 Limitations

The study does have some limitations which create issues surrounding the possibility to conclude anything firmly. The participating schools are both in from the same area in

Norway. The schools were chosen because they are a part of the ESIT project, which gathers its material from schools in the area. This means that the results found in this study can at most be claimed to be representative of the region of which the data was collected, perhaps only of these two schools, or these two classes. Secondly, only an excerpt of one task from the mock exams was used. This was a conscious choice made to control for the issues surrounding TTR and text length (Cobb & Horst, 2015, p. 192), as well as the possibility for more advanced words in a long text than in a short one. In addition, the texts were controlled for genre in order to avoid any differences in the richness of the vocabulary, as discussed above. Despite this, it limits the data to a portion of what the teacher has actually assessed. This may interfere with the correlation, as one may question whether the excerpt picked is representative of the entire mock exam. Thirdly, as stated in chapter 3, the students were allowed to use a dictionary to aid them in their writing, as well as a booklet with texts. As there is no way to know which words the students know themselves and which ones they found in the dictionary or booklet, this may have created false estimates of some students' knowledge. At most, this means that the high-level group may have scored better on lexical sophistication because they are better at utilizing the available aids, though as all the students had the same opportunity to use these resources, the difference is still evident. Besides, the aim of this study is to discover any correlation between lexical richness and teachers' holistic assessment, not the difference in lexical richness between students. Lastly, the measurements chosen to represent lexical richness may not represent the complex nature of the term. As stated above, lexical richness includes components such as diversity, density, and sophistication (Laufer & Nation, 1995, p. 309). In addition, originality, words hypernymy and word frequency are used by other studies when they aim to measure lexical richness (Crossley, 2010, p. 572). This entails that the measurements used in this study may only measure parts of lexical richness, not the entirety of it.

6.0 Conclusion

This study aimed to discover whether there is any correlation between the holistic assessment of EFL teachers and the lexical richness of written production as measured by lexical sophistication, diversity, and density. In order to investigate this thesis, statistical analyses were applied to the written production of 27 Norwegian lower secondary students. In addition, some texts were analysed more closely, in order to get a more detailed impression of how

lexical richness affects the text and whether other factors can overrule the vocabulary skills the student displays. The results revealed positive correlations between two of the three measures, sophistication and diversity, while lexical density was not found to have any impact on the teachers' assessments. As this is consistent with the results of similar research (Laufer & Nation, 1995; Vögelin et al., 2018; Vögelin et al., 2019; Crossley et al., 2010; Crossley et al., 2012; Zareva et al., 2005; Roessingh et al., 2015; Engberg, 1995; Olinghouse & Wilson; 2012), the correlation between lexical richness and the holistic quality of written production seems evident. This does not mean that lexical richness is the only defining factor of texts, as a closer analysis of selected texts revealed characteristics that differed from how the results of the analyses suggested they should look. One text scored high on measures of lexical richness but was given an intermediate grade (4). A second text was amongst the lowest scorers on all measures but received the same intermediate grade as the first student. The teacher's comments seemed to indicate some of the same issues with both texts, though the text with a high lexical diversity score had more lexical errors, which has been shown to create a halo effect (Vögelin et al., 2018, p. 9). In addition, the analysis of lexical errors showed a negative correlation between the number of errors and the grade level. These results indicate that other characteristics of written production can overshadow the lexical richness of the text. This is consistent with both research (Vögelin et al. 2018; Vögelin et al. 2019; Bestgen & Granger, 2011; Jarvis et al., 2003) and the national curriculum of English (Utdanningsdirektoratet, 2013, p. 9) which names other characteristics such as spelling, content, textual structure and grammar to be important in the assessment of written proficiency as well.

6.1 Pedagogical Implications

This study has found that Norwegian lower secondary school teachers of English seem to reward good vocabulary skills with good grades. Though this is not necessarily true for every text, the results show how compositions displaying a varied and advanced vocabulary are rated as high-level. This could be due to the fact that teachers get a better overall impression of these texts, or because highly proficient students tend to have a better vocabulary, in addition to being proficient in other parts of the language as well. Regardless, teachers should be aware of how a good vocabulary can affect their judgement on written production. It is important to see each characteristic for what it is, not in the light of something else. For example, as Vögelin et al. (2019) revealed, texts with a high degree of lexical sophistication and diversity are often judged to have better grammar and "frame of essay", even if these factors are equally good in two texts. It may be important to look beyond the first impression

and focus on the quality of each criterion before assessing the entire text put together. In conclusion, EFL teachers should be conscious of the impact a good, or bad, vocabulary can have on their holistic rating of student texts.

6.2 Suggestions for Future Research

First and foremost, it would be interesting for future research to expand the sample of the study, in order to investigate the thesis on a larger scale. It would also be interesting to interview teachers concerning their thoughts and impressions when it comes to how they assess written production. If the teachers' thoughts were compared to a statistical analysis of texts they have evaluated, the results could reveal how conscious the teachers are on which aspects of the written compositions that are most important to them when evaluating. As texts are graded according to the subjective opinion of the teacher, resulting in a low inter-rater reliability (Skar & Jølle, 2017; Midtbø et al. 2018), it may be informative for teachers to learn how their own view on how they rate texts may differ from what they actually do.

It may also be of interest to investigate differences in the importance of lexical richness as a predictor of holistic quality between age-groups and/or teachers. This study does not compare how the two teachers who assessed approximately half of the texts each differ when it comes to the lexical richness and the holistic grade they give their students' texts. It may, therefore, be interesting to take a closer look at whether some teachers seem to put more weight on vocabulary than others. This study does not investigate whether there are any differences in how well lexical richness correlates with holistic quality between age groups (lower secondary/primary school/upper secondary). Future studies could investigate whether e.g. lower secondary school students are rewarded more (in terms of higher grades) for utilizing a more advanced and varied vocabulary than upper secondary school students.

Another approach could be to do a qualitative evaluation of the comments the teachers give student texts in order to discover which linguistic features teachers seem to focus on in their assessment and whether vocabulary is prominent in such comments. As found in this study, lexical errors may be one such feature.

7.0 References

Bestgen, Y. & Granger, S. (2011). Categorizing spelling errors to assess L2 writing. *Inderscience Publishers 21*(2), pp. 235-252.

Breznia, V. & Gablasova, D. (2013). Is There a Core General Vocabulary? Introducing the New General Service List. *Applied Linguistics 36*(1), pp. 1–22. Doi: https://doi.org/10.1093/applin/amt018

Browne, C. (2014). A New General Service List: The Better Mousetrap We've Been Looking for. *Vocabulary Learning and Instruction 3*(2), pp. 1-92. Doi: http://dx.doi.org/10.7820/vli.v03.2.browne

Brown, C. Culligan, B. & Phillips, J. (2013). New Academic Word List 1.0. Retrieved at: https://www.newgeneralservicelist.org/nawl-new-academic-word-list

Cobb, T. *Compleat Web VP v.2* [computer program]. Accessed 06.02.19 at https://www.lextutor.ca/vp/comp/

Cobb, T. & Horst, M. (2015). Learner Corpora and Lexis. In S. Granger, G. Gilquin & F. Maunier (eds.) *Learner Corpus Research* (pp. 185-206). Cambridge: Cambridge University Press.

Council of Europe. (2001a). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press: Strasbourg. Retrieved at: https://rm.coe.int/1680459f97

Council of Europe. (2001b). Self-Assessment grid – Table 2 (CEFR 3.3): Common Reference Levels. Retrieved at: https://www.coe.int/en/web/common-european-framework-reference-languages/table-2-cefr-3.3-common-reference-levels-self-assessment-grid

Council of Europe. (2018). The CEFR Levels. Retrieved at: https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly, 34*(2), pp. 213-238. Doi: https://doi.org/10.2307/3587951

Crossley, S. A. & McNamara, D. S. (2012). Predicting second language writing proficiency: the roles of cohesion and linguistic sophistication. *Journal of Research in Reading 35*(2), pp. 115-135. Doi: 10.1111/j.1467-9817.2010.01449.x

Crossley, S. A., Salsbury, T., McNamara, D. S. & Jarvis, S. (2010). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing 28*(4), pp. 561-580. Doi: 10.1177/0265532210378031

Dirdal, H., Drange, E.-M., Graedler, A.-L., Guldal, T. M., Hasund, I. K., Nacey, S. L., & Rørvik, S. (2017). Tracking Written Learner Language (TRAWL): A longitudinal corpus of Norwegian pupils' written texts in second/foreign languages. Poster presented at *The 4th Learner Corpus Research Conference*, Bolzano / Bozen 5-7 October 2017.

Engberg, C. A. (1995). The Relationship of Lexical Proficiency to the Quality of ESL Compositions. *Journal of Second Language Writing 4*(2), pp. 139-155. Doi: https://doi.org/10.1016/1060-3743(95)90004-7

Garshol, L. (2019, Universitetet i Agder). *I Just Doesn't know. Agreement errors in English texts by Norwegian L2 learners: Causes and remedies*. (Doktorgradsavhandling). Retrieved at: https://uia.brage.unit.no/uia-xmlui/bitstream/handle/11250/2589044/Garshol%20Dissertation%20edit%20%28002%29.pdf?sequence=4&isAllowed=y

Goulden, R., Nation, P. & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics 11*(4), pp. 341-363. Doi: https://doi.org/10.1093/applin/11.4.341

Granger, S. & Wynne, M. (2000). Optimizing Measures of Lexical Variation in EFL Learner Corpora. In J. M Kirk (Ed.), *Corpora Galore: Analyses and Techniques in Describing English*, (pp. 249-57): Amsterdam: Rodopi

Hasselgreen, A. (2005). Assessing the language of young learners. *Language Testing 22*(3), pp. 337–354. Doi: 10.1191/0265532205lt312oa

Helness, H. L. (2012). A Comparative Study of the Vocabulary of 7th and 10th Graders in Scripts from the National Test of Writing in English. In A. Hasselgreen, I. Drew & B. Sørheim (eds.) *The Young Language Learner. Research-based Insights into Teaching and Learning* (pp. 145-158). Bergen: Fagbokforlaget

Horst, M. & Collins, L. (2006). From Faible to Strong: How Does Their Vocabulary Grow? *The Canadian Modern Language Review/La Revue canadienne des langues vivantes* 63(1), pp. 83-106. Doi:10.1353/cml.2006.0046

Jarvis, S. (2013). Defining and measuring lexical diversity. In S. Jarvis & M. Daller (eds.) *Vocabulary Knowledge. Human Ratings and Automated Measures*. Amsterdam/Philadelphia: John Benjamins Publishing Company. Retrieved at: https://ebookcentral.proquest.com/lib/agder/reader.action?docID=1394824

Jarvis, S., Grant, L., Bikowski, D. & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing 12*(4), pp. 377-403. Doi: https://doi.org/10.1016/j.jslw.2003.09.001

Langeland, A. S. (2012). Investigating Vocabulary Development in English from Grade 5 to

Grade 7 in a Norwegian Primary School. In A. Hasselgreen, I. Drew & B. Sørheim (eds.) *The Young Language Learner. Research-based Insights into Teaching and Learning* (pp. 131-143). Bergen: Fagbokforlaget

Laufer, B. (1998). The Development of Passive and Active Vocabulary in a Second Language: Same or Different? Applied Linguistics 19(2), pp. 255-271. Doi: https://doi.org/10.1093/applin/19.2.255

Laufer, B. & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics 16*(3), pp. 307-322. Doi: https://doi.org/10.1093/applin/16.3.307

Llach, M. P. A. (2007). Lexical errors as writing quality predictors. *Studia Linguistica 61*(1), pp. 1-19. Doi: https://doi.org/10.1111/j.1467-9582.2007.00127.x

McArthur, T. (2003). Teaching English Also the teaching of English and English teaching. In T. McArthur (ed.) *Concise Oxford Companion to the English Language*. Oxford University Press. Retrieved at: http://www.oxfordreference.com/view/10.1093/acref/9780192800619.001.0001/acref-9780192800619-e-1213

Meara, P. The Dimensions of Lexical Competence. (1996). In Brown, G., Malmkjær, K. & Williams, J. (Eds.), *Performance and Competence in Second Language Acquisition* (33-54). Cambridge University Press: Cambridge.

Midtbø, T., Rossow, A. & Sagbakken, B. (2018). Måling av sensorreliabilitet ved vurdering av norskprøve i skriftlig framstilling. *Acta Didactica Norge 12*(4), pp. 1-25. Doi: http://dx.doi.org/10.5617/adno.6358

Nation, I. S. P. (2001*). Learning Vocabulary in Another Language*. Cambridge University Press: Cambridge

Nation I. S. P. (2006). How Large a Vocabulary is needed for Reading and Listening? *The Canadian Modern Language Review 63*(1), pp. 59-81. Doi: http://dx.doi.org/10.3138/cmlr.63.1.59

Norsk senter for forskningsdata. (2019). Personverntjenester – Vanlige spørsmål. Retrieved at: https://nsd.no/personvernombud/hjelp/index.html

Olinghouse, N. G. & Wilson, J. (2012). The relationship between vocabulary and writing quality in three genres. *Reading and Writing: An Interdisciplinary Journal 26*(1), pp. 45 65. Doi: 10.1007/s11145-012-9392-5

Rasinger, S. M. (2013). *Quantitative Research in Linguistics – An Introduction*. Bloomsbury: London.

Read, J. (2000). *Assessing Vocabulary*. Cambridge University Press: Cambridge.

Richards, J. C. (2015). The Changing Face of Language Learning: Learning Beyond the Classroom. *RELC Journal 46(*1), pp. 5-22. Doi: 10.1177/0033688214561621

Robson, C. (2002). *Real World Research (2. edition).* Blackwell Publishing: Oxford

Roessingh, H., Elgie, S. & Kover, P. (2015). Using Lexical Profiling Tools to Investigate Children's Written Vocabulary in Grade 3: An Exploratory Study. *Language Assessment Quarterly 12*(1), 67-86. Doi: 10.1080/15434303.2014.936603

Sayer, P. & Ban, R. (2014). Young EFL students' engagements with English outside the classroom**.** *ELT Journal 68*(3), pp. 321–329. Doi: https://doi.org/10.1093/elt/ccu013

Schmitt, N. (2000). *Vocabulary in Language Teaching.* Cambridge University Press: Cambridge

Seidlhofer, B. (2005). English as a Lingua Franca. *ELT Journal 59*(4), pp. 339-341. Doi: 10.1093/elt/cci064

Simensen, A. M. (1998). *Teaching a Foreign Language. Principles and Procedures.* Fagbokforlaget: Bergen.

Skar, G. B. & Jølle, L. J. (2017). Teachers as raters: An investigation of a long-term writing assessment program. *L1-Educational Studies in Language and Literature 17,* pp. 1-30. https://doi.org/10.17239/L1ESLL-2017.17.01.06

Utdanningsdirektoratet. (2013). Lærerplan i engelsk (ENG1-03). Retrieved at: https://www.udir.no/kl06/ENG1-03

Utdanningsdirektoratet. (2018). Eksamensveiledning – om vurdering av eksamenssvar (ENG0012). Retrieved at:

https://sokeresultat.udir.no/eksamensoppgaver.html#?k=Engelsk&start=1&r1=%C7%82%C7%824772756e6e736b6f6c65&r1val=Grunnskole&r2=%C7%82%C7%82456b73616d656e737665696c65646e696e676572&r2val=Eksamensveiledninger

Vögelin, C., Jansen, T., Keller, S. D. & Möller, J. (2018). The impact of vocabulary and spelling on judgments of ESL essays: an analysis of teacher comments. *The Language Learning Journal,* pp. 1-17. Doi: 10.1080/09571736.2018.1522662

Vögelin, C., Jansen, T., Keller, S. D., Macht, N. & Möller, J. (2019). The influence of lexical features on teacher judgements of ESL argumentative essays. *Assessing Writing 39*(2019), pp. 50-63. Doi: https://doi.org/10.1016/j.asw.2018.12.003

Zareva A., Schwanenflugel P. & Nikolova Y. (2005). Relationship between Lexical Competence and Language Proficiency: Variable Sensitivity. *Studies in Second Language Acquisition 27*(4), 567-595. Doi: 10.1017/S027226310505025

8.0 Appendix

Appendix 1: 10th-Grade Exam Guidelines

Kjennetegn på måloppnåelse for ENG0012 ved sentralt gitt eksamen 2018

		Karakteren 2	Karakterene 3 og 4	Karakterene 5 og 6
Førsteinntrykk		**Eksamenssvaret** - kommuniserer til en viss grad ut fra oppgavene som er gitt	**Eksamenssvaret** - kommuniserer ut fra oppgavene som er gitt	**Eksamenssvaret** - kommuniserer helhetlig ut fra oppgavene som er gitt
Hovedområder i læreplanen for engelsk: Språklæring, Skriftlig kommunikasjon, Kultur, samfunn og litteratur	Innhold, fagkunnskap og bruk av kilder	**Eksamenssvaret** - samsvarer til en viss grad med oppgavens krav - viser noe faglig kunnskap - viser forståelse for noe av innholdet og gir enkelte eksempler - bruker enkelte kilder dersom oppgaven ber om det	**Eksamenssvaret** - samsvarer stort sett med oppgavens krav - viser faglig kunnskap - viser forståelse for og vurderer innhold og gir eksempler - bruker kilder dersom oppgaven ber om det	**Eksamenssvaret** - samsvarer med oppgavens krav - viser god faglig kunnskap - viser forståelse for og vurderer innhold, argumenterer og gir eksempler - velger relevante kilder dersom oppgaven ber om det
	Tekststruktur	**Eksamenssvaret** - er til en viss grad tilpasset til formål, mottaker og digitale formkrav ut fra oppgavebestillingen - har enkel tekstbinding på setnings- og/eller tekstnivå - er utformet med forståelig struktur og sammenheng	**Eksamenssvaret** - er stort sett tilpasset til formål, mottaker og digitale formkrav ut fra oppgavebestillingen - har ulike former for tekstbinding som stort sett fremhever innhold og lesbarhet på setnings- og tekstnivå - er stort sett utformet med struktur og sammenheng	**Eksamenssvaret** - er tilpasset til formål, mottaker og digitale formkrav ut fra oppgavebestillingen - har hensiktsmessige former for tekstbinding som fremhever innhold og øker lesbarheten på setnings- og tekstnivå - er utformet med struktur og sammenheng
	Språk	**Eksamenssvaret** - har et enkelt ordforråd om enkelte emner - har enkelte faglige begreper - har forståelig setningsbygning - er språklig til en viss grad tilpasset oppgavens formål	**Eksamenssvaret** - har et ordforråd som dekker kjente emner - har faglige begreper - har noe variasjon i setningsbygning - er språklig stort sett tilpasset oppgavens formål	**Eksamenssvaret** - har et generelt ordforråd om ulike emner - har relevante faglige begreper - har variasjon i setningsbygning - er språklig tilpasset oppgavens formål
	Formelle ferdigheter	**Eksamenssvaret** - har rettskriving som gjør teksten forståelig - har noe ordbøying - har til en viss grad engelsk setningsstruktur - oppgir benyttede kilder på en ufullstendig måte	**Eksamenssvaret** - har sentrale mønstre for rettskriving - har sentrale mønstre for ordbøying - har stort sett engelsk setningsstruktur - oppgir benyttede kilder på en stort sett etterprøvbar måte	**Eksamenssvaret** - har i hovedtrekk bruk av mønstre for rettskriving - har mønstre for ordbøying - har i hovedtrekk engelsk setningsstruktur - oppgir benyttede kilder på en etterprøvbar måte
Samlet vurdering		• Karakteren 2 uttrykker at eleven har **lav** kompetanse i faget.	• Karakteren 3 uttrykker at eleven har **nokså god** kompetanse i faget. • Karakteren 4 uttrykker at eleven har **god** kompetanse i faget.	• Karakteren 5 uttrykker at eleven har **meget god** kompetanse i faget. • Karakteren 6 uttrykker at eleven har **svært god** kompetanse i faget.
		Karakteren 1 uttrykker at eleven har **svært lav** kompetanse i faget, en lavere kompetanse enn det som er beskrevet ovenfor.		

Appendix 2: Mock Exam Tasks

School 1:

https://s-media-cache-ak0.pinimg.com/236x/2a/07/8b/2a078b4489f18a5c5f07b6b00d56f66c.jpg

SHADOWS OF THE PAST

TASKS
English Test for 9th Grade

Informasjon om prøven

Prøvetid	Prøven varer 5 timer
Tema	Shadows of the Past
Hjelpemidler	Alle hjelpemidler er tillat, bortsett fra internett og andre verktøy som kan brukes til kommunikasjon. Oversettelsesprogram er ikke tillat.
Bruk av kilder	Dersom du bruker kilder i svaret ditt, skal de alltid føres opp på en slik måte at leseren kan finne fram til dem. Det er fint om du bruker fotnote. Du skal føre opp forfatter, fullstendig tittel på både lærebøker og annen litteratur og sidetall. Dersom du bruker utskrift fra internett, skal du føre opp nøyaktig nettadresse og nedlastingsdato.
Informasjon om oppgaven	Prøven har 3 oppgaver: **Task 1**, **Task 2** og **Task 3**. Du skal svare på alle tre oppgaver, to som krever kortere svar og en langsvarsoppgave. **Task 1** og **2** (short answers), har en oppgave med utgangspunkt i tekstvedlegg (unknown text) som ikke er i forberedelsesmateriellet. Den vedlagte teksten er innenfor temaet for prøven. I **Task 3** (long answer) skal du velge en av oppgavene, husk å skrive tydelig hvilken oppgave du har valgt.
Praktiske opplysninger	Å presentere fagstoff, å bruke estetiske virkemidler og å utforme tekster er en del av kompetansekravene i engelskfaget. Du velger selv hvordan du best kan vise mottakerbevissthet og få fram hensikten med tekstene dine. Det gjelder også skrifttype og skriftstørrelse. **Bruk skriftstørrelse 12 og linjeavstand 1,5. (Arial, Calibri, Verdana og Times New Roman har god lesbarhet.)** **Husk også topptekst og sidetall x av y.**

Task 1 Short Answer

In the chapter *Shadows of the Past* we meet several brave people who fight for their rights.

Choose **ONE** of the two:

Rosa Parks

OR

Martin Luther King

- Why was the person you have chosen such a great inspiration to others?
- Who do you look to as a similar role model today?

Task 2 Short Answer

In the appendix on the next page, you will find an unknown text. Read the text. Answer the following questions:

- What is the message in the text?
- How does the language strengthen the message?

Task 3. Long Answer

Choose ONE of the tasks below.

a. Based on the text "The Ruin of a Culture" in Crossroads 9A, write a text about the first encounter between the Native Americans and the White settlers. Facts or fiction. Give your text a suitable title.

b. Look at the picture below. Write a text inspired by the picture about feeling like an outsider. Give your text a suitable title.


c. When you were clearing up the attic the other day, you came across an old letter. It was written by your great, great grandfather. He witnessed the *Trail of Tears* and he wrote a letter to his father about it. It is full of detailed descriptions of the tragedy. The letter is written in Norwegian. Write it in English. Give your text a suitable title.

d. You are a former slave who escaped to freedom two years ago and you are now going to write your autobiography. You might like to include information on where you were born, your daily routines, your longing for freedom, love and escape. Give your autobiography a suitable title.

e. You live in a country where you are being discriminated against because of the colour of your skin, your religion or your sex. Write an article for a newspaper / magazine. Discuss and reflect. Give your article a suitable title.

f. Look at the picture below. It is called Moving Day and painted by Norman Rockwell. Create a text where you tell the story from the point of view of one of the children. Make sure to include insight about American society at the time. Give your text a suitable title.

g. *Shadows of the Past* is based on periods of racism in history, do you see any similar situations in the present. Discuss. Give your text a suitable title.

School 2:

Here There and Everywhere
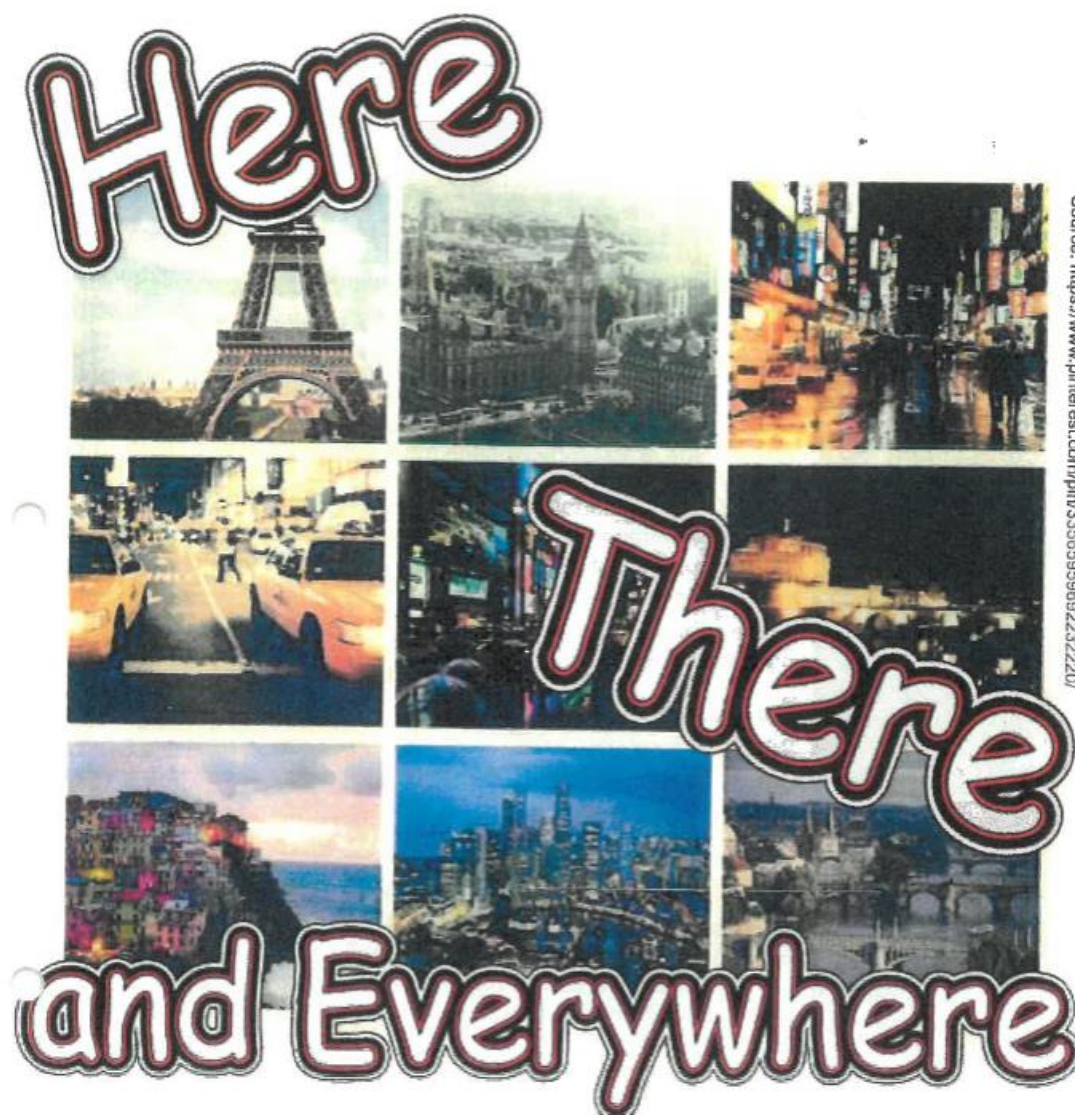
Tasks for English test
Spring 2015
Level B

Digital version not for print
Must be removed from student PC after use

Pedagogisk senter 2016, Ole Martin Kyllo og Sonya Eriksson

Informasjon om prøven

Prøvetid	Prøven varer i 5 timer
Tema	Here, There and Everywhere
Hjelpemidler	Alle hjelpemidler er tillatt, bortsett fra Internett og andre verktøy som kan brukes til kommunikasjon. Oversettelsesprogram er ikke tillatt.
Bruk av kilder	Dersom du bruker kilder i svaret ditt, skal de alltid føres opp på en slik måte at leseren kan finne fram til dem. Du skal føre opp forfatter og fullstendig tittel på både lærebøker og annen litteratur. Dersom du bruker utskrift fra Internett, skal du føre opp nøyaktig nettadresse og nedlastingsdato.
Informasjon om oppgaven	Prøven i engelsk har tre oppgaver. Du skal svare på oppgave 1 og 2. I oppgave 3 skal du velge en av fire oppgaver. Der det er oppgaver til valg, må du skrive nummer og bokstav på oppgaven du har valgt.
Vedlegg	Oppgaven har ett vedlegg: *Costa Rica*
Praktiske opplysninger	Å presentere fagstoff, å bruke estetiske virkemidler og å utforme tekster er en del av kompetansekravene i engelskfaget. Du velger selv hvordan du best kan vise mottakerbevissthet og få fram hensikten med tekstene dine. Det gjelder også skrifttype og skriftstørrelse. **Standard for brødtekster er skriftstørrelse 12 og linjeavstand 1,5. (Arial, Calibri og Times New Roman har god lesbarhet.)**
Vurdering	I vurderingen av denne prøven skal lærer bruke kjennetegnene på måloppnåelse (se neste side)

**There are three tasks in this test.
You must answer tasks 1 and 2 and ONE task
from 3a, 3b, 3c OR 3d.**

Task 1 Short answer

You have just visited a really interesting place when you went on holiday with your family.

Describe why the place was so interesting.

Task 2 Short answer - unknown text

In Appendix 1, you will find a travel poster on Costa Rica.

Write a short summary in your own words of the main points in the poster.

Husk å sjekke at du har svart på alt i oppgavene før du leverer.

Task 3 Long answer

Choose ONE of the tasks from 3a to 3d.

In your answer, you can use information, ideas, useful words and phrases from the texts you worked with on your preparation day and from other material that you have worked with in your English course. You may also use information from other sources.

Task 3A

In the booklet you read about Dublin's top five for teens.

Decide on a place of your choice and suggest your top five for teenagers. Give each of your paragraphs a suitable title.

Task 3B

In the letters Ask Alvin you read about some different places. Imagine you are going to spend your summer vacation at home.

Write a letter to a friend explaining what you could do to make sure you have a holiday that includes doing things indoors as well as outdoors.

Task 3C

Look at the artworks in the booklet. They have been made by artists inspired by a country they have visited or read about.

Pick one of the artworks and write a text inspired by it. Choose your own style of writing and give your text a suitable title.

Task 3D

The booklet describes places that are Here, There and Everywhere.

Compare a place from the booklet with a place you have read about or visited. Include in your comparison the differences and similarities between the two places.

Appendix 3: Lexical Errors in Student Texts

Student	Corrected/removed	Number of Errors	Grade
P60200	A sleep - asleep	1	4-
P60201	An-and x2 Pusch-push Down't-Do not Sayd-said Get's – gets	6	3+
P60202	Borrowing was removed, wrong use (carrying). Oh was removed	1	4-
P60203	Stock-stuck	1	4
P60204	To-too Shhh, removed	1	5
P60205	Quit-quiet Evan-even Thought-though Shud-Should x 2 Sais-says Read-red	7	4
P60206	Clothe-clothes Waternose-Waterhose Majestik-Majestic	3	4
P60207	Summing removed, used wrong (buzzing) Fuggy-foggy Breath-breathe Here-her x 2	5	4
P60208	Flyed removed, does not show that they know the correct form (flew) Jes-Yes Plase-Placed	3	4

P60209	Ehhh, removed Hey – hi now one – no one	2	4
P60211	Wright-write	1	6/5
P60212	No corrections	0	5
P60213	No corrections	0	6
P60215	No corrections	0	4/5
P60102	No corrections	0	5/6
P60104	jus-just	1	5-
P60105	During-Doing Veracious – various x2 cozy-cosy (BE) med-me	5	3-
P60106	hart-heart mounts-months flit-flight maschie-machine	4	4
P60107	closes-closest mal-mall was sent-was not	3	3 /4
P60108	no corrections	0	3+
P60109	nighttime – night time (Voc. counts two words)	0	6
P60110	removed oh sight sing – sightseeing	1	4
P60111	hallo-hello warn-warm God-good	3	4/3
P60112	th-the hes at-he sat h-he cam-came eplore-explore	4	3/ 4
P60113	No corrections	0	4/5
P60114 XX	t-shirt-tshirt (Voc. counts these as two words if the hyphen is not removed) sweeter-sweater Tex-tax Wes hall – we shall desert-dessert	4	3
P60115	stably-stable	1	3

Appendix 4: Student Texts

Student P60213, Rebekka

[**Student assistant's comment:** a separate "Self Assessment" form (filled out by the pupil) is also attached in this PDF file]

Task 1

Martin Luther King Junior was, and still is, a great inspiration for millions of people. His mission was to end the segregation in the United States, and he was to do so peacefully. He was very good at preaching, he was strong willed and brave, which makes a great leader. One of the reasonsIbelieve he became such a great inspiration, is because he "fought" for such a great cause, an important one that touched many people. All the African Americans needed someone to start a movement to stop segregation.

Leonardo Dicaprio is that kind of a role model to me. His cause is the environment. Not only is he a fantastic actor, but he also cares a lot for the environment and has participated on many boards where the discuss such things. I feel sort of bad for him since people doesn't take him seriously because he doesn't have a degree, and is only an actor, but he is pretty cool and very smart.

Commented [A1]: How sweet! I wouldn't have guessed!

Commented [A2]: could you be more specific?

Task 2

Commented [A3]: Well - you answer the task. Were you considering making it a two paragraph text with a bit more knowledge and information?

First of all, the song is about Moses when he freed the Israel people from the Egyptians. Since this is a negro spiritual, the African American slaves sings about them being freed and that they need to be freed.
"Let my people go" is repeated a great number of times and it helps strengthen the message because that is the message, to let someone go free.

1

Task 3

Wolves

Monster! You're a monstrosity! A wild animal, run!

These were all normal comments for Petrus as he would walk along the market. Everywhere he went, there would be a crowd of people waiting to throw their moldy food, rotten tomatoes and other unpleasant substances at him. Guards would even put him in the pillory, just for *fun* if they felt like it. And then they would just leave him there for the rest of the day, humiliating him. "I'd rather die" he always think when he see the guards at the end of the street. But still, he manages to live through every day. He stands in front of a pastry shop, not that he could afford any of it, or that the owner would even sell him anything, he just appreciates the smell. But he can also see if any guards or angry people are coming for him behind, because of the strange amount of mirrors in the pastry shop. The owner's name is Bewilda Garder, she looks like the offspring of a donkey, and sounds like a cold and sick bird. Somehow, her daughter looks like an angel, and smells like newly bloomed flowers. She is also very kind to him, but sadly, she is never home, because her goody-two-shoes of a "boyfriend", who reminds Petrus of someone, whom he cannot remember, always takes her on to "great adventures". Petrus just stands beside the pastry shop for as long as he can until Bewilda shouts him away, because he is *scaring* all her customers. Never thinking it might be herself who scares them away.

"Get off me shop, you blasted animal," she shouted at him as she noticed him from inside. And then comes a sneaky trick: "What if I don't want to?" he provoked. Bewilda's face fired up all red and grabbed the closest thing to her, bread, and started running after him while throwing the bread. Petrus snatched one of them, and ran away, with a smug expression on his face. He went home, or home and home, it was more like a prison, from which he could not escape. He took a sharp right turn in

2

to a narrow alley, beside a tavern called Three Hooks. And by narrow, the author means, "so narrow that the narrowest street in a narrow town would have to bow down to this narrowing narrowest alley for how narrow it really is" because Petrus could barely stretch an arm. He made sure his bread was as hidden as possible under his coat and pushed it as flat as needed to his chest, so that no one would suspect a thing. He smoothly went under some broken pieces in the wall on the left and entered an underground tunnel staircase, which light had never touched. At the end of the stairs, there was a door, a quite secure metal door with a little hatch. He knocked three times, waited, and then knocked two times more. The hatch opened and two bloodshot eyes could be seen from inside. Petrus heard sulking and the iron door opened up. The man with the bloodshot eyes was no other than Crookshanks, Petrus' *caretaker*. A man with great temper, comprehensible looks and an addiction for money. Crookshanks let him in without a word, he only looked at him as if he wondered if he should punish him or not. The room they were in hadn't been cleaned in ages. There was only one bed and on the floor beside it was a few carpets and pillows. In another end of the room was something that could have been a kitchen but it looked more or less bombed now. "Ya didn't happen to get some food now did ya?" Crookshanks asked. Petrus sighed: "No, no one wishes to trade silk carpets from me. If only you had provided me with those golden coins of yours, maybe the.." Crookshanks Interrupted him mid sentence. "Give ya money so that ya can run away? Don't ya dare think I'm that stupid, boy", his face was almost as red as his eyes and Petrus whimpered as Crookshanks hand hit him across the face. Petrus was on the verge of tears. He fell on the floor and looked down in shame. Crookshanks didn't seem to care, "Now now, the full moon starts tomorrow night, and ya know what that means? Dontcha?" Petrus shivered, and nodded slowly.

Petrus hated the full moon. He used to love it, but now, hatred is everything when someone mentions it. A full moon lasts for three nights, and along with it comes *werewolves*. Petrus, as well as many others, is a werewolf, but in his particular case, he is used to fighting. In a pit, with another werewolf, sometimes even another normal human, until one of them dies. Petrus doesn't remember anything from when he is a wolf, he can only remember the feeling, and it feels good. Which scares him. This has been going on for two years. Two years with 24 full moons. Each lasts for three nights. That is 62 nights, and 62 fights where Petrus has survived every single

Commented [A5]: Wow! I must say - narrow it is!

Commented [A6]: You describe the character well!

Commented [A7]: brilliant!

Commented [A8]: Hm... a bit incomplete?

3

one, meaning he has killed 62 beings, if not more, and is now about to begin year three. "Get some sleep, we head to the Vindales tomorrow morning", Crookshanks fell asleep almost instantly, but for Petrus, it would be a sleepless night.

Petrus was anxious, but Crookshanks was looking forwards for tonight's event, because Petrus was his prized animal, who never lost. "I've heard rumours that there will be a champion there tonight, *she* is apparently so fierce that she makes all the males fall down by looking at her. I don't know what kind of lusts ya have when you're an animal, but you better not fall for her tricks!" he was clearly excited to see Petrus beat a champion, but the excitement would be for nothing.
They walked in silence on a gravel road, when Petrus wished he owned a pair of shoes, the tiny rocks was so sharp that Petrus was convinced there was a trail of blood behind them. They entered a gate to a field of scrap metal. There was no wind, but an aura of death. It smelled dried blood and iron, Petrus knew the difference. There was no sound except the occasional tin cans being kicked around by Crookshanks. As they continued along the piles of metal, the sound of laughter and talking became higher. A big circle surrounded by stacks of metal was in sight and Petrus hesitated, but Crookshanks looked harshly at him and made him continue against his will. The circle was a sunken area so the sides was used as a tribune for a better view of the fight down in a ten feet deep and twice as wide pit. There were two queues, one where the audience paid for tickets and bet their money on the contestants, and the other queue was for volunteers and not so voluntary contestants. "Crookshanks!" one in the contestant queue shouted, it was a short man with thick and long hair. Crookshanks jumped in line, with Petrus on tow, over to the man, without caring about the angry reactions of the others behind them. "Fergus! Long time no see eyh? But what are ya doing here, I didn't know ya had a wolf?" Crookshanks said enthusiasticly, "Bugger, I ain't got no wolf, I'm going to fight for myselfme self, it's a bloody good way to be done with this." Fergus laughed a very sinister laugh. He seemed to be the very suicidal type, since no human have ever survived a no weapon combat against a wolf. "You're crazy, ya know that right?" and so the dialogue continued without Petrus paying attention, since he only wished he could put his hands around Crookshanks neck or even his own to end the suffering of being there.

Commented [A9]: How horrible!

All werewolves contestants has to wear an arrow badge so that everyone knows who to throw their food at. They are also forcefully thrown into a cage so that they don't escape. There they won't get any food, whatsoever, so that they will be more fierce and bloodthirsty as wolves. The ones who run the show, knows that it is illegal to have these types of betting fights and that it is inhumane to keep someone in a cage without food, but their excuse is that werewolves aren't humans, only animals, no matter if it is full moon or not. As Petrus and Crookshanks entered the "Cage room" he noticed there were more than there used to be. Crookshanks found an empty cage and pushed Petrus inside, he also took his clothes and his excuse for destroying that kind of privacy was "I don't want ya to rip them apart, besides you'll need them afterwards" which can be translated to "I do not want to spend the money you make on your own comfort.". Petrus was relieved, scared and humiliated. Relieved because he now had time for himself, to think and be alone, scared because he doesn't know if he will survive, and humiliated because he was without clothing where people could come and go as they please. None of the other werewolves had clothes either he soon noticed. For some strange reason it gave him some sort of feeling of solidarity and made him feel less uncomfortable.

There were a lot of strange people in the cages. There were old teenagers as himself, old men and women, and even children. Everyone was quiet, everyone sat still, waiting. In the cage closest to him, was a sleeping girl, around his age, perhaps a few years younger. She had scares all around her body and she was very beautiful. He felt sorry for her, to be so young and be used this way. She must have noticed he was watching her since she quickly woke up and breathed quickly as if she was scared. They made eye contact. "Where am I?" she looked around desperately and then noticed she didn't have any clothes and covered herself up, "And who the hell are you?" she almost shouted.
"Shhh! Calm down, I'm Petrus, what is your name?"
"Rosalia."
"Ok, Rosalia, you're a werewolf, right?"
Rosalia hesitated, "Yes, how did you know?"
"Because you are probably being used in betting fights if you're here."
Rosalia was shocked at first, but Petrus explained to her how it worked and she soon seemed to understand how it fit together.

5

"For a while now I've woken up in my dad's horse carriage, never remembering what happened before. Are you saying that my own father is drugging me down every full moon?"

"Yes probably, butIthink it is more like an amnesia drug, so you probably won't remember this either."

They talked for what seemed like ages, they both shared their history with each other and really got to know one another, for Petrus it felt like all bad was gone and that everything would be fine. But of course, it didn't last long. Petrus noticed something happening, to himself and all the others. The senses grew stronger, he saw every detail in the dark room, he heard the mice running inside the walls, he smelled meat from outside. Slowly everyone were wolves and began howling, the night had begun.

It was a room filled with blood thirsty and big wolves. Gustav Heimek was responsible for letting the wolves out in the pit. He picked two and two, without caring about who they actually were. First round: A brown one with raggy fur and one blind eye against a silver one with sleek fur. The silver wolf won.
A fully grown grey wolf against an old cub. The grey wolf won.
A human, Fergus, against an albino wolf. The wolf won.
And so it continued, and the crowd went wild for every death. Crookshanks had payed Gustav Heimek to pick Petrus to fight against the champion, the one everyone had been waiting for. A ragged silver fur against a black female. The cages had been right beside each other all the time, neither of them knew who the other really was.

Student P60209, Susanne

Shadows of the Past

Task 1.

Rosa Parks

Rosa Parks was a great inspiration to many others. She was a woman who was really tired of discrimination. When she was on the bus on her way home, the bus driver asked her to give up her seat to a white person, but she refused. The police arrested her. She was a really brave woman and I think that many others can learn from her. If you feel that you are discriminated because of your skin color, religion or sex you have to do something, like Rosa did.

Malala Yousafzai has fight for human rights and for girls rights for education. In some countries in the world, the women haven't got the same rights aslike men and this is just one of the things that has to changed. Malala is a very young girl, but she fights for equality. She has been given the (Nobel) peace prize for her work. She has been a model for girls right for education.

> Commented [A1]: How can you connect the two paragraphs?

Task 2.

Go down Moses

The Negro Spiritual songs are christan songs and it is about that they have been oppressed and discriminated and they have a hope to come to freedom.
I think the message in this text is that everyone is just as much worthwhile however you are black and white.
But they had hope to come to freedom or come to Christ and be free, so they continued walking.

> Commented [A2]: Yes!

> Commented [A3]: Suggestion: ..worth the same wether you are black or white.

The language strengthen the message because it is many sentence that are repeated like; Go down Moses, way down in Egypt's land, Tell old Pharaoh: Let my people go.
This strengthen the message because when you read it or hear it many times, you understand that it is more and more important.
The text is not straight forward, you have to think, what do they mean bywith that. So it is a lot of symbols in the text.

> Commented [A4]: Yes!

> Commented [A5]: Yes! Why is that? What do you call that?

1

Task 3.

B)

"Hey, you are welcome to my birthday. Heey, you are welcome to my birthday."
Sophia has birthday today and she has invited the whole class. I'm invited, but I don't
know if I'm want to go. Everyone has for sure buy a big surprise for her, but my
family has not that kind of money. I'm just glad if I get something to eat before I went
to sleep. Nobody knew that my family is poor and I want it to continue like that.

"Are you coming to my birthday, Tommy?", Sophia asked me. "Ehhh, I don't know
yet", I said but I was really sure that I doesn't want to go. "I hope you can come, all
the other boys in the class are coming", she said and walked away.

The boys in the class has fast cars, a lot of friends and, some of them has also have
a girlfriend. Me, howeverwhatever, haven't friends so than I will just stand for myself.
I'm a completely loserlooser. My parents has told be to be more social and make
some friends, but I guess now one will be my friend, but it's worth a try.
I should go to the party and make some friend. I should go to the party and make
some new friends. It feel that I has convinced myself.

At the end of the day, a boy from my class asked me if I'm going to Sophia's
birthday. "I haven't decided yet." "Not me either, but I think I will go and we can just
go home earlier." he said and look me in the eye. "Yes, I'll guess I see you then", I
said and walked away. I always be really nervous when someone looks me in the
eye. I feel that they can look through me.

"Hello, Tommy! How was your day?" "Fantastic, I made me a friend today and I will
meet him at Sophias birthday tonight." "Wow, great. What are you gonna wear
then?" "I have no idea. I think I just go in this clothes and I can maybe borrow dad's
jacket?" "You may borrow it if you are very careful because it cost a lot of money."

I walked upstairs in my room and started to fix my hair og look for dad's jacket. The
birthday starts 18.00 and now the clock is 16.45. I want every single hair to look
perfect so it will take some time.

It The clock is 17.45 and I am on my way to Sophia's house. It is really hot outside
and I began to sweat. Oh no, I hope I puttook on enough perfume on. I walked and
walked, my house is very long from Sophia's house. In front of me, I see many who
are dressed up and on their way for sure to Sophia. I hope the guy I talked to earlier

Commented [A6]: You will choose one of them, right!

Commented [A7]: Yes, but I dont know witch of them yet.

Commented [A8]: You start out in the present tense and then you switch to past. Remember to make up your mind and keep to one tense of the verb. :-)

Commented [A9]: https://painfulenglish.wordpress.com/2013/07/11/loose-or-lose-looser-or-loser/

Commented [A10]: https://painfulenglish.wordpress.com/2013/07/11/loose-or-lose-looser-or-loser/

Commented [A11]: Good description! And correct English!

Commented [A12]: Who says what? And how? Could some of the direct speech be referred to?

Commented [A13]: Now you're writing in the present tense again...

is there and not changed his mind. There is see the big white house at the end of the corner. I begin to feel insecure. Am I doing the right thing?

When I rap on the door, Sophia open and she looks very happy. She just stand in the door and looks me in the eyes. That makes me very uncomfortable. I asked her if I could come in and she mutter and say yes. When I come inside I saw a pile ofwith presents. When I come longer and longer inside I were more and more insecure that I went here. Yes, he I talked to earlier is here. He sit with the "cool" guys, but I went to him because he is the only one I know here.
Hello, I said and smiled to them. They gave me non response. I sat me down behind one of them in the sofa and they started staring at me. It is like they know that I be insecure when they do it. Everyone was quiet, very quiet and just staring. "So, I see you came here", I said to him I had talked to earlier. He doesn't look on me, just began to talk to one of the others. They talked and talked, and everytime I tried to say something they just look on me like I was a little kid.

"I think I am going home", I said. They didn't give me response on that either, like I am air. On my way out I hear Sophias voice that called out my name. "Tommy, you haven't give me a present yet!" "Oh sorry, I forgotten it", I said and tried to move. "If I can take you jaket, it is okay." "No, you can't", I shouted out. "It is not so expensive." She took it off me. I draw it but she get help from the other guys and it fissure. I started to run home. This has been the most awful day in my life. I will never in my life get some new friends.

Sources:

From task two, Malala: Written by Bjørn Johannesen and Marte Ericson Ryste. Publish 17. october 2017. *Malala Yousafzai*. Taken from:
https://snl.no/Malala_Yousafzai

Student P60205, Johanne:

Task 1

~~Martin Luther King~~

~~Martin Luther King~~He was a great inspiration for many because he stood up for his rights, and fought the system so that everyone would get equal rights. He never gave up, even when he got thousands of threats to his home and was attacked several times. He inspired people to stand up for their rights and to never give up on what they~~you~~ believed in.

Malala Yousafzai is a great role model for me and she has inspired many. She fights against oppression of kids and youth, for children's right to education and she also fights for woman rights for education. Even after she was shut in the head, she did not back~~ed~~ down, she just kept fighting and therefor she's a great inspiration for many to never give up and to always do what's right.

> **Commented [A1]:** how could you connect the two paragraphs even more? I guess a simple -...is also a great... would make it obvious that the two paragraphs are connected?

Task 2

It's a christen Negro Spiritual and the black people sings about being freed. I think the message is that everyone is equal worth and we have to stop acting like the black people is worth less. But I also think it was meant to give black people hope so they wouldn't give up and it had a hidden meaning, so people could escape from slavery.

They have several repetitions off Go down, Moses and Tell old Pharaoh: Let My people go and Let My people go. It emphasized the message, because you understand that it's important and then you understand the message better too. You have to read the text a couple times before you understand it, because they wrote the text quite mysteriously some places.

> **Commented [A2]:** The text has several literal devices to get the message across.....

> **Commented [A3]:** Can you give examples? Why do you think so? What effect does it have?

Task 3

(Text B)

Never again

The school bell rings and I see the other kids running inside, but all I wanna do is keep hiding. Keep hiding forever, never to be seen ever again. To be invisible has to be better than *this*, everything has to be better than *this*. There is no one left outside and it's so quit now. I know I'm gonna be late to class, but honestly, I would much rather go home. But I force my legs to go slowly towards the school door, and suddenly much faster then I wanted, I'm standing right outside the classroom. I'm both cold and warm and the last thing i want to do is to go inside that door, but that's exactly what I'm doing.

The door creaks and slowly opens. Everyone is staring at me, the teacher looks at me with a harsh look, but she doesn't say anything. That's almost worse then when she yells at me. I hurry down to my seat with my eyes fastened to the floor. Evan thought, I can feel Marius smiling and when I'm walking by him, he puts something in my hand. I sit down and sees that it's a piece of paper. Marius has drawn two dead adults and a kid, and written with read script *orphan*. My stomach hurts and I feel sick. I shud be angry, I shud be pissed, but all I feel is sadness.

The bell rings and it's break time. A cold hand grabs my shoulder, and before any normal person would react, I turn around in full alert with my pulse up in 220. "It's just me." sais Maira and my pulse calms down. "Why did you scare me like that?" I try to sound angry, but she can hear the sadness in my voice. "What's wrong, Mihow? Is it Marius again?" I just nod. "Well, let's focus on something else." She just keeps babbling, and with Maira by my side, the school day actually goes quite fast, even with Mariuses comments.

> **Commented [A4]:** Well described!

Maira is my only friend and she's the only one who treats me like a normal person. Because I'm an immigrant and an orphan, many threats me like a baby or like I'm not welcomed here. But still, I feel very lucky to have a friend and to be alive. Not many can say the same. Now, I live with my older brother, sister and my younger sister. My younger sister, Mani is only four and I'm often the one who has to watch over her, but that's fine by me.

The school day is finished and Maina asks if I want to go home to her place. "Of course, I just has to call Josef (my older brother) to check that it's ok." I borrow Mairas phone and call Josef. He sais it's fine, I just has to be home by 21.00. Me and Maira laughs and talks all the way to her house. Her parents isn't home and we can do whatever we want. "My parents is gone all day, so we can start with, what shall we make for dinner?" Maina smiles a sly smile. "I don't know. What do you have in mind? asks I. "Let's make waffles!"

> **Commented [A5]:** watch out for mix of verb tenses!

> **Commented [A6]:** Can you rephrase - I understand that it's oral - but it must be rewritten some how. Maybe go right from: My parents are gone all day! to the: "what should we make for dinner?"

I haven't had so much fun in ages, me and Maina makes horrible waffles, but it's so much fun and we're laughing so much that my stomach hurts. But then, I look at the clock, and it's 20.40. "Oh. Maina, I need to go. I had to be home at 21.00." "Oh, ja. That's right." I can hear the disappointment in her voice, and I feel the same. I give Maina a big hug and then I leave. "Bye."

> **Commented [A7]:** Remember both in Norwegian and English - put yourself last :-)

> **Commented [A8]:** What can replace "so much"?

> **Commented [A9]:** you don't need these small obvious direct speeches. Then I guess there should be something about the way she said bye...Are you with me?

It's cold outside and dark. I can't see anyone and it's so quiet, so very, very quiet. I got a bad feeling and even though it's cold, I start sweating. My eyes start looking wildly around and I'm almost running now. Then I freeze, I slowly turn around, and there it stands a guy with a black mask. It's two others too, they've surrounded me. My heart beats like a drum, fast, and I have a big lump in my throat. I start grasping, it doesn't feel like I can breathe properly. I'm scared, I'm death scared. What will they do with me?

They're getting closer and closer and I'm getting more and more panicked. "Please, don't hurt me. I'm just a normal guy." It doesn't looks like they heard me, but they're almost at me now. Suddenly, I'm so calm. I know what to do and I'm tired of letting people harass me because off my background. "Hallo, help! There is some guys attacking me!" In case no one heard me, I'm ready to fight back. I starts tripping and in that exact moment, they attack.

The highest guy tries to punch me, but I duck and kicks him in the stomach. He falls and grounds. Then the two others comes at me, but the silence is interrupted by a siren. I smile, and looks at the three boys. They try to run, but it's two police cars, and I know they'll catch them. I will never again let people harass me or discriminate me or others like that. I will fight for my rights and be proud of who I am, I will never be afraid again.

Sources
Task two, Malala: Johannesen, Bjørn and Ericsson, M.R(2017). *Malala Yousafzai*. Taken from: https://snl.no/Malala_Yousafzai